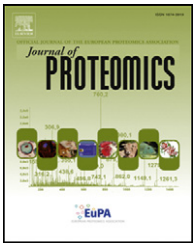
JOURNAL OF PROTEOMICS 79 (2013) 114–122

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot



In-depth proteomic analysis of the human sperm reveals complex protein compositions

Gaigai Wang[1], *Yueshuai Guo*[1], *Tao Zhou*[1], *Xiaodan Shi*[1], *Jun Yu*, *Ye Yang*, *Yibo Wu*, *Jing Wang*, *Mingxi Liu*, *Xin Chen*, *Wenjiao Tu*, *Yan Zeng*, *Min Jiang*, *Suying Li*, *Pan Zhang*, *Quan Zhou*, *Bo Zheng*, *Chunmei Yu*, *Zuomin Zhou*, *Xuejiang Guo*[*], *Jiahao Sha*[*]

State Key Laboratory of Reproductive Medicine, Department of Histology and Embryology, Nanjing Medical University, Nanjing 210029, China

ARTICLE INFO

Article history:
Received 2 October 2012
Accepted 7 December 2012

Keywords:
Sperm
Proteome
Mass spectrometry
Signaling pathway
Drug
Human

ABSTRACT

The male gamete (sperm) can fertilize an egg, and pass the male genetic information to the offspring. It has long been thought that sperm had a simple protein composition. Efforts have been made to identify the sperm proteome in different species, and only about 1000 proteins were reported. However, with advanced mass spectrometry and an optimized proteomics platform, we successfully identified 4675 human sperm proteins, of which 227 were testis-specific. This large number of identified proteins indicates the complex composition and function of human sperm. Comparison with the sperm transcriptome reveals little overlap, which shows the importance of future studies of sperm at the protein level. Interestingly, many signaling pathways, such as the IL-6, insulin and TGF-beta receptor signaling pathways, were found to be overrepresented. In addition, we found that 500 proteins were annotated as targets of known drugs. Three of four drugs studied were found to affect sperm movement. This in-depth human sperm proteome will be a rich resource for further studies of sperm function, and will provide candidate targets for the development of male contraceptive drugs.

1. Introduction

Approximately one in six couples experience difficulty in conceiving a child. Male infertility accounts for about half the cases in which assisted reproductive techniques are recommended [1]. Over 85% of infertile male can actually produce sperm [2]; however, for some reason, those sperm are often unable to fertilize an egg.

As the only cell performing its function outside the male human body, sperm is a highly specialized cell with distinct morphological and compositional differences compared with other somatic and germ cells [3]. It was long believed that the function of sperm was only to deliver the paternal genomes to the egg. However, recent studies have shown that sperm can deliver a complex set of RNAs to the egg [4]. In addition, the entire cell, including the midpiece and tail, enters the egg in most species [5]. Current studies have suggested that sperm defects can disrupt embryo development, even if the genome carried by the cells is perfectly normal [6]. Thus, characterization of the protein composition of sperm can help better understand sperm function.

The recently proposed Chromosome-Centric Human Proteome Project (C-HPP) aims to define the full set of proteins encoded in each chromosome. The initial goal of the C-HPP is to identify at least one representative protein encoded by each of the approximately 20,300 human genes [7]. The genes and

* *Corresponding authors*. Tel.: +86 25 86862038; fax: +86 25 86862908.
E-mail addresses: guo_xuejiang@njmu.edu.cn (X. Guo), shajh@njmu.edu.cn (J. Sha).
[1] These authors contributed equally to the work.

1874-3919/$ – see front matter.
http://dx.doi.org/10.1016/j.jprot.2012.12.008

proteins had tissue-dependent expression. According to previous analyses, the human testis, the male gonad producing sperm, contains the largest number of tissue-specific genes across the 31 human tissues [8]. Thus, in order to characterize all the proteins, proteins in the testis and/or testicular cells should be carefully studied. The in-depth proteomic analysis of human sperm will produce data important for C-HPP.

To date, some efforts have been made to identify the human sperm proteome, and in fact, a list of 1056 proteins has been reported [9]. Additionally, Johnston et al. [10] claimed identification of 1760 proteins in human sperm, but this protein list is not available. Efforts to identify the sperm proteome in other species, including drosophila and mammals, such as rat and mouse [11,12], have also been made, and the numbers of proteins identified were all around or below one thousand. In the present study, using the advanced mass spectrometry and an optimized proteomics platform, we successfully identified 4675 unique proteins from human sperm, which showed the complex composition of human sperm.

2. Materials and methods

2.1. Sperm collection

This human study was ratified by the Ethics Committee of Nanjing Medical University, and was in accordance with National and International guidelines. Before initiating the study, consent was obtained from all participants. The sperm subjected to proteomics analysis were from 32 healthy male volunteers with a mean age of 30±4 years old (mean±standard deviation). These men had proven fertility and normal semen quality, as assessed by World Health Organization criteria (1999).

The semen samples were obtained by masturbation after at least 3 days of abstinence. The samples were ejaculated into sterile containers and allowed to liquefy for at least 30 min before being processed by centrifugation in a 60% Percoll gradient (GE Healthcare, Waukesha, WI, USA) to remove seminal plasma, immature germ cells, and non-sperm cells (mainly epithelial cells), as described by Loredana-Gandini et al. [13,14]. The purified sperms were then washed in PBS three times before subsequent proteomics analysis. For purity evaluation, the sperm were resuspended in PBS and stained with Hoechst H33342 (Sigma, St Louis, MO) for 30 s; 1000 cells were counted by light microscopy.

2.2. Sample preparation for mass spectrometry

Human sperm were dissolved in 7 M urea, 2 M thiourea, 65 mM DTT, and 1% (v/v) protease inhibitor cocktail, and the extracted proteins from different men were mixed for subsequent proteomic studies. Proteins of 240 μg were reduced, alkylated and sequentially digested with modified trypsin (sequencing grade, Promega, Madison, WI). These in-solution digests were loaded onto a strong-cation exchange column (1 mm ID×10 cm, packed with Poros 10S, Dionex, Sunnyvale, CA) for fractionation. A linear salt gradient ammonium formate in 5% acetonitrile (ACN) was applied at a flow-rate of 50 μl/min. Mobile phase A=95:5 H2O:ACN, 5 mM ammonium formate buffer pH=2.7, mobile phase B=mobile phase A+800 mM ammonium formate, pH=2.7. The gradient used was 0–56% B for 20 min, 56% to 100% B for 1 min, 100% B for 5 min, 100% to 0% B for 1 min, and 0% B for 20 min before the next run. In each series of experiments, 100 μl fractions were collected every 2 min, and 20 fractions were obtained. The experiments were repeated for 3 times.

2.3. Mass spectrometric analysis and database search

For capillary reverse-phase liquid chromatography (LC) and mass spectrometric analysis, each fraction was directly loaded onto a μ-precolumn™ cartridge (0.3×5 mm, 5 μm, 100 Å; Dionex) at a flow rate of 20 μl/min. The trap column effluent was then transferred to a reverse-phase microcapillary column (0.075×150 mm, Acclaim® PepMap100 C18 column, 3 μm, 100 Å; Dionex). The reverse-phase separation of peptides was performed using the following buffers: 2% ACN, 0.5% acetic acid (buffer A) and 80% ACN, 0.5% acetic acid (buffer B); a 122 or 82-min ACN gradient (4% to 7% buffer B for 3 min, 7% to 33% buffer B for 102 min or 62 min, 33% to 50% buffer B for 10 min, 50% to 100% buffer B for 3 min, 100% buffer B for 3 min, 100% to 4% buffer B for 1 min) was used. Peptide analysis was performed using a LTQ Orbitrap Velos (ThermoFisher Scientific, San Jose, CA) coupled directly to an LC column. An MS survey scan was obtained for the m/z range 350–1800, and MS/MS spectra were acquired from the survey scan for the 20 most intense ions (as determined by Xcalibur mass spectrometer software in real time). Dynamic mass exclusion windows of 60 s were used, and siloxane (m/z 445.120025) was used as an internal standard.

RAW files for LC–MS/MS identification were processed by MaxQuant (v1.2.2.5), and identified with the Andromeda search engine according to standard workflow [15]. The peak lists were searched against the UniProtKB human proteome sequence database (2012/04/18), which contains 86,770 entries. Carbamidomethylation of cysteine (+57 Da) was set as a fixed modification, and oxidized methionine (+16 Da) was set as a variable modification. The initial mass tolerances for protein identification on MS and MS/MS peaks were 20 ppm and 0.5 Da, respectively. Two missed cleavages were permitted, and full cleavage by trypsin was used. The false discovery rate (FDR) of the identified peptides and proteins was estimated by searching against the database with the reversed amino acid sequence [15]. Only peptides that were a minimum of six amino acids in length and had a FDR of 1% were considered for identification.

2.4. Human sperm proteome annotation

For bioinformatics analysis, the international protein index (IPI) accession number was converted to an Entrez Gene ID or Ensembl Gene ID. All Ensembl Gene IDs were loaded onto the Database for Annotation, Visualization and Integrated Discovery (DAVID) [16] to identify the enriched biological themes, including Gene Ontology. A FDR of less than 0.05 was considered statistically enriched. The Entrez Gene IDs were loaded onto a Web-based Gene Set Analysis Toolkit (http://bioinfo.vanderbilt.edu/webgestalt/) [17] to identify the hyper-represented WikiPathways; a FDR of less than 0.05