Deep Coverage Proteomics Identifies More Low-Abundance Missing Proteins in Human Testis Tissue with Q-Exactive HF Mass Spectrometer

Wei Wei, Weijia Luo, Feilin Wu, Xuehui Peng, Yao Zhang, Manli Zhang, Yan Zhao, Na Su, Yingzi Qi, Lingsheng Chen, Yangjun Zhang, Bo Wen, Fuchu He, and Ping Xu

J. Proteome Res., Just Accepted Manuscript • DOI: 10.1021/acs.jproteome.6b00390 • Publication Date (Web): 18 Aug 2016

Downloaded from http://pubs.acs.org on August 25, 2016
Deep Coverage Proteomics Identifies More Low-Abundance Missing Proteins in Human Testis Tissue with Q-Exactive HF Mass Spectrometer

Wei Wei1*, Weijia Luo1,2*, Feilin Wu1,3*, Xuehui Peng1,4*, Yao Zhang1,5, Manli Zhang1, Yan Zhao1, Na Su1, YingZi Qi1, Lingsheng Chen1, Yangjun Zhang1, Bo Wen6, Fuchu He*, Ping Xu1,2,4*

1State Key Laboratory of Proteomics, National Center for Protein Sciences Beijing, Beijing
Proteome Research Center, National Engineering Research Center for Protein Drugs, Beijing
Institute of Radiation Medicine, Beijing 102206, China

2Graduate School, Anhui Medical University, Hefei 230032, China

3Life Science College, Southwest Forestry University, Kunming, 650224, China

4Key Laboratory of Combinatorial Biosynthesis and Drug Discovery of the Ministry of Education,
School of Pharmaceutical Sciences, Wuhan University, Wuhan 430072, China

5Institute of Microbiology, Chinese Academy of Science, Beijing 100101, China

6BGI-Shenzhen, Shenzhen 518083, China

*These authors contributed equally to this work.

*To whom correspondence should be addressed:

Ping Xu
Beijing Proteome Research Center, 38 Science Park Road, Changping District, Beijing 102206, China.
Tel and Fax: 86-10-80705066; E-mail: xuping_bpbc@126.com.
Fuchu He

Beijing Institute of Radiation Medicine, 27 Taiping Road, Beijing 100850, China

Tel and Fax: 86-10-68171208; E-mail: hefc@nic.bmi.ac.cn

ABSTRACT

Since 2012, missing proteins (MPs) investigation has been one of the critical missions of Chromosome-Centric Human Proteome Project (C-HPP) through various biochemical strategies. Based on our previous testis MPs study, faster scanning and higher resolution mass spectrometry-based proteomics might be conducive to MPs exploration, especially for low-abundance proteins. In this study, Q-Exactive HF (HF) was used to survey proteins from the same testis tissues separated by two separating methods (Tricine- and Glycine-SDS-PAGE) as previously. A total of 8,526 proteins were identified, of which more low-abundance proteins were uniquely detected in HF data but not in our previous LTQ Orbitrap Velos (Velos) reanalysis data. Further transcriptomics analysis showed that these uniquely identified proteins by HF also had lower expression at the mRNA level. Of total identified 81 MPs, 74 and 39 proteins were listed as MPs in HF and Velos datasets, respectively. Among the above MPs, 47 proteins (43 neXtProt PE2 and 4 PE3) were ranked as confirmed MPs after the verifying with the stringent spectra match, isobaric and single amino acid variants (SAAVs) filtering. Functional investigation of these 47 MPs revealed that 11 MPs were testis-specific proteins and 7 MPs involved in spermatogenesis process. Therefore, we concluded that higher scanning speed and resolution of HF might be one of factors for improving the low-abundance MPs identification in future C-HPP studies. All mass spectrometry data from this study have been deposited in the ProteomeXchange with identifier
PXD004092.

**KEYWORDS**

Chromosome-Centric Human Proteome Project, Missing proteins, testis, Q-Exactive HF, LTQ Orbitrap Velos

#### INTRODUCTION

Following steps in the Human Proteome Project (HPP), the main mission of the Chromosome-Centric Human Proteome Project (C-HPP) is still to explore the protein evidence of “missing proteins” (MPs)\(^1\). According to the neXtProt database (2016-02), there are 2,949 protein-coding genes that have not reliable evidence at the protein level. Previous research showed that various factors led to MPs difficult identification. Firstly, the intrinsic properties of MPs might be the major reasons, including its expressing abundance, molecular weight (MW) and hydrophobicity\(^3\). Therefore, various analytical and experimental methods were optimized to solubilize, capture, and characterize special proteins, especially low-molecular-weight (LMW), membrane proteins and post-translational modification (PTM) proteins \(^4,5\).

In MPs searching project, mass spectrometry has been a common way following by antibody-based validation, bioinformatics and RNC-mRNA (the mRNA attached to ribosome-nascent chain complex) strategies. With the advance of technology, new generation of mass spectrometer has been optimized in many properties, such as scanning rate, resolution, sensitivity and dynamic range. Nowadays, Q-Exactive HF (HF), a widely used mass spectrometer in all over the world, is a new segmented quadrupole mass analyzer with smaller ultra-high-field Orbitrap and allows increasing in the electric field and detection frequency. The compact Orbitrap
of HF allows stronger field strength and shorter acquisition time or transient length, which theoretically provides 1.8 higher resolution at the same transient length. At the same time, the capability of parallel acquisition of HF results in faster-scan cycles, which significantly increases the peptide and protein identification. Additionally, new segmented quadrupole provides an double-fold improvement of transmission speed, which also contributes to the scanning-rate of HF. In HeLa cell phosphoproteomics, Kelstrup found that the sequencing speed of HF routinely exceeds 10 peptide spectrum per second. Therefore, the high resolution HF might be a power tool in MPs study.

Secondly, tissue-specific expression or special modification proteins are also the rich resources for MPs searching. Following the pace of Human Protein Atlas with antibodies, various samples from different tissues were used (e.g., the liver, brain, kidney, pancreas, skin and adipose tissue et al.) to discover MPs. Interestingly, tissue transcript analysis showed that the testis tissue endows plentiful tissue-specific genes and abundant transcripts of the MPs. The gene expression analysis showed that 77% of protein-code genes were expressed in the testis with more than 1,000 genes showing an at least 5-fold higher expression than other tissues. Fagerberg found that 8% of all genes were uniquely detected in the testis but not in any others tissues. All these researches suggest that testis is a high gene expression tissue, which may endows more MPs. Last year, Zhang and Jumeau reported 166 and 89 MPs from human testis and spermatozoa samples, respectively, corresponding to 5.6% and 3% of total 2,948 MPs. Thus, we believe that the testis is still a reservoir for MPs searching.

In this study, we efficiently identified more MPs by HF in human testis samples isolated by Glycine- and Tricine-SDS-PAGE. A total of 8,526 proteins were identified by HF, while 7,537
proteins in our previous Velos datasets. Total 1,230 proteins had low abundance, which were uniquely identified by HF. Transcriptome analysis also found that these uniquely identified proteins had low mRNA abundance property. Total 81 MPs were detected, in which 74 and 39 MPs were identified by HF and Velos, respectively. Through strict manual spectra checking, verifying with synthesized-peptide matching, isobaric filtering and SAAVs filtering, 47 of 81 MPs were confidently verified. Function investigation suggested that most MPs were testis-specific and related with spermatogenesis. HF, an advanced mass spectrometer, might be a powerful tool for finding more low-abundance MPs.

- MATERIALS AND METHODS

Samples Used in this Study

The same human testis samples (IRB approval number BGI-IRB15076) were processed as previously described. Briefly, three testis tissue samples (50 mg) were put in liquid nitrogen and sonicated in lysis buffer (8 M urea, 5 mM Iodoacetamide (IAA), 50 mM NH₄HCO₃, 1×protease cocktail) on ice. The cell debris was removed by centrifugation at 13,300g in 4°C for 15 min. Then extracted proteins were separated by Glycine- (10%) and Tricine-SDS-PAGE (12%), respectively. Each lane of Glycine SDS-PAGE was excised into 28 fractions and Tricine SDS-PAGE into 22 fractions based on their MW and the protein abundance in specific regions followed by trypsin digestion.

LC-MS/MS Analysis and Database Searching

Prior to LC-MS/MS analysis, digested peptides were resuspended in loading buffer (1% acetonitrile and 1% formic acid (FA) in ddH₂O) and analyzed by an LC (Ultimate 3000, Thermo
Scientific) equipped with self-packed capillary column (150 μm i.d. × 12cm, 1.9 μm C18 reverse-phase fused-silica, Trap), using a 90 min nonlinear gradient at a flow rate of 600 nL/min. The elution gradient was as follows: 6-9% B for 8 min, 9-14% B for 16 min, 14-30% B for 36 min, 30-40% B for 15 min, 40-95% for 10 min, 95-6% for 5 min (Buffer A, 0.1% FA in ddH2O; Buffer B, 0.08% FA and 80% ACN in ddH2O). Eluted samples were analyzed by Q-Exactive HF (Thermo Fisher Scientific). MS full scans were performed in the ultra-high-field Orbitrap mass analyzer in ranges m/z 300-1,400 with a resolution of 120,000 at m/z 200, the maximum injection time (MIT) was 80 ms and the automatic gain control (AGC) was set to 3×10⁶. The top 20 intense ions were subjected to Orbitrap for further fragmentation via high energy collision dissociation (HCD) activation over a mass range between m/z 200 and 2,000 at a resolution of 15,000 with the intensity threshold kept at 8.3×10³. We selected ions with charge state from 2+ to 6+ for screening. Normalized collision energy (NCE) was set at 27²¹. For each scan, the AGC was set at 2 × 10⁴ and the MIT was 60 ms. The following dynamic exclusion of precursor ion masses over a time window of 12s was used to suppress repeated peak fragmentation²².

To evaluate and compare the identification of HF and Velos, MS/MS raw files produced by HF in this study and Velos in last year (PXD002179)⁹ were processed in Proteome Discoverer 2.0 (v2.0.0.802, Matrix Science Mascot 2.3.01) against overlap queries (20,055 entries, including 2,949 MPs) between the Swiss-Prot database (release 2015.12) and the neXtProt database (release 2016.02). The same parameters were set for database searching as follows: cysteine carbamidomethyl was specified as a fixed modification and oxidation of methionine was set as variable modifications. For HF dataset, the tolerances of precursor and fragment ions were set at 15 ppm and 20 mmu, respectively. For Velos dataset, the tolerances of precursor and fragment ions
were set at 20 ppm and 0.5 Da, respectively. For digestion, trypsin was set as protease with two missed cleavage permitted. Only the proteins’ identification satisfying the following criteria were considered: (1) the peptide length ≥ 7; (2) the FDR ≤ 1% at peptide level; (3) the FDR ≤ 1% at protein level; (4) at least two different peptides (both unique and shared peptides were considered) as protein identification. The peptides were quantified by the peak area in Proteome Discoverer. For protein quantification, only the top 3 unique peptides were used for calculation. The FDRs at the PSM, peptide, and protein levels were calculated by dividing the number of target identifications by the number of decoy identifications.

Bioinformatics Analysis of Identified Proteins

To evaluate the performance of different mass spectrometers, the peptides and proteins identified by HF and Velos were compared, including the protein number, protein abundance, and MPs’ properties, etc. In addition, RNA-seq data was used to further evaluate and verify the protein abundance identified by HF and Velos at the mRNA level. Individual contribution to testis proteome was inspected by the peptide and protein number and saturation curve from HF and Velos.

Verification of MPs with Synthesized Peptides

To evaluate the authenticity of MPs, the spectra referring to MPs were manually inspected by observing base peak intensity and b/y ions matching assisted by pFind and pBuild softwares. Considering alternate explanations of PSMs which passed the manually check, open search was performed. The spectrum conflicting to previous identification was filtered. Relatively high quality peptides (unique peptides ≥ 2, peptide length ≥ 9 aa) were selected and synthesized for further analysis. Both m/z and ion intensity were used to calculate the cosine similarity score, and
specific formula was similar as previously described\textsuperscript{10}. Only when the similarity matching score is higher than 0.9, its corresponding peptides was considered as very high confidence peptide of MPs. Considering isobaric substitutions could change the mapping of the peptide from MPs to a commonly-observed protein, the isobaric filtering was performed by evaluating whether I=L, Q[Deamidated]=E, GG=N in its protein database. Considering the possibility of single amino acid variants (SAAVs) could also turn an extraordinary result into an ordinary result, the SAAVs filtering was performed according to the Swiss-Prot and RefSeq database.

**Functional Investigation of MPs**

The function of verified MPs were searched against UniProt\textsuperscript{25} (http://www.uniprot.org/) online, including biological function and process, expression, subcellular location, etc.

### RESULTS and DISCUSSION

**HF Achieves the Deeper Proteome Coverage of Human Testis Tissue**

To further improve the proteome coverage for the testis tissue, the digested peptides were detected by HF in this study on the basis of Velos detection in last year. After database searching, total 7,049,718 MS\textsubscript{2} scans were obtained from three testis tissues. For increasing the trust in data reliability, strict criteria were considered, such as FDR value for PSM-(0.25%), peptide-(0.33%), and protein-level (0.17%). In addition, the total number of the expected true positives and false positives at each level (PSM, peptide and protein) were calculated (Table S-1). As a result, 8,526 proteins were identified confidently in HF datasets.

To test the resolution of the Glycine- and Tricine-gels on the separation of testis proteins, we drew the accumulation curves with identified proteins and peptides from each fractions (Figure 1A&B).
The number of non-redundant peptides and proteins consistently increased with the addition of individual fractions from lower band to top. Compared with Glycine-SDS-PAGE, Tricine-SDS-PAGE identification curve were steeper on the lower part. However, it is also easier to achieve saturation level. In addition, we found that the Tricine gel had higher resolution on LMW proteins (Figure S-1). The total number of identified proteins rose to 8,526 when two separation methods were combined, in which 598 proteins (7.6%) and 160 proteins (1.3%) were uniquely identified either from Glycine- or Tricine-SDS-PAGE (Figure 1C). This result suggested that the different protein separation methods can increase the number of proteins.

Although samples’ pooling strategy has been a common way in proteomics, individual contribution was also considered in this study similar to our previous research. Total 326, 123 and 77 proteins were uniquely identified in three individuals, and 805 proteins were identified in at least two individuals, which all accounts for about 16% of total identified proteins (Figure 1D). Total 7,195 proteins were overlapped in three samples, which accounts for 84% of the whole testis proteins. The similar conclusion also found in two separation methods (Figure S-2 A&B). These results implied that HF can identifies more individual expressed proteins and improves the coverage of the testis proteome. To comprehensively evaluate the contribution of proteins isolation methods and individual difference on testis sample, the accumulation curves of proteins and peptides were drawn (Figure 1E&F). The accumulated curve approximately achieved saturation of 125,792 peptides for 8,526 protein groups with three individual testis tissues. This resulted in the largest tissue proteome datasets with high mass accuracy on both of MS1 and MS2 through HF under the stringent criteria.
HF Has Higher Coverage than Velos

As the HF has higher sensitivity, resolution and scanning speed than Velos\textsuperscript{36}, it may help us to make a further exploration to the testis proteome. To understand the characteristic of these two mass generations, we analyzed these two datasets generated from HF and Velos with the same searching engine and criteria. As shown in Figure 2A, the number of identified peptides from Glycine-SDS-PAGE by HF was about 17\% higher than that from the Velos in all of the fractions, which resulted in 14\% more identified proteins. Following that, we also found that HF produced 67\% more MS\textsubscript{2} spectra, 24\% more PSMs, 19\% more peptides and 13\% more protein groups than Velos (Figure 2B). These advantages might due to the faster scanning speed, which further result in more MS\textsubscript{2} spectra. Another 1,230 and 240 proteins were uniquely identified in HF and Velos data, respectively (Figure 2C). Therefore, HF can identified more proteins than Velos in shorter time (10,500 min vs 13,500 min) (Table S-1). These phenomena were also found in each individual testis samples with any of separation method (Figure S-3 A&B). This result suggested that the HF can still detect more specific proteins from the same human testis samples.

The whole sequence coverage of proteins identified by HF or Velos were also compared (Figure 2D). The sequence coverage of HF mainly distributed in 25-50\%, which was higher than that of the Velos with distribution of 10-25\%. This may be contributed by the high intrinsic resolution of HF which allows short transients and short cycle times in topN methods\textsuperscript{27}.

HF Identifies More Low-Abundance Proteins than Velos

To estimate the detection sensitivity of these two different mass spectrometers used in this study, we compared the intensity of shared proteins in HF and Velos datasets (Figure 3A). The protein abundance of HF was highly correlated with that of the Velos (Pearsom’s correlated coefficient
was 0.94), which suggested the good reproducibility of these two mass spectrometers. Consistently, we saw the higher sequence coverage of proteins identified by HF than that of Velos, especially for the lower coverage proteins (Figure 3B). This result illustrated that HF may have an advantage on identifying low sequence coverage proteins. To verify this result, we compared the abundance and sequence coverage between the commonly and uniquely identified proteins from HF (Figure 3C). As expected, the uniquely identified proteins showed lower abundance and lower sequence coverage than that of commonly identified ones. Thus, the higher sensitivity of HF benefits the identification of low abundance proteins with limited protein length. In addition of sensitivity and resolution of the instrument, pre-fractionation of the sample might also the parameter for enhancing the identification of low-abundance proteins.

The dynamic range of proteins abundance also affects the number of identified proteins. Thus, protein identification may be improved by wider dynamic range of mass spectrometer. Normalized intensities were used to compare the intensity of total proteins identified by HF and Velos, result showed that HF has wider dynamic range than Velos (Figure 3D). We also compared the intensity of uniquely identified proteins by HF and Velos (Figure 3E). Although they had the similar dynamic range expanded from 4 to 8 for the mainly identified proteins, the HF could detect more low abundance proteins in the range of 0 to 4. This result was consistent with what we described previously. As the same, RNA-Seq data consistently verified that these low-abundance proteins had lower expression level at the mRNA level as well (Figure 3F). Compared with the abundance of overlapped proteins at the mRNA level, uniquely identified proteins by HF and Velos had low-abundance at the transcriptional level. Moreover, more uniquely identified proteins were detected by HF compared with Velos data. Thus, the high resolution and wide dynamic range of
HF could achieve deeper proteome coverage in human testis tissue, which potentially benefits the MPs study in the same sample.

**HF Identified More MPs with Lower-Abundance**

In our previous research, testis was regarded as an ideal target tissue for MPs study through comparing the abundance distribution of the corresponding gene of MPs in total genes from 27 tissues at the mRNA level. Total 81 MPs were detected, in which 32 MPs were commonly identified by HF and Velos, while another 42 MPs and 7 MPs were uniquely identified by HF and Velos, respectively (Figure 4A). Of them, total 51 proteins were new identified MPs. (Figure S-4A). The MPs distribution suggested that common identified MPs tended to distribute in multiple samples and had relatively higher abundance (Figure 4B). However, uniquely identified MPs by HF or Velos tend to distribute in single sample, which may be due to the individual difference from different donors. The status of spermatogenesis might be a factor of individual differences. Various age (53, 35 and 52 years old) and native place (Beijing and Hebei) of three donors might led to different spermatogenesis status. The normalized intensity distribution comparison showed that MPs abundance were much lower than common identified proteins’ abundance from HF and Velos, , and the abundance of HF MPs were lower than that of Velos (Figure 4C).

As the HF can achieve higher sequence coverage than Velos because of higher sequencing speed, we compared MPs sequence coverage of HF and Velos (Figure 4D). Each dot represents one MP and the sizes of the dot varies with the number of identified unique peptides. Interestingly, almost all of MPs identified by HF have at least two unique peptides, which accounts for 78.3% of total MPs (Figure S-4B). This result illustrated that HF detected more convincing MPs than that of
Velos.

The reliability of MPs identified in this study was further verified by spectra quality checking manually. The total 56 MPs were retained by filtering with at least two unique-mapping peptides whose sequence length 9 or more aa. These above 56 proteins were listed as credible MPs, their spectra have higher base peak intensity and continuous b/y ion matches. In addition, the matching of synthesized peptides with scanning spectra of these 56 MPs showed that they had high similarity, the score were 0.9 or higher based on comparing the matches for b1+, y1+, b2+, and y2+ ions and the pattern for peak intensity of these match (Figure 4E&Figure S-6&Table S-3). In addition, we also evaluated whether the isobaric substitutions and SAAVs could influence the authenticity of MPs identified. For example, the unique peptide “ISNIFVIGNGNKPWISLPR” mapped to the MP Q8TD47 might be isobaric substituted to the “LSNIFVIGNKNPWISSLPR” peptide mapped to the commonly identified protein P22090(40S ribosomal protein S4, Y isoform 1) in our data. After isobaric and SAAVs filtering, 47 proteins were confirmed MPs with authentic MS evidence finally. The high verification rate for all of these MPs implied that the high quality of our MS data and reliability of MPs.

**The MPs Identified by HF Has High Density in Lower-MW and Higher-pl Properties**

Previous studies showed that low MW, high hydrophobicity, and high pl were the main properties of MPs. The analysis of MPs characteristics showed that HF could identify more MPs with even LMW (Figure 5A) or high pl compared with Velos (Figure 5B). These results illustrated that HF can detect more MPs which were difficultly identified by previous generations of mass spectrometer. This may be caused by higher sensitivity of HF compared with that Velos for LMW proteins identification (Figure S-5A). The similar result was obtained for MPs with higher pl
The 47 confirmed MPs in UniProt (http://www.uniprot.org/) showed that 26 MPs (55.3%) were functionally annotated (Figure 5C). Most of these MPs take part in mRNA process like RNA splicing\textsuperscript{29-31}. Total 11 MPs were testis-specific proteins, among which, Q9BWV3, P0C5Z0, A6NEQ0, and Q14598 are related with spermatogenesis process (Figure 5C). Q9BWV3, P0C5Z0 and A6NEQ0 are all bonding proteins which are bond with zinc ion, DNA, and RNA respectively. Testis-specific basic protein Y1 (Q14598) may play a role in sex ratio distortion\textsuperscript{32}. Zinc finger MYND domain-containing protein 15 is predicted to be a candidate in some disease-forming process, might interact with histone deacetylases (HDACs) and play an important role in differentiation, spermatogenesis, transcription and transcription regulation. Mutation of this gene may cause spermatogenic failure \textsuperscript{14}(SPGF14)\textsuperscript{33}. In addition, Q5BKT4 and A6NNE9 may participate in protein modification. Dol-P-Glc:Glc(2)Man(9)GlcNAc(2)-PP-Dol alpha-1,2-glucosyltransferase (Q5BKT4) is involved in the protein glycosylation pathway, which adds the third glucose residue to the lipid-linked oligosaccharide precursor for N-linked glycosylation\textsuperscript{34}. While E3 ubiquitin-protein ligase MARCH11 (A6NNE9) may play a role in ubiquitin-dependent protein sorting in the development of spermatids\textsuperscript{35}. The known biological functions of these MPs indicate that most of these identified MPs with low-abundance might take part in various important biological processes especially in spermatogenesis. These functions are ongoing by our other lab teams and worthwhile researching thoroughly.

\section*{Conclusion}

To achieve deeper coverage, HF was used to analyze the same three individual human testis
samples by two different gel-separations based on our previous work. Total 8,526 proteins were identified with high confidence, in which more than 1,000 proteins were uniquely identified with low-abundance when compared with our Velos datasets. The deeper coverage might also contribute to tandem protein separating methods and individual diversity. Among them, total 81 MPs were identified, of which 74 MPs and 39 MPs were identified from HF and Velos datasets, respectively. HF alone detected more MPs with lower abundance and higher sequence coverage from the same human testis tissues. After strict filtering, 47 MPs were selected as high-credible MPs by manual checking of spectrum quality, isobaric and SAAVs filtering. The higher sensitivity and resolution of MS instrument can explore more low-abundance MPs in human testis tissue and could be applied in other tissues and in the near future for C-HPP study on different proteomics samples.

**ASSOCIATED CONTENT**

**Supporting Information**

Supplementary Figure 1. Molecular weight (MW) distribution of proteins identified by Glycine- and Tricine-SDS-PAGE. Supplementary Figure 2. Comparison of identified proteins from three individual testis tissues based on two gel-separating methods. Supplementary Figure 3. Comparison of identified proteins from three individual testis tissues based on two separation ways by HF and Velos. Supplementary Figure 4. The number and sequence coverage of testis MPs. Supplementary Figure 5. The MW and pI of identified proteins by HF and Velos. Supplementary Figure 6. The verification of MPs by its synthesized peptide and previous MS spectra matching. Supplementary Table 1. The searching criteria and results summary of HF and Velos.
Supplementary Table 2. Detailed information about the identified proteins/genes for three samples by HF and Velos based on proteomics and transcriptomics. Supplementary Table 3. The identification and verification of the total MPs in human testis.

- AUTHOR INFORMATION

Corresponding Authors

* P.X.: Tel/Fax: 86-10-80705066. E-mail:xuping_bprc@126.com.

* F.H.: Tel/Fax: 86-10-68171208. E-mail:hefc@nic.bmi.ac.cn.

Author Contributions

*These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

- ACKNOWLEDGMENTS

We thank Wantao Ying, Pengyuan Yang, Siqi Liu and Qinyu He for their valuable suggestions. We also appreciate lab members for their discussion and efforts. This work was funded by the Chinese National Basic Research Programs (2013CB911201 & 2015CB910700), the International Collaboration Program (2014DFB30020), the National Natural Science Foundation of China (Grant No. 31470809, 31400698 & 31400697), the National High-Tech Research and Development Program of China (2014AA020900 & 2014AA020607), the National Natural Science Foundation of Beijing (Grant No. 5152008), Beijing Training Project for The Leading Talents in S&T, National Megaprojects for Key Infectious Diseases (2013zx10003002 &
2016ZX10003003), the Key Projects in the National Science & Technology Pillar Program (2012BAF14B00), the Unilevel 21th Century Toxicity Program (MA-2014-02409) and the Foundation of State Key Lab of Proteomics (SKLPYB201404).

- **ABBREVIATIONS**

HPP, Human Proteome Project; C-HPP, ChromosomeCentric Human Proteome Project; MPs, missing proteins; HF, Q-Exactive HF; Velos, LTQ-Orbitrap Velos; MW, molecular weight; LMW, low molecular-weight proteins; RNC-mRNA, the mRNA attached to ribosome-nascent chain complex; RPKM, reads per kilobase per million mapped reads; PSM, peptide–spectrum; MIT, maximum injection time; AGC, automatic gain control; NCE, normalized collision energy match; FDR, false discovery rate;

- **REFERENCE**


proteins. *Journal of proteome research* 2015, 14, (9), 3441-3451.


For TOC Only

In order to improve the proteome coverage and find more specific MPs in human testis, the three testis samples were detected by HF. HF have many advantages in large-scale proteomics research, including high resolution, wide dynamic range and quick scanning speed, etc. The comparison of HF datasets and Velos dataset reveals that we used less machine time (10,500 min vs 13,500 min) to get more proteins (8,526 vs 7,537). What’s more, the detection sensitivity of HF is of higher than that of Velos (3.22 × 10^9 vs 3.59 × 10^8). Therefore, we identified more MPs from HF than that of Velos (74 vs 39). This study suggests that more MPs can be efficiently identified by new generation of mass spectrometer such as HF in the near future in C-HPP.

Figure Legends

Figure 1. Deep proteomics study for human testis tissues by HF. (A-B) The accumulation curves of identified proteins and peptides of three testis samples resolved by two gel-separation methods.
Figure 2. HF resulted in higher proteome coverage than Velos. (A) Comparison of identified peptides in each fraction by Velos and HF from three testis samples (B) Comparison of proteome identification of human testis tissues by Velos and HF, including MS/MS, PSM, peptide, protein, etc. (C) Venn diagram of identified proteins based on Velos and HF. (D) Comparison of the whole sequence coverage from commonly identified proteins by Velos and HF.

Figure 3. HF identifies more low-abundance proteins than Velos. (A) Correlation of the commonly identified protein abundance of two mass spectrometers. The different shades of colors represent the density of proteins. (B) Comparison of sequence coverage of commonly identified proteins by HF and Velos. (C) Comparison of protein abundance and sequence coverage of unique and shared proteins in HF datasets. (D) The dynamic range of normalized intensity in two mass spectrometers. (E) The normalized intensity distribution of uniquely identified proteins in two mass spectrometers. (F) Comparison of abundance distribution of mRNA based on RNA-seq of HF and Velos.

Figure 4. HF identifies more MPs with low-abundance in human testis. (A) Venn diagram of HF and Velos MPs. (B) The distribution of MPs in each samples. Gray represents not identified. (C) Comparison of the abundance distribution of MPs and the commonly identified proteins from HF and Velos. (D) The sequence coverage of identified MPs from HF and Velos. (E) Example of verification of the identified peptide for MPs by using the synthesized peptide.
Figure 5. MPs from HF behaves high density in lower-MW and higher-pl. (A) The molecular weight distribution of MPs. (B) The isoelectric-point distribution of MPs. (C) The function information of verified MPs in UniProt.
Abstract

147x79mm (300 x 300 DPI)
Figure 1

303x344mm (300 x 300 DPI)
Figure 2

275x326mm (300 x 300 DPI)
Figure 3

286x355mm (300 x 300 DPI)
Figure 4

283x360mm (300 x 300 DPI)
Figure 5

151x150mm (300 x 300 DPI)