

# Exploring the Human Seminal Plasma Proteome: An Unexplored Gold Mine of Biomarker for Male Infertility and Male Reproduction Disorder

Kambiz Gilany <sup>1\*</sup>, Arash Minai-Tehrani <sup>2</sup>, Elham Savadi-Shiraz <sup>3</sup>, Hassan Rezadoost <sup>4</sup>, Niknam Lakpour <sup>1</sup>

1- Reproductive Biotechnology Research Center, Avicenna Research Institute, ACECR, Tehran, Iran

2- Nanobiotechnology Research Center, Avicenna Research Institute, ACECR, Tehran, Iran

3- Department of Urology, Pediatric Urology and Andrology, Section Molecular Andrology, Justus Liebig University, Giessen, Germany

4- Medicinal Plants and Drugs Research Institute, Shahid Beheshti University, Tehran, Iran

## Abstract

**Background:** The human seminal fluid is a complex body fluid. It is not known how many proteins are expressed in the seminal plasma; however in analog with the blood it is possible up to 10,000 proteins are expressed in the seminal plasma. The human seminal fluid is a rich source of potential biomarkers for male infertility and reproduction disorder.

**Methods:** In this review, the ongoing list of proteins identified from the human seminal fluid was collected. To date, 4188 redundant proteins of the seminal fluid are identified using different proteomics technology, including 2-DE, SDS-PAGE-LC-MS/MS, MudPIT. However, this was reduced to a database of 2168 non-redundant protein using UniProtKB/Swiss-Prot reviewed database.

**Results:** The core concept of proteome were analyzed including pI, MW, Amino Acids, Chromosome and PTM distribution in the human seminal plasma proteome. Additionally, the biological process, molecular function and KEGG pathway were investigated using DAVID software. Finally, the biomarker identified in different male reproductive system disorder was investigated using proteomics platforms so far.

**Conclusion:** In this study, an attempt was made to update the human seminal plasma proteome database. Our finding showed that human seminal plasma studies used to date seem to have converged on a set of proteins that are repeatedly identified in many studies and that represent only a small fraction of the entire human seminal plasma proteome.

**Keywords:** Biomarker, Human seminal plasma, Proteome, Proteomics.

**To cite this article:** Gilany K, Minai-Tehrani A, Savadi-Shiraz E, Rezadoost H, Lakpour N. Exploring the Human Seminal Plasma Proteome: An Unexplored Gold Mine of Biomarker for Male Infertility and Male Reproduction Disorder. *J Reprod Infertil*. 2015;16(2):61-71.

\* Corresponding Author:  
Kambiz Gilany,  
Reproductive  
Biotechnology Research  
Center, Avicenna  
Research Institute,  
ACECR, Tehran, Iran,  
P.O. Box: 19615-1177  
E-mail:  
k.gilany@avicenna.ac.ir

Received: Aug, 6, 2014

Accepted: Dec, 8, 2014

## Introduction

Infertility is one of the major health issues and it is estimated that 15-20% of couples are infertile in the world. Moreover, males and females contribute equally to the problem. Clinical semen analysis, based on sperm factors such as morphology, concentration, motility, etc., fails to identify the causes in 30%-50% of infertility cases (1, 2). Male reproduction disorder is another major health issue including varicocele, prostate cancer and prostatitis. Varicocele affects

15-25% of the male population as a time-dependent disease that begins at puberty and is considered the major treatable cause of male factor infertility (3, 4). Prostate cancer is the second most frequently diagnosed cancer and the sixth leading cause of cancer death in males worldwide. The introduction of serum prostate specific antigen screening led to a significant increase in the number of diagnosed cases but failed to demonstrate a statistically significant prostate cancer mortality

benefit (5, 6). Prostatitis (inflammation of the prostate gland) is a very common condition, with symptoms affecting approximately 10% of all men. Diagnosing prostatitis remains confusing and frustrating to urologists, as many of the symptoms overlap (7, 8). Therefore, a new diagnosis tool is urgently needed for diagnosing male infertility and male reproduction disorder.

The spermatozoa are bathed in a continuously and progressively changing medium of fluid proteins and chemical compositions. The constituents of the human seminal plasma include secretions originated from the testis, epididymis, and male accessory glands such as seminal vesicles, prostate and Cowper's gland. Seminal plasma provides a safe surrounding for the spermatozoa and serves as a vehicle for ejaculated spermatozoa to the female genital tract. Due to its buffer capacities, it also protects the spermatozoa from acidic environment of the vagina (9, 10).

The human seminal plasma is a rich source of potential biomarkers. It is estimated that the human seminal plasma protein concentration is 35-55 mg/ml, which makes it a rich and easily accessible source for protein identification (11). The proteome is defined as the protein complement of the genome. This old definition has a new face based on the development of proteomics technology. Nowadays, the proteome is defined as the sum and the time dynamics of all protein species occurring during the life-time of an individual. By this definition, the proteome includes the expression level of the individual protein, a protein isoform and the post-translation modification (12).

Despite remarkable advances in proteomics technology, a limited number of studies have focused on the human seminal plasma proteome (HSPP) (13-19). A search on the PubMed database shows the word "human seminal plasma AND proteomics" has only 98 hits compared to the word "blood AND proteomics" which has 6189 hits till 2014. This is the case despite the fact that male infertility and male reproduction disorder is a sensitive health issue. Therefore, it is urgent to investigate the underlying biology of the HSPP in order to provide a better understanding of male infertility causes and to fully utilize the potential of proteomics technology.

In this review, core concepts in HSPP were investigated and burgeoning list of proteins identified in the human seminal plasma was outlined. Bioinformatics and the literature on the collected HSPP including isoelectric point (pI), post-trans-

lational modifications (PTMs), amino acids distribution, chromosome distribution, molecular and biological function and enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway have been analyzed. Additionally, the major proteomics findings of the HSPP in different male infertility and male reproductive disorders were highlighted.

## Methods

**Database search:** To find all relevant studies based on HSPP, a search was done according to keywords "human seminal plasma", "proteome" and "proteomics" in PubMed and Google. Only the data regarding the human seminal plasma were included in this study.

**Bioinformatics:** In order to have a non-redundant and an accurate HSPP database, only proteins which correspond to UniProtKB/Swiss-Prot accession number (2013/05/01 release) were included in the protein database. The theoretical human proteome was downloaded from UniProtKB/Swiss-Prot database. Information regarding post-translation modifications, chromosome and tissue origin were extracted from UniProtKB/Swiss-Prot. The pI and MW were calculated using the tools on the Expsay website (<http://expasy.org/>). For calculation of amino acids distribution, a program was written in the Python language. Molecular function, biological function and KEGG enriched pathway of the collected HSPP were analyzed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) software (20).

## Results

**Techniques:** A wide range of proteomics technology platforms are available such as gel-based applications including sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and two-dimensional gel electrophoresis (2-DE). Gel-free high-throughput analysis methods are equally available including multidimensional protein identification technology (MudPIT) and filter-aided sample preparation (FASP) technology (21-23). However, because of the limited study on the HSPP, only a few proteomics technology platforms have been applied that include 2-DE, sodium dodecyl sulfate polyacrylamide gel electrophoresis, liquid chromatography tandem mass spectrometry (SDS-PAGE-LC-MS/MS), reverse-phase tandem mass spectrometry analysis (1D-LC-MS/MS), and MudPIT (11, 13, 14, 24-29). Using these proteomics platforms, 4188 redundant

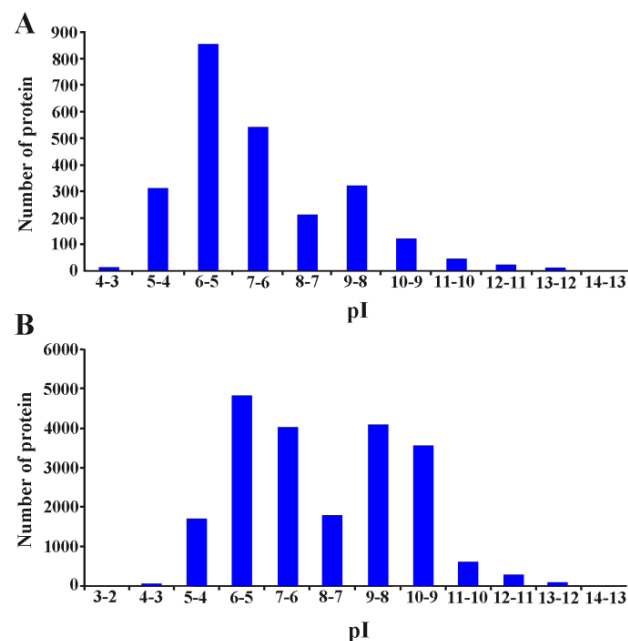
proteins from the human seminal plasma fluids were identified. 75 of mentioned proteins were identified by 2-DE, 2182 were identified by different SDS-PAGE- LC-MS/MS studies, 118 were identified by 1D-LC-MS/MS and finally 1822 were identified by MudPIT technology. Using UniProtKB/Swiss-Prot database, the protein list was reduced to 2168 non-redundant databases. To the best of our knowledge, this is the largest protein database identified in the human seminal plasma fluid (supplementary).

By looking at the performance of the applied techniques, it is not surprising that the techniques are complements (30). However, it is clear from different HSPP profiling that 2-DE (15 unique proteins) has the worst performance (supplementary). This is caused by the nature of the human seminal plasma fluid, which is a highly viscous sample. An introduction of cold acetone/trichloroacetic acid or chloroform/methanol precipitation might improve the quality of 2-DE of the HSPP. It has been demonstrated that under optimal conditions, a large scale 2-DE is able to detect more than 10,000 spots (31). However, protein identification from 2-DE is still challenging and in comparison, MudPIT technology had the best performance in identification of unique proteins (supplementary).

**pI:** The human seminal plasma fluid is a relative basic environment. The pH of the human seminal plasma fluid is 8 (measured by universal indicator pH=0-14). This means that the HSPP should contain mostly proteins with the pI lower than 8. Figure 1 shows the pI distribution of collected HSPP compared to the theoretical human proteome (HP).

As shown in figure 1, the pattern of the pI distribution of the HSPP does not follow the HP which is a biphasic distribution. The majority of identified proteins from the HSPP have pI lower than 8. There are, however, some proteins with pI over 8, but they do not follow the biphasic form of a typical proteome such as sperm (12). Figure 1 shows that 2-DE technology should have a high performance for the HSPP since the majority of the HSPP have pI lower than 8. Therefore, in this region, it is a well known fact that the 2-DE has a better performance.

**Post-translation modifications:** The first generation of proteomics technology was the mapping of the proteome of the cell, tissue or fluid. The second generation of the proteomics technology was the quantification of the proteome (21). It becomes obvious by the development of the prote-

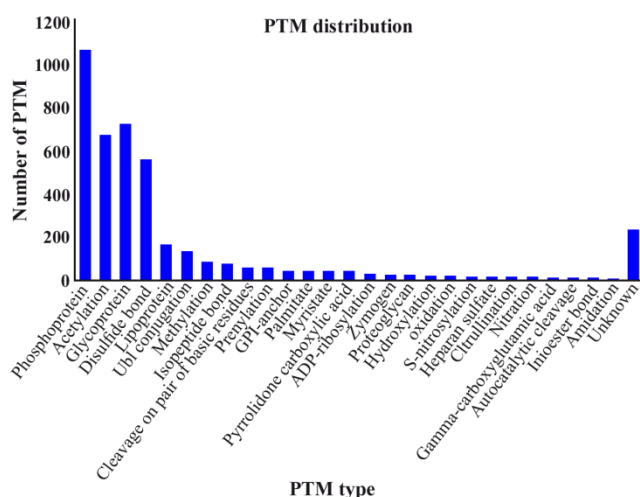


**Figure 1.** A: pI distribution of collected human seminal plasma proteome. B: pI distribution of theoretical human proteome

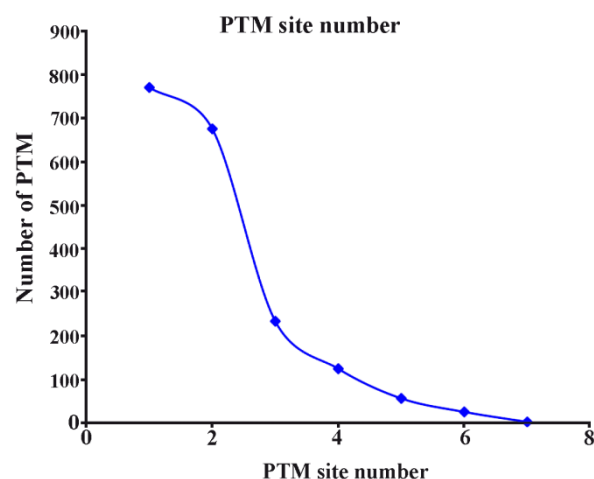
omics platform that the post-translation modification (PTM) plays an important role in the proteomics. The third generation of proteomics technology is investigation of PTM (32, 33). It has at least two very important functions. Firstly, after expression of proteins, they mostly become post-translationally modified for *e.g.* translocation. Secondly, PTM plays an important role in the signaling pathways *e.g.* phosphorylation.

In this study, by reviewing the literature, PTM in the collected HSPP was investigated. Figure 2 shows PTM distribution of the collected HSPP.

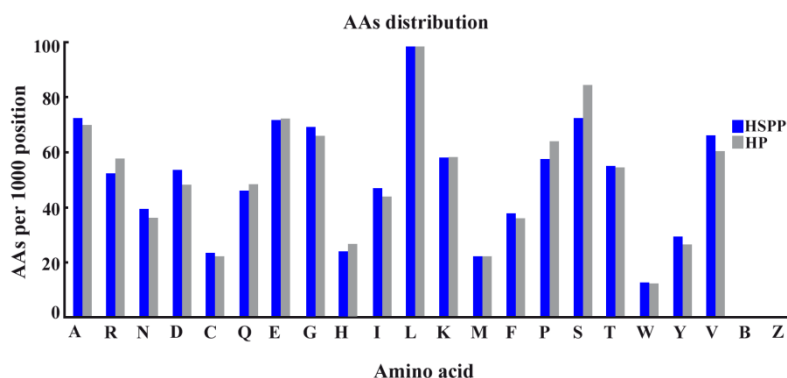
As shown in figure 2, the majority of the collected HSPP post-translation modifications are: phosphorylation, acetylation, glycosylation and disulfide bond. Totally, 27 types of the PTMs were observed in the collected HSPP. However, the types of functions these PTMs play in the human seminal plasma fluid were not well investigated. To the best of our knowledge, no proteomics study has been made on the PTM in the human seminal plasma fluid. Comparing to the word "blood", the "PubMed" searching for the word "post-translation modification AND blood" gives 4670 hits. This means a plenty of works still lie ahead that investigate what function these PTMs have in male fertility or male reproduction disorder that may include activation or protection of sperm, *etc.*



**Figure 2.** Number of post-translation types in the collected human seminal plasma proteome



**Figure 3.** Number of post-translation modification (PTM) site on the collected human seminal plasma



**Figure 4.** Amino acid distribution of cataloged the human seminal plasma proteome (HSPP) compared to the theoretical human proteome (HP)

Another way of looking at the PTM is to observe how many sites in a protein become post-translationally modified *e.g.* HIF-1- $\alpha$  (34). Figure 3 shows the number of sites in a protein that were post-translationally modified in the collected HSPP. As it is shown in the figure, the majority of proteins have at least one site of modification while some proteins have 6-7 sites of modification.

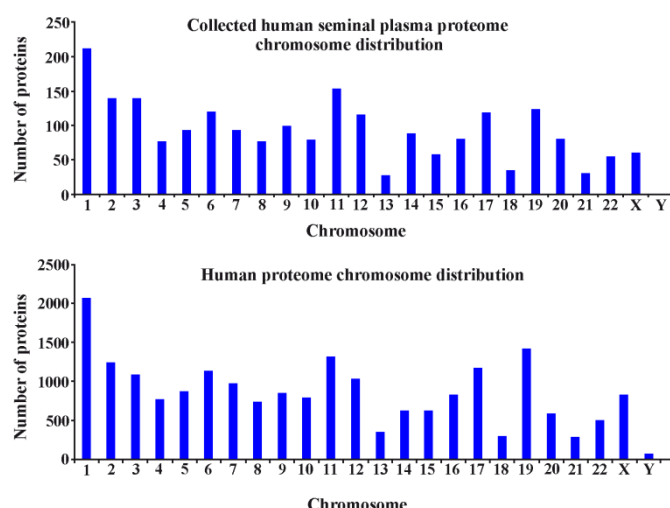
**Amino Acids Distributions:** Also, amino acids distribution of the established HSPP was investigated and compared to the HP. Figure 4 shows the amino acid distribution of the HSPP and the HP. As shown in figure 4, for most amino acids of the catalog, HSPP follows the theoretical HP. However, in case of proline (P) and serine (S), these amino acids are underrepresented in the HSPP. These amino acids usually are presented at the site of phosphorylation. It is possible that the HSPP is

not a heavily phosphorylated proteome. To the best of our knowledge, no phosphorylation work has been done on the HSPP, and it is difficult to suggest why P and S are underrepresented in the HSPP.

**Chromosome distribution:** The near-complete sequencing of the human genome has yielded a total of 25-30,000 genes. However, it is unknown how many proteins are expressed in the human. An estimation about 2 million proteins have been suggested (12). To define each of these proteins, the Chromosome-Centric Human Proteome Project (C-HPP) has been designed to map the entire human proteome in a systematic effort. To date, 20 international teams are involved in the mapping of 18 different chromosomes (35).

Figure 5 shows the chromosome distribution of the HSPP compared to the theoretical HP. In Iran,





**Figure 5.** Chromosome distribution of the human seminal plasma versus the theoretical human proteome

current effort is to map the Y chromosome proteome (36, 37). The HSPP is an excellent source for detection of Y chromosome proteins. However, as it is shown in figure 5, no Y chromosome proteins were detected by the used proteomics platform. This could be caused by the mass spectrometry instrument used in the studies. A recent published paper of the sperm proteome, in which a LTQ Orbitrap Velos mass spectrometer was used, the authors were able to identify more than 4675 unique proteins of which 4 were Y chromosome proteins (38). This means that by an improved proteomics platform, it is possible to identify some of the low-abundance proteins, in which Y chromosome proteins were included.

**Tissue origin:** It is not possible to know the tissue origin of the identified proteins of the HSPP using

mass-spectrometry-based proteomics analysis. A tissue origin of identified protein can be found in the UniProtKB/Swiss-Prot database. A significant number of identified proteins from the HSPP belonged to testis (25%). Additionally, proteins from the prostate (11%), the epididymis (1%) and a few seminal vesicles were identified. According to the literature, the rest of the proteins were widely expressed in other tissues. These proteins can be considered as "housekeeping" proteins.

**Biological processes and molecular function:** The collected HSPP were functionally categorized based on Gene Ontology (GO) terms and annotations using the Database for Annotation, and Visualization and Integrated Discovery (DAVID) program package (<http://david.abcc.ncifcrf.gov/>) (20). For any given protein list, DAVID tools are able to identify enriched biological terms, particularly GO terms, discover enriched functional-related protein groups, visualize proteins on BioCarta & KEGG pathway maps, etc.

Table 1 shows the ten most important catalog outputs for biological function analysis by DAVID software. DAVID software was able to catalog 1690 (83%) of the submitted proteins. This means that the biological function of almost 17% of submitted proteins is still unknown. As shown in table 1, the most important biologically functional proteins in the HSPP belong to proteolysis (11.3%, p-value: 3.13E-19) and carbohydrate catabolic process (2.5%, p-value: 1.13E-18).

Further down the biologically functional list, additional interesting groups of proteins were identified. These groups belong to the oxygen and reactive oxygen species metabolomics processes (1.2%, p-value: 1.25E-5) and regulation of oxygen and

**Table 1.** Tabulated are the ten important biological functions with the greatest statistical significance for enrichment in the collected proteome data set of the HSPP (GOTERM: level ALL)

Biological functions	%	p-value
Proteolysis	11.3	3.13E-19
Carbohydrate catabolic process	2.5	1.13E-18
Positive regulation of ubiquitin-protein ligase activity	2	5.85E-18
Proteasomal protein catabolic process	2.4	8.71E-18
Proteasomal ubiquitin-dependent protein catabolic process	2.4	8.7E-18
Positive regulation of ubiquitin-protein ligase activity during mitotic cell cycle	1.9	1.35E-17
Negative regulation of cellular protein metabolic process	3.3	1.37E-17
Negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle	1.9	1.57E-17
Anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process	1.9	1.57E-17
Response to organic substance	8.4	1.9E-17

The percentage is calculated as: involved proteins divided by the total number of proteins multiplied by one-hundred. The enrichment p-value (compared to the theoretical human proteome) is calculated based on EASE Score, a modified Fisher's Exact Test and ranges from 0 to 1. Fisher's Exact p-value=0 represent perfect enrichment. Usually the p-value must be equal to or smaller than 0.05 to be considered strongly enriched in the annotation categories. The closer the value is to zero, the more enriched is the category

**Table 2.** Tabulated are the top ten important molecular functions with the greatest statistical significance for enrichment in the collected proteome data set of the HSPP (GOTERM: level ALL)

Molecular functions	%	p-value
Peptidase activity	7.5	2.99E-20
Peptidase inhibitor activity	3.1	5.25E-19
GTPase activity	3.8	1.4E-18
Endopeptidase inhibitor activity	3	2.77E-18
Peptidase activity, acting on L-amino acid peptides	7	5.16E-18
Unfolded protein binding	2.4	1.06E-15
Carbohydrate binding	5	1.82E-15
Enzyme inhibitor activity	4	8.02E-15
Exopeptidase activity	1.9	3.33E-14
Endopeptidase activity	4.9	9.95E-14

Explanations for the percentage and p-values can be found in table 1

reactive species metabolomics processes (0.24%, p-value: 0.039). It is shown that reactive oxygen species (ROS) plays an important role in male infertility, especially in the case of the asthenospermic men, a deregulation of ROS proteins has been reported (13, 39). Other interesting biological functional groups are fertilization proteins (1%, p-value: 0.001), binding of sperm to zona pellucid proteins (0.4%, p-value: 0.003), and sperm-egg recognition proteins (0.4%, p-value: 0.003).

Table 2 shows the ten most important molecular functional groups of the HSPP analyzed by DAVID software. DAVID software was able to categorize 1636 (80%) of the submitted proteins. As shown in table 2, the most important molecular function of identified proteins belongs to peptidase activity (7.5%, p-value: 2.99E-20) and peptidase inhibitor activity (3.1%, p-value: 5.35E-19). The increase of peptidase activity is because of high energy consummation of sperm. Peptidase activity causes catabolism of proteins and thereby production of amino acids and an activation of catabolism pathways such as glucogenesis. This will lead to production of energy in the form of ATP. A high level of energy supply in the form of ATP is needed for sperm movement.

Additionally, interesting molecular functional groups with lower p-value were the calcium binding proteins (8.5%, p-value: 2.22E-8) and the insulin-like growth factor binding proteins (0.5%, p-value: 0.0025).

**Enriched Pathway:** One of the functions of the DAVID software is to show the enriched KEGG pathways. The following pathways were enriched in the HSPP: lysosome (3%, p-value: 2.09E-20),

proteasome (1.7%, p-value: 1.18E-16), pentose phosphate pathway (1%, p-value: 1.22E-11), amino sugar and nucleotide sugar metabolism (1.2%, p-value: 2.36E-8), glycolysis/gluconeogenesis (1.5%, p-value: 3.01E-8), glutathione metabolism (1.2%, p-value: 5.6E-7), fructose and mannose metabolism (0.9%, p-value: 2.48E-6), galactose metabolism (0.7%, p-value: 1.39E-4), and pyruvate metabolism (0.8%, p-value 0.0019). It is not surprising that the majority of known energy catabolism pathways are enriched in the human seminal fluid since sperm have a high consumption of ATP. An enrichment of the pentose phosphate pathway causes a production of reducing equivalents. It is well established that human seminal plasma is a natural reservoir of antioxidants. It is known that an imbalance in oxidative system causes infertility (40-42). Additionally, Pathogenic *Escherichia coli* infection pathway (1.3%, p-value: 2.67E-6) was enriched in the HSPP.

**The human seminal plasma proteome and biomarker discovery:** In the first study, using gel-based proteomics technology in male infertility with the human seminal fluid as source for biomarker discovery, several potential candidates for spermatogenesis impairment were observed. Using 2-DE, several groups of spots were detected which were deregulated or disappeared in the proteome profile of the human seminal fluid of fertile compared to infertile men. However, none of the deregulated spots were identified (43). Based on the development of 2-DE and mass spectrometry, another attempt was done to better understand the spermatogenesis impairment in infertile men (44). In the mentioned study, the seminal plasma of four different groups of men including normozoospermic, asthenozoospermic, oligozoospermic and azoospermic were compared using two-dimensional differential in-gel electrophoresis (2D DIGE) followed by matrix-assisted laser desorption mass spectrometry (MALDI-TOF-MS). Eight proteins showed significant increased expression level in the azoospermic men compared to at least one of other groups. These proteins were fibronectin, prostatic acid phosphatase (PAP), proteasome subunit alpha type-3, beta-2-microglobulin, galectin-3-binding protein, prolactin-inducible protein and cytosolic nonspecific dipeptidase. Specifically, PAP was upregulated in azoospermic men compared to all other groups (44).

In another gel-based technology study, the human seminal fluid of prostate cancer patients was optimized and analyzed by 2-DE followed by

MALDI-TOF-MS. The result showed that proteins kallikrein 3 (prostate specific antigen), PAP, Zinc  $\alpha$ 2-glycoprotein and progastricsin were up-regulated in prostate cancer patients compared to the normal seminal plasma (45).

In the most recent study of the human seminal plasma using gel-based technology, the human seminal plasma of adolescents with and without varicocele was analyzed through 2-DE followed by electrospray mass spectrometry (ESI-Quad-TOF-MS). Forty-seven of interested spots were subjected to mass spectrometry analysis. In the study, adolescents with varicocele and normal semen quality showed an overexpression of spermatogenesis proteins, whereas adolescents with varicocele and abnormal semen quality showed an overexpression of apoptosis regulated proteins compared to adolescents without varicocele (17).

The limitation of the gel-based technology encouraged using the gel-free technology proteomics platform to study the seminal plasma as a source of biomarker for male reproductive system disorder.

To the best of our knowledge, the first "semi" gel-free based proteomics study of the seminal plasma used SDS-PAGE-LC-MS/MS. In the mentioned study, the seminal plasma of the asthenozoospermic patients were compared to the normozoospermic men. More than 700 proteins were identified. Of these, 45 proteins were upregulated and 56 proteins were downregulated in the asthenozoospermic men compared to the normozoospermic men. The most deregulated proteins belonged to the regulation of reactive oxygen species. Specifically, DJ-1 protein which is involved in the oxidative stress was shown to be significantly downregulated in asthenozoospermic men (13).

Another popular gel-free based proteomics platform is MudPIT technology. Most recently, two studies have used MudPIT technology for analysis of the seminal plasma in order to search for biomarker (18, 46).

Batrach et al. used the MudPIT technology to examine the seminal plasma of the non-obstructive azoospermia for identification of potential biomarkers of male infertility. More than 2000 proteins were identified. Of these, 34 proteins were upregulated and 18 proteins were downregulated in control relative to non-obstructive azoospermia. The upregulated proteins are involved in reproduction, carbohydrate catabolic process and gly-

colytic pathway. The downregulated proteins belong to the glutathione metabolism pathway and the glycolytic pathway (46).

In the second MudPIT technology study, the seminal plasma in men with prostatitis was compared to the men without prostatitis. More than 1700 proteins were identified. The authors generated a list of 59 candidates of prostatitis biomarkers, of which 33 proteins were significantly upregulated in prostatitis compared to the control group, and 26 of which were downregulated. The most significant upregulated prostatitis candidates' proteins are involved in enzyme regulated activity and in the defensive response. The downregulated proteins are involved in the development, regulation of biological processes and transport (18).

The proteome of the seminal plasma is complex and contains some high abundance proteins including semenogelins and kallikrein 3 (also known as prostate specific antigen). Progress in biological mass spectrometry has facilitated the identification of thousand proteins from different biological samples. However, routine quantification by mass spectrometry, especially for low abundance proteins in a complex mixture, is still challenging. Quantitative selected reaction monitoring (SRM)- also called multiple reaction monitoring-assays were introduced as a means to supplement antibody-based enzyme-linked immunosorbent assay (ELISA). Quantification and verification by SRM assays is an emerging field of proteomics technology (47, 48). In a novel study using SRM assays by Drabovich et al., 31 proteins of the seminal plasma of individuals with non-obstructive azoospermia were analyzed. In this study, testis specific proteins-LDHC, TEX101 and SPAG11B-showed absolute specificities and sensitivity. Additionally, cell-specific classification of protein expression indicated that Sertoli or germ cell dysfunction, but not Leydig cell dysfunction, was observed in seminal plasma from patients with non-obstructive azoospermia (49). Although SRM assays are excellent tools for diagnosis, the main disadvantage is that the candidate biomarker proteins have to be known in advance through other gel-free based proteomics technology.

Although different proteomics platforms have been used for biomarker discovery of male infertility and male reproduction disorder, it is possible to generate a common biomarker protein list. The

following proteins have been identified to be deregulated in several studies: semenogelins 1 (SEMG1), semenogelins 2 (SEMG2), prolactin-inducible protein (PIP), fibronectin (FN1), prostatic acid phosphatase (ACPP), kallikrein 3 (KLK3) and epididymal secretory protein E1 (NPC) (13, 17, 18, 44-46, 49). To our current knowledge, no quantification study has been done on the level of the protein in human seminal plasma. However, Batruch et al. have used the label-free quantification spectra counting method to analyze the deregulated proteins. Using Batruch et al.'s data, the common biomarker protein list is amongst most abundant proteins in the human seminal plasma (46). PeptideAtlas was used to evaluate which of these highly abundant biomarker proteins are expressed in the blood (50). It turned out that PIP, FN1, KLK3 and NPC are highly expressed in the blood and are identified with several distinct peptides.

### Discussion

Seminal plasma is a complex body fluid, containing a large diversity of proteins. It is not known how many proteins are expressed in the seminal plasma; however, possibly up to 10,000 proteins can be expressed in the seminal plasma. Seminal plasma is an excellent source of protein biomarkers because it circulates through and comes in contact with the male reproductive system. Consequently, seminal plasma proteomics has great potential for the discovery of biomarkers to improve diagnosis or classification of a wide range of male reproductive system disorders, including prostate cancer. However, seminal plasma is one of the most complex human proteomes with considerable difference in the concentration of individual proteins. The analytical challenge for biomarker discovery arises from the high variability in the concentration of some seminal plasma proteins. This observation was based on the Batruch et al.'s relative quantification of the human seminal plasma proteins using label-free quantification by the spectra counting method. The average spectra count value changed from 4500 to 0.3 (46). Semenogelin is a protein of very high abundance in seminal plasma which can be a prime candidate for a complete selective removal prior to performing a proteomics analysis of lower abundance proteins. The presence of higher abundance proteins interferes with the identification and quantification of lower abundance proteins. Complexity and dynamic range of protein concen-

trations can be addressed with a combination of prefractionation techniques that deplete highly abundant proteins and fractionate. Protein prefractionation by immunodepletion and reversed-phase separation of the depleted seminal plasma provides methods compatible with gel-free based proteomics technology analysis.

Another main problem using seminal plasma as a source for biomarker discovery is the interindividual variations that exist in the seminal plasma compared to other biological samples. Using both gel and gel-free based proteomics technology, this observation has been reported (51, 52). However, no protein lists are available from these two studies. A recently developed depletion method that mixes 14 high-specificity polyclonal antibodies (MARS) to remove the top 14 proteins in the blood in a single purification step is commercially available (53). By creating such column for the human seminal fluid, an improvement of the interindividual variations will be observed.

### Conclusion

Biomarker discovery remains a very challenging task due to the complexity of the samples and the wide dynamic range of protein concentrations (54). Most of the human seminal plasma biomarker studies performed to date seem to have converged on a set of proteins that are repeatedly identified in many studies and that represent only a small fraction of the entire HSPP. Processing and analysis of proteomics data is indeed a very complex multistep process (55, 56). The consistent and transparent analysis of LC/MS and LC-MS/MS data requires multiple stages (57) and this process remains the main bottleneck for many larger proteomics studies. To overcome these issues, effective sample preparation (to reduce complexity and to enrich for lower abundance components while depleting the most abundant ones), state-of-the-art mass spectrometry instrumentation, and extensive data processing and data analysis are required.

### Acknowledgement

The authors would like to thank Mrs. Mohtaram Vafakhah for critical reading of manuscript. Additionally, we would like to thank Elisabeth Noergaard Nielsen for proofreading of the manuscript.

### Conflict of Interest

The authors declare no conflict of interest.



## References

- Brugh VM 3rd, Lipshultz LI. Male factor infertility: evaluation and management. *Med Clin North Am*. 2004;88(2):367-85.
- Tomar AK, Sooch BS, Singh S, Yadav S. Differential proteomics of human seminal plasma: A potential target for searching male infertility marker proteins. *Proteomics Clin Appl*. 2012;6(3-4):147-51.
- Steen O, Knops J, Declerck L, Adimoelja A, van de Voorde H. Prevention of fertility disorders by detection and treatment of varicocele at school and college age. *Andrologia*. 1976;8(1):47-53.
- Steen OP. Varicocele in the adolescent. *Adv Exp Med Biol*. 1991;286:295-321.
- Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. *CA Cancer J Clin*. 2009;59(4):225-49.
- Ilic D, O'Connor D, Green S, Wilt TJ. Screening for prostate cancer: an updated Cochrane systematic review. *BJU Int*. 2011;107(6):882-91.
- Nickel JC, Downey J, Hunter D, Clark J. Prevalence of prostatitis-like symptoms in a population based study using the National Institutes of Health chronic prostatitis symptom index. *J Urol*. 2001;165(3):842-5.
- Nickel JC. Recommendations for the evaluation of patients with prostatitis. *World J Urol*. 2003;21(2):75-81.
- Wolters-Everhardt E, Dony JM, Lemmens WA, Doesburg WH, De Pont JJ. Buffering capacity of human semen. *Fertil Steril*. 1986;46(1):114-9.
- Owen DH, Katz DF. A review of the physical and chemical properties of human semen and the formulation of a semen simulant. *J Androl*. 2005;26(4):459-69.
- Pilch B, Mann M. Large-scale and high-confidence proteomic analysis of human seminal plasma. *Genome Biol*. 2006;7(5):R40.
- Gilany K, Lakpour N, Vafakhah M, Sadeghi MR. The profile of human sperm proteome; A mini-review. *J Reprod Infertil*. 2011;12(3):193-9.
- Wang J, Wang J, Zhang HR, Shi HJ, Ma D, Zhao HX, et al. Proteomic analysis of seminal plasma from asthenozoospermia patients reveals proteins that affect oxidative stress responses and semen quality. *Asian J Androl*. 2009;11(4):484-91.
- Batruch I, Lecker I, Kagedan D, Smith CR, Mullen BJ, Grober E, et al. Proteomic analysis of seminal plasma from normal volunteers and post-vasectomy patients identifies over 2000 proteins and candidate biomarkers of the urogenital system. *J Proteome Res*. 2011;10(3):941-53.
- Cross NL. Human seminal plasma prevents sperm from becoming acrosomally responsive to the agonist, progesterone: cholesterol is the major inhibitor. *Biol Reprod*. 1996;54(1):138-45.
- Veveris-Lowe TL, Kruger SJ, Walsh T, Gardiner RA, Clements JA. Seminal fluid characterization for male fertility and prostate cancer: kallikrein-related serine proteases and whole proteome approaches. *Semin Thromb Hemost*. 2007;33(1):87-99.
- Zylbersztejn DS, Andreoni C, Del Giudice PT, Spaine DM, Borsari L, Souza GH, et al. Proteomic analysis of seminal plasma in adolescents with and without varicocele. *Fertil Steril*. 2013;99(1):92-8.
- Kagedan D, Lecker I, Batruch I, Smith C, Kaploun I, Lo K, et al. Characterization of the seminal plasma proteome in men with prostatitis by mass spectrometry. *Clin Proteomics*. 2012;9(1):2.
- da Silva BF, Souza GH, lo Turco EG, Del Giudice PT, Soler TB, Spaine DM, et al. Differential seminal plasma proteome according to semen retrieval in men with spinal cord injury. *Fertil Steril*. 2013;100(4):959-69.
- Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57.
- Gilany K, Moens L, Dewilde S. Mass spectrometry based proteomics in the life sciences: a review. *JPS*. 2010;1(1):53-78.
- Chandramouli K, Qian PY. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Hum Genomics Proteomics*. 2009;2009.
- Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nat Methods*. 2009;6(5):359-62.
- Starita-Geribaldi M, Roux F, Garin J, Chevallier D, Fenichel P, Pointis G. Development of narrow immobilized pH gradients covering one pH unit for human seminal plasma proteomic analysis. *Proteomics*. 2003;3(8):1611-9.
- Fung KY, Glode LM, Green S, Duncan MW. A comprehensive characterization of the peptide and protein constituents of human seminal fluid. *Prostate*. 2004;61(2):171-81.
- Kumar V, Hassan MI, Tomar AK, Kashav T, Nautiyal J, Singh S, et al. Proteomic analysis of heparin-binding proteins from human seminal plasma: a step towards identification of molecular markers of male fertility. *J Biosci*. 2009;34(6):899-908.

27. Poliakov A, Spilman M, Dokland T, Amling CL, Mobley JA. Structural heterogeneity and protein composition of exosome-like vesicles (prostasomes) in human semen. *Prostate*. 2009;69(2):159-67.
28. Thimon V, Frenette G, Saez F, Thabet M, Sullivan R. Protein composition of human epididymosomes collected during surgical vasectomy reversal: a proteomic and genomic approach. *Hum Reprod*. 2008;23(8):1698-707.
29. Utleg AG, Yi EC, Xie T, Shannon P, White JT, Goodlett DR, et al. Proteomic analysis of human prostasomes. *Prostate*. 2003;56(2):150-61.
30. Gilany K, Van Elzen R, Mous K, Coen E, Van Dongen W, Vandamme S, et al. The proteome of the human neuroblastoma cell line SH-SY5Y: an enlarged proteome. *Biochim Biophys Acta*. 2008;1784(7-8):983-5.
31. Klose J. From 2-D electrophoresis to proteomics. *Electrophoresis*. 2009;30 Suppl 1:S142-9.
32. Jensen ON. Interpreting the protein language using proteomics. *Nat Rev Mol Cell Biol*. 2006;7(6):391-403.
33. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol*. 2003;21(3):255-61.
34. Gilany K, Vafakhah M. Hypoxia: a review. *J Paramedical Sci*. 2010;1(2):43-60.
35. HUPO: Human Proteome Organization. Human Proteome Project (HPP) [Internet]. Santa Fe, New Mexico, USA: HUPO [updated 2015; cited 2015 Feb 24]. Available from: <http://www.hupo.org/initiatives/human-proteome-project/>
36. Jangravi Z, Alikhani M, Arefnezhad B, Sharifi Tabar M, Taleahmad S, Karamzadeh R, et al. A fresh look at the male-specific region of the human Y chromosome. *J Proteome Res*. 2013;12(1):6-22.
37. Salekdeh GH. Human Y chromosome proteome project. *Int J Fertil Steril*. 2010;4(1):49.
38. Wang G, Guo Y, Zhou T, Shi X, Yu J, Yang Y, et al. In-depth proteomic analysis of the human sperm reveals complex protein compositions. *J Proteomics*. 2013;79:114-22.
39. Kovac JR, Pastuszak AW, Lamb DJ. The use of genomics, proteomics, and metabolomics in identifying biomarkers of male infertility. *Fertil Steril*. 2013;99(4):998-1007.
40. Kullisaar T, Turk S, Kilk K, Ausmees K, Punab M, Mandar R. Increased levels of hydrogen peroxide and nitric oxide in male partners of infertile couples. *Andrology*. 2013;1(6):850-8.
41. Pahune PP, Choudhari AR, Muley PA. The total antioxidant power of semen and its correlation with the fertility potential of human male subjects. *J Clin Diagn Res*. 2013;7(6):991-5.
42. Sharma R, Agarwal A, Mohanty G, Du Plessis SS, Gopalan B, Willard B, et al. Proteomic analysis of seminal fluid from men exhibiting oxidative stress. *Reprod Biol Endocrinol*. 2013;11:85.
43. Starita-Geribaldi M, Poggioli S, Zucchini M, Garin J, Chevallier D, Fenichel P, et al. Mapping of seminal plasma proteins by two-dimensional gel electrophoresis in men with normal and impaired spermatogenesis. *Mol Hum Reprod*. 2001;7(8):715-22.
44. Davalieva K, Kiprijanovska S, Noveski P, Plaseski T, Kocavska B, Broussard C, et al. Proteomic analysis of seminal plasma in men with different spermatogenic impairment. *Andrologia*. 2012;44(4):256-64.
45. Hassan MI, Kumar V, Kashav T, Alam N, Singh TP, Yadav S. Proteomic approach for purification of seminal plasma proteins involved in tumor proliferation. *J Sep Sci*. 2007;30(12):1979-88.
46. Batruch I, Smith CR, Mullen BJ, Grober E, Lo KC, Diamandis EP, et al. Analysis of seminal plasma from patients with non-obstructive azoospermia and identification of candidate biomarkers of male infertility. *J Proteome Res*. 2012;11(3):1503-11.
47. Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol*. 2008;4:222.
48. Pan S, Aebersold R, Chen R, Rush J, Goodlett DR, McIntosh MW, et al. Mass spectrometry based targeted protein quantification: methods and applications. *J Proteome Res*. 2009;8(2):787-97.
49. Drabovich AP, Jarvi K, Diamandis EP. Verification of male infertility biomarkers in seminal plasma by multiplex selected reaction monitoring assay. *Mol Cell Proteomics*. 2011;10(12):M110.004127.
50. Farrah T, Deutsch EW, Hoopmann MR, Hallows JL, Sun Z, Huang CY, et al. The state of the human proteome in 2012 as viewed through PeptideAtlas. *J Proteome Res*. 2013;12(1):162-71.
51. Yamakawa K, Yoshida K, Nishikawa H, Kato T, Iwamoto T. Comparative analysis of interindividual variations in the seminal plasma proteome of fertile men with identification of potential markers for azoospermia in infertile patients. *J Androl*. 2007;28(6):858-65.
52. Milardi D, Grande G, Vincenzoni F, Messana I, Pontecorvi A, De Marinis L, et al. Proteomic approach in the identification of fertility pattern in seminal plasma of fertile men. *Fertil Steril*. 2012;97(1):67-73.

53. Johansen E, Schilling B, Lerch M, Niles RK, Liu H, Li B, et al. A lectin HPLC method to enrich selectively-glycosylated peptides from complex biological samples. *J Vis Exp*. 2009;(32).
54. Seibert V, Ebert MP, Buschmann T. Advances in clinical cancer proteomics: SELDI-ToF-mass spectrometry and biomarker discovery. *Brief Funct Genomic Proteomic*. 2005;4(1):16-26.
55. Kearney P, Thibault P. Bioinformatics meets proteomics--bridging the gap between mass spectrometry data analysis and cell biology. *J Bioinform Comput Biol*. 2003;1(1):183-200.
56. Listgarten J, Emili A. Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. *Mol Cell Proteomics*. 2005;4(4):419-34.
57. Baldwin MA. Protein identification by mass spectrometry: issues to be considered. *Mol Cell Proteomics*. 2004;3(1):1-9.