



HAL
open science

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.

1
2
3 **Validating missing proteins in human sperm cells by targeted mass spectrometry- and**
4 **antibody-based methods**
5
6

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

12
13
14 ¹ Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), IPHC, Université de
15 Strasbourg, CNRS UMR7178, 25 Rue Becquerel, Strasbourg F-67087, France
16

17
18 ² CALIPHO Group, SIB-Swiss Institute of Bioinformatics, CMU, rue Michel-Servet 1, CH-
19 1211 Geneva 4, Switzerland
20

21 ³ H2P2 Core facility, UMS BioSit, University of Rennes 1, Rennes F-35040, France
22

23 ⁴ Protim, Inserm U1085, Irset, Campus de Beaulieu, Rennes F-35042, France
24

25 ⁵ Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala
26 University, Uppsala, Sweden
27

28 ⁶ Service de Médecine de la Reproduction, CHU de Nantes, 38 boulevard Jean Monnet,
29 Nantes cedex F-44093, France
30

31 ⁷ Inserm UMR1064, Nantes F-44093, France
32

33 ⁸ CEA, DRF, BIG, Laboratoire de Biologie à Grande Echelle, 17, rue des Martyrs, Grenoble
34 F-38054, France
35

36 ⁹ Inserm U1038, Grenoble F-38054, France
37

38 ¹⁰ Grenoble-Alpes University, Grenoble F-38054, France
39

40 ¹¹ Department of Human Protein Sciences, Faculty of medicine, University of Geneva, 1, rue
41 Michel-Servet, 1211 Geneva 4, Switzerland
42

43 ***: To whom correspondence should be addressed**
44

45 Christine Carapito: ccarapito@unistra.fr tel: +33 (0)3 68 85 27 30
46

47 Charles Pineau: charles.pineau@inserm.fr tel: +33 (0)2 23 23 52 79
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 **The mass spectrometry proteomics data have been deposited in PeptideAtlas**
5 **with the dataset identifier PASS01013**

6
7 *Reviewers, please access our data using the following account details:*

8
9
10 *Submission details:*

11
12 **Identifier:** PASS01013

13
14 **Dataset type:** PRM

15
16 **Dataset tag:** HPP_Testis_PRM_2017

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

19
20
21 *Reviewer account details:*

22
23
24 **Username:** PASS01013

25
26 **Password:** JU2578wb

27
28
29
30
31 **Keywords:** human proteome project, spermatozoon, missing proteins, Parallel Reaction
32 Monitoring, targeted proteomics, immunohistochemistry, immunocytochemistry,
33 bioinformatics, data mining

34
35
36
37 **Email addresses:**

38 CC: ccarapito@unistra.fr

39 LL: lydie.lane@sib.swiss

40 YV: yves.vandenbrouck@cea.fr

41 PD: paula.duek@sib.swiss

42 MS: marine.seffals@univ-rennes1.fr

43 KR: karine.rondel@univ-rennes1.fr

44 CM: c.macron@unistra.fr

45 FD: delaland@unistra.fr

46 TF: thomas.freour@chu-nantes.fr

47 CL: cecilia.lindskog@igp.uu.se

48 CP: charles.pineau@inserm.fr

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

1
2
3 including proliferation and differentiation of spermatogonia taking place in the basal
4 region of the seminiferous epithelium, meiotic division of primary and secondary
5 spermatocytes, and differentiation of haploid spermatids, leading to the production of
6 spermatozoa - commonly called sperm - in the luminal region of the seminiferous
7 epithelium (for a review see ⁹). Interestingly, the selective pressure on most genes
8 involved in spermatogenesis implies a high degree of germ cell-specific expression
9 with approximately half of these only expressed in the meiotic and post meiotic germ
10 cell lineages ¹⁰. A first trans-chromosome-based pilot study demonstrated the
11 potential of isolated spermatozoa for the validation of a first set of 89 missing proteins
12 ¹¹. More recently, we conducted an in-depth analysis of human sperm in which we
13 could validate 206 missing proteins (PE2-PE4) that fully comply with current HPP
14 guidelines ¹². In order to complete the annotation of the proteins from chromosome 2
15 and 14 that are expressed in human sperm, we turned to a targeted mass
16 spectrometry (MS) approach allowing for high sensitivity ¹³. Based on information
17 available in the literature and in the main C-HPP resources (Human Protein Atlas,
18 PeptideAtlas and neXtProt), a subset of 40 theoretically detectable missing proteins
19 from chromosome 2 and 14 was proposed for a targeted search in the human sperm
20 ¹⁴.

21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43 This work lead to the launch of the so-called “neXt50 challenge”, a coordinated effort
44 across all chromosome teams to identify the 50 most tractable missing proteins for
45 their chromosome. The present study is the contribution of the Swiss and French
46 teams to this new C-HPP challenge. We report the targeted search of 38 theoretically
47 detectable missing proteins in sperm fractions following a Triton X-100 fractionation
48 procedure previously employed by our consortium ¹². Using Liquid Chromatography-

1
2
3 Parallel Reaction Monitoring (LC-PRM) assays with heavy labeled synthetic peptides
4 targeting 92 proteotypic peptides, we were able to unambiguously detect 12 of the 38
5
6 selected MPs with 2 or more peptides and 3 with 1 peptide. Antibody-based
7
8 validation was done for 7 of the proteins identified by targeted-MS, and for C14orf177
9
10 for which MS evidence was weak and non-conclusive, and functional hypotheses
11
12 suggested for some of them, notably for LRRC9, FAM181A, ADAM20, C14orf177
13
14 and DNAJC5G/ CSPgamma.
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 collaborators ¹⁵ and used in our previous study ¹². Briefly, three “swim up” spermatozoa fractions in 300 μ L 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 collapse ¹⁵. 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: 300 μ L of sample was added to 300 μ L methanol and 150 μ L of chloroform. After vortexing and centrifugation for 2 minutes at 10000 x g, the upper phase was removed and 225 μ L 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

1
2
3 The Triton X-100 soluble fraction was resuspended in Laemmli buffer containing
4
5
6 Tris HCl pH 6.8, EDTA 1mM, 5% β -mercaptoethanol, 5% SDS and protease
7
8
9 inhibitors while Triton X-100 insoluble fraction was diluted in the same, but two
10
11
12 times concentrated Laemmli buffer. For both samples, protein concentrations were
13
14
15
16 assayed using Bradford protein assay (Bradford Assay; Biorad, Hercules, USA)
17
18 according to manufacturer's instructions. 150 μ g proteins from both fractions were
19
20
21 loaded into an in-house poured 10-wells 4-10% acrylamide gel. Electrophoresis was
22
23 performed under a continuous voltage of 50 V during 20 minutes and then 110 V
24
25 during 75 minutes. Gel was stained with Coomassie Blue and each lane was
26
27 manually cut into 23 slices of similar size. In parallel, stacking gels of 50 μ g proteins
28
29 were run for both fractions. Samples were loaded into an in-house poured 10-wells 4-
30
31 10% acrylamide gel and electrophoresis was performed under a continuous voltage
32
33 of 50 V during 23 minutes. The gel was stained with Coomassie Blue and stacking
34
35 bands were manually cut. Proteins in the gel slices were reduced, alkylated and
36
37 digested overnight at 37°C with modified trypsin at a 1:100 enzyme:protein ratio
38
39 (Promega, Madison, USA). Peptides were extracted during 45 minutes with 100 μ L of
40
41 60% acetonitrile, 0.1% formic acid (FA) and 15 minutes with a solution of 100%
42
43 acetonitrile.
44
45
46
47
48
49

50 **Targeted LC-PRM Analyses**

51
52 For the 38 selected missing proteins, 92 proteotypic peptides that passed the
53
54 neXtProt uniqueness criteria ¹⁶ were chosen based on prediction of MS-friendliness
55
56

1
2
3 (SRM Atlas resource and manual selection) and synthetic heavy labeled peptides
4
5 were purchased (crude SpikeTides L, JPT Peptide Technologies GmbH, Berlin,
6
7 Germany). Selected peptide sequences are listed in Supplementary Table 1. The
8
9 heavy labeled peptides were mixed together in a hand-adjusted concentration-
10
11 balanced mixture to equilibrate individual peptides signals and spiked in a complex
12
13 matrix (pool of stacking bands extracts). The peptide mixture was analyzed by
14
15 nanoLC-MS/MS (Q Exactive Plus, Thermo Fisher Scientific) to acquire reference
16
17
18 higher energy collisional dissociation (HCD) fragmentation spectra (analytical details
19
20
21 are provided in Supplementary Material file). Following this step, targeted assays
22
23
24 using PRM acquisitions were developed to target all 92 peptides and their modified
25
26
27 forms (with/without oxidized methionines), first in the unfractionated stacking gel
28
29
30 extracts and subsequently in gel bands obtained from 1D SDS-PAGE separation of
31
32
33 both Triton X-100 soluble and insoluble fractions. LC-PRM analyses of spermatozoa
34
35
36 extracts were performed on a NanoAcquity LC-system (Waters, Milford, MA, USA)
37
38
39 coupled to a Q-Exactive Plus Orbitrap (Thermo Fisher Scientific, Waltham, MA, USA)
40
41
42 mass spectrometer equipped with a nanoelectrospray ion source. The HPLC system
43
44
45 consisted of a solvent degasser nanoflow pump, a thermostated column oven kept at
46
47
48 60°C and a thermostated autosampler kept at 10°C. Mobile phase A (99.9% water
49
50
51 and 0.1% FA) and mobile phase B (99.9% acetonitrile and 0.1%FA) were delivered
52
53
54 at 450 nL/min. Samples were loaded into a Symmetry C18 precolumn (0.18x20mm,
55
56
57 5µm particle size, Waters) over 3 minutes in 1% buffer B at a flow rate of 5 µL/min.
58
59
60 This step was followed by reverse-phase separation at a flow rate of 450 nL/min

1
2
3 using an ACQUITY UPLC® BEH130 C18 separation column (250mmx75µm id,
4
5 1.7µm particle size, Waters). Peptides were separated using the following gradient:
6
7 from 1% B to 5% B in 2 minutes, from 5% B to 25% B in 58 minutes, from 25% B to
8
9 35% B in 19 minutes, from 35% B to 90% B in 1 minute, maintained at 90% B for 5
10
11 minutes and the column was reconditioned at 1% B for 20 minutes.

12
13
14 Peptides were detected by directly eluting them from the HPLC column into the
15
16 electrospray ion source of the mass spectrometer. An electrospray ionization (ESI)
17
18 voltage of 2 kV was applied to the HPLC buffer using the liquid junction provided by
19
20 the nanoelectrospray ion source, and the ion transfer tube temperature was set to
21
22 250°C. The Q-Exactive Plus Orbitrap instrument was operated as follows: after a full
23
24 MS scan, twelve peptide ions defined in the inclusion list were selected for MS/MS
25
26 (all 212 targets and time scheduling options are listed in Supplementary Table 1 and
27
28 details on the PRM time-scheduled method development are provided in
29
30 Supplementary Material file). Full scan MS spectra (mass range 300-1800m/z) were
31
32 acquired with a resolution of 70,000 at 200 m/z with an automatic gain control (AGC)
33
34 fixed at 3×10^6 ions and a maximum injection time set at 50 ms. Targeted MS/MS
35
36 spectra were acquired with a resolution of 17,500 at 200 m/z with an AGC fixed at $1 \times$
37
38 10^5 and with the maximum injection time set at 60 ms. LC-PRM .raw and Skyline
39
40 results files were submitted to the PeptideAtlas/PASSEL repository (PASS01013),
41
42 dot-product and rdot-product values were calculated with Skyline, all results were
43
44 manually checked and validated peptides and traces are provided in Figure 1, Table
45
46 1 and Supplementary Figure 1.

54 **In situ validation of expression data by immunohistochemistry**

1
2
3 To confirm the germline expression of proteins of interest, immunohistochemistry
4 experiments were performed on human testes, fixed in 4% paraformaldehyde and
5 embedded in paraffin, as previously described ¹⁷. Normal human testes were
6
7
8
9
10 collected at autopsy at Rennes University Hospital from HIV-1-negative cadavers.
11
12
13
14 Paraffin-embedded tissues were cut into 4 µm thick slices, mounted on positively
15
16
17 charged slides, and dried at 58°C for 60 min. Immunohistochemical staining was
18
19
20 performed on the Discovery Automated IHC stainer using the Ventana DABMap or
21
22
23 OMNIMap detection kit (Ventana Medical Systems, Tucson, USA). Following
24
25 deparaffinization with Ventana EZ Prep solution at 75°C for 8 minutes, antigen
26
27 retrieval was performed using Tris-based buffer solution CC1 (Ventana Medical
28
29 Systems) at 95-100°C for 36 minutes. Endogen peroxidase was blocked with
30
31 Inhibitor-D 3% H₂O₂ (Ventana Medical Systems) for 8 minutes at 37°C. After rinsing
32
33
34 with reaction buffer (Roche), slides were incubated at 37°C for 60 minutes with
35
36 polyclonal rabbit antibodies specific for the selected missing proteins (Human Protein
37
38 Atlas Antibodies). The antibody dilutions used are listed in Supplementary Table 2.
39
40 After rinsing, signal enhancement was performed at 37°C using the Ventana
41
42
43 DABMap Kit or OMNIMap and secondary antibody: Anti-Rabbit HRP (pre-diluted
44
45 commercial solution, Roche) for 16 minutes or Biotinylated horse anti Rabbit (1:700)
46
47 for 32 minutes incubation. Slides were then counterstained for 16 minutes with
48
49 hematoxylin (commercial solution, Microm), 4 minutes with bluing reagent
50
51 (commercial solution, Microm), and rinsed with Milli-Q water. After removal from the
52
53
54 instrument, slides were manually dehydrated, mounted in Eukitt (Labnord, Villeneuve
55
56
57
58
59
60

1
2
3 d'Ascq, France) and coverslipped. Finally, digital immunohistochemistry images were
4
5 obtained using NDP.Scan acquisition software (v2.5, Hamamatsu) and visualized
6
7 with NDP.View2 software (Hamamatsu).
8
9

10 11 **Immunocytochemistry on sperm smears**

12
13
14 Immunocytochemical staining was performed on the Discovery Automated IHC
15
16 stainer using the Ventana DABMap or OMNIMap detection kit (Ventana Medical
17
18 Systems, Tucson, Ariz). After cell fixation and permeabilization in methanol, sperm
19
20 smears were treated using the same procedure as the one used for
21
22 immunohistochemistry (see above). Immunocytochemistry images were visualized
23
24 under an Eclipse NI-E microscope (Nikon, Champigny sur Marne, France).
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

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

1
2
3 unfractionated samples. The second attempt including a prior extensive 1D-SDS-
4
5 PAGE fractionation in 23 bands of each of the soluble and insoluble fractions
6
7 revealed to be far more successful. Indeed, signals were detected for 29 out of the
8
9 92 tracked peptides enabling 15 out of the 38 targeted MPs to be validated (Table 1).
10
11 Complementary results were obtained thanks to the Triton X-100 fractionation as 17
12
13 and 14 peptides were validated in the soluble and insoluble fraction, respectively
14
15 (Table 1). It is of note that even with thorough fractionation, the measured signals
16
17 were overall of low intensity and for most peptides close to detection limits
18
19 (Supplementary Figure 1). All chromatographic traces were manually checked and
20
21 the following criteria needed to be met for a peptide to be validated: a good co-elution
22
23 between light and heavy transition groups, high calculated dot-product for
24
25 endogenous peptide (all >0.85, average value=0.93) and rdot-product (all >0.94,
26
27 average value=0.99) and repeated detection of the endogenous transition group
28
29 (mostly in adjacent gel bands or in replicate injections). Finally, signals of unlabeled
30
31 peptide in the crude heavy labeled peptides were checked to be below detection
32
33 limits and transition groups unspecifically detected in multiple gel bands were
34
35 excluded. Altogether, these results indicate that the 12 MPs detected with 2 or more
36
37 peptides and the 3 observed with a single peptide using targeted MS may have
38
39 escaped prior detection because they are expressed at very low copy numbers.
40
41
42
43
44
45

46 ***Targeted MS results support data mining prioritization strategy***

47
48 The prioritization strategy presented by Duek and collaborators ¹⁴ was based on
49
50 information retrieval in the literature and from the three main C-HPP resources
51
52 (Human Protein Atlas, PeptideAtlas and neXtProt). Proteins that were experimentally
53
54
55

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

18 According to neXtProt, RNF212B is predicted to act in the synaptonemal complex,
19 which is disassembled during sperm maturation. Therefore, it might be present only
20 in germ cells but not in mature sperm. RNASE11 is highly expressed in testis and
21 epididymis, according to recent RNAseq data. However, since it is predicted to be
22 secreted, it might be easier to detect it in seminal plasma than in sperm cells.
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 Based on RNA-Seq data from HPA, six MPs for which we obtained mass
34 spectrometry evidence in sperm (FAM181A, DNAJC5G, LRRC9, ALLC, C2orf78,
35
36

1
2
3 MAP3K19) correspond to genes showing a medium expression level in testis
4
5 (between 10 and 100 transcripts per million, TPM), seven others (ADAM20, TEX22,
6
7 EVX2, SLC35F4, FER1L5, KIAA2012 and ANKRD30BL) correspond to genes with
8
9 a low expression level (less than 10 TPM) and two others to genes for which the
10
11 expression in testis was validated by quantitative RT-PCR while no RNA-Seq data
12
13 was reported (HERVs) ¹⁴.

14
15
16 Although they were selected according to their preferential expression in testis, the
17
18 23 MPs that could not be validated may be present at levels below the PRM
19
20 detection limit in sperm. Indeed, the expression level of a transcript does not
21
22 necessarily indicate that the corresponding protein will be readily detectable in a
23
24 specific organ/cell. As described in the pioneer study by Cagney and collaborators ¹⁹
25
26 there is a low degree of correlation between gene and protein expression in the testis
27
28 (*i.e.*, 0.138 versus 0.432 for the liver, which displayed the strongest correlation). This
29
30 prominent uncoupling between transcription and translation was further studied by
31
32 Gan and collaborators in the mouse testis ²⁰. The authors observed that several
33
34 regulation mechanisms including transcript degradation, translation repression and
35
36 de-repression and protein degradation affected most genes and may account for the
37
38 low correlation between mRNAs and proteins during spermatogenesis ²⁰. This
39
40 original feature of spermatogenesis is mainly linked to the chromatin condensation in
41
42 elongating spermatid and the complete cessation of transcription from mid-
43
44 spermiogenesis onwards. As a consequence, many genes are transcribed much
45
46 earlier during germ cell differentiation, then sequestered in chromatoid bodies,
47
48 translationally repressed and finally translated several days after the mRNA
49
50 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

1
2
3 in agreement with the lack of identification by mass spectrometry in sperm from other
4
5 model organisms such as mouse and rat. Taken together, this information suggests
6
7 that LRRC9 could play a role in flagellum biogenesis.
8
9

10 **EVX2** is a homeobox transcription factor homolog of *Drosophila* eve (even-skipped)
11
12 and *Caenorhabditis elegans* vab-7, required for embryonic patterning and neuronal
13
14 identity in these model species ^{25 26}. In vertebrates EVX2 is required for the
15
16 development of limbs, genitalia and spinal cord ^{27 28}. In humans, a microdeletion
17
18 removing HOXD9-HOXD13 and EVX2 causes synpolydactyly ²⁹. Using RNA-Seq,
19
20 EVX2 transcripts were shown to be among the population of abundant and stable
21
22 RNAs stored in human sperm that encode proteins that may serve a role in the final
23
24 stages of spermiogenesis or at fertilization ³⁰. This is in accordance with our
25
26 observation that EVX2 immunoreactivity is intense in the nuclei of pachytene
27
28 spermatocytes but discrete in elongated spermatids. However, the high sensitivity of
29
30 mass spectrometry allowed us to identify residual EVX2 protein in sperm extracts
31
32 with 2 peptides.
33
34
35
36

37 **ALLC** is highly conserved from bacteria to vertebrates. In mouse, rat, bovine and
38
39 human, ALLC transcripts were shown to be enriched in testes ³¹. In mouse testes, the
40
41 expression seems finely regulated: it is detected from 2-5 to 29 days postpartum,
42
43 with upregulation from 17 to 23 days postpartum and is detected again in adults older
44
45 than 2 months ³². Another study shows that in 22-day-old mice the mRNA is rendered
46
47 translationally inactive by sequestration at ribonucleoproteins, and released in adult
48
49 mice ³³. The timing of expression in testis indicates that the gene would be expressed
50
51 from meiosis to elongating spermatids (steps 7-8), downregulated during late
52
53 spermiogenesis and re-expressed in mature sperm, in accordance with what we
54
55
56
57
58
59
60

1
2
3 observed at the protein level by immunohistochemistry and by the identification of 3
4
5 unique peptides in the insoluble fraction using PRM analysis. In many species, ALLC
6
7 is performing allantoicase activity (EC 3.5.3.4) as a final step of the uric acid
8
9 degradation pathway, but in mammals this pathway is incomplete and allantoicase
10
11 activity is supposed to be absent ³⁴. Therefore, the role of ALLC in mammals is
12
13 intriguing and our observations suggest a specific role in spermatogenesis that will
14
15 be interesting to investigate.

16
17
18
19 TEX22 and FAM181A displayed much more specific immunoreactivity patterns in
20
21 human testis sections than LRRC9, EVX2 and ALLC. The staining intensity for
22
23 **TEX22** gradually increased from premeiotic and meiotic germ cells onwards with a
24
25 strong immunoreactivity observed in elongated spermatids, where it accumulated in
26
27 cytoplasmic lobes/ residual bodies (Figure 2D). At spermiation - the process by which
28
29 mature spermatids are released from Sertoli cells into the seminiferous tubule lumen,
30
31 numerous germ cell proteins and organelles concentrate into the cytoplasmic lobes
32
33 that will condense to form the residual bodies. In humans, most of these residual
34
35 bodies will be left behind after disengagement and phagocytosed by Sertoli cells, but
36
37 a few of them can be found in ejaculated sperm (for a review see ³³). TEX22
38
39 immunoreactivity could not be seen on ejaculated spermatozoa. This may indicate
40
41 that most copies of the protein are eliminated during through residual bodies and that
42
43 the TEX22 protein that was identified with two distinct peptides by PRM in sperm cell
44
45 extracts only corresponds to a few copies, below the limits of detection of
46
47 immunochemistry techniques. Another explanation for the loss of recognition of
48
49 human TEX22 by antibodies in ejaculated sperm is that specific post-translational
50
51 modifications occurring in mature sperm could affect recognition by the antibody.
52
53
54
55
56
57
58
59
60

1
2
3 Such post-translational modifications have been suggested for Tep22, the mouse
4
5 TEX22 ortholog ³⁵. The observed patterns of expression and localization of TEX22
6
7 and Tep22 are not strictly identical: Tep22 transcripts were detected in 2-week-old
8
9 mouse testes onwards ³⁶, while Tep22 protein was detected by mass spectrometry in
10
11 male haploid germ cells ³⁷. Tep22 displayed a dynamic localization during
12
13 spermatogenesis, *i.e.*, in the acrosome of elongating spermatids at early
14
15 spermiogenesis, in the cytoplasm later, then in the midpiece, and both in the
16
17 midpiece and mitochondrial sheath in mature spermatozoa ³⁸. The function of the
18
19 mouse protein is not clear: Tep22 knock-out mice are fertile suggesting that this gene
20
21 is not essential for mouse male fertility ³⁶. Data obtained from mouse cannot be
22
23 entirely extrapolated to humans as TEX22 is quite highly evolving in eutherians ³⁶, so
24
25 a function of TEX22 at early stages of spermatogenesis cannot be excluded.
26
27
28
29

30 **FAM181A** immunoreactive signal was intense in pachytene spermatocytes and weak
31
32 in postmeiotic germ cells with a discrete staining in the cytoplasmic lobes of
33
34 elongated spermatids (Figure 2E). Considering the intense expression of the protein
35
36 in pachytene spermatocytes, it is surprising that FAM181A could not be identified in
37
38 proteomics studies performed on total testis extracts ³⁹ ⁴⁰. FAM181A was
39
40 unambiguously identified by PRM in sperm cell extracts thanks to two peptides, but
41
42 no FAM181A immunoreactivity could be seen on ejaculated spermatozoa (data not
43
44 shown), meaning expression levels were below the limits of detection of
45
46 immunochemistry techniques. FAM181A is conserved in Vertebrates. In mouse,
47
48 Fam181a expression appears mainly limited to neural tissues where it shows
49
50 extensive overlap with its paralog Fam181b. Both murine Fam181 tagged proteins
51
52 localize to the nucleus in transiently transfected cells. Additionally, Fam181a is found
53
54
55
56
57
58
59
60

1
2
3 as weak speckles in the cytoplasm ⁴¹. On the other hand, FAM181A interacts with
4
5 MAPRE1/EB1, a microtubule plus-end tracking protein ⁴² which was shown to interact
6
7 with RABL2, a protein required for sperm intra-flagellar transport ⁴³. These
8
9 observations suggest that FAM181A could function during early meiosis at the
10
11 nucleus and could regulate cytoskeletal dynamics at later stages. Cytoskeletal
12
13 dynamics are crucial for the series of events occurring from meiosis and onwards
14
15 including the Sertoli cell-germ cell dialog and spermatid maturation (for review see
16
17
18 ⁴⁴).

20
21 **ADAM20** immunoreactive signal is very strong in the flagella of elongated spermatids
22
23 (arrow) at stage prior spermiation (Figure 3A) and this regionalized immunoreactivity
24
25 was confirmed on the flagella of ejaculated spermatozoa (Figure 3B). ADAM20
26
27 belongs to the disintegrin and metalloprotease (ADAM) transmembrane family of
28
29 proteases conserved in various Metazoa, where they play functions in development,
30
31 inflammation, adhesion, migration and fertilization. In mammals, more than half of
32
33 these proteins are expressed specifically or predominantly in testes and epididymis,
34
35 suggesting a role in male reproduction ^{45 46}. According to neXtProt, there are 21
36
37 ADAM protein-coding genes in human. Except ADAM20 that has only been validated
38
39 at transcript level (PE2) and ADAM5 that is classified as a possible pseudogene
40
41 (PE5), all of them have been validated at protein level (PE1). According to HPA, nine
42
43 human ADAM genes are preferentially expressed in testis or epididymis: ADAM2,
44
45 ADAM7, ADAM18, ADAM20, ADAM21, ADAM28, ADAM29, ADAM30 and ADAM32.

46
47
48 Usually, ADAM proteins are present as precursors in the testis, and processed to
49
50 mature forms lacking the prodomain or both the prodomain and metalloproteinase
51
52 domain between the stages of testicular to cauda epididymal sperm ⁴⁵. Interestingly,
53
54
55

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

1
2
3 one of the two peptides from ADAM20 that were validated by PRM is located inside
4 the metalloproteinase domain, indicating that there was no processing further than
5
6
7 cleavage of the prodomain.
8

9
10 The involvement of ADAM proteins in fertilization largely derives from gene knockout
11 studies in mice. However, the orthology between mouse and human genes varies
12 considerably especially for the testis enriched expressing ADAM genes. Adam1, 3, 4,
13
14 6, 24 and 26 are pseudogenes or absent in humans while ADAM20 is absent in
15
16 mouse. Mouse Adam1, 2, 4, 6, 24, 26, and 30 localize in the head of mature sperm ⁴⁷
17
18
19 ⁴⁸, Adam3 in the acrosome of spermatids and mature sperm ⁴⁸, and Adam15 in the
20
21 acrosome and flagellum ⁴⁹. Mouse Adam7 is synthesized in the epididymis, secreted
22
23 into the epididymal lumen and then transferred to sperm membranes ⁵⁰. Mice lacking
24
25 Adam1, Adam2 and Adam3 show deficiency in sperm migration from the uterus to
26
27 the oviduct, reduced sperm binding to egg zona pellucida, and sperm-egg membrane
28
29 adhesion and fusion ^{51 52 53}. Mice lacking Adam7 show decreased sperm motility,
30
31 flagellum deformation and modified epididymal epithelium ⁵⁴. The very strong signal
32
33 observed for ADAM20 in flagella suggests a role for this protein in sperm motility.
34
35
36
37
38

39 LC-PRM detection of **C14orf177** was inconclusive as the signal of a unique 8 amino-
40 acids long peptide did not meet detection quality criteria required to be MS-validated.
41
42 Nevertheless, a cross-validation by Ab-based techniques had been initiated and
43
44 successfully pursued. C14orf177 immunoreactivity was observed in the cytoplasm of
45
46 germ cells at all stages of development up to the cytoplasmic lobes of elongated
47
48 spermatids (Figure 3C). In ejaculated spermatozoa, a ring-shaped signal was visible
49
50 in the headpiece (Figure 3D). This peculiar staining corresponds to the *Annulus*, a
51
52 structure located beneath the plasma membrane that connects the midpiece and the
53
54
55
56
57
58
59
60

1
2
3 principal piece of mature sperm flagellum (for a review see ⁵⁵). Its biogenesis and
4
5 functional significance remain unclear but the *Annulus* might behave as a
6
7 morphological organizer, guiding flagellum assembly and preventing caudal
8
9 displacement of mitochondria during spermiogenesis ^{56 57}, and as a diffusion barrier,
10
11 confining proteins to particular compartments of the mature sperm tail ⁵⁸.
12
13 Interestingly, only a few proteins have been shown to date to localize to the *Annulus*
14
15 in the mouse and human (for a review, see ⁵⁵). C14orf177 immunoreactivity was also
16
17 observed in the midpiece of ejaculated sperm (Figure 3D). C14orf177 is an
18
19 uncharacterized protein conserved in Primates with no homology to known domains
20
21 or proteins, except in the first 35 amino acids with the C-terminal terminal part of
22
23 SPACA9. In mouse Spaca9k localizes at the sperm head, midpiece and flagellum ⁵⁹,
24
25 and was proposed as a candidate for ciliary function ⁶⁰. Although the function of
26
27 C14orf177 remains unknown, its observed localization in the headpiece and
28
29 midpiece of ejaculated sperm suggests it could play a role in the flagellum biology.
30
31
32

33
34 **DNAJC5G** was validated by PRM with 2 unique peptides. DNAJC5G
35
36 immunoreactivity was observed in all germ cells up to the cytoplasmic lobes of
37
38 elongated spermatids (Figure 3E). As for C14orf177, DNAJC5G staining was also
39
40 observed as ring-shaped in the headpiece and in the midpiece of ejaculated
41
42 spermatozoa (Figure 3F), suggesting a role in the flagellum biology. DNAJC5G/CSP-
43
44 gamma is a member of the DnaJ/Hsp40 family of chaperones, which localize in
45
46 various subcellular compartments (mitochondria, endoplasmic reticulum, nucleus,
47
48 and cytosol) and have many housekeeping and stress-associated functions in
49
50 folding, translocation and degradation of proteins, as well as in gene expression and
51
52 translational initiation. There are 45 entries for human DnaJ protein-coding genes in
53
54
55
56

60

1
2
3 neXtProt. Except for DNAJB7, DNAJC22 and DNAJC5G, they all have been
4
5 validated at protein level (PE1). According to HPA, the majority (75%) of DnaJ
6
7 proteins are broadly expressed. Eight of them are preferentially expressed in testis
8
9 (DNAJB7, DNAJB8, DNAJB13, DNAJC5B, DNAJC5G, DNAJC18, DNAJC27 and
10
11 DNAJC28).

12
13
14 Based on their architecture, DnaJ/Hsp40 proteins are classified in three main
15
16 subfamilies. Several type I and type II HSP40 family members have been shown to
17
18 be involved in male reproduction. DNAJB13 is involved in axoneme formation ⁶¹ and
19
20 mutations cause primary ciliary dyskinesia and male infertility in humans ⁶². DnajA1
21
22 regulates androgen receptor signaling and spermatogenesis in mouse ⁶³. DNAJB3/
23
24 MSJ1 is highly expressed in haploid male germ cells in mouse and *Macaca fuscata* ⁶⁴
25
26 ⁶⁵ and associates with the acrosome and centriolar region in mouse germ cells ^{66 67 68}.
27
28 DnaJB1 was identified as an acrosome and flagella associated component in mouse
29
30 and rat spermatozoa ⁶⁹. DNAJC5G/ CSP-gamma is a type III DnaJ/Hsp40 protein
31
32 and its closest homologs are DNAJC5/CSP and DNAJC5B/CSP-beta. DNAJC5/CSP,
33
34 which is conserved in invertebrates such as *Drosophila* and *Caenorhabditis elegans*,
35
36 is widely expressed and has a general role in exocytosis and in presynaptic function
37
38 ^{70 71}. DNAJC5B/CSP-beta and DNAJC5G/CSP-gamma are not conserved outside of
39
40 Vertebrates and are less characterized. They are localized in nerve terminals and
41
42 secretory organelles in mouse brain, where they may contribute to regulated
43
44 secretory events ^{72 73}. DNAJC5B/ CSP-beta was found in rat testis, in a palmitoylated
45
46 form at the plasma membrane ⁷⁴ and as a non-palmitoylated form in the trans-Golgi
47
48 network ⁷⁵, but its potential role in male reproduction is still unclear. The presence of
49
50 DNAJC5G/ CSPgamma in the headpiece and the midpiece of mature spermatozoa
51
52
53
54
55
56
57
58
59
60

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

31
32 We hope that the present data will be used to validate the existence of 12 missing
33
34 proteins in PeptideAtlas and neXtProt. Our consortium continues to look for the
35
36 remaining missing proteins whose expression may be preferentially observable in the
37
38 male gamete cells. The present work reveals we have certainly reached technical
39
40 limits due to the low level of expression of the targeted proteins. To go beyond this
41
42 limit, we will mine the most recent RNA-Seq data to select a larger set of MPs from
43
44 chromosomes 2/14 expected to be present in sperm, and apply targeted-MS assays
45
46 to a collection of enriched subcellular fractions (*i.e.*, head, midpiece and flagellum) of
47
48 human sperm samples. In case of unsuccessful detection in sperm fractions, the
49
50 developed LC-PRM assays will be used to detect specific MPs in cell line or tissue
51
52
53
54
55
56
57
58
59
60

1
2
3 extracts in which mRNA expression was also demonstrated. We suggest that option
4
5 to become the step two of the neXt50 challenge for our teams.
6

7
8 As a preliminary work in that direction, we mined literature and The Global Proteome
9
10 Machine Database (<http://gpmdb.thegpm.org/index.html>) to find identified proteotypic
11
12 peptides that did not meet the required quality or length criteria to be considered for
13
14 protein validation or had not been submitted to ProteomeXchange, in order to identify
15
16 other tissues or cell lines in which these proteins could be present. We found that 14
17
18 missing proteins might be present in commonly used cell lines, and 3 others
19
20 (RGP3, C2orf27 and GPR148) in the brain. COX8C, UNC79, TM4SF20 and
21
22 GPR148 that have transmembrane domains (1, 2, 4 and 7, respectively) may require
23
24 the development of specific solubilization and/or enrichment protocols. For the three
25
26 other proteins, the use of a detergent other than Triton X-100 may be required due to
27
28 the presence of more than one transmembrane domain. Indeed, the use of Triton X-
29
30 100 at 1% final concentration is not a specific membrane solubilization method and
31
32 only facilitates the solubilization of membrane-associated hydrophobic proteins.
33
34
35
36
37

38
39 Additionally, to help other C-HPP teams in the global neXt50 challenge, we propose
40
41 to select new sets of MPs from other chromosomes based on preferential expression
42
43 in testis and perform targeted-MS searches in sperm extracts.
44
45

46
47 The observed decline in male reproductive potential worldwide calls for dedicated
48
49 efforts to characterize male-factor defects. From the standpoint of understanding
50
51 sperm physiopathology, the characterization and identification of new structural and
52
53 functional sperm proteins in a systematic manner merits further research. The
54
55

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

1
2
3 the cytoplasmic lobes of elongated spermatids (C; square). TEX22 immunoreactivity
4
5 is intense in germ cells at all development stages and concentrated in the cytoplasm
6
7 of elongated spermatids (D; arrows). FAM181A immunoreactivity was intense in
8
9 pachytene spermatocytes (arrows) and a discrete staining was observed in the
10
11 cytoplasmic portion of elongated spermatids (square) (E). Scale bars = 20 μ m.
12
13

14
15
16 **Figure 3:** Antibody staining for ADAM20 (O43506), C14orf177 (Q52M58), and
17
18 DNAJC5G (Q8N7S2) in adult human testis (A,C,E) and in isolated spermatozoa
19
20 (B,D,F). Proteins were detected in transverse testis sections at stages IV to VI of the
21
22 seminiferous epithelium ⁷⁹ and on sperm smears using polyclonal antibodies from the
23
24 HPA specific for ADAM20 (HPA059377) (A), C14orf177 (HPA018091) (C), and
25
26 DNAJC5G (HPA041445) (E). ADAM20 immunoreactivity is very strong in the flagella
27
28 of elongated spermatids (arrow) at stage prior spermiation (A). The regionalized
29
30 immunoreactivity of ADAM20 is confirmed on the flagella of ejaculated spermatozoa
31
32 (arrows) whereas the midpiece is clearly unstained (arrowhead) (B). C14orf177
33
34 immunoreactivity is visible in the cytoplasm of germ cells at all stages of development
35
36 up to the cytoplasmic lobes of elongated spermatids (arrow), whereas it appeared
37
38 stronger in the nucleus of premeiotic and meiotic germ cells (arrowhead) (C). In
39
40 ejaculated spermatozoa, a ring-shaped signal was visible for C14orf177 in the
41
42 headpiece (arrows) and the midpiece (arrowhead) (D). DNAJC5G immunoreactivity
43
44 was intense in germ cells at all stages of their development up to the cytoplasmic
45
46 lobes of elongated spermatids (arrow), whereas a stronger signal was observed in
47
48 the cytoplasm of spermatogonia and preleptotene spermatocytes (arrowhead) (E).
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 DNAJC5G staining was observed in the headpiece (arrow) and midpiece
4
5 (arrowhead) of ejaculated spermatozoa (F).
6
7

8
9
10 **Table 1:** List of 12 MPs identified with 2 or more peptides and 3 candidate MPs
11 identified with 1 peptide by targeted LC-PRM. Gene names, protein IDs, protein
12 names, peptide sequences and fraction (Triton X-100 soluble and/or insoluble
13 fraction) in which these 15 missing proteins were identified are provided.
14
15
16
17

20 **Supporting information**

21
22 This material is available free of charge via <http://pubs.acs.org/>.
23
24
25
26

27 **Supplementary Material:** Analytical details on the acquisition of reference MS/MS
28 spectra for the 92 proteotypic peptides corresponding to the 38 targeted missing
29 proteins and on the targeted LC-PRM time-scheduled method development.
30
31
32
33
34
35

36 **Supplementary Table 1:** Targeted LC-PRM inclusion list. List of the 212 targets
37 followed during LC-PRM analyses: precursor m/z ratio, charge state, polarity,
38 scheduling start and end times, normalized collision energy (CE), amino acid
39 sequence including modifications and precursor type (light form and heavy labeled
40 form), are provided for each target.
41
42
43
44
45
46
47
48

49 **Supplementary Table 2:** List of missing proteins for which immunohistochemistry or
50 immunocytochemistry was successful, HPA antibody names and dilutions used.
51
52
53
54
55
56

1
2
3 **Supplementary Figure 1:** LC-PRM data of validated peptides. For all validated
4 peptides, traces of selected transition groups were extracted for the endogenous and
5 heavy labeled peptides and dot-product and rdot-product values were calculated
6 using Skyline.
7
8
9

10
11
12
13
14 **Notes:**

15
16 The authors declare no competing financial interest.
17
18
19

20
21 **Author's contribution:**

22
23 CC, LL and CP co-coordinated the study. TF and KR performed spermatozoa
24 preparation. CC, CP, LL and YV conceived and designed the experiments and
25 analyses. CC, FD and CM conducted LC-PRM experiments and validated LC-PRM
26 data and peptides. PD, LL and CP performed extensive bioinformatics analysis and
27 data/literature mining. CL, CP, KR and MS designed and performed immuno-
28 histochemistry and -cytochemistry studies. CC, PD, LL, CP and YV wrote the
29 manuscript. All the authors approved the final version of the manuscript.
30
31
32
33
34
35
36
37
38
39

40
41 **Acknowledgements:**

42
43 This work was partially funded through the French National Agency for Research
44 (ANR) (grant ANR-10-INBS-08; ProFI project, "Infrastructures Nationales en Biologie
45 et Santé"; "Investissements d'Avenir" call). This work was also supported by grants
46 from Biogenouest, *Infrastructures en Biologie Santé et Agronomie* (IBiSA) and
47 *Conseil Régional de Bretagne* awarded to C.P. We are grateful to Mathias Uhlen
48 (Human Protein Atlas, Sweden) for the gift of antibodies, to Alain Gateau for help
49
50
51
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

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 Rockberg, J.; Nilsson, P.; Schwenk, J. M.; Hamsten, M.; von Feilitzen, K.; Forsberg,
4 M.; Persson, L.; Johansson, F.; Zwahlen, M.; von Heijne, G.; Nielsen, J.; Ponten, F.,
5 Proteomics. Tissue-based map of the human proteome. *Science* 2015, 347 (6220),
6 1260419.
7

8
9
10 8. Uhlen, M.; Hallstrom, B. M.; Lindskog, C.; Mardinoglu, A.; Ponten, F.; Nielsen,
11 J., Transcriptomics resources of human tissues and organs. *Molecular systems*
12 *biology* 2016, 12 (4), 862.
13

14 9. Chocu, S.; Calvel, P.; Rolland, A. D.; Pineau, C., Spermatogenesis in
15 mammals: proteomic insights. *Systems biology in reproductive medicine* 2012, 58
16 (4), 179-90.
17

18
19 10. Djureinovic, D.; Fagerberg, L.; Hallstrom, B.; Danielsson, A.; Lindskog, C.;
20 Uhlen, M.; Ponten, F., The human testis-specific proteome defined by transcriptomics
21 and antibody-based profiling. *Molecular human reproduction* 2014, 20 (6), 476-88.
22

23 11. Jumeau, F.; Com, E.; Lane, L.; Duek, P.; Lagarrigue, M.; Lavigne, R.; Guillot,
24 L.; Rondel, K.; Gateau, A.; Melaine, N.; Guevel, B.; Sergeant, N.; Mitchell, V.;
25 Pineau, C., Human Spermatozoa as a Model for Detecting Missing Proteins in the
26 Context of the Chromosome-Centric Human Proteome Project. 2015, 14 (9), 3-0--20.
27
28
29
30
31

32
33 12. Vandenbrouck, Y.; Lane, L.; Carapito, C.; Duek, P.; Rondel, K.; Bruley, C.;
34 Macron, C.; Gonzalez de Peredo, A.; Coute, Y.; Chaoui, K.; Com, E.; Gateau, A.;
35 Hesse, A. M.; Marcellin, M.; Mear, L.; Mouton-5arbosa, E.; Robin, T.; Burlet-Schiltz,
36 O.; Cianferani, S.; Ferro, M.; Freour, T.; Lindskog, C.; Garin, J.; Pineau, C., Looking
37 for Missing Proteins in the Proteome of Human Spermatozoa: An Update. 201-, 15
38 (11), 3998-4019.
39
40

41
42 13. Gallien, S.; Kim, S. Y.; Domon, B., Large-Scale Targeted Proteomics Using
43 Internal Standard Triggered-Parallel Reaction Monitoring (IS-PRM). *Molecular &*
44 *cellular proteomics : MCP* 2015, 14 (6), 1630-44.
45
46

47 14. Duek, P.; Bairoch, A.; Gateau, A.; Vandenbrouck, Y.; Lane, L., Missing Protein
48 Landscape of Human Chromosomes 2 and 14: Progress and Current Status. 201-,
49 15 (11), 3971-3978.
50
51

52 15. Baker, M. A.; Reeves, G.; Hetherington, L.; Muller, J.; Baur, I.; Aitken, R. J.,
53 Identification of gene products present in Triton X-100 soluble and insoluble fractions
54
55

1
2
3 of human spermatozoa lysates using LC-MS/MS analysis. *Proteomics. Clinical*
4 *applications* 2007, 1 (5), 524-32.

5
6 16. Schaeffer, M.; Gateau, A.; Teixeira, D.; Michel, P. A.; Zahn-Zabal, M.; Lane,
7 L., The neXtProt peptide uniqueness checker: a tool for the proteomics community.
8 *Bioinformatics* 2017 May 17. doi: 10.1093/bioinformatics/btx318. [Epub ahead of
9 print].

10
11
12 17. Com, E.; Rolland, A. D.; Guerrois, M.; Aubry, F.; Jegou, B.; Vallet-Erdtmann,
13 V.; Pineau, C., Identification, molecular cloning, and cellular distribution of the rat
14 homolog of minichromosome maintenance protein 7 (MCM7) in the rat testis.
15 *Molecular reproduction and development* 2006, 73 (7), 866-77.

16
17
18 18. Kusebauch, U.; Campbell, D. S.; Deutsch, E. W.; Chu, C. S.; Spicer, D. A.;
19 Brusniak, M. Y.; Slagel, J.; Sun, Z.; Stevens, J.; Grimes, B.; Shteynberg, D.;
20 Hoopmann, M. R.; Blattmann, P.; Ratushny, A. V.; Rinner, O.; Picotti, P.; Carapito,
21 C.; Huang, C. Y.; Kapousouz, M.; Lam, H.; Tran, T.; Demir, E.; Aitchison, J. D.;
22 Sander, C.; Hood, L.; Aebersold, R.; Moritz, R. L., Human SRMATlas: A Resource of
23 Targeted Assays to Quantify the Complete Human Proteome. *Cell* 2016, 166 (3),
24 766-78.

25
26
27 19. Cagney, G.; Park, S.; Chung, C.; Tong, B.; O'Dushlaine, C.; Shields, D. C.;
28 Emili, A., Human tissue profiling with multidimensional protein identification
29 technology. 2005, 4 (5), 1757-&'A

30
31
32 20. Gan, H.; Cai, T.; Lin, X.; Wu, Y.; Wang, X.; Yang, F.; Han, C., Integrative
33 proteomic and transcriptomic analyses reveal multiple post-transcriptional regulatory
34 mechanisms of mouse spermatogenesis. *Molecular & cellular proteomics : MCP*
35 2013, 12 (5), 1144-57.

36
37
38 21. Hecht, N. B., Regulation of 'haploid expressed genes' in male germ cells.
39 *Journal of reproduction and fertility* 1990, 88 (2), 679-93.

40
41
42 22. Treutlein, B.; Brownfield, D. G.; Wu, A. R.; Neff, N. F.; Mantalas, G. L.;
43 Espinoza, F. H.; Desai, T. J.; Krasnow, M. A.; Quake, S. R., Reconstructing lineage
44 hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 2014, 509
45 (7500), 371-5.

46
47
48 23. Pazour, G. J.; Agrin, N.; Leszyk, J.; Witman, G. B., Proteomic analysis of a
49 eukaryotic cilium. *The Journal of cell biology* 2005, 170 (1), 103-13.

- 1
2
3 24. Kubo, T.; Yanagisawa, H. A.; Liu, Z.; Shibuya, R.; Hirono, M.; Kamiya, R., A
4 conserved flagella-associated protein in *Chlamydomonas*, FAP234, is essential for
5 axonemal localization of tubulin polyglutamylase TLL9. *Molecular biology of the cell*
6 2014, 25 (1), 107-17.
7
- 8
9 25. Doe, C. Q.; Smouse, D.; Goodman, C. S., Control of neuronal fate by the
10 *Drosophila* segmentation gene *even-skipped*. *Nature* 1988, 333 (6171), 376-8.
11
- 12 26. Ahringer, J., Posterior patterning by the *Caenorhabditis elegans* *even-skipped*
13 homolog *vab-7*. *Genes & development* 1996, 10 (9), 1120-30.
14
- 15 27. Herault, Y.; Hraba-Renevey, S.; van der Hoeven, F.; Duboule, D., Function of
16 the *Evx-2* gene in the morphogenesis of vertebrate limbs. *The EMBO journal* 1996,
17 15 (23), 6727-38.
18
- 19 28. Tarchini, B.; Huynh, T. H.; Cox, G. A.; Duboule, D., *HoxD* cluster scanning
20 deletions identify multiple defects leading to paralysis in the mouse mutant *Ironside*.
21 *Genes & development* 2005, 19 (23), 2862-76.
22
- 23 29. Goodman, F. R.; Majewski, F.; Collins, A. L.; Scambler, P. J., A 117-kb
24 microdeletion removing *HOXD9-HOXD13* and *EVX2* causes synpolydactyly.
25 *American journal of human genetics* 2002, 70 (2), 547-55.
26
- 27 30. Sandler, E.; Johnson, G. D.; Mao, S.; Goodrich, R. J.; Diamond, M. P.;
28 Hauser, R.; Krawetz, S. A., Stability, delivery and functions of human sperm RNAs at
29 fertilization. *Nucleic acids research* 2013, 41 (7), 4104-17.
30
- 31 31. Vigetti, D.; Binelli, G.; Monetti, C.; Prati, M.; Bernardini, G.; Gornati, R.,
32 Selective pressure on the allantoicase gene during vertebrate evolution. *Journal of*
33 *molecular evolution* 2003, 57 (6), 650-8.
34
- 35 32. Ike, A.; Tokuhira, K.; Hirose, M.; Nozaki, M.; Nishimune, Y.; Tanaka, H.,
36 Comprehensive analysis of gene expression in testes producing haploid germ cells
37 using DNA microarray analysis. *International journal of andrology* 2007, 30 (5), 462-
38 75.
39
- 40 33. Iguchi, N.; Tobias, J. W.; Hecht, N. B., Expression profiling reveals meiotic
41 male germ cell mRNAs that are translationally up- and down-regulated. *Proceedings*
42 *of the National Academy of Sciences of the United States of America* 2006, 103 (20),
43 7712-7.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 34. Leulliot, N.; Quevillon-Cheruel, S.; Sorel, I.; Graille, M.; Meyer, P.; Liger, D.;
4 Blondeau, K.; Janin, J.; van Tilbeurgh, H., Crystal structure of yeast allantoicase
5 reveals a repeated jelly roll motif. *The Journal of biological chemistry* 2004, 279 (22),
6 23447-52.
7
8
9
10 35. Choi, E.; Lee, J.; Oh, J.; Park, I.; Han, C.; Yi, C.; Kim, D. H.; Cho, B. N.; Eddy,
11 E. M.; Cho, C., Integrative characterization of germ cell-specific genes from mouse
12 spermatocyte UniGene library. *BMC genomics* 2007, 8, 256.
13
14 36. Miyata, H.; Castaneda, J. M.; Fujihara, Y.; Yu, Z.; Archambeault, D. R.;
15 Isotani, A.; Kiyozumi, D.; Kriseman, M. L.; Mashiko, D.; Matsumura, T.; Matzuk, R.
16 M.; Mori, M.; Noda, T.; Oji, A.; Okabe, M.; Prunskaitė-Hyyryläinen, R.; Ramirez-Solis,
17 R.; Satouh, Y.; Zhang, Q.; Ikawa, M.; Matzuk, M. M., Genome engineering uncovers
18 54 evolutionarily conserved and testis-enriched genes that are not required for male
19 fertility in mice. *Proceedings of the National Academy of Sciences of the United*
20 *States of America* 2016, 113 (28), 7704-10.
21
22
23
24
25
26 37. Guo, X.; Shen, J.; Xia, Z.; Zhang, R.; Zhang, P.; Zhao, C.; Xing, J.; Chen, L.;
27 Chen, W.; Lin, M.; Huo, R.; Su, B.; Zhou, Z.; Sha, J., Proteomic analysis of proteins
28 involved in spermiogenesis in mouse. 2010, 9 (3), 124--56A
29
30
31
32
33 38. Neesen, J.; Hartwich, T.; Brandhorst, G.; Aumüller, G.; Glaser, B.; Burfeind,
34 P.; Mendoza-Lujambio, I., Tep22, a novel testicular expressed gene, is involved in
35 the biogenesis of the acrosome and the midpiece of the sperm tail. *Biochemical and*
36 *biophysical research communications* 2002, 297 (4), 737-48.
37
38
39 39. Wei, W.; Luo, W.; Wu, F.; Peng, X.; Zhang, Y.; Zhang, M.; Zhao, Y.; Su, N.;
40 Qi, Y.; Chen, L.; Zhang, Y.; Wen, B.; He, F.; Xu, P., Deep Coverage Proteomics
41 Identifies More Low-Abundance Missing Proteins in Human Testis Tissue with Q-
42 Exactive HF Mass Spectrometer. *2A1-*, 15 (11), 3988-
43 3997.
44
45
46
47 40. Zhang, Y.; Li, Q.; Wu, F.; Zhou, R.; Qi, Y.; Su, N.; Chen, L.; Xu, S.; Jiang, T.;
48 Zhang, C.; Cheng, G.; Chen, X.; Kong, D.; Wang, Y.; Zhang, T.; Zi, J.; Wei, W.; Gao,
49 Y.; Zhen, B.; Xiong, Z.; Wu, S.; Yang, P.; Wang, Q.; Wen, B.; He, F.; Xu, P.; Liu, S.,
50 Tissue-5ased Proteogenomics Reveals that Human Testis Endows Plentiful Missing
51 Proteins. *2A15*, 14 (9), 3583-94.
52
53
54
55
56
57
58
59
60

- 1
2
3 41. Marks, M.; Pennimpede, T.; Lange, L.; Grote, P.; Herrmann, B. G.; Wittler, L.,
4 Analysis of the Fam181 gene family during mouse development reveals distinct
5 strain-specific expression patterns, suggesting a role in nervous system development
6 and function. *Gene* 2016, 575 (2 Pt 2), 438-51.
7
8
9 42. Jiang, K.; Toedt, G.; Montenegro Gouveia, S.; Davey, N. E.; Hua, S.; van der
10 Vaart, B.; Grigoriev, I.; Larsen, J.; Pedersen, L. B.; Bezstarosti, K.; Lince-Faria, M.;
11 Demmers, J.; Steinmetz, M. O.; Gibson, T. J.; Akhmanova, A., A Proteome-wide
12 screen for mammalian SxIP motif-containing microtubule plus-end tracking proteins.
13 *Current biology : CB* 2012, 22 (19), 1800-7.
14
15
16 43. Lo, J. C.; Jamsai, D.; O'Connor, A. E.; Borg, C.; Clark, B. J.; Whisstock, J. C.;
17 Field, M. C.; Adams, V.; Ishikawa, T.; Aitken, R. J.; Whittle, B.; Goodnow, C. C.;
18 Ormandy, C. J.; O'Bryan, M. K., RAB-like 2 has an essential role in male fertility,
19 sperm intra-flagellar transport, and tail assembly. *PLoS genetics* 2012, 8 (10),
20 e1002969.
21
22
23 44. Lie, P. P.; Mruk, D. D.; Lee, W. M.; Cheng, C. Y., Cytoskeletal dynamics and
24 spermatogenesis. *Philosophical transactions of the Royal Society of London. Series*
25 *B, Biological sciences* 2010, 365 (1546), 1581-92.
26
27
28 45. Cho, C., Testicular and epididymal ADAMs: expression and function during
29 fertilization. *Nature reviews. Urology* 2012, 9 (10), 550-60.
30
31
32 46. Moreno, R. D.; Urriola-Munoz, P.; Lagos-Cabre, R., The emerging role of
33 matrix metalloproteases of the ADAM family in male germ cell apoptosis.
34 *Spermatogenesis* 2011, 1 (3), 195-208.
35
36
37 47. Han, C.; Choi, E.; Park, I.; Lee, B.; Jin, S.; Kim, D. H.; Nishimura, H.; Cho, C.,
38 Comprehensive analysis of reproductive ADAMs: relationship of ADAM4 and ADAM6
39 with an ADAM complex required for fertilization in mice. *Biology of reproduction*
40 2009, 80 (5), 1001-8.
41
42
43 48. Kim, E.; Nishimura, H.; Iwase, S.; Yamagata, K.; Kashiwabara, S.; Baba, T.,
44 Synthesis, processing, and subcellular localization of mouse ADAM3 during
45 spermatogenesis and epididymal sperm transport. *The Journal of reproduction and*
46 *development* 2004, 50 (5), 571-8.
47
48
49 49. Pasten-Hidalgo, K.; Hernandez-Rivas, R.; Roa-Espitia, A. L.; Sanchez-
50 Gutierrez, M.; Martinez-Perez, F.; Monrroy, A. O.; Hernandez-Gonzalez, E. O.;
51
52
53
54
55
56
57
58
59
60

1
2
3 Mujica, A., Presence, processing, and localization of mouse ADAM15 during sperm
4 maturation and the role of its disintegrin domain during sperm-egg binding.
5 *Reproduction* 2008, 136 (1), 41-51.
6

7
8 50. Oh, J. S.; Han, C.; Cho, C., ADAM7 is associated with epididymosomes and
9 integrated into sperm plasma membrane. *Molecules and cells* 2009, 28 (5), 441-6.
10

11 51. Cho, C.; Bunch, D. O.; Faure, J. E.; Goulding, E. H.; Eddy, E. M.; Primakoff,
12 P.; Myles, D. G., Fertilization defects in sperm from mice lacking fertilin beta. *Science*
13 1998, 281 (5384), 1857-9.
14

15 52. Nishimura, H.; Kim, E.; Nakanishi, T.; Baba, T., Possible function of the
16 ADAM1a/ADAM2 Fertilin complex in the appearance of ADAM3 on the sperm
17 surface. *The Journal of biological chemistry* 2004, 279 (33), 34957-62.
18

19 53. Yamaguchi, R.; Muro, Y.; Isotani, A.; Tokuhira, K.; Takumi, K.; Adham, I.;
20 Ikawa, M.; Okabe, M., Disruption of ADAM3 impairs the migration of sperm into
21 oviduct in mouse. *Biology of reproduction* 2009, 81 (1), 142-6.
22

23 54. Choi, H.; Han, C.; Jin, S.; Kwon, J. T.; Kim, J.; Jeong, J.; Kim, J.; Ham, S.;
24 Jeon, S.; Yoo, Y. J.; Cho, C., Reduced Fertility and Altered Epididymal and Sperm
25 Integrity in Mice Lacking ADAM7. *Biology of reproduction* 2015, 93 (3), 70.
26

27 55. Dirami, T.; Rode, B.; Wolf, J. P.; Gacon, G.; Dulioust, E.; Toure, A.,
28 Assessment of the frequency of sperm annulus defects in a large cohort of patients
29 presenting asthenozoospermia. *Basic and clinical andrology* 2015, 25, 10.
30

31 56. Fawcett, D. W., A comparative view of sperm ultrastructure. *Biology of*
32 *reproduction* 1970, 2, Suppl 2:90-127.
33

34 57. Phillips, D. M., Mitochondrial disposition in mammalian spermatozoa. *Journal*
35 *of ultrastructure research* 1977, (2), 144-54.
36

37 58. Cesario, M. M.; Bartles, J. R., Compartmentalization, processing and
38 redistribution of the plasma membrane protein CE9 on rodent spermatozoa.
39 Relationship of the annulus to domain boundaries in the plasma membrane of the
40 tail. *Journal of cell science* 1994, 107 (Pt 2), 561-70.
41

42 59. Bhattacharya, R.; Devi, M. S.; Dhople, V. M.; Jesudasan, R. A., A mouse
43 protein that localizes to acrosome and sperm tail is regulated by Y-chromosome.
44 *BMC cell biology* 2013, 14, 50.
45
46
47
48
49
50
51
52
53
54

- 1
2
3 60. Stauber, M.; Weidemann, M.; Dittrich-Breiholz, O.; Lobschat, K.; Alten, L.; Mai,
4 M.; Beckers, A.; Kracht, M.; Gossler, A., Identification of FOXJ1 effectors during
5 ciliogenesis in the foetal respiratory epithelium and embryonic left-right organiser of
6 the mouse. *Developmental biology* 2017, 423 (2), 170-188.
- 7
8
9 61. Li, W.; Liu, G., DNAJB13, a type II HSP40 family member, localizes to the
10 spermatids and spermatozoa during mouse spermatogenesis. *BMC developmental*
11 *biology* 2014, 14, 38.
- 12
13 62. El Khouri, E.; Thomas, L.; Jeanson, L.; Bequignon, E.; Vallette, B.;
14 Duquesnoy, P.; Montantin, G.; Copin, B.; Dastot-Le Moal, F.; Blanchon, S.; Papon, J.
15 F.; Lores, P.; Yuan, L.; Collot, N.; Tissier, S.; Faucon, C.; Gacon, G.; Patrat, C.; Wolf,
16 J. P.; Dulioust, E.; Crestani, B.; Escudier, E.; Coste, A.; Legendre, M.; Toure, A.;
17 Amselem, S., Mutations in DNAJB13, Encoding an HSP40 Family Member, Cause
18 Primary Ciliary Dyskinesia and Male Infertility. *American journal of human genetics*
19 2016, 99 (2), 489-500.
- 20
21 63. Terada, K.; Yomogida, K.; Imai, T.; Kiyonari, H.; Takeda, N.; Kadomatsu, T.;
22 Yano, M.; Aizawa, S.; Mori, M., A type I DnaJ homolog, DjA1, regulates androgen
23 receptor signaling and spermatogenesis. *The EMBO journal* 2005, 24 (3), 611-22.
- 24
25 64. Yu, S. S.; Takenaka, O., Molecular cloning, structure, and testis-specific
26 expression of MFSJ1, a member of the DNAJ protein family, in the Japanese monkey
27 (*Macaca fuscata*). *Biochemical and biophysical research communications* 2003, 301
28 (2), 443-9.
- 29
30 65. Berruti, G.; Martegani, E., MSJ-1, a mouse testis-specific DnaJ protein, is
31 highly expressed in haploid male germ cells and interacts with the testis-specific heat
32 shock protein Hsp70-2. *Biology of reproduction* 2001, 65 (2), 488-95.
- 33
34 66. Berruti, G.; Martegani, E., mUBPy and MSJ-1, a deubiquitinating enzyme and
35 a molecular chaperone specifically expressed in testis, associate with the acrosome
36 and centrosome in mouse germ cells. *Annals of the New York Academy of Sciences*
37 2002, 973, 5-7.
- 38
39 67. Meccariello, R.; Berruti, G.; Chianese, R.; De Santis, R.; Di Cunto, F.; Scarpa,
40 D.; Cobellis, G.; Zucchetti, I.; Pierantoni, R.; Altruda, F.; Fasano, S., Structure of msj-
41 1 gene in mice and humans: a possible role in the regulation of male reproduction.
42 *General and comparative endocrinology* 2008, 156 (1), 91-103.
- 43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 68. Meccariello, R.; Chianese, R.; Ciaramella, V.; Fasano, S.; Pierantoni, R.,
4 Molecular chaperones, cochaperones, and ubiquitination/deubiquitination system:
5 involvement in the production of high quality spermatozoa. *BioMed research*
6 *international* 2014, 2014, 561426.
7
8
9 69. Doiguchi, M.; Kaneko, T.; Urasoko, A.; Nishitani, H.; Iida, H., Identification of a
10 heat-shock protein Hsp40, DjB1, as an acrosome- and a tail-associated component
11 in rodent spermatozoa. *Molecular reproduction and development* 2007, 74 (2), 223-
12 32.
13
14 70. Burgoyne, R. D.; Morgan, A., Cysteine string protein (CSP) and its role in
15 preventing neurodegeneration. *Seminars in cell & developmental biology* 2015, 40,
16 153-9.
17
18 71. Fernandez-Chacon, R.; Wolfel, M.; Nishimune, H.; Tabares, L.; Schmitz, F.;
19 Castellano-Munoz, M.; Rosenmund, C.; Montesinos, M. L.; Sanes, J. R.;
20 Schneggenburger, R.; Sudhof, T. C., The synaptic vesicle protein CSP alpha
21 prevents presynaptic degeneration. *Neuron* 2004, 42 (2), 237-51.
22
23 72. Gundersen, C. B.; Kohan, S. A.; Souda, P.; Whitelegge, J. P.; Umbach, J. A.,
24 Cysteine string protein beta is prominently associated with nerve terminals and
25 secretory organelles in mouse brain. *Brain research* 2010, 1332, 1-11.
26
27 73. Sammut, S. J.; Feichtinger, J.; Stuart, N.; Wakeman, J. A.; Larcombe, L.;
28 McFarlane, R. J., A novel cohort of cancer-testis biomarker genes revealed through
29 meta-analysis of clinical data sets. *Oncoscience* 2014, 1 (5), 349-59.
30
31 74. Gorleku, O. A.; Chamberlain, L. H., Palmitoylation and testis-enriched
32 expression of the cysteine-string protein beta isoform. *Biochemistry* 2010, 49 (25),
33 5308-13.
34
35 75. Boal, F.; Le Pevelen, S.; Cziepluch, C.; Scotti, P.; Lang, J., Cysteine-string
36 protein isoform beta (Cspbeta) is targeted to the trans-Golgi network as a non-
37 palmitoylated CSP in clonal beta-cells. *Biochimica et biophysica acta* 2007, 1773 (2),
38 109-19.
39
40 76. Boehme, S. A.; Franz-Bacon, K.; DiTirro, D. N.; Ly, T. W.; Bacon, K. B.,
41 MAP3K19 Is a Novel Regulator of TGF-beta Signaling That Impacts Bleomycin-
42 Induced Lung Injury and Pulmonary Fibrosis. *PloS one* 2016, 11 (5), e0154874.
43
44
45
46
47
48
49
50
51
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

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

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

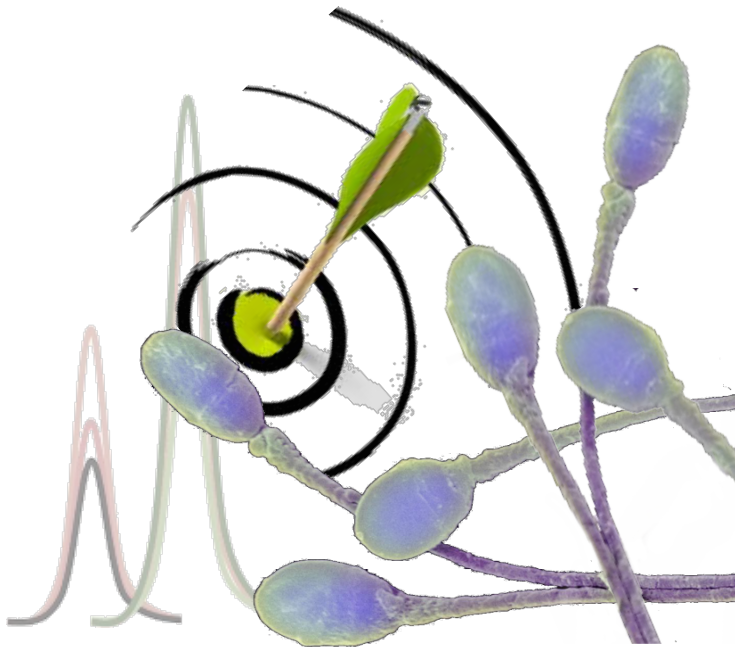
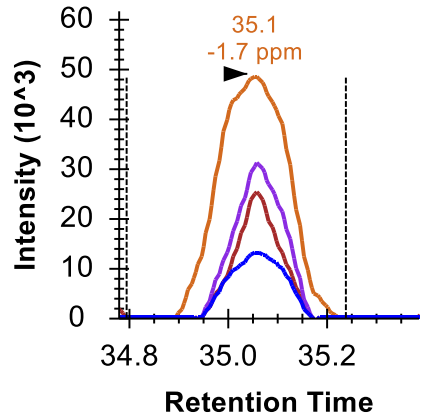


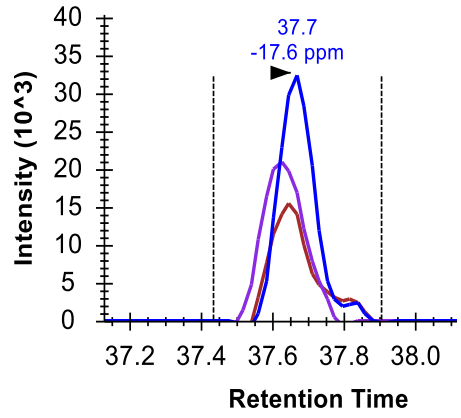
Figure 18

Transitions extracted for 2 peptides validating Protein FAM181A (Q8N9Y4)

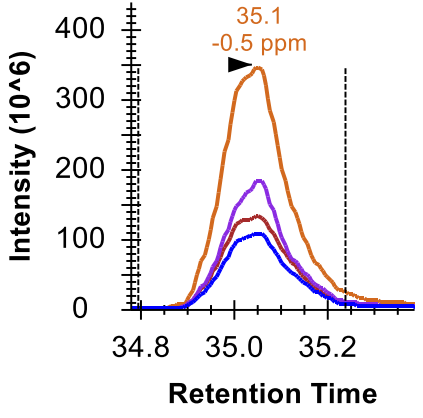
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42



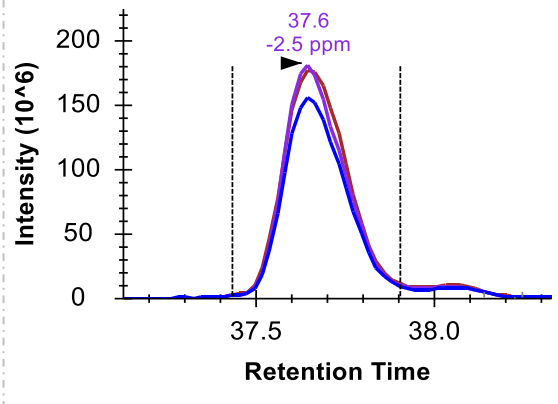
ASCSGPLVMASDSDVK (dotp 0,96)
rdotp 0,99



MPGVSLVGR (dotp 0,94)
rdotp 0,96



ASCSGPLVMASDSDVK dotp 0,99

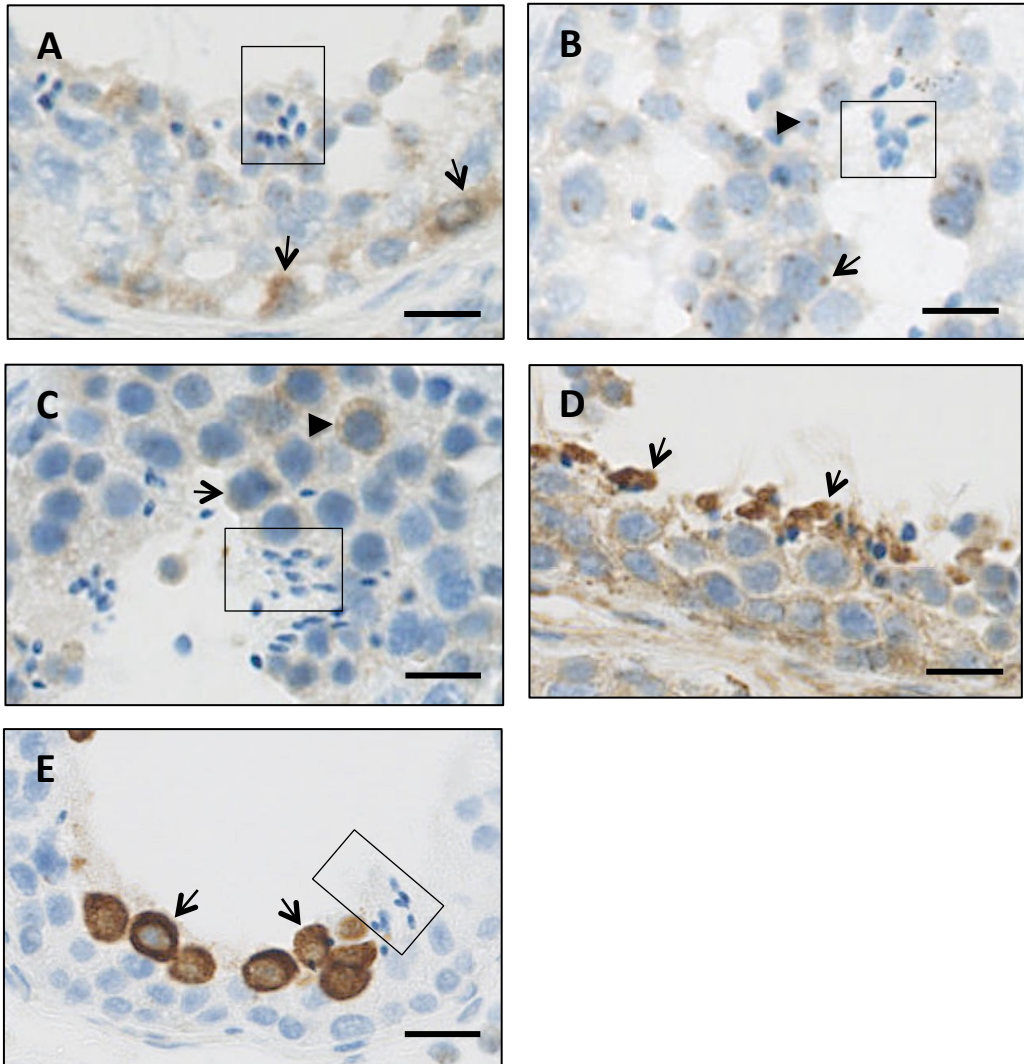


MPGVSLVGR dotp 0,92

- y12 - 1218.6035+
- y11 - 1161.5820+
- y9 - 951.4452+
- y8 - 852.3768+

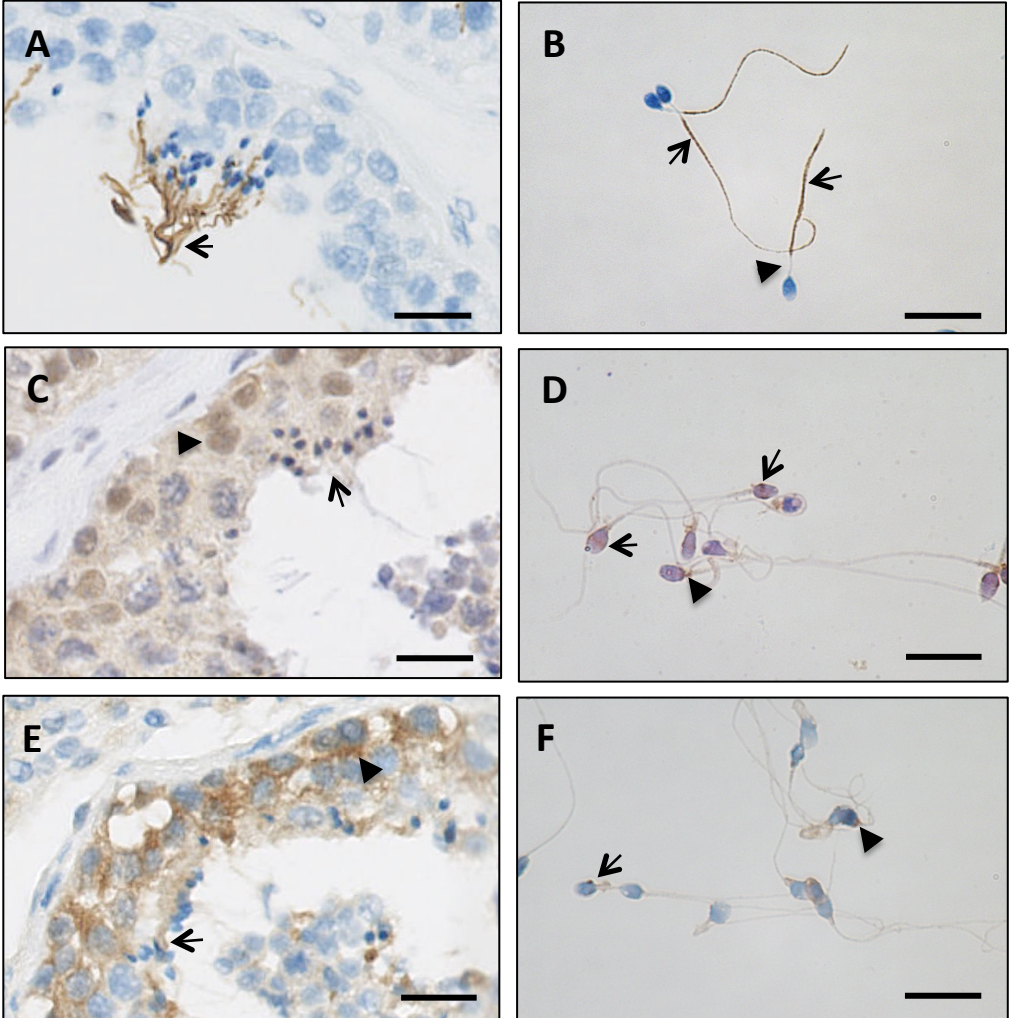
- y8 - 784.4676+
- y5 - 531.3249+
- y8 - 392.7374++

Figure 2:



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

Figure 3:



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

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	Q43506	Disintegrin and metalloproteinase domain-containing protein 20	LQHDVAHLFIK FGHCIGIVGTTYVK	X X	
FAM181A	Q8N9Y4	Protein FAM181A	ASCSGPLVMASDSDVK MPGVSLVGR	X X	
SLC35F4	A4IF30	Solute carrier family 35 member F4	YTQVTNFLGTSR IFGEDGLTLK	X	X
TEX22	C9J3V5	Testis-expressed protein 22	ERPGAAGTQLHCR KLESHLSQEHR		X X
LRRC9	Q6ZRR7	Leucine-rich repeat-containing protein 9	VPSSGYGQQGISK LDEIEAIYHIEVK		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 ELPVPLMLTLPQK	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