The Advent of Sperm Proteomics has Arrived

Anthony H. Kashou[1], David J. Benjamin[1], Ashok Agarwal[1] and Stefan S. du Plessis[*,2]

[1]*Center for Reproductive Medicine, Cleveland Clinic, Cleveland, Ohio, USA*

[2]*Division of Medical Physiology, Stellenbosch University, Tygerberg, South Africa*

Abstract: Sperm proteomics is the identification and functional study of sperm proteins. It is based on the separation of proteins to generate a sample suitable for mass spectrometry and subsequent protein identification. Various proteomic approaches can be employed to study sperm proteins. Currently it has led to the identification and cataloging of thousands of sperm proteins. Ultimately, the goal is to apply sperm proteomics not only as a research method, but also as a clinical and diagnostic tool in the field of male infertility. This manuscript aims to review proteomics and the approaches used to analyze sperm proteins as well as put its application in context with some of the current findings.

Keywords: Spermatozoa, proteomics, proteins, mass spectrometry.

I. INTRODUCTION

Male gametes are highly specialized cells that get produced during the process of spermatogenesis in the testis. They represent a unique subtype of cells and differ dramatically from somatic cells in general. Their crucial function remains the successful delivery of the paternal complement of genome to the oocyte. Despite spermatozoa being highly accessible cells, more in-depth studies still remain to be done on these gametes in order to elucidate their sub-cellular composition and activities in relation to their activities and function.

Innovative methods and original technologies act as catalysts and driving forces for the expansion of knowledge with regards to systems and biological studies. One such emergent group of novel technologies that can be applied to study large sets of biological molecules in order to micro-measure the performance of cells at a given time is collectively referred to as omics. The term omics encompass the study of genes (genomics), transcript (transcriptomics), proteins (proteomics) and metabolites (metabolomics) [1]. These technologies allow for the identification and quantification of cellular components in a spatiotemporal fashion. What researchers once envisioned is now a reality; omics now allows for a transformation from once only genomic analysis to proteomic analysis. This approach offers an opportunity to investigate the relationship between an organism's genotype and resulting phenotype.

The specific field of proteomics allows for the measurement of the production of protein levels and describes the changes in all proteins expressed and translated from a single genome in a tissue or cell [2]. It aims to provide the expression levels of all proteins of one functional state in a biological system [3-5]. This enables researchers to gain

further insight on how proteins function in and around cells to fully comprehend biological processes and dissect the intracellular pathways. These advances may aid in clinical applications for disease diagnosis and potential therapy development [6, 7].

As reports continue to indicate that nearly 15% of reproductively active couples are infertile and that greater tendencies are due to male-factor defects [8, 9], the developing proteomic field offers much hope. Several studies have linked abnormal sperm parameters such as motility, morphology, and concentration, as well as DNA fragmentation and genetic defects to suboptimal sperm quality. However, current knowledge pertaining to spermatozoa and how such pathological and physiological effects are induced is limited. A deeper understanding of how these modifications arise may provide answers to some of the complications that bring about poor sperm quality.

Studying of the sub-cellular components at a molecular level might offer evidence to areas as to why phenotypical and functional defects may develop in spermatozoa. Since the spermatozoon is a highly accessible and intricate cell, proteomic analysis can provide information concerning not only sperm abnormalities, but also genetic defects that may arise in offspring. For these reasons, proteomics offers a promising approach that may ultimately uncover the underlying metabolic pathways causing male infertility. While findings remain limited, current advances in the field of mass spectrometry (MS) to better correlate the unique masses of amino acids to their corresponding protein sequence, have offered some insight in the characterization and identification of both the structural and functional proteins of spermatozoa [3-5, 7, 10-14].

A general understanding of proteomics and its potential will make for a greater comprehension of the promising outlook as many scientists and clinicians believe that it might be applied in sperm dysfunction diagnosis and clinical application. This review intends to compliment recent reviews and research articles in the field of sperm proteomics

*Address correspondence to this author at the Division of Medical Physiology, Stellenbosch University, P.O. Box 19063, Tygerberg, 7505, South Africa; Tel: +27(0)219389388; Fax: +27(0)219389476; E-mail: ssdp@sun.ac.za

by providing a glimpse into the novelty and potential of proteomic analysis, while exploring the advent of proteomics and its astonishing advancements and contributions made to sperm biology.

II. A SYNOPSIS OF PROTEOMICS

With an abundant amount of information attained, the complexity of interpreting findings and solving problems has reached a new level. The human genome, which codes for 35,000 genes, can encode for at least ten times as many proteins. In extreme cases a single gene alone may code for more than 1,000 proteins [2]. Moreover it is believed that the number of proteins that is currently needed to be identified has multiplied; yet, the resolving capacities of microscopes are still incapable of viewing proteins on a peptide level.

Innovative approaches have led molecular biologists to make extraordinary discoveries. From gene cloning to sequencing to expressional analysis, the human genome was decoded. This groundbreaking news marked a turning point in the field of genomics. However, as research continued, investigators began to realize that the human genome was only the first step in deciphering the biochemical functions of each gene. To add to the challenges, comprehending human disease was much more complex than expected.

Amidst the mid-1990s, scientists attempted to draw a fine relationship between protein and gene expression of cells (Fig. **1**). Their inquisitiveness brought about novel advances in high-throughput techniques that allowed for successful sequencing analysis, reversible approaches that primarily focused on phenotype instead of genotype, and the development of gene sequencing libraries. It was believed that if a link between genotype and phenotype could be established, it might be possible to examine a number of proteins simultaneously. The tremendous amount of information gained would provide a deeper understanding of not only protein

structure and function, but also their biochemical role throughout the body.

Ever since this initial drive to advance the study of proteins, the field of proteomics evolved rapidly. This transformation from genomics to proteomics brought much excitement [15]. Scientists would use genomic data to compile a complete profile of proteins expressed in a biological system at a given time. It is now thought to be one of the most promising fields in providing analysis of an individual's health. Scrutiny of proteins in the blood can reflect a healthy or disease condition; hence, examining biological fluids at the site of disease may enhance efforts in therapy development. Nevertheless, as with any new innovation, proteomics brought its own challenges and disadvantages. The human proteome, estimated at over a million proteins presented researchers with an abundance of both diverse and dynamic proteins and more work than imagined. Alternative splicing of pre-mRNA (introns and exons), post-translation modifications (phosphorylation, glycosylation, hydroxylation, etc.) of proteins, as well as cell-age and health are only a few of the several tasks posed.

There are predominantly two basic proteomic approaches to focus on proteins expressed by the genome of the sperm cell namely: (i) Two dimensional (2D) electrophoresis for separation of proteins followed by matrix assisted laser desorption ionisation MS (MALDI-MS) or liquid chromatography (LC) followed by tandem MS (LC-MS/MS); (ii) initial digestion of proteins to generate peptides followed by LC-MS/MS analysis [3, 7, 16]. All of these methodologies consist basically of two principle steps, i.e.: 1) separation of proteins to provide a sample with decreased complexity and 2) MS used for protein identification.

The basic process begins with protein extraction from the samples of interest (Fig. **2**). This initial step is critical as preparation is sample specific and often less than 1mg of protein is attained. The protein mixture is then separated by
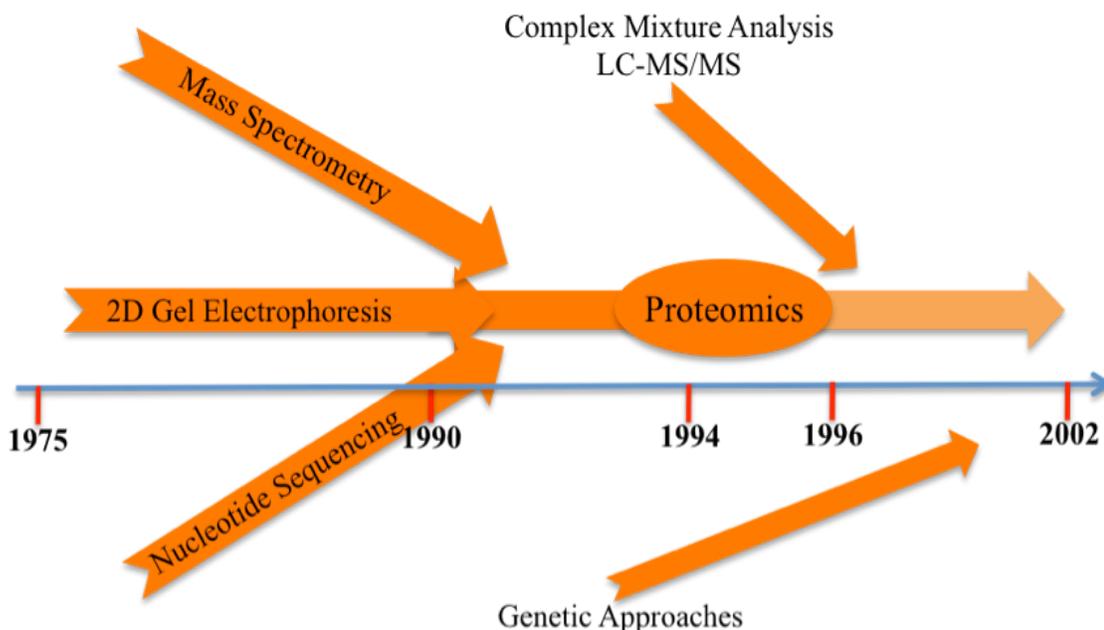

Fig. (1). The onset of proteomics. This timeline illustrates the evolution of major proteomic techniques that revolutionized the field.

electrophoresis or chromatography. High-resolution separation is essential due to the complex nature of the samples. Two-dimensional gel electrophoresis (2-DE) is the most common separation technique utilized and has led to the separation and subsequent identification of many proteins present in spermatozoa [17, 18]. It allows for the highest-resolution protein separation and is based on differences in intrinsic charge and size. Variations in spot intensities between control and test sample gels can also be used to indicate different levels of protein expression. The resolved protein spots are then excised and subjected to "in-gel" tryptic digestion, where after they can then be quantified MS. This is a key step in the characterization of the protein. MS analysis provides a surrogate measure of protein abundance, thereby permitting determination of the elemental composition and chemical structure of the sample. Alternatively, LC-MS/MS may be employed for complex mixture analysis. In LC-MS/MS, the entire protein mixture is digested, fractionated, and subjected to mass spectrometry. The separation of proteins or peptides by HPLC is based on unique properties such as charge, size, hydrophobicity and presence of specific tags or amino acids [3].
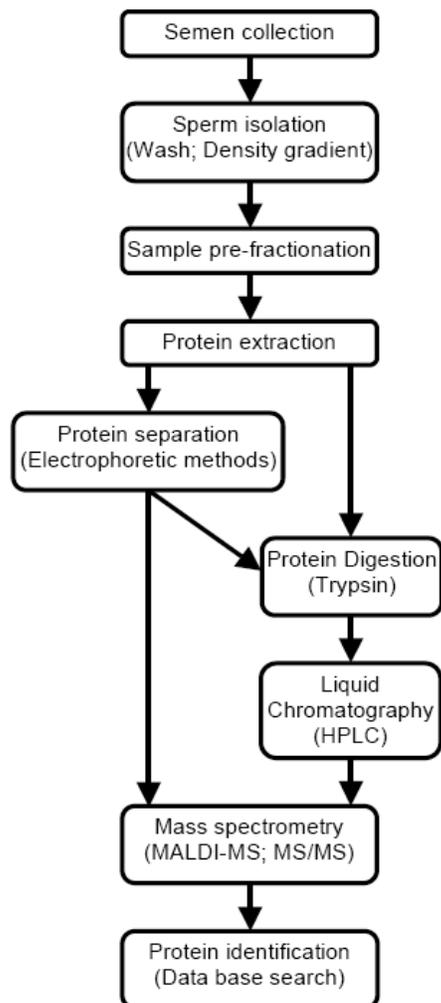

Fig. (2). The basic proteomic process from protein extraction to analysis. (HPLC = high pressure liquid chromatography; MS = mass spectrometry; MALDI-MS = matrix assisted laser desorption ionization mass spectrometry).

Each of these different approaches and procedures has its own benefits and disadvantages. The preservation of the epigenetic code is an advantage of 2-D separation, which is lost when the crude protein extract is digested to peptides. Conversely LC-MS/MS methods are more robust and allow for a higher throughput [3, 6, 7]. Subsequently the data retrieved can then be entered into a data base to allow protein identification by means of bioinformatics. Two-dimensional difference in gel electrophoresis (2-D-DIGE) provides for an additional and relatively newer method. It is based on differential labeling of proteins extracted from control and experimental samples with fluorochromes [3, 19].

III. SPERM PROTEOMICS

Sperm proteomics intends to identify the proteins that make up the sperm cell in order to study their function [3]. The spermatozoon is a highly intricate yet accessible cell. Its high purification capacity has allowed for it to be particularly suitable for proteomic analysis. As an extremely differentiated cell resulting from the process of spermatogenesis, the spermatozoon is marked with a number of cellular, genetic, chromatin, and functional changes [7]. Furthermore, since male defects have been associated with lowered fecundity, the importance of the studying sperm and their role in infertility cannot be understated.

Catalogs of thousands of human spermatozoal proteins as well as proteins from different model species are becoming available [6]. The protein inventory lists being generated have shed light on transmembrane proteins, kinases and chaperones never previously recognized [20]. Recent advances in proteomic techniques have provided deeper insight into mature human spermatozoa and their role in fertilization (e.g. transcription factors, DNA-binding proteins, proteins involved in chromatin metabolism), as well as the capability of identifying the underlying proteins contributing to sperm dysfunction [6, 7, 20, 21]. A recent study by Martinez-Heredia *et al.* identified 98 different protein spots in which 23% of them had yet to be recorded [22].

Several proteomic studies have revealed that sperm proteins take part in crucial processes such as glycolysis, oxidative phosphorylation, and sperm capacitation [20, 23]. Proteomic studies are being undertaken to characterizing tyrosine phosphorylated proteins in order to understand the processes involved in sperm capacitation. A recent study by Kota and colleagues found that glycerol-3-phosophate dehydrogenase (GP2D) is one of the proteins that enables tyrosine phosphorylation during sperm capacitation in hamster spermatozoa [24]. Future proteomic studies must be done to fully characterize the human sperm proteome in order to understand the differences between the various stages of spermatozoa, and to identify post-translational modifications that may lead to suboptimal sperm quality.

The ability to quantify the differences between cells of two different populations gives proteomic analysis of spermatozoa a real chance to solve for age-old questions [25]. Subsequently comparative studies, using proteomic approaches, between the spermatozoa of infertile and fertile men have determined biomarkers which represent proteins of interest that may be involved in suboptimal sperm function. In one of the latest studies to be published, Siva and co-

workers compared men with asthenozoospermia to healthy normozoospermic men who served as controls [26]. The study identified 75 different sperm proteins, while comparative proteome analysis revealed eight sperm proteins with either increased or decreased frequency in the asthenozoospermic samples. The proteins could be grouped into three functional groups: proteins involved in sperm energy and metabolism, sperm movement and organization, and protein turnover, folding and stress response. Another recent comparative study by Kreigel *et al.* evaluated differences in the sperm proteome between type 1 diabetics, non-diabetic obese individuals and healthy fertile donors [27]. The study, which employed DIGE and MS, identified 20 spots of significant increased or decreased frequency in patients with either diabetes or who were obese; 8 of these protein spots were linked to type 1 diabetes, while 12 spots were related to obesity. Such comparative studies currently undertaken can contribute appreciably to the understanding of sperm motility, and may help identify biomarkers to better understand and diagnose male infertility. The identification of differentially expressed proteins between fertile and infertile patients could provide an opportunity to study candidate genes in search for mutations, while it in turn can be clinically related to DNA repair and integrity as well as oxidative stress [7, 18, 28-34].

Another study, employing a differential proteomic approach aimed at identifying proteins acquired by rat spermatozoa during epididymal transit, noted a total of 335 protein spots. Of these spots 140 could be identified by MS; nine of these were novel. Most of the proteins identified were primarily involved in metabolic processes. Such comprehensive data of domain specific epididymal sperm proteins can be useful in understanding the development of fertilizing ability as well as in the development of newer targets for post testicular contraception and diagnostic markers for infertility [35]. Similarly protein expression profiles of sperm from high and low fertility bulls were compared. Comparative proteomic analysis showed that expression of at least nine different proteins was significantly related to fertility. These identified proteins are involved in sperm-egg interactions and cell cycle regulation. Studies of this nature are critical for the identification of fertility protein biomarkers [36].

Although much is still unknown about the proteome of human spermatozoa, ongoing and future proteomic studies may help to fully and completely characterize the human sperm proteome. Future areas of interest include better understanding the number of post-translational modifications (PTMs) that sperm proteins undergo, as well as conducting comparative studies between fertile and subfertile men in order to identify biomarkers [19]. Proteomic approaches for the study of spermatozoa holds much promise to better understand and treat male-factor infertility.

IV. CONCLUSION

As infertility rates are on the increase globally, there has been increased focus on male-factor defects and in particular, abnormal sperm function. In recent years, proteomic studies of human spermatozoa, using multi-dimensional techniques, have helped to better understand sperm by identifying several thousands of proteins. Some of these proteins have been identified as potential causes of suboptimal sperm quality. Several sperm protein databases are already available and provide important reference for further studies [7]. This represents only the beginning of a new era as future studies will continue to try and determine the entire sperm proteome. While research help identify biomarkers with the aid of bioinformatics, there is hope that novel therapies and clinical tests can be developed to better treat male-factor infertility due to sperm dysfunction. Further studies will help to confirm the advantageous application of sperm proteomics testing as a diagnostic tool for male factor infertility. The challenge is to correlate gene/transcript/protein/metabolite function and regulation with specific events in spermatozoa. While each technique is powerful in itself, the combination of these omics techniques will revolutionize sperm biology.

V. EXPERT COMMENTARY

Proteomic-based approaches have evolved from 1D separation techniques to 2D techniques, as well as the use of mass spectrometry and other tools. Study of the sperm proteome provides promise to clinicians and researchers by determining which metabolic pathways may affect sperm quality. Although the proteomic approach faces numerous challenges, including the laborious nature of such techniques as well as the difficulty in protein identification, new tools are being developed to aid in proteomic studies. The field of bioinformatics can provide the statistical support to identify proteins of interest so that novel treatments and drugs can be developed to improve sperm function.

VI. FIVE-YEAR REVIEW

There has been an increase in study of the human spermatozoa proteome using a proteomics-based approach in the past few years. Although recent studies have identified up to several thousands of protein spots using 2D techniques, much of the human sperm proteome remains unknown. Comparative studies have also helped to identify proteins of interest that may be involved in suboptimal sperm quality. A similar approach may help to identify proteins that contribute to male-factor infertility. Future proteomic studies should be undertaken to determine the entire human sperm proteome, as well as to provide a better understanding of the effects that PTMs have on sperm quality. Furthermore, future comparative studies may help to identify proteins that contribute to male-factor infertility. With a better understanding of the protein makeup of the human spermatozoa, there is the possibility of improved diagnosis and treatment therapies for male-factor infertility.

VII. KEY POINTS

- Male-factor defects, including abnormal sperm quality due to morphology, motility and concentration, significantly contribute to infertility among couples of reproductive age.

- Proteomics is the study of proteins encoded by a genome; recently, proteomic techniques have been

applied to characterize and better understand the human sperm proteome.

- Several techniques that are primarily used in proteomic research include 2-DE and MS.

- The emerging field of bioinformatics provides statistical analyses in comprehending and making sense of the vast amount of data generated through proteomic-based approaches, and hence, a better understanding of the sperm proteome.

- While early and ongoing proteomic studies into the human spermatozoa have characterized several thousand sperm proteins, a large number of proteins have yet to be identified.

- Studies of human sperm using proteomic approaches have highlighted the importance and impact of PTMs on sperm quality.

- Comparative studies between fertile and infertile men have identified proteins that may prove to be biomarkers linked to suboptimal sperm quality.

- As infertility rates continue to increase, the field of proteomics offers tremendous promise in better understanding the causes of male-factor infertility and potential therapeutic development.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

None Declared.

REFERENCES

[1]   Hood L, Heath JR, Phelps ME, Lin B. Systems biology and new technologies enable predictive and preventative medicine. Science (New York) 2004; 306(5696): 640-3.

[2]   Venter JC, Adams MD, Myers EW, *et al.* The sequence of the human genome. Science (New York) 2001; 291(5507): 1304-51.

[3]   Oliva R, De Mateo S, Castillo J, Azpiazu R, Oriola J, Ballesca JL. Methodological advances in sperm proteomics. Hum fertil (Cambridge, England) 2010 Dec**;**13(4): 263-7.

[4]   Domon B, Aebersold R. Mass spectrometry and protein analysis. Science (New York) 2006; 312(5771): 212-7.

[5]   Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 2003; 422(6928): 198-207.

[6]   Oliva R, Castillo J. Proteomics and the genetics of sperm chromatin condensation. Asian J Androl 2011; 13(1): 24-30.

[7]   Oliva R, de Mateo S, Estanyol JM. Sperm cell proteomics. Proteomics 2009; 9(4): 1004-17.

[8]   Stephen EH, Chandra A. Updated projections of infertility in the United States: 1995-2025. Fertil Steril 1998; 70(1): 30-4.

[9]   Thonneau P, Marchand S, Tallec A, *et al.* Incidence and main causes of infertility in a resident population (1,850,000) of three French regions (1988-1989). Hum Reprod (Oxford, England) 1991; 6(6): 811-6.

[10]   Bailey JL, Tardif S, Dube C, *et al.* Use of phosphoproteomics to study tyrosine kinase activity in capacitating boar sperm. Kinase activity and capacitation. Theriogenology 2005; 63(2): 599-614.

[11]   Lefievre L, Barratt CL, Harper CV, *et al.* Physiological and proteomic approaches to studying prefertilization events in the human. Reprod Biomed Onlin**e** 2003; 7(4): 419-27.

[12]   Barratt CL. The human sperm proteome: the potential for new biomarkers of male fertility and a transformation in our understanding of the spermatozoon as a machine: commentary on the article 'Identification of proteomic differences in asthenozoospermic sperm samples' by Martinez *et al.* Hum Reprod (Oxford, England) 2008; 23(6): 1240-1.

[13]   Oliva R, Martinez-Heredia J, Estanyol JM. Proteomics in the study of the sperm cell composition, differentiation and function. Syst Biol Reprod Med 2008; 54(1): 23-36.

[14]   Aitken RJ, Baker MA. The role of proteomics in understanding sperm cell biology. Int J Androl 2008; 31(3): 295-302.

[15]   Brewis IA, Gadella BM. Sperm surface proteomics: from protein lists to biological function. Mol Hum Reprod 2009; 16(2): 68-79.

[16]   Baker MA, Aitken RJ. Proteomic insights into spermatozoa: critiques, comments and concerns.   Expert Rev Proteomics 2009; 6(6): 691-705.

[17]   Com E, Evrard B, Roepstorff P, Aubry F, Pineau C. New insights into the rat spermatogonial proteome: identification of 156 additional proteins. Mol Cell Proteomics 2003; 2(4): 248-61.

[18]   Pixton KL, Deeks ED, Flesch FM, *et al.* Sperm proteome mapping of a patient who experienced failed fertilization at IVF reveals altered expression of at least 20 proteins compared with fertile donors: case report. Hum Reprod (Oxford, England) 2004; 19(6): 1438-47.

[19]   Baker MA, Witherdin R, Hetherington L, Cunningham-Smith K, Aitken RJ. Identification of post-translational modifications that occur during sperm maturation using difference in two-dimensional gel electrophoresis. Proteomics 2005; 5(4):1003-12.

[20]   Baker MA, Reeves G, Hetherington L, Muller J, Baur I, Aitken RJ. Identification of gene products present in Triton X-100 soluble and insoluble fractions of human spermatozoa lysates using LC-MS/MS analysis. Proteomics Clin Appl 2007; 1(5): 524-32.

[21]   Oliva R. Protamines and male infertility. Hum Reprod Updat 2006; 12(4): 417-35.

[22]   Martinez-Heredia J, Estanyol JM, Ballesca JL, Oliva R. Proteomic identification of human sperm proteins. Proteomics 2006; 6(15): 4356-69.

[23]   Johnston DS, Wooters J, Kopf GS, Qiu Y, Roberts KP. Analysis of the human sperm proteome. Ann N Y Acad Sci 2005; 1061: 190-202.

[24]   Kota V, Rai P, Weitzel JM, Middendorff R, Bhande SS, Shivaji S. Role of glycerol-3-phosphate dehydrogenase 2 in mouse sperm capacitation. Mol Reprod Develop 2010; 77(9): 773-83.

[25]   Baker MA. The 'omics' revolution and our understanding of sperm cell biology. Asian j androl 2011; 13(1): 6-11.

[26]   Siva AB, Kameshwari DB, Singh V, *et al.* Proteomics-based study on asthenozoospermia: differential expression of proteasome alpha complex. Mol Hum Reprod 2010; 16(7): 452-62.

[27]   Kriegel TM, Heidenreich F, Kettner K, Pursche T, Hoflack B, Grunewald S, *et al.* Identification of diabetes- and obesity-associated proteomic changes in human spermatozoa by difference gel electrophoresis. Reprod Biomed Online 2009; 19(5): 660-70.

[28]   Seli E, Sakkas D. Spermatozoal nuclear determinants of reproductive outcome: implications for ART. Hum Reprod Update 2005; 11(4): 337-49.

[29]   de Mateo S, Martinez-Heredia J, Estanyol JM, *et al.* Marked correlations in protein expression identified by proteomic analysis of human spermatozoa. Proteomics 2007; 7(23): 4264-77.

[30]   Martinez-Heredia J, de Mateo S, Vidal-Taboada JM, Ballesca JL, Oliva R. Identification of proteomic differences in asthenozoospermic sperm samples. Hum Reprod (Oxford, England) 2008; 23(4): 783-91.

[31]   Huo R, He Y, Zhao C, Guo XJ, Lin M, Sha JH. Identification of human spermatogenesis-related proteins by comparative proteomic analysis: a preliminary study. Fertil Steril 2008; 90(4): 1109-18.

[32]   Baarends WM, van der Laan R, Grootegoed JA. DNA repair mechanisms and gametogenesis. Reproduction (Cambridge, England) 2001; 121(1): 31-9.

[33]   Oehninger S, Morshedi M, Weng SL, Taylor S, Duran H, Beebe S. Presence and significance of somatic cell apoptosis markers in human ejaculated spermatozoa. Reprod Biomed Online 2003; 7(4): 469-76.

[34]   Aitken RJ, Krausz C. Oxidative stress, DNA damage and the Y chromosome. Reproduction (Cambridge, England) 2001; 122(4): 497-506.

[35] Suryawanshi AR, Khan SA, Gajbhiye RK, Gurav MY, Khole VV. Differential proteomics leads to identification of domain specific epididymal sperm proteins. J Androl 2011; 32(3):240-59.

[36] Gaviraghi A, Deriu F, Soggiu A, *et al.* Proteomics to investigate fertility in bulls. Vet Res Commun 2010; 34 (Suppl 1): 33-6.