



**HAL**  
open science

# The functionally unannotated proteome of human male tissues: A shared resource to uncover new protein functions associated with reproductive biology

Yves Vandembrouck, Charles Pineau, Lydie Lane

## ► To cite this version:

Yves Vandembrouck, Charles Pineau, Lydie Lane. The functionally unannotated proteome of human male tissues: A shared resource to uncover new protein functions associated with reproductive biology. *Journal of Proteome Research*, 2020, 19 (12), pp.4782-4794. 10.1021/acs.jproteome.0c00516 . hal-03003302

**HAL Id: hal-03003302**

**<https://hal.science/hal-03003302>**

Submitted on 20 Nov 2020

**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  
4  
5  
6  
7 **The functionally unannotated proteome of**  
8  
9  
10  
11 **human male tissues: a shared resource to**  
12  
13  
14  
15 **uncover new protein functions associated with**  
16  
17  
18  
19 **reproductive biology.**  
20  
21  
22  
23

24 *Yves Vandenbrouck<sup>1,\*</sup>, Charles Pineau<sup>2</sup>, Lydie Lane<sup>3</sup>*  
25  
26  
27  
28  
29

- 30 1. Univ. Grenoble Alpes, INSERM, CEA, IRIG-BGE, U1038, 38000, Grenoble, France  
31  
32 2. Univ. Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et  
33 travail) - UMR\_S 1085, F-35042 Rennes cedex, France  
34  
35 3. SIB Swiss Institute of Bioinformatics and Department of microbiology and  
36 molecular medicine, Faculty of medicine, University of Geneva, CMU, Michel Servet  
37 1, 1211 Geneva 4, Switzerland  
38  
39  
40  
41

42 \* To whom correspondence should be addressed: [yves.vandenbrouck@cea.fr](mailto:yves.vandenbrouck@cea.fr)  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**ABSTRACT**

In the context of the Human Proteome Project, we built an inventory of 412 functionally unannotated human proteins for which experimental evidence at the protein level exist (uPE1) and which are highly expressed in tissues involved in human male reproduction. We implemented a strategy combining literature mining, bioinformatics tools to collate annotation and experimental information from specific molecular public resources, and efficient visualization tools to put these unknown proteins into their biological context (protein complexes, tissue and subcellular location, expression pattern). The gathered knowledge allowed pinpointing five uPE1 for which a function has recently been proposed and which should be updated in protein knowledge bases. Furthermore, this bioinformatics strategy allowed to build new functional hypotheses for five other uPE1s in link with phenotypic traits that are specific to male reproductive function such as ciliogenesis/flagellum formation in germ cells (CCDC112 and TEX9), chromatin remodeling (C3orf62) and spermatozoon maturation (CCDC183). We also discussed the enigmatic case of MAGEB proteins, a cancer/testis antigen subtype poorly documented. Tools used and computational outputs produced during this study are freely accessible via ProteoRE (<http://www.proteore.org>), a Galaxy-based instance, for re-use purposes. We propose these five uPE1s should be investigated in priority by expert laboratories and hope that this inventory and shared resources will stimulate the interest of the community of reproductive biology.

**KEYWORDS:** human proteome project, proteomics, male reproduction, functionally unannotated proteins, bioinformatics, biocuration, data mining, protein complexes, protein interaction networks, ciliogenesis, spermatogenesis, acrosome, chromatin remodeling, cancer/testis antigens.

## INTRODUCTION

The HUPO Human Proteome Project (HPP) recently launched a new initiative, nextCP50, aiming to fill the gaps in the functional annotation of the human proteome<sup>1</sup>. Thanks to the continuous effort of the life science community, the number of proteins devoid of functional annotations - derived from bioinformatics prediction or from experimental studies, is constantly decreasing. Although they are functionally unannotated, these proteins can have other types of annotations such as subcellular locations, protein-protein interactions, tissue expression, association with diseases or 3D structure. This information can be examined using the advanced search functionality of neXtProt<sup>2</sup>. In addition, neXtProt provides a "protein existence" (PE) status to each entry, which indicates the type of evidence that supports the existence of the protein. The value 'Experimental evidence at protein level' (PE1) indicates that there is clear experimental evidence for the existence of at least one proteoform in the entry. The value 'Experimental evidence at transcript level' (PE2) indicates that the existence of a protein has not been proven but that expression data indicate the existence of a transcript. The value 'Protein inferred by homology' (PE3) indicates that the existence of a protein is probable because clear orthologs exist in closely related species. The value 'Protein predicted' (PE4) is used for entries without evidence at protein, transcript, or homology levels. Finally, the value 'Protein uncertain' (PE5) indicates that the existence of the protein is unsure and that it most probably corresponds to an incorrect in silico translation of a non-coding element. In neXtProt 2020-01-17, there are 1,813 PE1-4 proteins with no function annotated. Among them, 559 have not been confidently detected at protein level in a human sample and are considered as "missing" (PE2-4), whereas 1254 have been experimentally verified (PE1) and are called "uPE1"<sup>3</sup>. The present study focuses on the uPE1 proteins whose expression is elevated in the male reproductive system. We included in our analysis the testis, epididymis, vas (ductus) deferens, seminal vesicle and prostate. The function of the male reproductive system is to produce male gamete - spermatozoa

1  
2  
3 - and transfer them to the female reproductive tract. The testis is a crucial component in this  
4  
5 process, as it produces both spermatozoa and androgens that support male reproductive  
6  
7 physiology. Accessory organs and ducts also contribute to sperm maturation and transport  
8  
9 mainly through the production of the seminal plasma, a complex biological fluid that originates  
10  
11 from the testes (1-2%), the epididymis (2-4%), the seminal vesicles (65-75%), the prostate  
12  
13 gland (25-30%) and the bulbourethral or Cowper's gland (<1%). The seminal plasma, in  
14  
15 addition to its obvious role in sperm transportation, provides a nutritive and protective milieu  
16  
17 and is involved in the final maturation of the male gametes<sup>4</sup>. Previous transcriptomics and  
18  
19 proteomics studies revealed that the human testis displays the largest number of tissue-enriched  
20  
21 genes followed by the brain and liver<sup>5-8</sup>. The large number of enriched genes in testis is  
22  
23 considered to be due to the highly specialized processes occurring during spermatogenesis that  
24  
25 are linked to the presence of the germ cell lineage. Conversely a significant number of enriched  
26  
27 genes is present in the epididymis, and to a lesser extent some can also be found in seminal  
28  
29 vesicle and prostate; organs that are known to be involved in the post-testicular maturation of  
30  
31 spermatozoa. In the present study we established an inventory of 412 uPE1 proteins enriched  
32  
33 in the male reproductive system by collating information (e.g., experimental data, annotation,  
34  
35 facts from literature) from publicly available resources. Then, we built functional hypotheses  
36  
37 on a small set of these functionally unannotated protein entries by: i. combining contextual  
38  
39 information (molecular complexes, cellular localization, expression level), ii. visualizing the  
40  
41 resulting protein complexes, and iii. documenting their likelihood by cross-checking with the  
42  
43 literature and our own knowledge. The datasets and tools used in this study are available in  
44  
45 ProteoRE, a freely accessible Galaxy-based instance<sup>9</sup> dedicated to biomedical research<sup>10,11</sup>.  
46  
47 Likewise, results and computational outputs have been shared for re-use by experts from the  
48  
49 community of reproductive biologists.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## MATERIALS AND METHODS

### Public resources used: content and release information

Expression profiles for transcripts and proteins in human male tissues (testis, epididymis, ductus deferens, seminal vesicle and prostate) based on RNA sequencing analysis and immunohistochemistry, respectively, were downloaded from Human Protein Atlas (HPA) (<https://www.proteinatlas.org/humanproteome/tissue>; version 19.2 and Ensembl version 92.38). Lists of functionally unannotated and missing protein entries were retrieved from neXtProt using SPARQL queries NXQ\_000022 and NXQ\_00204 respectively (February 2020). Human protein complex map files (hu.MAP 2.0) were downloaded from <http://hu.proteincomplexes.org/download>. Literature survey was performed using Pubmed, GoogleScholar and the GenCLiP3 web application (<http://ci.smu.edu.cn/genclip3/analysis.php>).

### Bioinformatics analyses

The tools implemented and used in this study are part of the freely accessible ProteoRE platform: <http://www.proteore.org>. Specifically, data retrieval, mapping identifiers from Ensembl to neXtProt, results manipulation and Venn diagram were performed using the following tools of ProteoRE: “Data manipulation and visualization” section: “ID Converter (Human, Mouse, Rat)”, “Filter by keywords and/or numerical value”, “Venn diagram [JVenn]”; “Get Features/Annotation” section: “Add expression data (RNAseq or Immuno-assays) [Human Protein Atlas]”, “Add protein features [neXtProt]”; “Pathways analysis” section: “Build Protein interaction network [BioGRID, BioPlex, HuMAP]”. For ease of use, a detailed user documentation is provided for each tool at the bottom of the central panel of the Galaxy workflow. This documentation describes what the tool does, the input it requires, parameters to be tuned, and the output produced and lists data sources, along with their release date when applicable.

### Mapping from Ensembl to neXtProt

Each ENSG identifier was converted into a neXtProt identifier using mapping tables from neXtProt ([ftp://ftp.nextprot.org/pub/current\\_release/mapping/](ftp://ftp.nextprot.org/pub/current_release/mapping/) - March 2020). 126 HPA entries with ENSG identifiers that do not map to any neXtProt ID were removed from the study. Most of these unmapped ENSG identifiers correspond to “overlapping locus” in which exon(s) overlap exon(s) of a readthrough transcript or a transcript belonging to another locus. Such locus are not represented in UniProtKB/Swiss-Prot or neXtProt because their biological relevance at protein level is unsure. In addition, we manually verified and corrected cases where neXtProt ID map to several ENSG ID (e.g. NX\_A1L429 <-> ENSG00000237671; ENSG00000227488; ENSG00000216649) and reciprocally (e.g. ENSG00000212710 <=> NX\_Q96RT6; NX\_Q9HC47).

### Protein complex map visualization

The tabular files with network and nodes attributes were generated by the “Build Protein interaction network [BioGRID, BioPlex, HuMAP]” tool of ProteoRE by selecting hu.MAP as resource (see above). These two files were uploaded in Cytoscape software (Version 3.7.2) (<http://www.cytoscape.org>) to visualize the protein complex map including the functionally unannotated proteins (uPE1). In this network, nodes represent proteins (unknown and known function) with their expression level in testis and their first interactors, whereas edges between nodes represent protein–protein physical interaction along with their interaction score according to hu.MAP. A visual style (e.g., node shape, continuous mapping color, edge width) was mapped to node and edge attributes, highlighting uPE1 in their cellular and molecular context (see Figure 4 and 5 legends).

### History, datasets and availability of analyses

Datasets (testis “tissue\_category\_rna\_testis\_Tissue.tsv” ; epididymis “tissue\_category\_rna\_epididymis\_Tissue.tsv” ; ductus deferens

1  
2  
3 "tissue\_category\_rna\_ductus.tsv" ; seminal vesicle "tissue\_category\_rna\_seminal" ; prostate  
4  
5 "tissue\_category\_rna\_prostate\_Tissue"), used and results (network and node attributes files  
6  
7 (.tsv) and network (.cys format)) produced in this study, have all been shared and made  
8  
9 available at <http://www.proteome.org> (via shared data menu: "Data libraries" -> "Human male  
10  
11 reproduction uPE1 datasets") (available after free registration). History can also be accessed  
12  
13 and imported via the following link:  
14  
15 <http://www.proteome.org/u/yvdb/h/buildingunknownproteomemaletissue2020> (name:  
16  
17 "Building\_UnknownProteome\_MaleTissue\_2020").  
18  
19  
20  
21

## 22 RESULTS AND DISCUSSION

### 23 Overall workflow

24  
25 The overall workflow for establishing the functionally unannotated proteome of human male  
26  
27 tissue, and to build functional hypotheses for uPE1 proteins discussed herein, is illustrated in  
28  
29 Figure 1 and described in the Material and Methods, with full details about results availability.  
30  
31 Briefly, the core proteome of the human male reproductive system was built using the Human  
32  
33 Protein Atlas (HPA) database by selecting genes/proteins flagged as "elevated" in at least one  
34  
35 of the five human male tissues available. These tissues are: testis, epididymis, ductus deferens,  
36  
37 seminal vesicle and prostate (Figure 2). This set of 2714 entries was compared to the unknown  
38  
39 proteome and the latest release of the missing proteome (February 2020, SPARQL query  
40  
41 NXQ\_00204) enabling to select 412 protein entries which are PE1 with no function annotation  
42  
43 (Figure 3). Then we performed a survey literature and identified five uPE1 (TEX33, CCDC38,  
44  
45 CFAP70, CFAP58, C3orf62 and ANKDR55) for which a functional role was recently reported  
46  
47 (Table 1). In parallel, this set of 412 uPE1 was placed in their cellular and molecular context  
48  
49 (Table 1). In parallel, this set of 412 uPE1 was placed in their cellular and molecular context  
50  
51 by mapping them to the human protein complex map (hu.MAP), a public resource integrating  
52  
53 over 9000 protein interaction datasets<sup>12</sup> (Figure 4). Mapping co-complex membership of uPE1  
54  
55 proteins revealed five protein complexes for which we have developed functional hypotheses  
56  
57  
58  
59  
60



1  
2  
3 by considering additional information such as co-expression and co-localization (from HPA  
4 and neXtProt), and experimental evidence from literature mining. This strategy enabled us to  
5  
6 propose a potential functional role for five uPE1 (CCDC112, TEX9, CCDC183, C3orf62 and  
7  
8 MAGEB4) (Figure 5) out of seven (Table 3), in link with human male reproduction.  
9

### 10 11 12 ***Building a core proteome of the human male reproductive system***

13  
14  
15 The core proteome of the male reproductive system was built using the Human Protein Atlas  
16 (HPA) database<sup>5,13</sup>. To the best of our knowledge, this database is one of the most exhaustive  
17  
18 resources on human protein localization and expression. Its most recent version includes five  
19  
20 tissues that are parts of the male reproductive system: testis, epididymis, ductus deferens,  
21  
22 seminal vesicle and prostate. We restricted our study to genes/proteins flagged as "elevated"  
23  
24 according to the tissue specificity category defined by HPA (i.e., belonging to the HPA  
25  
26 categories "tissue enriched", "group enriched" and "tissue enhanced") in male tissues<sup>5</sup>, thus  
27  
28 possibly related to male reproductive physiology. Tissue specificity category is defined based  
29  
30 on the total set of all normalized expression (NX) in 37 tissues, 18 blood cell types or 64 cell  
31  
32 lines and using a cutoff value of 1 NX as a limit for detection across all tissues or cell  
33  
34 types. According to HPA definition, "tissue enriched" corresponds to a NX level in a particular  
35  
36 tissue/region/cell type at least four times any other tissue/region/cell type, "group enriched" to  
37  
38 NX levels of a group (of 2-5 tissues or 2-10 cell types) at least four times any other  
39  
40 tissue/region/cell type and "tissue enhanced" to NX levels of a group (of 1-5 tissues or 1-10 cell  
41  
42 types) at least four times the mean of other tissue/region/cell types. As expected, the number of  
43  
44 genes showing an elevated expression in testis is far higher compared to the four other male  
45  
46 tissues (Figure 2, lower part), while the number of genes with elevated expression in two or  
47  
48 three male tissues is rather low: 71 genes are elevated both in testis and epididymis and 50  
49  
50 genes are elevated in both ductus deferens and seminal vesicle (Figure 2, upper part), including  
51  
52 semenogelins 1 (SEMG1) and 2 (SEMG2), which are predominant in semen. These two  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 proteins have antibacterial activity and are involved in sperm motility<sup>14</sup>. Seventeen genes are  
4 elevated in ductus deferens, seminal vesicle and epididymis (Figure 2, upper part). Very few  
5 genes have elevated expression in other combinations of three tissues. One of them is  
6 KIAA1210, a uPE1 protein that is elevated in testis, epididymis and seminal vesicle and  
7 reported as enhanced in cancer prostate in HPA. The mouse ortholog of KIAA1210 has been  
8 suggested to be an essential cell junction protein interacting with topoisomerase TOP2B to  
9 regulate the dynamic change of chromatin structures during spermiogenesis<sup>15</sup>. Spermiogenesis  
10 is the final postmeiotic phase of spermatogenesis (for a review see <sup>16</sup>). We also noticed the  
11 presence of MSMB, a uPE1 protein whose expression is restricted to the prostate at a very high  
12 level (Table S1, normalized expression value ~437). The MSMB gene encodes beta-  
13 microseminoprotein (MSP), a member of the immunoglobulin binding factor family  
14 synthesized by epithelial cells of the prostate and secreted into seminal plasma. MSP has been  
15 proposed as a serum marker for early detection of high grade prostate cancer<sup>17</sup>.

### 33 ***Establishing the functionally unannotated human male tissue proteome***

34  
35 These five male tissue datasets were then concatenated, and duplicated entries sorted out  
36 resulting in 2,883 ENSG entries, among which 2,757 were successfully mapped to 2,714  
37 neXtProt entries constituting “the human male proteome”. Out of these 2714 neXtProt entries,  
38 25 are flagged as PE5 and probably correspond to non-coding elements, whereas 2134 were  
39 validated at protein level (PE1). 555 are only validated at transcript level (PE2) and await  
40 confident detection at protein level. Specific mass spectrometry-based workflows aiming at  
41 identifying these proteins in testis or sperm have been conducted by several HPP teams<sup>6,7,18–20</sup>.

42  
43 As the neXtProt list also contains functionally unannotated proteins with no experimental  
44 evidence at the protein level (i.e. missing proteins), we restricted our inventory to the  
45 functionally unannotated proteins assigned with a PE1 status indicating that there is  
46 experimental evidence for the existence of at least one proteoform in the entry. The comparison  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 between the human male proteome, the unknown proteome and the latest release of the missing  
4 proteome (February 2020, SPARQL query NXQ\_00204) showed there are currently 583  
5 neXtProt functionally unannotated entries having elevated expression level in the Human male  
6 tissue (Figure 3). Out of these, 412 are PE1 validated proteins with no function annotated and  
7 thus constitute the current uncharacterized human male tissue proteome. Experimental  
8 information, annotations and features associated with these 412 proteins matching to 428  
9 Ensembl entries were retrieved from HPA and neXtProt and gathered in Table S1. A quarter of  
10 the 428 Ensembl entries (101) have detectable mRNA expression only in testis (Table S1, in  
11 blue), and 165 others are mainly expressed in testis but also detectable in other tissues (Table  
12 S1, in light blue). For these entries, we suggest a role in spermatogenesis or testis development.  
13 For 14 Ensembl entries, expression was restricted to male tissues but not only to testis,  
14 suggesting a role in male reproduction that might be different from spermatogenesis or testis  
15 development (Table S1, in green). For 26 Ensembl entries, high expression levels were found  
16 in both testis and Fallopian tubes. As Fallopian tubes are ciliated organs, high levels in both  
17 testis and Fallopian tube may indicate a role in cilia biogenesis or function (Table S1, in yellow).  
18 The remaining proteins are expressed in male tissues but also in other tissues such as the brain,  
19 blood or skin, suggesting that they can perform functions not specific to the male reproductive  
20 system.  
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 ***Updating the functionally unannotated male tissue proteome***

46  
47 The set of 412 uPE1 was first compared to the consolidated list of uncharacterized proteins  
48 published by Duek et al., in their seminal article<sup>21</sup>. We noticed eight proteins for which a  
49 functional hypothesis has been proposed and extensively discussed by the authors. Six may be  
50 involved in ciliogenesis (CASC1, CCDC33, CCDC172, CFAP45, ODF3L1, ODF3L2), one in  
51 late spermatogenesis (TEX44), and one in the gut-brain-microbiota axis (NXPE2).  
52 Interestingly, in the testis NXPE2 is moderately and uniformly expressed in the germ cell  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 lineage, whereas a strong expression was observed in the head of the epididymis with typical  
4 secretion patterns in the pseudostratified columnar epithelium. There, the so-called principal  
5 cells possess microvilli that serve to resorb the fluid originating from the testicular secretion<sup>22</sup>.  
6  
7 Fluid resorption is a function somewhat similar to that of cells in the salivary gland acini that  
8 also display a strong NXPE2 immunoreactivity. We also noticed the presence of PNMA6E, a  
9 protein only expressed in the germ cell lineage up to meiosis in testis, and to a lower extent in  
10 the ovary and Fallopian tubes, for which a role in gametogenesis was recently proposed<sup>23</sup>. We  
11 performed a survey of the literature published on these 412 uPE1 proteins using various search  
12 engines and could find a publication reporting the location of TEX33 in the mouse male germ  
13 cell lineage<sup>24</sup>. We also found publications reporting functional roles for CCDC38<sup>25</sup>, CFAP70<sup>26</sup>,  
14 CFAP58<sup>27</sup> and C3orf62<sup>28</sup> (further discussed in this study) in spermatogenesis, and for  
15 ANKDR45 in cell proliferation<sup>29</sup> (Table 1). The five latter proteins should no longer be  
16 considered as “uPE1” according to neXtProt criteria. The function of CFAP70 has just been  
17 annotated in UniProt and neXtProt (Release 2020-07-17). Update requests were sent to the  
18 UniProt team for the four other ones.

### 37 ***Putting the male tissue uPE1 proteins in their cellular and molecular context***

38  
39  
40 The vast majority of cellular functions are carried out by molecular machines made up of non-  
41 covalent protein–protein interactions (PPIs) and called protein complexes. Putting uPE1  
42 proteins in their molecular context as well as considering their cellular localization can be a  
43 valuable approach for predicting their functional role. In line with this idea, Drew et al.,<sup>12</sup>  
44 constructed a global human protein complex map (hu.MAP) by re-analyzing over 9,000 large-  
45 scale human protein complex mass spectrometry experimental datasets obtained on cell lines.  
46 They showed that integrating these datasets provided substantial improvement on  
47 discriminating true and false protein interactions, and allowed to highlight both known and  
48 novel complexes relevant to ciliopathies. For instance, the global map predicted an interaction  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 between the uPE1 ANKRD55 (Ankyrin repeat domain-containing protein 55) and the IFT-B  
4 complex, that is one of the two sub-complexes of the IFT (Intraflagellar Transport) system,  
5 which link cargos to microtubule motors for transport along ciliary axonemes<sup>30</sup>. This interaction  
6 was validated in vivo by confocal imaging and genetic perturbation<sup>12</sup>. Inspired by this approach,  
7 we looked for publicly available interactomics data on the 412 uPE1 proteins with an elevated  
8 expression in the human male tissues. As we were unable to find publicly available data from  
9 experiments performed in human testis (or sperm) samples, we constructed a network using the  
10 hu.MAP resource to map co-complex membership of uPE1 proteins with an elevated expression  
11 in the human male tissue, and to explore their potential for revealing consistent functional  
12 relationships within protein complexes. Whereas 22% of all the uPE1 proteins (286/1254) have  
13 data in Hu.MAP, only 10% of the uncharacterized proteins that are enriched in male tissues  
14 (41/412) do so. This indicates that a number of the 412 uPE1 proteins enriched in male tissues  
15 may have specific interactors in male tissues that cannot be studied in cell lines. The 41  
16 uncharacterized proteins that are enriched in male tissues and have data in hu.MAP mapped  
17 onto 37 protein complexes. We hypothesized that most of these interactions would be conserved  
18 in male tissues. Using graphical functionalities of Cytoscape, a global view of these complexes  
19 was built highlighting uPE1, their expression level, their interacting partners and the predicted  
20 interaction score (Figure 4).  
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 ***Building functional hypothesis by combining predicted complexes, specialized database***  
46 ***annotations and literature survey***  
47  
48

49 We elaborated functional hypothesis on some uPE1 by using graphical properties and node  
50 attributes of this protein complex network (Figure 4). Since more and more proteins are found  
51 to have different roles depending on their cellular context<sup>31</sup>, it is important to check that the  
52 interactions highlighted in the hu.MAP network can have functional relevance in male tissues.  
53  
54  
55  
56  
57  
58  
59  
60 Therefore, we made sure to select protein pairs that are found in the same cellular compartments

1  
2  
3 and have shared expression in male tissues. Protein-protein interaction scores (edge thickness),  
4 shared expression level in male tissue (node color gradient) and protein features such as  
5 subcellular co-localization (nodes attributes) were used to select the most likely protein  
6 complexes. According to Drew et al., an interaction score greater than 0.27 is considered as of  
7 very high confidence<sup>12</sup>. Using this score threshold and visual properties, five protein complexes  
8 were selected (Figure 5). Annotation and experimental data associated with each protein from  
9 these five complexes are provided in Table 2 (detailed information in Table S2).

10  
11 ANKRD55 has an elevated expression in testis and is one of the 412 uPE1 proteins from our  
12 set. Unsurprisingly, one of the five selected complexes is the previously reported complex  
13 formed by ANKRD55 and three components of the IFT-B complex (IFT52, TTC30A and  
14 TTC30B)<sup>32-34</sup> (Figure 5A).

15  
16 With the aim of unraveling possible physiological roles for uPE1s in male germ cells and sperm  
17 maturation, each protein component of the four other selected complexes was investigated  
18 further by mining the literature. We first focused on two uPE1 members of the CCDC (coiled-  
19 coil domain containing) protein superfamily<sup>35</sup>, CCDC112 and CCDC183. Dimeric coiled coils  
20 are structural motifs known to mediate protein-protein interaction driving dynamic assembly of  
21 protein complexes<sup>36</sup>, and some coiled-coil domain containing proteins have been shown to be  
22 involved in the biology of cilia and flagella (e.g., CCDC114, CCDC169, CCDC34, CCDC40)<sup>37</sup>.  
23 Secondly, we investigated C3orf62, recently studied using knockout (KO) male mice<sup>28</sup> for a  
24 possible role in chromatin remodeling via an interaction with an histone acetyltransferase. At  
25 last, we discussed a complex composed of several cancer/testis (CT) antigens, a family of  
26 proteins only expressed in germ cells and aberrantly expressed in various human cancers.

### 27 ***CCDC112 and TEX9 are centriolar satellite proteins regulating ciliogenesis***

28  
29 CCDC112 was confidently predicted to interact with TEX9, another uPE1 protein (interaction  
30 score = 0.95), and less confidently (interaction score = confidence score 0.15) with NOL4 and  
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 NOL4L , two nucleolar-localized proteins of unknown function (Figure 5B). According to the  
4 interaction score threshold considered in hu.MAP, NOL4 and NOL4L are likely false positives.  
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

CCDC112 and TEX9 are not only expressed in the five male tissues (at mRNA level) but also in all, or many other examined tissues examined by HPA respectively (Table S2). These two proteins were reported to interact in an independent study investigating the dynamic changes in the centrosome-cilium protein interaction landscape during ciliogenesis<sup>38</sup>. They were shown to co-localize in centriolar satellites, which are membraneless granules that localize and move around centrosomes and cilia, and are key regulators of cilia, flagella and centrosome biogenesis<sup>39-41</sup>. During the maturation of spermatids, centrioles are important to form the sperm flagellum and for sperm movement. Failure to properly regulate centrosome function could be associated with male infertility<sup>42,43</sup>. Considering that satellite proteins have been shown to regulate ciliogenesis<sup>44,45</sup> and centriole assembly<sup>40,46</sup>, refining the role of CCDC112 and TEX9 in the assembly, maintenance and remodeling of the centrosome/cilium complex, could be particularly informative with respect to male infertility.

### ***CCDC183 as essential in sperm development and maturation?***

41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

According to HPA, CCDC183 is only expressed in testis (Table S2). The protein expression pattern of immunohistochemistry (IHC) data from HPA for CCDC183 is assigned with an “uncertain” reliability score in HPA (Table 2), but single-cell RNA sequencing data reveals that CCDC183 is upregulated during the transition from the secondary spermatocytes to round spermatids, and downregulated during late spermiogenesis<sup>47</sup>. A proteomic analysis (SWATH-MS) of good- vs. poor-quality human sperm in a human cohort showed that CCDC183 is highly abundant in the good-quality spermatozoa<sup>48</sup>. Of note is that according to the World Health Organization, and among other features, good-quality sperm display at least 80% forward progressive motility upon Percoll gradient density centrifugation. In this study poor-quality sperm fractions are asthenozoospermic with <60% forward progressive motility and correlated

1  
2  
3 with lower levels of CCDC183<sup>48</sup>. We thus investigated for clues supporting a role of CCDC183  
4  
5 in sperm motility. The predicted module containing CCDC183 revealed seven protein-protein  
6  
7 interactions, each having a high interaction score (ranging from 0.64 to 0.96) (Figure 5C).  
8  
9 According to HPA, these seven interactors are expressed in all examined tissues including  
10  
11 testis, and IHC data for FOXP1, RABEP1, RUNDC1, TEDC1 and TMF1 proteins show an  
12  
13 immunostaining ranging from moderate to strong in seminiferous tubules. The interaction  
14  
15 between CCDC183 and FOXP1, a transcriptional repressor, could be relevant for the  
16  
17 acquisition of motility by sperm cells, as FOXP1 was suggested to play a role in the regulation  
18  
19 of ciliogenesis<sup>49</sup>. Likewise, TEDC1 has been shown to be required for centriole stability<sup>50</sup>.  
20  
21 TMF1 (TATA element modulatory factor, also known as androgen receptor-associated protein  
22  
23 of 160 kDa (ARA160)) is not known to be involved in flagella formation but plays a role in  
24  
25 sperm development. Indeed its murine ortholog was characterized as a Golgi-associated protein  
26  
27 essential for acrosome formation during sperm development<sup>51</sup>; it was also observed in post  
28  
29 meiotic germ cells (i.e., early/round spermatids) in human and found to be mainly localized in  
30  
31 the Golgi apparatus in cell lines. The early steps of spermiogenesis are characterized by re-  
32  
33 orientation of the Golgi towards the cell nucleus, and tethering of trans-Golgi-derived vesicles  
34  
35 onto the acroplaxome to initiate acrosome formation in early/round spermatids<sup>52,53</sup>. In *Tmf*  
36  
37 mutant mice spermatids, the Golgi apparatus has abnormal orientation and the pro-acrosomal  
38  
39 vesicles lose their guidance towards the nucleus<sup>54</sup>. These results suggest the involvement of  
40  
41 this protein in the spatial re-orientation of the Golgi apparatus during the so-called Golgi-phase  
42  
43 during which the organelle produces and releases lytic enzymes that will accumulate inside the  
44  
45 forming acrosome in maturing spermatids. Furthermore, RABEP1 and RABGEF1, both  
46  
47 localized in early endosomes, play a role in endocytic trafficking between endosome and trans-  
48  
49 Golgi networks<sup>55</sup>. These proteins might also be involved in acrosome biogenesis. Investigating  
50  
51 the functional links that potentially exist between CCDC183, FOXP1, TEDC1, TMF1,  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 RABEP1 and RABGEF1, might be of high interest, particularly with respect to its potential  
4 role during the biogenesis of flagellum and acrosome in haploid male germ cells.  
5  
6  
7

8 ***C3orf62 may be involved in spermatogenesis by modulating the chromatin remodeling***  
9

10  
11 Chromatin remodeling is an essential process for normal sperm development<sup>56,57</sup> and  
12 fertilization<sup>58,59</sup>. During spermiogenesis, most of the germ cell cytoplasm, intracellular  
13 organelles and presumably a number of cytoplasmic proteins, are eliminated through residual  
14 bodies. Concomitantly, the chromatin undergoes extensive remodeling as most of the histones  
15 are replaced by highly basic protamines<sup>60</sup>. As a result, transcriptional activity is greatly  
16 repressed due to the packaging of sperm chromatins. Chromatin remodeling involves a dynamic  
17 interchange between open (transcriptionally active) and closed (transcriptionally repressed or  
18 silenced) chromatin structures, and is controlled by enzymatic modification of the DNA and/or  
19 histone proteins, such as DNA methylation and histone acetylation, respectively. With this  
20 respect, the interaction of C3orf62 with EP300, a histone acetyltransferase that regulates  
21 transcription via chromatin remodeling<sup>61</sup> (confidence score ~0.8) caught our attention (Figure  
22 5D). C3orf62 mRNA is expressed in testis and in blood while EP300 mRNA is expressed in all  
23 HPA tissues (Table 2, Table S2). There is no data available for C3orf62 protein expression in  
24 tissues, but this protein was detected in nuclear speckles in human cancer cell lines (A-431, U-2  
25 OS and U-251 MG). Nuclear speckles are located in the interchromatin regions of the  
26 nucleoplasm of mammalian cells<sup>62</sup>. They are dynamic structures, and their constituents  
27 exchange continuously with the nucleoplasm and other nuclear locations<sup>63</sup>. As EP300 was  
28 reported to be located in the nucleoplasm, deeper experimental investigations are required to  
29 determine if EP300 and C3orf62 may co-localize either in the nucleoplasm or in speckles, and  
30 in which conditions. Our literature survey revealed a recent study<sup>28</sup> about the role of the murine  
31 ortholog of C3orf62 highly expressed in spermatocytes. Using a 1700102P08Rik deficient  
32 mouse model generated by CRISPR-Cas9, the study showed that the disruption of  
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 1700102P08Rik causes a germ cell maturation arrest at the spermatocyte stage, indicating that  
4 this gene is essential for spermatogenesis and that its dysfunction leads to male infertility<sup>28</sup>. The  
5 authors also found that C3orf62 mRNA is downregulated in infertile patients with a  
6 spermatocyte maturation arrest. Together, these experimental observations could support the  
7 hypothesis that C3orf62 interacts with EP300, by modulating its activity and consequently  
8 chromatin remodeling in haploid germ cells.  
9  
10  
11  
12  
13  
14  
15

### 16 17 ***The enigmatic case of the MAGE-B protein family***

18  
19  
20 Cancer/testis antigens (CTA) are encoded by genes that are normally expressed only in the  
21 human germ cell lineage, but are also expressed in various tumor types. The aberrant expression  
22 of germline genes in cancers reflects the activation of the silenced gametogenic program in  
23 somatic cells, and suggests that this programmatic acquisition is one of the driving forces of  
24 tumorigenesis<sup>64</sup>. Here we focused on MAGE (melanoma-associated antigen) B, a CTA family  
25 composed of ten genes located on Xp21–p22 chromosome. The human complex map reveals  
26 that three MAGEB (i.e., B3, B4 and B6) proteins form an intricate network composed of eleven  
27 proteins with multiple interactions (Figure 3). When restricting this complex by selecting  
28 protein-protein interactions of high confidence (i.e., score > 0.3), we obtained a smaller network  
29 composed of MAGEB3, MAGEB4, MAGEB6, MAGEA1 (another CTA encoded by a gene  
30 located on Xq28), and ROCK1 (Rho-associated protein kinase 1). MAGEA1 interacts with  
31 MAGEB3, B4 and B6, while MAGEB4 interacts with ROCK1 (Figure 5E). Gene expression  
32 of MAGEA1 and MAGEB6 are restricted to testis (detected in single), while MAGEB3 and  
33 MAGEB4 are both expressed in testis and epididymis, and ROCK1 is expressed in all tissues.  
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  
Visual inspection of IHC data performed on human testis sections (HPA) confirmed MAGEA1,  
MAGEB3, MAGEB6 and ROCK1 to be expressed in germ cells at all stages of  
spermatogenesis, with higher expression levels in spermatogonia and meiotic spermatocytes.  
MAGEB4 is expressed at very low levels at all stages. MAGEA1 has been proposed to act as a

1  
2  
3 potent transcriptional repressor and participate in the setting of specific gene expression patterns  
4  
5 in spermatogenesis<sup>65</sup>. Interestingly, data from single-cell RNA sequencing analysis of human  
6  
7 male germ cell transitions revealed that both MAGEA1 and MAGEB6 were expressed in early  
8  
9 primary spermatocytes, with a similar differential expression pattern during spermatogenesis  
10  
11 (i.e., up from differentiated spermatogonia to preleptotene spermatocytes stages and then down  
12  
13 past meiosis)<sup>47</sup>. No single-cell RNA sequencing is available for MAGEB3 and MAGEB4. In  
14  
15 the developing mouse testis, the *Mageb4* gene was shown to be highly expressed in  
16  
17 proliferating gonocytes until they arrest at the G0/G1 phase and thus likely maintains gonocytes  
18  
19 in a non-proliferative state. When *Mage-b4* gene is down regulated in the adult testis, the germ  
20  
21 cells are allowed to enter the first meiotic division<sup>66</sup>.

22  
23  
24  
25  
26 Rho kinases ROCK1 and ROCK2 are effectors of the small GTPase Rho. ROCKs are important  
27  
28 cytosolic nodes whose activation leads to numerous downstream events such as actin  
29  
30 polymerization, focal adhesion, and stress fiber formation (for a review see <sup>67</sup>). It has been  
31  
32 suggested that, in the testis, ROCKs and Rho GTPase play a role in the assembly of Sertoli-  
33  
34 germ cell actin-based adherens junction (AJ)<sup>68</sup>. However, no experimental data could be found  
35  
36 on the possible interaction between MAGEB4 and ROCK1. One may hypothesize that both  
37  
38 proteins could be functionally related during the transformation of primordial germ cells (PGC)  
39  
40 into gonocytes and their further maintenance in a non-proliferative state. The small MAGEB3,  
41  
42 MAGEB4, MAGEB6, MAGEA1 and ROCK1 network highlighted in this study could thus be  
43  
44 of major importance to explain failed gonocytes differentiation into carcinoma *in situ* and/or  
45  
46 germ cell maturation arrests at later stages of spermatogenesis.  
47  
48  
49  
50

## 51 52 53 **CONCLUSIONS**

54  
55 This study is a first attempt to establish an inventory of 412 proteins expressed in the human  
56  
57 male reproductive system for which a functional role is still unknown (uPE1). Messenger RNA  
58  
59 and protein expression are strictly regulated during spermatogenesis and the process involves a  
60

1  
2  
3 myriad of coordinated events (for review see<sup>4</sup>). As a consequence, the simple assessment of  
4  
5 expression levels during spermatogenesis can be misleading in the absence of a more detailed  
6  
7 knowledge on the regulation of isoforms and differences in the fates of products<sup>69</sup>, but also on  
8  
9 interactions between proteins and their involvement into functional complexes. Assigning one  
10  
11 or more biological functions to functionally unannotated proteins expressed in the human male  
12  
13 reproductive system can help to discover potential biomarkers for spermatogenesis and male  
14  
15 reproductive pathologies, among which are testicular cancers and infertilities.  
16  
17

18  
19  
20 How to prioritize these proteins and upon which criteria for a wet biologist is a non-trivial  
21  
22 question. Most testicular proteins with unknown function are expressed in the germ cell lineage,  
23  
24 and it may be difficult for scientists other than those with extensive experience on  
25  
26 spermatogenesis to access the relevant biological samples for further characterization of uPE1.  
27  
28 Importantly, access to human testes is limited and requires ethical authorizations. Furthermore,  
29  
30 reproductive biologists regret the impossibility of using some of the classic approaches of cell  
31  
32 and molecular biology. As a matter of fact, quality immortalized germ cell lines do not really  
33  
34 exist, especially because immortalization is not compatible with meiosis. For direct studies on  
35  
36 the germ cell lineage, it is therefore mandatory to use freshly purified germ cells from human  
37  
38 testis, taking into consideration that the more differentiated they are, the lower their survival  
39  
40 rate upon isolation. Moreover transfection of meiotic and post meiotic germ cells is still  
41  
42 challenging, if not impossible, making this approach unsuitable to study the function and  
43  
44 regulation of genes or gene products. Regarding the 85% of proteins that have orthologs in  
45  
46 rodents, a reasonable option is to use mouse KO models. The Mouse Genome Informatics  
47  
48 (MGI) database<sup>70,71</sup> is a reliable resource for narrowing down candidate genes, the inactivation  
49  
50 of which affects male fertility. As of June 2020, entries in the MGI mammalian phenotype  
51  
52 browser such as “male infertility”, “reduced male fertility” and “abnormal sperm” corresponded  
53  
54 respectively to 1280, 406 and 949 phenotypes. Of course, also of potential interest will be any  
55  
56  
57  
58  
59  
60

1  
2  
3 protein identified in a protein–protein interaction experiment for which the referenced partner  
4  
5 in the MGI knowledgebase would match with an infertility phenotype. Fortunately, we can rely  
6  
7 on the International Mouse Phenotyping Consortium (IMPC;  
8  
9 <https://www.mousephenotype.org>), a global effort to identify the function of every protein-  
10  
11 coding gene in the mouse genome<sup>72</sup> to order animals at reasonable cost. At best, mouse strains  
12  
13 generated by the IMPC can be obtained as a conditional ready allele that allows tissue restricted  
14  
15 knockouts via either the Cre-Lox or FLP-FRT recombination systems. In order to perform a  
16  
17 thorough analysis of a reproductive phenotype, a large colony must be obtained so as to get  
18  
19 about 30-40 knockout males reaching adulthood. Parameters to be measured will be fertility in  
20  
21 terms of sexual behavior and mating outcome, LH, FSH and testosterone concentrations in  
22  
23 serum, reproductive organ weights and anatomy, testicular quantitative histology, sperm  
24  
25 reserves, and if sperm is produced, a complete evaluation of sperm quantitative and qualitative  
26  
27 traits (concentration, morphology, mobility...). A complete integration of the data will  
28  
29 conclude whether the candidate gene is an essential player in the control of male reproductive  
30  
31 function in the mouse. However, the cost and the timeline of such an approach have significant  
32  
33 limitations that only allow a handful of candidate genes/proteins to be functionally studied by  
34  
35 a single laboratory. As soon as testicular germ cell proteins focus attention, uPE1 which are of  
36  
37 most interest are those linked to centrosomes, cilia and acrosome development that are likely  
38  
39 to be involved in the formation and functioning of the sperm flagella, the acrosome biogenesis,  
40  
41 or the acrosome reaction. These uPE1 proteins potentially represent quality traits and could help  
42  
43 in delineating sperm quality apart from semen analysis, especially in the case of idiopathic male  
44  
45 infertility. Novel information on ciliary or acrosomal proteins will help understand sperm  
46  
47 biology and idiopathic infertilities, ultimately offering perspectives on the counselling of male  
48  
49 patients about their fertility status and about possible clinical treatments required for  
50  
51 conceiving.  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 This is typically illustrated by the case of C2orf62. This protein is a former uPE1 protein  
4 produced in large amounts in the testes of zebrafish, rat and human that was shown to regulate  
5 actin polymerization and ciliogenesis and proposed to be involved in processes occurring before  
6 or during the formation of mature spermatozoa<sup>73</sup>. Recently a recessive mutation of C2orf62  
7 (CATIP) was described in humans that may contribute to asthenozoospermia through its  
8 involvement in actin polymerisation and on the actin cytoskeleton in sperm<sup>74</sup>, confirming the  
9 study published six years before.

10  
11 Here we implemented a knowledge-driven approach assisted by computing tools that allowed  
12 us to propose functional hypotheses related to male reproduction physiology (ciliogenesis,  
13 spermatogenesis and sperm maturation) for five uPE1 (Table 3). Based on the above data,  
14 CCDC112, TEX9, C3orf62 and CCDC183 are supposed to be involved in the biology of mature  
15 germ cells. Their further study using knockout mice models of ortholog genes is relevant.  
16 Through a registered account, the IMPC website offers the possibility to indicate interest in a  
17 gene and get regular updates on its status (e.g., mouse production, phenotyping in process,  
18 phenotype data available). Interestingly, for all four candidates embryonic stem (ES) cells have  
19 already been produced by IMPC consortium partners. When mice are produced, phenotyping  
20 is organized into pipelines, which assess the mouse biology at different life stages from embryo  
21 to aged adults. A broad range of phenotyping tests are carried out among which that of the  
22 reproductive system. However, fertility phenotyping is gross since it is understandable that a  
23 consortium like IMPC cannot go into an in-depth evaluation of a testicular phenotype, as  
24 reproductive biologists would do. The following step is thus to order the mice and develop a  
25 colony to produce enough KO males in house to evaluate the role of these four genes into the  
26 biology of germ cells. The role of the MAGEB3, MAGEB4, MAGEB6, MAGEA1 and ROCK1  
27 network in gonocytes differentiation could be advantageously studied using organotypic culture  
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 of mouse fetal gonads<sup>75</sup>, a model successively developed to get new insights into fetal testis  
4  
5 development.  
6  
7

8 The functional characterization of human proteins could be accelerated by a better integration  
9  
10 of existing knowledge resources and by initiating large collaborative projects involving  
11  
12 specialists of different biology fields. To this end, bioinformatics coupled with biocuration is  
13  
14 more than ever required to bridge annotation gaps and we hope this activity will receive more  
15  
16 attention and support from the community. As we believe that this effort must be continued  
17  
18 collectively, all results presented herein have been shared via ProteoRE for researchers who  
19  
20 wish to reuse them to guide further experimentation about the possible function of these proteins  
21  
22 highly enriched in the human male reproductive system.  
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

**TABLES:****Table 1:** List of uPE1 entries with elevated expression in human male tissues for which localization and/or functional characterization papers were found. (Chr.: Human chromosome number).

Gene	NextprotID	Year of publication	Literature	Main conclusion	Functional role	Chr.	RNA tissue specific NX
ANKRD55	NX_Q5TZF3	2019	<a href="#">PMID: 31208154</a>	"Ankrd45 is required for cell proliferation in both zebrafish embryos and mammalian cells."	Cell proliferation	1	brain: 14.6;fallopian tube: 24.8;parathyroid gland: 16.8;pituitary gland: 7.9;testis: 20.8
CCDC38	NX_Q502W7	2016	PMID: 27278724	"Ccde38 is mainly localized in the nuclei of spermatogonia and spermatocytes of the mice seminiferous tubules and interacted with ubiquitinated histone H2A in mouse testis."	Spermatogenesis	12	testis: 37.0
CFAP58 (CCDC147)	NX_Q5T655	2020	<a href="#">PMID: 31904090</a>	"Cfap58 is a testis-enrichment protein that exhibits similar localization to Odf2/Cenexin proteins and is required for the elongation of the primary cilium and sperm midpiece via modulation of the Notch signaling pathway."	Cilia/Flagella elongation	10	fallopian tube: 10.9;testis: 17.5
CFAP70	NX_Q5T0N1	2019	<a href="#">PMID: 31621862</a>	"Loss of CFAP70 function causes MMAF (multiple morphological abnormalities of the sperm flagella) and that ODA-related proteins may be crucial for the assembly and/or stability of the flagellum axoneme in addition to its motility."	Flagellum Assembly	10	pituitary gland: 27.2;testis: 22.3
C3orf62	NX_Q6ZUJ4	2020	PMID: 31930642	"We report that the testis-specific gene 1700102P08Rik is highly expressed in spermatocytes...we show that the disruption of 1700102P08Rik causes male infertility as a result of spermatocyte arrest."	Spermatogenesis	3	blood: 25.0;testis: 76.9
TEX33	NX_Q43247	2018	<a href="#">PMID: 30381129</a>	"Tex33 gene was expressed in spermatids and sperms of adult mouse testis. And Tex33 was located on the acrosome and flagellum of spermatozoa."	Unknown	22	testis: 49.1



**Table 2:** hu.MAP protein complexes containing uPE1. RNA tissue specificity: according to HPA data; Tissue RNA testis [NX]: normalized expression value in testis tissue according to HPA; Reliability score (IHC): score associated to immunohistochemistry data (detailed information in Table S2). \*: uPE1 protein entries.

ANKRD55 protein complex						
Gene Name (official symbol)	NextProt_ID	RNA tissue specificity	Tissue RNA testis [NX]	Reliability score (IHC)	Subcellular location (HPA)	Subcellular location (neXtProt)
ANKRD55*	NX_Q3KP44	Group enriched	16.7	Uncertain	Nucleoplasm,Nuclear speckles,Cytosol	NA
IFT27	NX_Q9BW83	Low tissue specificity	20.1	Uncertain	Nucleoplasm,Mitochondria	Nucleoplasm;Cytoplasm;Mitochondrion;Flagellum;Cilium
IFT52	NX_Q9Y366	Low tissue specificity	36.8	Enhanced	Cytosol	Cytosol;Cilium
TTC30A	NX_Q86WT1	Tissue enhanced	24.5	Supported	Nucleoplasm,Centrosome	Cilium
TTC30B	NX_Q8N4P2	Low tissue specificity	11.5	Supported	Nucleoplasm,Centrosome	Cilium
CCDC112 protein complex						
Gene Name (official symbol)	NextProt_ID	RNA tissue specificity	Tissue RNA testis [NX]	Reliability score (IHC)	Subcellular location (HPA)	Subcellular location (neXtProt)
CCDC112*	NX_Q8NEF3	Tissue enhanced	40.4	Uncertain	Nucleoplasm,Plasma membrane,Cytosol	Nucleoplasm;Cytosol;Cell membrane
TEX9*	NX_Q8N6V9	Tissue enhanced	19.3	Uncertain	Plasma membrane,Cytosol	Cytosol;Cell membrane
NOL4	NX_O94818	Group enriched	14.5	NA	Nucleoplasm	Nucleolus;Nucleus
NOL4L	NX_Q96MY1	Low tissue specificity	33.2	Enhanced	Nucleoplasm	Nucleoplasm
CCDC183 protein complex						
Gene Name (official symbol)	NextProt_ID	RNA tissue specificity	Tissue RNA testis [NX]	Reliability score (IHC)	Subcellular location (HPA)	Subcellular location (neXtProt)

CCDC183*	NX_Q5T5S1	Tissue enriched	26.4	Uncertain	NA	NA
FOXP1	NX_Q9H334	Low tissue specificity	11.7	Supported	Nucleoplasm	Nucleoplasm;Nucleus
RABEP1	NX_Q15276	Low tissue specificity	11.8	Supported	Vesicles	Cytoplasmic vesicle;Cytoplasm;Recycling endosome;Early endosome
RABGEF1	NX_Q9UJ41	Low tissue specificity	21.5	NA	Nucleoli,Cytosol	Cytosol;Cytoplasm;Recycling endosome;Nucleolus;Early endosome
RUNDC1	NX_Q96C34	Low tissue specificity	11.9	Uncertain	Nuclear speckles,Cytosol	Cytosol;Nucleus speckle
TMF1	NX_P82094	Low tissue specificity	20.9	Approved	Golgi apparatus	Nucleus;Golgi apparatus membrane;Golgi apparatus;Cytoplasm
C14orf80	NX_Q86SX3	Low tissue specificity	9.1	Uncertain	Nucleoplasm,Golgi apparatus,Cytosol	Cytosol;Nucleoplasm;Golgi apparatus
CEBPA	NX_P49715	Tissue enhanced	1.2	Approved	Nucleoplasm,Vesicles	Cytoplasmic vesicle;Nucleoplasm;Nucleolus;Nucleus
<b>C3orf62 protein complex</b>						
Gene Name (official symbol)	NextProt_ID	RNA tissue specificity	Tissue RNA testis [NX]	Reliability score (IHC)	Subcellular location (HPA)	Subcellular location (neXtProt)
C3orf62*	NX_Q6ZUJ4	Group enriched	76.9	NA	Nuclear speckles	Nucleus speckle
EP300	NX_Q09472	Low tissue specificity	20.6	Supported	Nucleoplasm	Nucleoplasm;Cytoplasm;Nucleus;Chromosome
<b>MAGEB3/B4/B6 proteins complex</b>						
Gene Name (official symbol)	NextProt_ID	RNA tissue specificity	Tissue RNA testis [NX]	Reliability score (IHC)	Subcellular location (HPA)	Subcellular location (neXtProt)
MAGEA1	NX_P43355	Tissue enriched	13.5	Enhanced	Nuclear and cytoplasmic expression of spermatogonia in seminiferous ducts in testis.	Cytosol;Cytoplasm;Nucleus
MAGEB3*	NX_O15480	Group enriched	7.8	Approved	Nuclear expression in testis and epididymis	NA
MAGEB4*	NX_O15481	Group enriched	18.1	Enhanced	Cytoplasmic expression in testis and epididymis	Cytoplasm

MAGEB6*	NX_Q8N7X4	Tissue enriched	8.6	Enhanced	Cytoplasmic expression in testis, mainly in spermatocytes	NA
ROCK1	NX_Q13464	Low tissue specificity	6.6	Approved	NA	Ruffle;Cytoplasm;Cell membrane;Centriole;Golgi apparatus membrane;Bleb;Lamellipodium;Cytoskeleton

**Table 3:** Functional hypotheses on five uPE1 related to the male reproductive system. (Chr.: Human chromosome number; Reliability: score; RNA in each male tissue expressed in normalized expression value (NX) from HPA).

Gene Name	NextprotID	Chr	Supporting literature (PMID)	Functional hypothesis	Antibody	Reliability (IH)	Subcellular location (neXtProt)	RNA ductus deferens	RNA epididymis	RNA prostate	RNA seminal vesicle	RNA testis
CCDC112	NX_Q8NEF3	5	26638075	Regulation of cilia and flagella formation, centriolar satellite protein	HPA045120, HPA050621	Uncertain	Nucleoplasm;Cytosol;Cell membrane	4.3	7.3	7.7	8.6	40.4
TEX9	NX_Q8N6V9	15	26638075	Regulation of cilia and flagella formation, centriolar satellite protein	HPA039415	Uncertain	Cytosol;Cell membrane	6.0	11.6	5.4	4.5	19.3
CCDC183	NX_Q5T5S1	9	<a href="#">29228106</a>	Spermiogenesis, acrosome formation, sperm motility	HPA043812	Uncertain	N.A.	0.0	0.4	0.6	0.0	26.4
C3orf62	NX_Q6ZUJ4	3	<a href="#">31930642</a>	Chromatin remodelling, male germ cells differentiation	HPA043328	N.A.	Nucleus speckles	4.4	4.9	4.1	2.2	76.9
MAGEB4	NX_O15481	X	<a href="#">10706124</a>	Male germ cells differentiation	HPA030456	Enhanced	Cytoplasm	0.8	26.3	0.0	0.0	18.1

**FIGURES:**

**Figure 1:** Flowchart illustrating the strategy used to establish the functionally unannotated human male tissue proteome and to build functional hypotheses.

**Figure 2:** Venn diagram showing the number of genes with an elevated expression in each human male tissue (testis, epididymis, ductus deferens, seminal vesicle and prostate) and the number of shared genes between each of them.

**Figure 3.** Venn diagram illustrating overlap between the set of proteins showing an elevated expression in the five human male tissue (testis, epididymis, ductus deferens, seminal vesicle and prostate), the set of unknown and missing protein entries from neXtProt (February 2020).

**Figure 4.** uPE1 with an elevated expression in the human male tissues mapped to the protein complex map hu.MAP. uPE1 were highlighted (hexagon nodes represent uPE1 proteins, while ellipse nodes represent their interactors ). Normalized expression (NX) values in testis are represented along a color gradient from white (no expression) to dark red (highest expression value). Gradient edge width is proportional to the hu.MAP interaction score (the thicker the edge, the higher the score).

**Figure 5.** A detailed view of uPE1 proteins interacting molecular assemblies. A. ANKRD55. B. CCDC183. C. CCDC112. D. C3orf62. E. MAGEB3, MAGEB4, MAGEB6 (graphical legends as in Figure 3).

**ASSOCIATED CONTENT****Supporting Information.**

**Table S1: List of 412 uPE1 neXtprot entries with an elevated expression in the human male reproductive system (detailed information) (XLSX).** Official gene name (gene), Ensembl identifier (Ensembl) was converted into a neXtProt identifier (NextprotID) using

1  
2  
3 mapping tables from neXtProt ([ftp://ftp.nextprot.org/pub/current\\_release/mapping/](ftp://ftp.nextprot.org/pub/current_release/mapping/) - March  
4 2020). Protein class(es) of the gene product (Protein class), RNA tissue specificity, RNA tissue  
5 distribution, consensus normalized expression (NX) value for each gene (RNA tissue specific  
6 NX), antibody product name (Antibody), reliability score of each protein expression pattern  
7 measured by IHC (Reliability (IH)), normalized transcript abundance measured in each male  
8 tissue considered in this study (Tissue RNA - ductus deferens [NX] (column J), Tissue RNA -  
9 epididymis [NX] (column K), Tissue RNA - prostate [NX] (column L), Tissue RNA - seminal  
10 vesicle [NX] (column M), Tissue RNA - testis [NX] (column N)) were retrieved from HPA.  
11 Protein sequence length (SeqLength), molecular weight MW), isoelectric point (IsoPoint),  
12 number of transmembrane domain (TMDomains), chromosome location (Chr), subcellular  
13 location (SubcellLocations (neXtProt Gold/Silver)), and disease (Diseases) information were  
14 retrieved from neXtProt.  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

31 **Table S2: The five hu.MAP protein complexes containing uPE1 with an expression**  
32 **elevated in male tissues presented in this study (detailed information) (XLSX).** Gene Name  
33 (official symbol), chromosome location (Chr), neXtProt identifier (NextprotID), protein name  
34 (Protein name), subcellular location (SubcellLocations (neXtProt)), entry name of the protein  
35 (Entry.name), (Protein.names), description of protein function (Function.[CC]), information  
36 about the protein quaternary structure and interaction(s) with other proteins or protein  
37 complexes (Subunit.structure.[CC]) were retrieved from neXtProt and Uniprot; RNA tissue  
38 specificity, RNA tissue distribution, consensus normalized expression (NX) value for each gene  
39 (RNA tissue specific NX), RNA cancer specificity, RNA cancer distribution, antibody product  
40 name (Antibody), reliability score of each protein expression pattern measured by IHC  
41 (Reliability (IH)), normalized transcript abundance measured in each male tissue considered in  
42 this study (Tissue RNA - ductus deferens [NX], Tissue RNA - epididymis [NX], Tissue RNA  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 - prostate [NX], Tissue RNA - seminal vesicle [NX], Tissue RNA - testis [NX]) were retrieved  
4  
5 from HPA  
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

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: yves.vandenbrouck@cea.fr

### ORCID

Yves Vandenbrouck: 0000-0002-1292-373X

Charles Pineau: 0000-0002-7461-5433

Lydie Lane: 0000-0002-9818-3030

### Notes

The authors declare no competing financial interest.

### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

## ACKNOWLEDGEMENTS

This work was partly supported by the “Investissement d’Avenir Infrastructures Nationales en Biologie et Santé” grants ANR-10-INBS-08 (Proteomics French Infrastructure - ProFI), ANR-11-INBS-0013 (French Institute of Bioinformatics - IFB) (YV) and by the Swiss State Secretariat for Education, Research and Innovation SERI funding to the SIB Swiss Institute of Bioinformatics (LL). This work was also supported by grants from Biogenouest, IBiSA and the Conseil Régional de Bretagne (CP).

## ABBREVIATIONS

CTA, Cancer/testis antigens; HPA, Human Protein Atlas; HPP, Human Proteome Project, IHC, immunohistochemistry; KO, knockout; NP, neXtProt; NX, normalized expression; uPE1, unknown protein evidence 1.



## REFERENCES

- 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) Paik, Y.-K.; Lane, L.; Kawamura, T.; Chen, Y.-J.; Cho, J.-Y.; LaBaer, J.; Yoo, J. S.; Domont, G.; Corrales, F.; Omenn, G. S.; Archakov, A.; Encarnación-Guevara, S.; Lui, S.; Salekdeh, G. H.; Cho, J.-Y.; Kim, C.-Y.; Overall, C. M. Launching the C-HPP neXt-CP50 Pilot Project for Functional Characterization of Identified Proteins with No Known Function. *J. Proteome Res.* **2018**, *17* (12), 4042–4050.

(2) Zahn-Zabal, M.; Michel, P.-A.; Gateau, A.; Nikitin, F.; Schaeffer, M.; Audot, E.; Gaudet, P.; Duek, P. D.; Teixeira, D.; Rech de Laval, V.; Samarasinghe, K.; Bairoch, A.; Lane, L. The NeXtProt Knowledgebase in 2020: Data, Tools and Usability Improvements. *Nucleic Acids Res.* **2019**. <https://doi.org/10.1093/nar/gkz995>.

(3) Zahn-Zabal, M.; Lane, L. What Will NeXtProt Help Us Achieve in 2020 and Beyond? *Expert Rev. Proteomics* **2020**, *17* (2), 95–98.

(4) Rolland, A. D.; Lavigne, R.; Dauly, C.; Calvel, P.; Kervarrec, C.; Freour, T.; Evrard, B.; Rioux-Leclercq, N.; Auger, J.; Pineau, C. Identification of Genital Tract Markers in the Human Seminal Plasma Using an Integrative Genomics Approach. *Hum. Reprod.* **2013**, *28* (1), 199–209.

(5) Uhlén, M.; Fagerberg, L.; Hallström, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; Olsson, I.; Edlund, K.; Lundberg, E.; Navani, S.; Szigyaró, C. A.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Edqvist, P.; Berling, H.; Tegel, H.; Mulder, J.; Rockberg, J.; Nilsson, P.; Schwenk, J. M.; Hamsten, M.; Feilitzén, K. Von; Forsberg, M.; Persson, L.; Johansson, F.; Zwahlen, M.; Heijne, G. Von; Nielsen, J.; Pontén, F. Tissue-Based Map of the Human Proteome. *Science* (80-. ). **2015**, *347* (6220), 1260419–1260419.

(6) Vandenbrouck, Y.; Lane, L.; Carapito, C.; Duek, P.; Rondel, K.; Bruley, C.; Macron, C.; Gonzalez de Peredo, A.; Couté, Y.; Chaoui, K.; Com, E.; Gateau, A.; Hesse, A.-M.; Marcellin, M.; Méar, L.; Mouton-Barbosa, E.; Robin, T.; Burlet-Schiltz, O.; Cianferani, S.; Ferro, M.; Fréour, T.; Lindskog, C.; Garin, J.; Pineau, C. Looking for Missing Proteins in the Proteome of Human Spermatozoa: An Update. *J. Proteome Res.* **2016**, *15* (11), 3998–4019.

(7) Carapito, C.; Duek, P.; Macron, C.; Seffals, M.; Rondel, K.; Delalande, F.; Lindskog, C.; Fréour, T.; Vandenbrouck, Y.; Lane, L.; Pineau, C. Validating Missing Proteins in Human Sperm Cells by Targeted Mass-Spectrometry- and Antibody-Based Methods. *J. Proteome Res.* **2017**, *16* (12), 4340–4351.

(8) 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. *J. Proteome Res.* **2016**, *15* (11), 3988–3997.

(9) Jalili, V.; Afgan, E.; Gu, Q.; Clements, D.; Blankenberg, D.; Goecks, J.; Taylor, J.; Nekrutenko, A. Corrigendum: The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2020 Update. *Nucleic Acids Res.* **2020**, *48* (14), 8205–8207.

(10) Nguyen, L.; Brun, V.; Combes, F.; Loux, V.; Vandenbrouck, Y. Designing an In Silico Strategy to Select Tissue-Leakage Biomarkers Using the Galaxy Framework. *Methods Mol. Biol.* **2019**, *1959*, 275–289.

- 1  
2  
3 (11) Vandembrouck, Y.; Christiany, D.; Combes, F.; Loux, V.; Brun, V. Bioinformatics  
4 Tools and Workflow to Select Blood Biomarkers for Early Cancer Diagnosis: An  
5 Application to Pancreatic Cancer. *Proteomics* **2019**, *19* (21–22), e1800489.
- 6  
7 (12) Drew, K.; Lee, C.; Huizar, R. L.; Tu, F.; Borgeson, B.; McWhite, C. D.; Ma, Y.;  
8 Wallingford, J. B.; Marcotte, E. M. Integration of over 9,000 Mass Spectrometry  
9 Experiments Builds a Global Map of Human Protein Complexes. *Mol. Syst. Biol.* **2017**,  
10 *13* (6), 932.
- 11  
12 (13) Thul, P. J.; Åkesson, L.; Wiking, M.; Mahdessian, D.; Geladaki, A.; Ait Blal, H.; Alm,  
13 T.; Asplund, A.; Björk, L.; Breckels, L. M.; Bäckström, A.; Danielsson, F.; Fagerberg,  
14 L.; Fall, J.; Gatto, L.; Gnann, C.; Hober, S.; Hjelmare, M.; Johansson, F.; Lee, S.;  
15 Lindskog, C.; Mulder, J.; Mulvey, C. M.; Nilsson, P.; Oksvold, P.; Rockberg, J.;  
16 Schutten, R.; Schwenk, J. M.; Sivertsson, Å.; Sjöstedt, E.; Skogs, M.; Stadler, C.;  
17 Sullivan, D. P.; Tegel, H.; Winsnes, C.; Zhang, C.; Zwahlen, M.; Mardinoglu, A.;  
18 Pontén, F.; von Feilitzen, K.; Lilley, K. S.; Uhlén, M.; Lundberg, E. A Subcellular Map  
19 of the Human Proteome. *Science* **2017**, *356* (6340), eaal3321.
- 20  
21 (14) de Lamirande, E. Semenogelin, the Main Protein of the Human Semen Coagulum,  
22 Regulates Sperm Function. *Semin. Thromb. Hemost.* **2007**, *33* (1), 60–68.
- 23  
24 (15) Iwamori, T.; Iwamori, N.; Matsumoto, M.; Ono, E.; Matzuk, M. M. Identification of  
25 KIAA1210 as a Novel X-Chromosome-Linked Protein That Localizes to the Acrosome  
26 and Associates with the Ectoplasmic Specialization in Testes. *Biol. Reprod.* **2017**, *96*  
27 (2), 469–477.
- 28  
29 (16) O'Donnell, L. Mechanisms of Spermiogenesis and Spermiation and How They Are  
30 Disturbed. *Spermatogenesis* **2014**, *4* (2), e979623.
- 31  
32 (17) Thomas, G.; Jacobs, K. B.; Yeager, M.; Kraft, P.; Wacholder, S.; Orr, N.; Yu, K.;  
33 Chatterjee, N.; Welch, R.; Hutchinson, A.; Crenshaw, A.; Cancel-Tassin, G.; Staats, B.  
34 J.; Wang, Z.; Gonzalez-Bosquet, J.; Fang, J.; Deng, X.; Berndt, S. I.; Calle, E. E.;  
35 Feigelson, H. S.; Thun, M. J.; Rodriguez, C.; Albanes, D.; Virtamo, J.; Weinstein, S.;  
36 Schumacher, F. R.; Giovannucci, E.; Willett, W. C.; Cussenot, O.; Valeri, A.; Andriole,  
37 G. L.; Crawford, E. D.; Tucker, M.; Gerhard, D. S.; Fraumeni, J. F.; Hoover, R.; Hayes,  
38 R. B.; Hunter, D. J.; Chanock, S. J. Multiple Loci Identified in a Genome-Wide  
39 Association Study of Prostate Cancer. *Nat. Genet.* **2008**, *40* (3), 310–315.
- 40  
41 (18) Jumeau, F.; Com, E.; Lane, L.; Duek, P.; Lagarrigue, M.; Lavigne, R.; Guillot, L.;  
42 Rondel, K.; Gateau, A.; Melaine, N.; Guével, B.; Sergeant, N.; Mitchell, V.; Pineau, C.  
43 Human Spermatozoa as a Model for Detecting Missing Proteins in the Context of the  
44 Chromosome-Centric Human Proteome Project. *J. Proteome Res.* **2015**, *14* (9), 3606–  
45 3620.
- 46  
47 (19) Wei, W.; Luo, W.; Wu, F.; Peng, X.; Zhang, Y.; Zhang, M.; Zhao, Y.; Su, N.; Qi, Y.;  
48 Chen, L.; Zhang, Y.; Wen, B.; He, F.; Xu, P. Deep Coverage Proteomics Identifies  
49 More Low-Abundance Missing Proteins in Human Testis Tissue with Q-Exactive HF  
50 Mass Spectrometer. *J. Proteome Res.* **2016**, *15* (11), 3988–3997.
- 51  
52 (20) Zhang, Y.; Li, Q.; Wu, F.; Zhou, R.; Qi, Y.; Su, N.; Chen, L.; Xu, S.; Jiang, T.; Zhang,  
53 C.; Cheng, G.; Chen, X.; Kong, D.; Wang, Y.; Zhang, T.; Zi, J.; Wei, W.; Gao, Y.;  
54 Zhen, B.; Xiong, Z.; Wu, S.; Yang, P.; Wang, Q.; Wen, B.; He, F.; Xu, P.; Liu, S.  
55 Tissue-Based Proteogenomics Reveals That Human Testis Endows Plentiful Missing  
56 Proteins. *J. Proteome Res.* **2015**, *14* (9), 3583–3594.
- 57  
58  
59  
60

- 1  
2  
3 (21) Duek, P.; Gateau, A.; Bairoch, A.; Lane, L. Exploring the Uncharacterized Human  
4 Proteome Using NeXtProt. *J. Proteome Res.* **2018**, *17* (12), 4211–4226.  
5 <https://doi.org/10.1021/acs.jproteome.8b00537>.  
6
- 7 (22) Dacheux, J.-L.; Dacheux, F. New Insights into Epididymal Function in Relation to  
8 Sperm Maturation. *REPRODUCTION* **2014**, *147* (2), R27–R42.  
9
- 10 (23) Melaine, N.; Com, E.; Bellaud, P.; Guillot, L.; Lagarrigue, M.; Morrice, N. A.; Guével,  
11 B.; Lavigne, R.; Velez de la Calle, J.-F.; Dojahn, J.; Pineau, C. Deciphering the Dark  
12 Proteome: Use of the Testis and Characterization of Two Dark Proteins. *J. Proteome*  
13 *Res.* **2018**, *17* (12), 4197–4210.  
14
- 15 (24) Xia, M.; Xia, J.; Yang, D.; Liu, M.; Niu, C.; Shen, X.; Sun, H.; Zheng, Y. [Preparation  
16 and Application of Rabbit Polyclonal Antibody against Mouse Tex33]. *Chinese J. Cell.*  
17 *Mol. Immunol.* **2018**, *34* (7), 643–649.  
18
- 19 (25) Lin, S.-R.; Li, Y.-C.; Luo, M.-L.; Guo, H.; Wang, T.-T.; Chen, J.-B.; Ma, Q.; Gu, Y.-  
20 L.; Jiang, Z.-M.; Gui, Y.-T. Identification and Characteristics of the Testes-Specific  
21 Gene, Ccdc38, in Mice. *Mol. Med. Rep.* **2016**, *14* (2), 1290–1296.  
22
- 23 (26) Beurois, J.; Martinez, G.; Cazin, C.; Kherraf, Z.-E.; Amiri-Yekta, A.; Thierry-Mieg,  
24 N.; Bidart, M.; Petre, G.; Satre, V.; Brouillet, S.; Touré, A.; Arnoult, C.; Ray, P. F.;  
25 Coutton, C. CFAP70 Mutations Lead to Male Infertility Due to Severe Astheno-  
26 Teratozoospermia. A Case Report. *Hum. Reprod.* **2019**, *34* (10), 2071–2079.  
27
- 28 (27) Li, Z.-Z.; Zhao, W.-L.; Wang, G.-S.; Gu, N.-H.; Sun, F. The Novel Testicular  
29 Enrichment Protein Cfap58 Is Required for Notch-Associated Ciliogenesis. *Biosci.*  
30 *Rep.* **2020**, *40* (1). <https://doi.org/10.1042/BSR20192666>.  
31
- 32 (28) Wu, X.-L.; Yun, D.-M.; Gao, S.; Liang, A.-J.; Duan, Z.-Z.; Wang, H.-S.; Wang, G.-S.;  
33 Sun, F. The Testis-Specific Gene 1700102P08Rik Is Essential for Male Fertility. *Mol.*  
34 *Reprod. Dev.* **2020**, *87* (2), 231–240.  
35
- 36 (29) Kang, Y.; Xie, H.; Zhao, C. Ankrd45 Is a Novel Ankyrin Repeat Protein Required for  
37 Cell Proliferation. *Genes (Basel)*. **2019**, *10* (6). <https://doi.org/10.3390/genes10060462>.  
38
- 39 (30) Taschner, M.; Lorentzen, E. The Intraflagellar Transport Machinery. *Cold Spring*  
40 *Harb. Perspect. Biol.* **2016**, *8* (10), a028092.  
41
- 42 (31) Chapple, C. E.; Robisson, B.; Spinelli, L.; Guien, C.; Becker, E.; Brun, C. Extreme  
43 Multifunctional Proteins Identified from a Human Protein Interaction Network. *Nat.*  
44 *Commun.* **2015**, *6* (May), 7412.  
45
- 46 (32) Deane, J. A.; Cole, D. G.; Seeley, E. S.; Diener, D. R.; Rosenbaum, J. L. Localization  
47 of Intraflagellar Transport Protein IFT52 Identifies Basal Body Transitional Fibers as  
48 the Docking Site for IFT Particles. *Curr. Biol.* **2001**, *11* (20), 1586–1590.  
49
- 50 (33) Fan, Z.-C.; Behal, R. H.; Geimer, S.; Wang, Z.; Williamson, S. M.; Zhang, H.; Cole, D.  
51 G.; Qin, H. Chlamydomonas IFT70/CrDYF-1 Is a Core Component of IFT Particle  
52 Complex B and Is Required for Flagellar Assembly. *Mol. Biol. Cell* **2010**, *21* (15),  
53 2696–2706.  
54
- 55 (34) Ou, G.; Koga, M.; Blacque, O. E.; Murayama, T.; Ohshima, Y.; Schafer, J. C.; Li, C.;  
56 Yoder, B. K.; Leroux, M. R.; Scholey, J. M. Sensory Ciliogenesis in *Caenorhabditis*  
57 *Elegans*: Assignment of IFT Components into Distinct Modules Based on Transport  
58 and Phenotypic Profiles. *Mol. Biol. Cell* **2007**, *18* (5), 1554–1569.  
59  
60

- 1  
2  
3 (35) De Lima Morais, D. A.; Fang, H.; Rackham, O. J. L.; Wilson, D.; Pethica, R.; Chothia,  
4 C.; Gough, J. SUPERFAMILY 1.75 Including a Domain-Centric Gene Ontology  
5 Method. *Nucleic Acids Res.* **2011**, *39* (SUPPL. 1), 427–434.  
6  
7 (36) Rose, A.; Meier, I. Scaffolds, Levers, Rods and Springs: Diverse Cellular Functions of  
8 Long Coiled-Coil Proteins. *Cell. Mol. Life Sci.* **2004**, *61* (16), 1996–2009.  
9  
10 (37) van Dam, T. J.; Wheway, G.; Slaats, G. G.; Huynen, M. A.; Giles, R. H. The  
11 SYSCILIA Gold Standard (SCGSv1) of Known Ciliary Components and Its  
12 Applications within a Systems Biology Consortium. *Cilia* **2013**, *2* (1), 7.  
13  
14 (38) Gupta, G. D.; Coyaud, É.; Gonçalves, J.; Mojarad, B. A.; Liu, Y.; Wu, Q.;  
15 Gheiratmand, L.; Comartin, D.; Tkach, J. M.; Cheung, S. W. T.; Bashkurov, M.;  
16 Hasegan, M.; Knight, J. D.; Lin, Z. Y.; Schueler, M.; Hildebrandt, F.; Moffat, J.;  
17 Gingras, A. C.; Raught, B.; Pelletier, L. A Dynamic Protein Interaction Landscape of  
18 the Human Centrosome-Cilium Interface. *Cell* **2015**, *163* (6), 1484–1499.  
19  
20 (39) Nigg, E. A.; Raff, J. W. Centrioles, Centrosomes, and Cilia in Health and Disease. *Cell*  
21 **2009**, *139* (4), 663–678.  
22  
23 (40) Tollenaere, M. A. X.; Mailand, N.; Bekker-Jensen, S. Centriolar Satellites: Key  
24 Mediators of Centrosome Functions. *Cell. Mol. Life Sci.* **2015**, *72* (1), 11–23.  
25  
26 (41) Odabasi, E.; Batman, U.; Firat-Karalar, E. N. Unraveling the Mysteries of Centriolar  
27 Satellites: Time to Rewrite the Textbooks about the Centrosome/Cilium Complex. *Mol.*  
28 *Biol. Cell* **2020**, *31* (9), 866–872.  
29  
30 (42) Avidor-Reiss, T.; Khire, A.; Fishman, E. L.; Jo, K. H. Atypical Centrioles during  
31 Sexual Reproduction. *Front. cell Dev. Biol.* **2015**, *3*, 21.  
32  
33 (43) Avidor-Reiss, T.; Mazur, M.; Fishman, E. L.; Sindhvani, P. The Role of Sperm  
34 Centrioles in Human Reproduction – The Known and the Unknown. *Front. Cell Dev.*  
35 *Biol.* **2019**, *7*. <https://doi.org/10.3389/fcell.2019.00188>.  
36  
37 (44) Lopes, C. A. M.; Prosser, S. L.; Romio, L.; Hirst, R. A.; O’Callaghan, C.; Woolf, A. S.;  
38 Fry, A. M. Centriolar Satellites Are Assembly Points for Proteins Implicated in Human  
39 Ciliopathies, Including Oral-Facial-Digital Syndrome 1. *J. Cell Sci.* **2011**, *124* (Pt 4),  
40 600–612.  
41  
42 (45) Tang, Z.; Lin, M. G.; Stowe, T. R.; Chen, S.; Zhu, M.; Stearns, T.; Franco, B.; Zhong,  
43 Q. Autophagy Promotes Primary Ciliogenesis by Removing OFD1 from Centriolar  
44 Satellites. *Nature* **2013**, *502* (7470), 254–257.  
45  
46 (46) Kodani, A.; Yu, T. W.; Johnson, J. R.; Jayaraman, D.; Johnson, T. L.; Al-Gazali, L.;  
47 Sztriha, L.; Partlow, J. N.; Kim, H.; Krup, A. L.; Dammermann, A.; Krogan, N. J.;  
48 Walsh, C. A.; Reiter, J. F. Centriolar Satellites Assemble Centrosomal Microcephaly  
49 Proteins to Recruit CDK2 and Promote Centriole Duplication. *Elife* **2015**, *4*.  
50 <https://doi.org/10.7554/eLife.07519>.  
51  
52 (47) Guo, J.; Grow, E. J.; Mlcochova, H.; Maher, G. J.; Lindskog, C.; Nie, X.; Guo, Y.;  
53 Takei, Y.; Yun, J.; Cai, L.; Kim, R.; Carrell, D. T.; Goriely, A.; Hotaling, J. M.; Cairns,  
54 B. R. The Adult Human Testis Transcriptional Cell Atlas. *Cell Res.* **2018**, *28* (12),  
55 1141–1157.  
56  
57 (48) Netherton, J. K.; Hetherington, L.; Ogle, R. A.; Velkov, T.; Baker, M. A. Proteomic  
58 Analysis of Good- and Poor-Quality Human Sperm Demonstrates That Several  
59

- 1  
2  
3 Proteins Are Routinely Aberrantly Regulated. *Biol. Reprod.* **2018**, *99* (2), 395–408.
- 4  
5 (49) Marley, A.; von Zastrow, M. A Simple Cell-Based Assay Reveals That Diverse  
6 Neuropsychiatric Risk Genes Converge on Primary Cilia. *PLoS One* **2012**, *7* (10),  
7 e46647.
- 8  
9 (50) Breslow, D. K.; Hoogendoorn, S.; Kopp, A. R.; Morgens, D. W.; Vu, B. K.; Kennedy,  
10 M. C.; Han, K.; Li, A.; Hess, G. T.; Bassik, M. C.; Chen, J. K.; Nachury, M. V. A  
11 CRISPR-Based Screen for Hedgehog Signaling Provides Insights into Ciliary Function  
12 and Ciliopathies. *Nat. Genet.* **2018**, *50* (3), 460–471.
- 13  
14 (51) Lerer-Goldshtein, T.; Bel, S.; Shpungin, S.; Pery, E.; Motro, B.; Goldstein, R. S.; Bar-  
15 Sheshet, S. I.; Breitbart, H.; Nir, U. TMF/ARA160: A Key Regulator of Sperm  
16 Development. *Dev. Biol.* **2010**, *348* (1), 12–21.
- 17  
18 (52) Berruti, G.; Paiardi, C. Acrosome Biogenesis: Revisiting Old Questions to Yield New  
19 Insights. *Spermatogenesis* **2011**, *1* (2), 95–98.
- 20  
21 (53) Abou-Haila, A.; Tulsiani, D. R. Mammalian Sperm Acrosome: Formation, Contents,  
22 and Function. *Arch. Biochem. Biophys.* **2000**, *379* (2), 173–182.
- 23  
24 (54) Elkis, Y.; Bel, S.; Rahimi, R.; Lerer-Goldstein, T.; Levin-Zaidman, S.; Babushkin, T.;  
25 Shpungin, S.; Nir, U. TMF/ARA160 Governs the Dynamic Spatial Orientation of the  
26 Golgi Apparatus during Sperm Development. *PLoS One* **2015**, *10* (12), e0145277.
- 27  
28 (55) Galvez, T.; Gilleron, J.; Zerial, M.; O’Sullivan, G. A. SnapShot: Mammalian Rab  
29 Proteins in Endocytic Trafficking. *Cell* **2012**, *151* (1), 234–234.e2.
- 30  
31 (56) Ward, W. S.; Coffey, D. S. DNA Packaging and Organization in Mammalian  
32 Spermatozoa: Comparison with Somatic Cells. *Biol. Reprod.* **1991**, *44* (4), 569–574.
- 33  
34 (57) Rousseaux, S.; Reynoird, N.; Escoffier, E.; Thevenon, J.; Caron, C.; Khochbin, S.  
35 Epigenetic Reprogramming of the Male Genome during Gametogenesis and in the  
36 Zygote. *Reprod. Biomed. Online* **2008**, *16* (4), 492–503.
- 37  
38 (58) McLay, D.; Clarke, H. Remodelling the Paternal Chromatin at Fertilization in  
39 Mammals. *Reproduction* **2003**, 625–633.
- 40  
41 (59) Ajduk, A.; Yamauchi, Y.; Ward, M. A. Sperm Chromatin Remodeling after  
42 Intracytoplasmic Sperm Injection Differs from That of In Vitro Fertilization I. *Biol.*  
43 *Reprod.* **2006**, *75* (3), 442–451.
- 44  
45 (60) Amaral, A.; Castillo, J.; Ramalho-Santos, J.; Oliva, R. The Combined Human Sperm  
46 Proteome: Cellular Pathways and Implications for Basic and Clinical Science. *Hum.*  
47 *Reprod. Update* **2014**, *20* (1), 40–62.
- 48  
49 (61) Tropberger, P.; Pott, S.; Keller, C.; Kamieniarz-Gdula, K.; Caron, M.; Richter, F.; Li,  
50 G.; Mittler, G.; Liu, E. T.; Bühler, M.; Margueron, R.; Schneider, R. Regulation of  
51 Transcription through Acetylation of H3K122 on the Lateral Surface of the Histone  
52 Octamer. *Cell* **2013**, *152* (4), 859–872.
- 53  
54 (62) Spector, D. L.; Lamond, A. I. Nuclear Speckles. *Cold Spring Harb. Perspect. Biol.*  
55 **2011**, *3* (2). <https://doi.org/10.1101/cshperspect.a000646>.
- 56  
57 (63) Phair, R. D.; Misteli, T. High Mobility of Proteins in the Mammalian Cell Nucleus.  
58 *Nature* **2000**, *404* (6778), 604–609.
- 59  
60

- 1  
2  
3 (64) Simpson, A. J. G.; Caballero, O. L.; Jungbluth, A.; Chen, Y.-T.; Old, L. J.  
4 Cancer/Testis Antigens, Gametogenesis and Cancer. *Nat. Rev. Cancer* **2005**, *5* (8),  
5 615–625.  
6  
7 (65) van der Bruggen, P.; Traversari, C.; Chomez, P.; Lurquin, C.; De Plaen, E.; Van den  
8 Eynde, B.; Knuth, A.; Boon, T. A Gene Encoding an Antigen Recognized by Cytolytic  
9 T Lymphocytes on a Human Melanoma. *Science* **1991**, *254* (5038), 1643–1647.  
10  
11 (66) Osterlund, C.; Töhönen, V.; Forslund, K. O.; Nordqvist, K. Mage-B4, a Novel  
12 Melanoma Antigen (MAGE) Gene Specifically Expressed during Germ Cell  
13 Differentiation. *Cancer Res.* **2000**, *60* (4), 1054–1061.  
14  
15 (67) Amano, M.; Nakayama, M.; Kaibuchi, K. Rho-Kinase/ROCK: A Key Regulator of the  
16 Cytoskeleton and Cell Polarity. *Cytoskeleton (Hoboken)*. **2010**, *67* (9), 545–554.  
17  
18 (68) Lui, W.-Y.; Lee, W. M.; Cheng, C. Y. Rho GTPases and Spermatogenesis. *Biochim.*  
19 *Biophys. Acta* **2003**, *1593* (2–3), 121–129.  
20  
21 (69) Com, E.; Melaine, N.; Chalmel, F.; Pineau, C. Proteomics and Integrative Genomics  
22 for Unraveling the Mysteries of Spermatogenesis: The Strategies of a Team. *J.*  
23 *Proteomics* **2014**, *107*, 128–143.  
24  
25 (70) Blake, J. A.; Bult, C. J.; Eppig, J. T.; Kadin, J. A.; Richardson, J. E.; Mouse Genome  
26 Database Group. The Mouse Genome Database Genotypes::Phenotypes. *Nucleic Acids*  
27 *Res.* **2009**, *37* (Database issue), D712–9.  
28  
29 (71) Eppig, J. T.; Blake, J. A.; Bult, C. J.; Kadin, J. A.; Richardson, J. E. The Mouse  
30 Genome Database (MGD): Comprehensive Resource for Genetics and Genomics of the  
31 Laboratory Mouse. *Nucleic Acids Res.* **2012**, *40* (D1), D881–D886.  
32  
33 (72) Dickinson, M. E.; Flenniken, A. M.; Ji, X.; Teboul, L.; Wong, M. D.; White, J. K.;  
34 Meehan, T. F.; Weninger, W. J.; Westerberg, H.; Adissu, H.; Baker, C. N.; Bower, L.;  
35 Brown, J. M.; Caddle, L. B.; Chiani, F.; Clary, D.; Cleak, J.; Daly, M. J.; Denegre, J.  
36 M.; Doe, B.; Dolan, M. E.; Edie, S. M.; Fuchs, H.; Gailus-Durner, V.; Galli, A.;  
37 Gambadoro, A.; Gallegos, J.; Guo, S.; Horner, N. R.; Hsu, C.-W.; Johnson, S. J.;  
38 Kalaga, S.; Keith, L. C.; Lanoue, L.; Lawson, T. N.; Lek, M.; Mark, M.; Marschall, S.;  
39 Mason, J.; McElwee, M. L.; Newbigging, S.; Nutter, L. M. J.; Peterson, K. A.;  
40 Ramirez-Solis, R.; Rowland, D. J.; Ryder, E.; Samocho, K. E.; Seavitt, J. R.; Selloum,  
41 M.; Szoke-Kovacs, Z.; Tamura, M.; Trainor, A. G.; Tudose, I.; Wakana, S.; Warren, J.;  
42 Wendling, O.; West, D. B.; Wong, L.; Yoshiki, A.; Wurst, W.; MacArthur, D. G.;  
43 Tocchini-Valentini, G. P.; Gao, X.; Flicek, P.; Bradley, A.; Skarnes, W. C.; Justice, M.  
44 J.; Parkinson, H. E.; Moore, M.; Wells, S.; Braun, R. E.; Svenson, K. L.; de Angelis,  
45 M. H.; Herault, Y.; Mohun, T.; Mallon, A.-M.; Henkelman, R. M.; Brown, S. D. M.;  
46 Adams, D. J.; Lloyd, K. C. K.; McKerlie, C.; Beaudet, A. L.; Bućan, M.; Murray, S. A.  
47 High-Throughput Discovery of Novel Developmental Phenotypes. *Nature* **2016**, *537*  
48 (7621), 508–514.  
49  
50 (73) Bontems, F.; Fish, R. J.; Borlat, I.; Lembo, F.; Chocu, S.; Chalmel, F.; Borg, J.-P.;  
51 Pineau, C.; Neerman-Arbez, M.; Bairoch, A.; Lane, L. C2orf62 and TTC17 Are  
52 Involved in Actin Organization and Ciliogenesis in Zebrafish and Human. *PLoS One*  
53 **2014**, *9* (1), e86476.  
54  
55 (74) Arafat, M.; Harlev, A.; Har-Vardi, I.; Levitas, E.; Priel, T.; Gershoni, M.; Searby, C.;  
56 Sheffield, V. C.; Lunenfeld, E.; Parvari, R. Mutation in CATIP (C2orf62) Causes  
57  
58  
59

1  
2  
3 Oligoteratoasthenozoospermia by Affecting Actin Dynamics. *J. Med. Genet.* **2020**,  
4 jmedgenet-2019-106825.  
5

- 6 (75) Livera, G.; Delbes, G.; Pairault, C.; Rouiller-Fabre, V.; Habert, R. Organotypic  
7 Culture, a Powerful Model for Studying Rat and Mouse Fetal Testis Development. *Cell*  
8 *Tissue Res.* **2006**, 324 (3), 507–521.  
9

TOC graphic

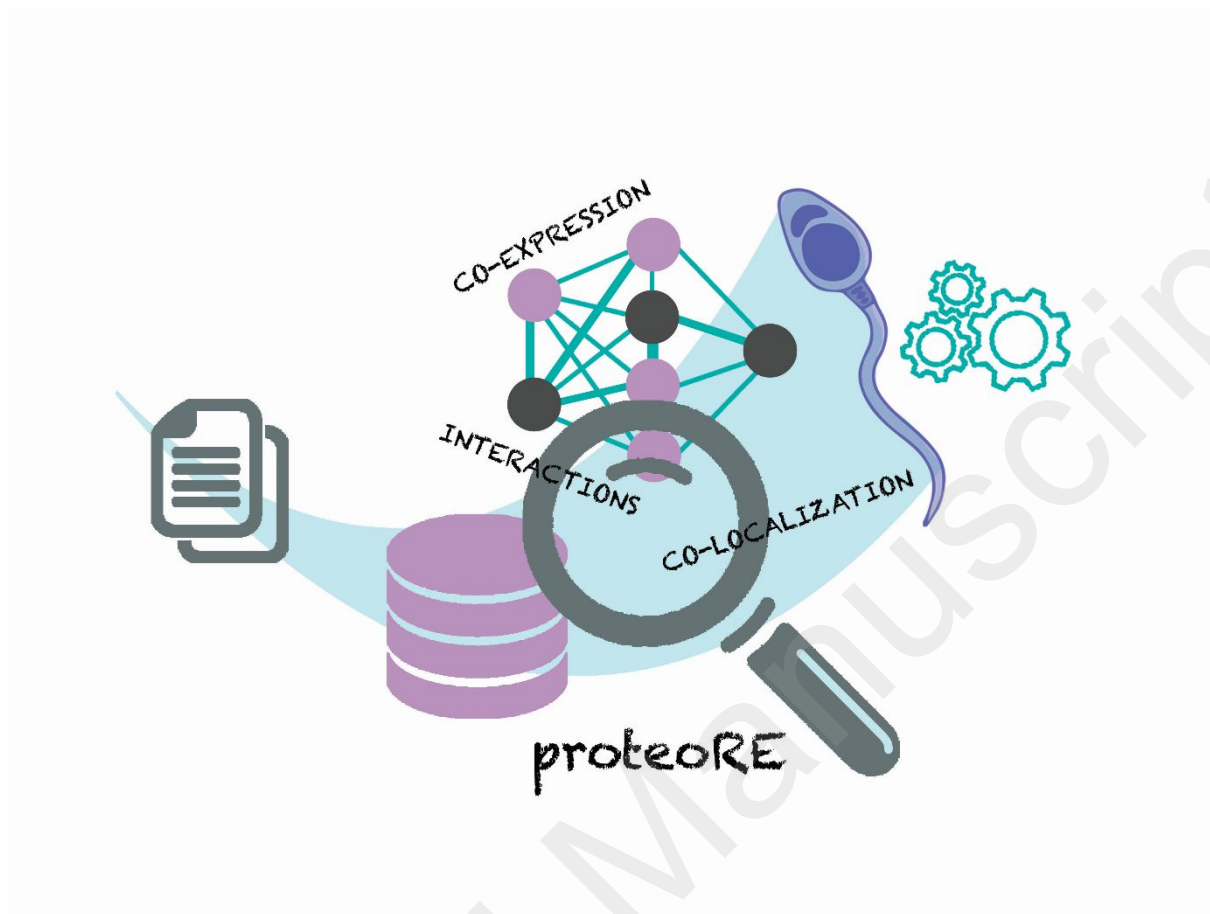
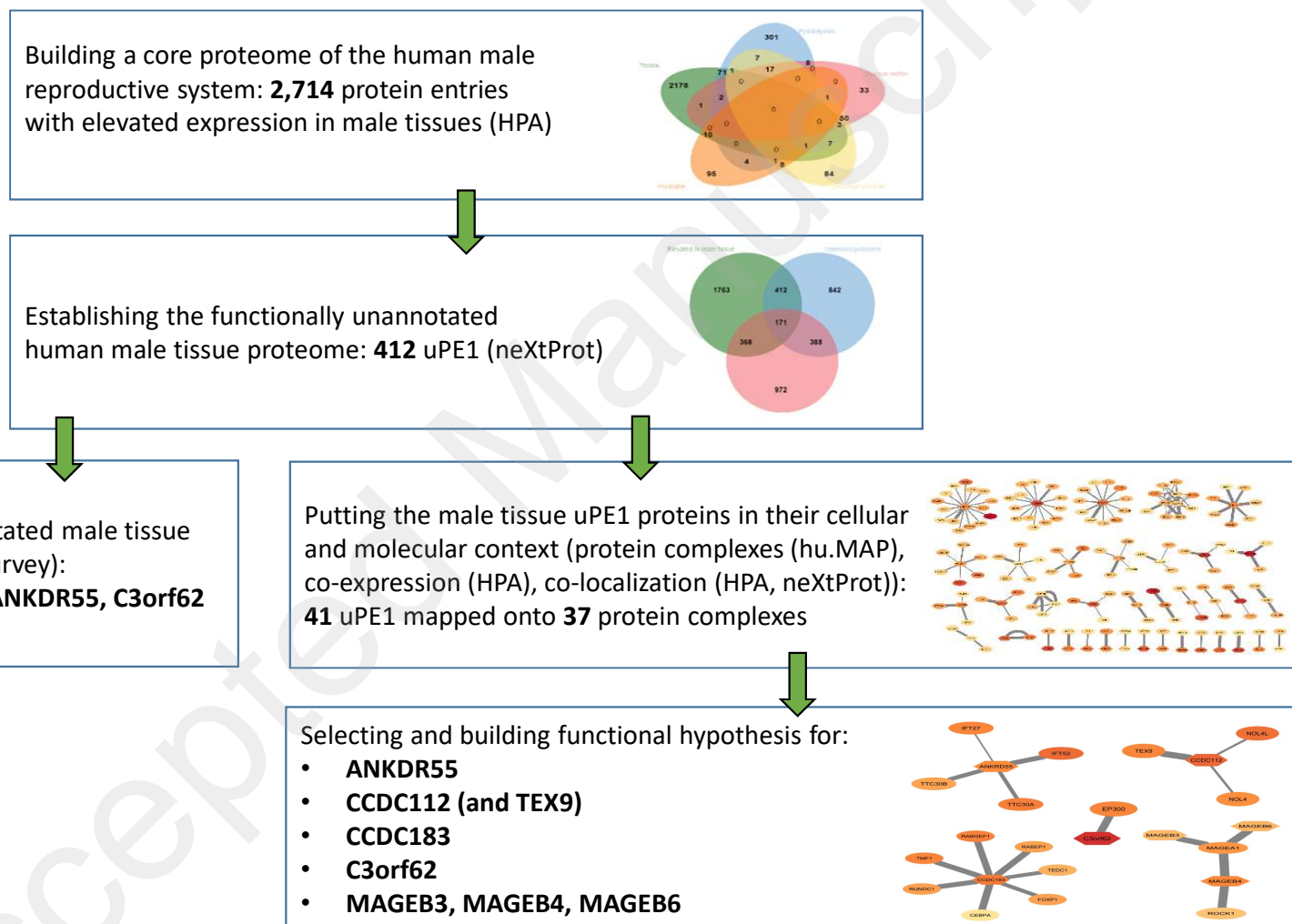


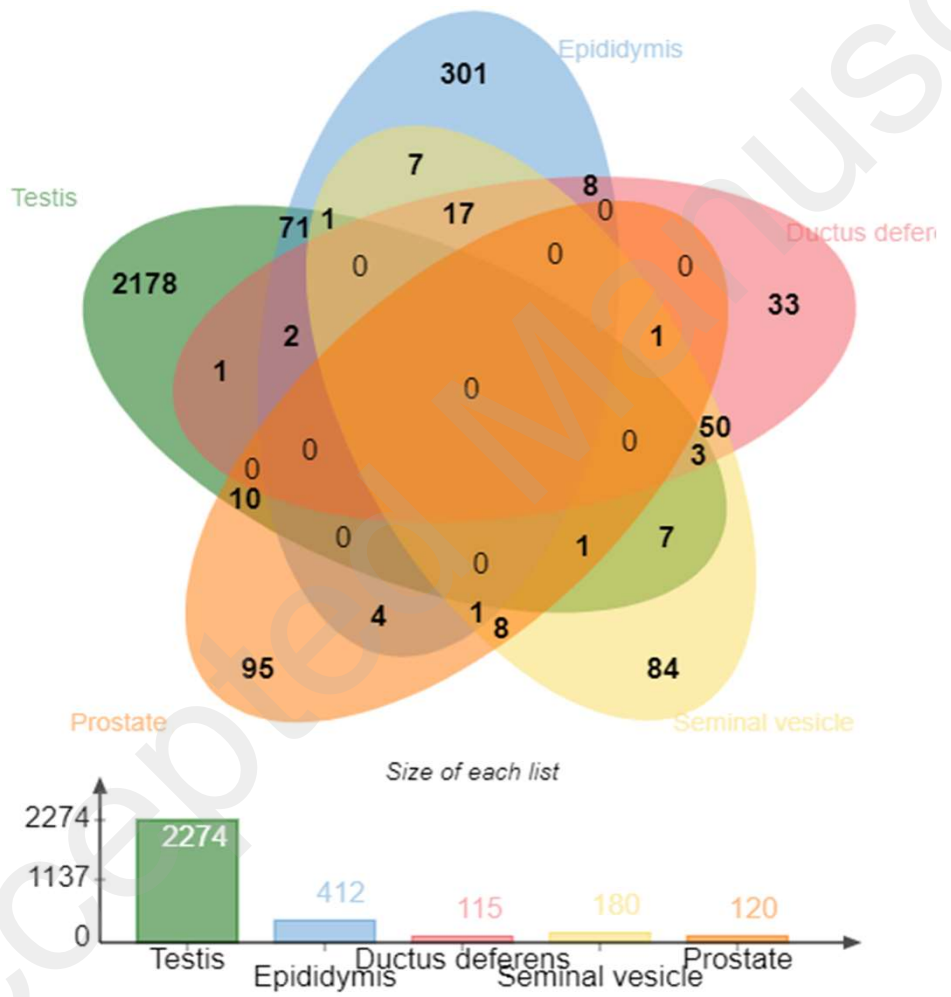


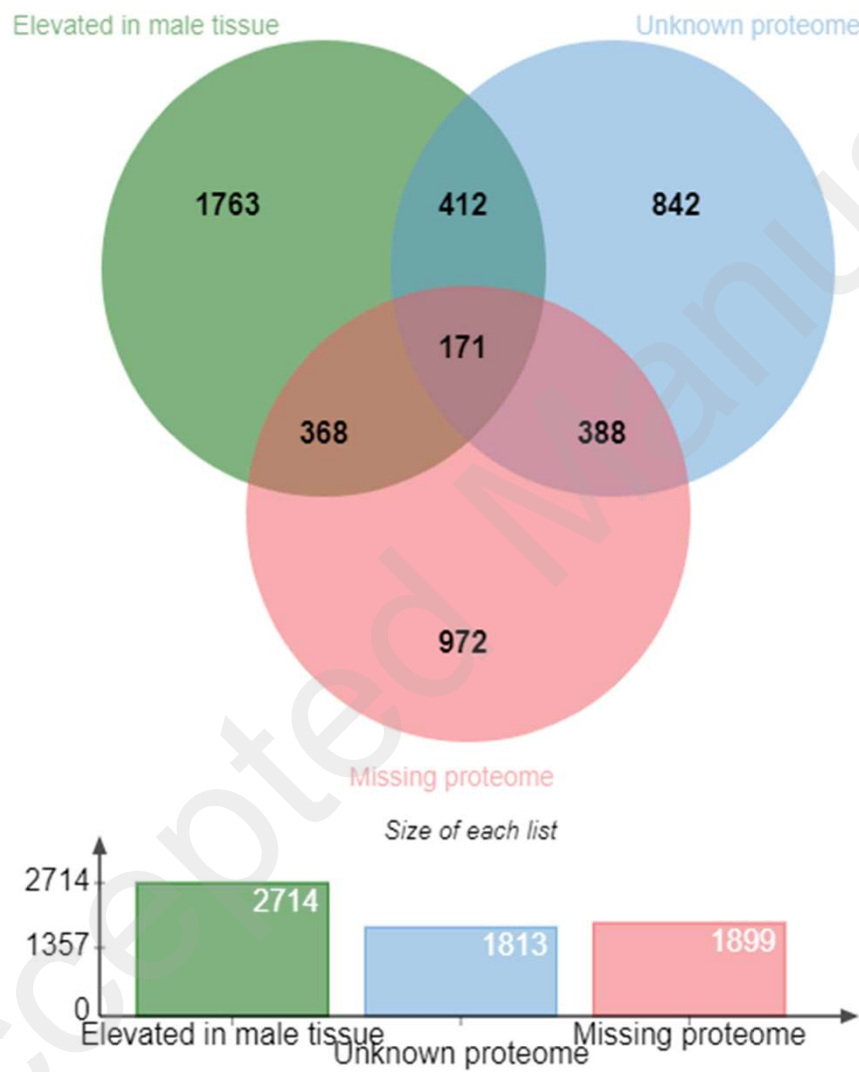
Figure 1.



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

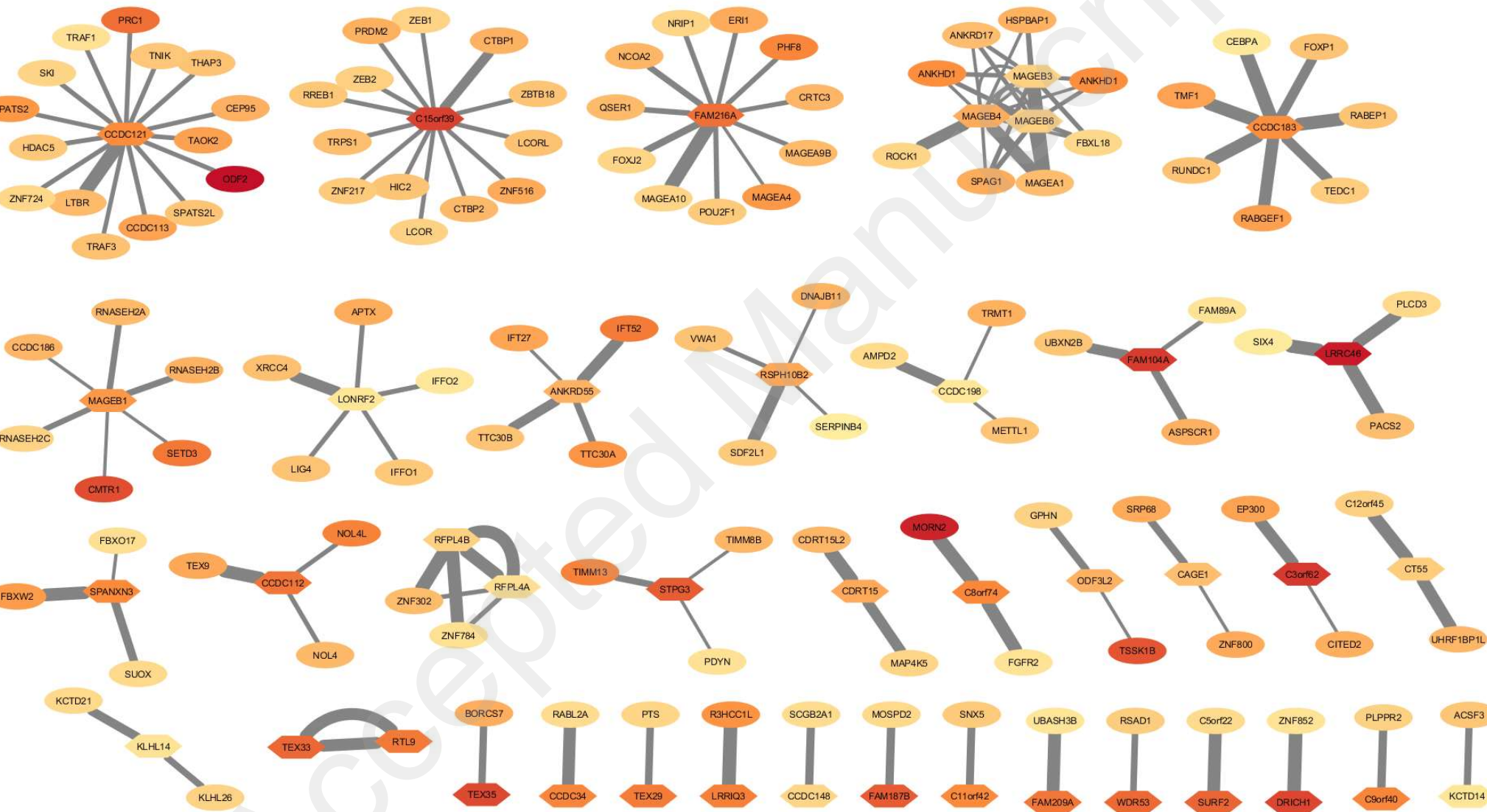
Figure 2.



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**Figure 3.**

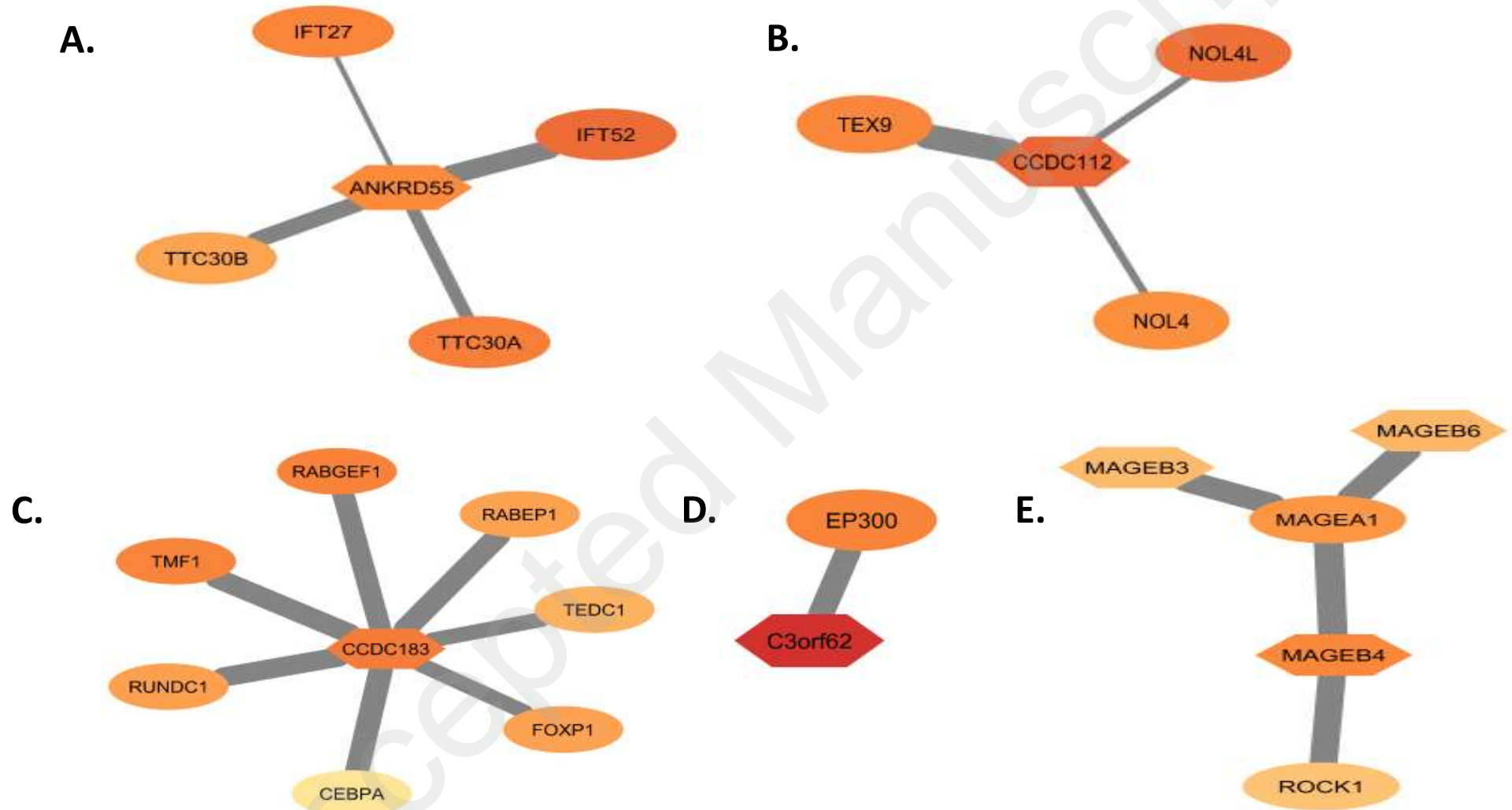
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

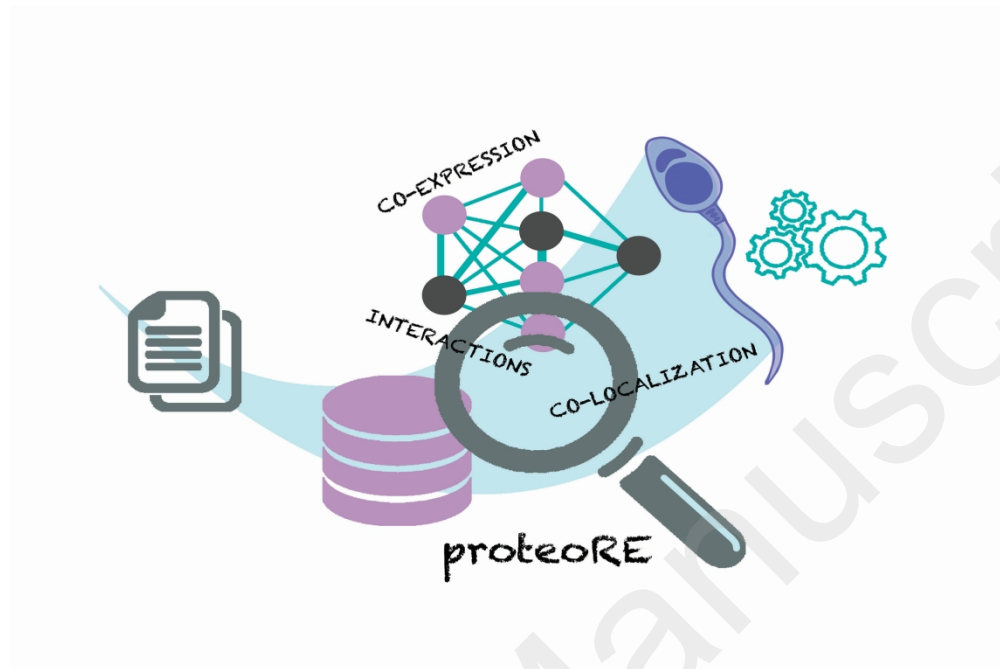
Figure 4.



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

Figure 5.





604x403mm (96 x 96 DPI)