



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

# Extensive dataset of boar seminal plasma proteome displaying putative reproductive functions of identified proteins

Cristina Perez-Patiño<sup>a</sup>, Isabel Barranco<sup>a</sup>, Inmaculada Parrilla<sup>a</sup>,  
Emilio A. Martinez<sup>a</sup>, Heriberto Rodriguez-Martinez<sup>b</sup>,  
Jordi Roca<sup>a,\*</sup>

<sup>a</sup> Department of Medicine and Animal Surgery, Faculty of Veterinary Science, University of Murcia, Spain

<sup>b</sup> Department of Clinical & Experimental Medicine (IKE), Linköping University, Sweden

## ARTICLE INFO

## Article history:

Received 9 May 2016

Received in revised form

6 July 2016

Accepted 19 July 2016

Available online 29 July 2016

## Keywords:

Porcine

Seminal plasma

Proteome

Reproductive functionality

## ABSTRACT

A complete proteomic profile of seminal plasma (SP) remains challenging, particularly in porcine. The data reports on the analysis of boar SP-proteins by using a combination of SEC, 1-D SDS PAGE and NanoLC-ESI-MS/MS from 33 pooled SP-samples (11 boars, 3 ejaculates/boar). A complete dataset of the 536 SP-proteins identified and validated with confidence  $\geq 95\%$  (Unused Score  $> 1.3$ ) and a false discovery rate (FDR)  $\leq 1\%$ , is provided. In addition, the relative abundance of 432 of them is also shown. Gene ontology annotation of the complete SP-proteome complemented by an extensive description of the putative reproductive role of SP-proteins, providing a valuable source for a better understanding of SP role in the reproductive success. This data article refers to the article entitled “Characterization of the porcine seminal plasma proteome comparing ejaculate portions” (Perez-Patiño et al., 2016) [1].

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

DOI of original article: <http://dx.doi.org/10.1016/j.jprot.2016.04.026>

\* Corresponding author.

E-mail address: [roca@um.es](mailto:roca@um.es) (J. Roca).

<http://dx.doi.org/10.1016/j.dib.2016.07.037>

2352-3409/© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Specifications Table

Subject area	<i>Reproductive biology</i>
More specific subject area	<i>Proteomics of boar seminal plasma</i>
Type of data	<i>Excel file</i>
How data was acquired	<i>Samples were analyzed using a NanoLC Ultra 1-D plus Eksigent (Eksigent Technologies, Dublin, CA, USA) directly connected to an AB SCIEX TripleTOF 5600 mass spectrometer (AB SCIEX, Framingham, MA, USA)</i>
Data format	<i>Processed, analyzed</i>
Experimental factors	<i>Boar seminal plasma from twice centrifuged ejaculates in order to remove sperm and obtain cell-free seminal plasma</i>
Experimental features	<i>Boar seminal plasma proteome description and relative quantification. The sample proteins were fractionated using SEC and 1-D SDS PAGE followed by trypsin digest to analyze less abundant proteins and in-solution digestion to analyze more abundant proteins</i>
Data source location	<i>Murcia (Spain)</i>
Data accessibility	<i>Data are available within this article and via the PRIDE partner repository with the dataset identifier PRIDE: PXD003579. <a href="http://dx.doi.org/10.6019/PXD003579">http://dx.doi.org/10.6019/PXD003579</a></i>

## Value of the data

- The data provides the so far largest proteomic profile of boar seminal plasma.
- The experimental approach used, SEC and 1-D SDS PAGE followed by NanoLC-ESI-MS/MS, is useful for identifying proteins in samples with a highly complex mixture of proteins such as seminal plasma.
- Comprehensive information and references of putative reproductive functionality of proteins identified in boar seminal plasma.
- This dataset can be used as a primary guide to characterize protein biomarkers for sperm quality and fertility in pig seminal plasma.

## 1. Data

A unique dataset is presented resulting from a qualitative and quantitative proteomic analysis of boar seminal plasma (SP), with more than 500 proteins listed and showing the relative abundance of a total of 432 proteins (Supplementary Table 1). Furthermore, an extensive description about their putative reproductive function is also provided, including appropriate references. This available information could help to a better understanding of the role of SP-proteins on boar sperm reproductive success.

## 2. Experimental design, materials and methods

In order to describe the boar SP-proteome, 33 entire ejaculates were collected, by using the semi-automatic collection device Collectis<sup>®</sup>, from 11 healthy and sexually mature Landrace and Large White boars (3 ejaculates per boar). Immediately after collection, ejaculated samples were centrifuged twice (1,500 × g 10 min) to obtain SP sperm-free samples, which were stored at –80 °C until proteomic analysis. The 33 SP-samples were pooled and analyzed using a combination of SEC, 1-D SDS PAGE followed by NanoLC-ESI-MS/MS. The proteomics data and result-files from the analysis have been deposited to the ProteomeXchange Consortium [2] via the PRIDE partner repository, with the dataset identifier PRIDE: PXD003579 and <http://dx.doi.org/10.6019/PXD003579>. Single SP-pools

from each boar were analyzed by LC-SWATH-MS acquisition for determination of protein relative abundance.

### 2.1. Sample preparation

Seminal plasma samples were thawed at room temperature and ultracentrifuged ( $16,100 \times g$ ,  $4^\circ\text{C}$ , 1 min). Before starting the proteome analysis, the 33 SP-samples were split each one in two aliquots. One of them was mixed in a single pool for characterization the pig SP-proteome. Simultaneously, the second aliquots from each boar ( $n=3$ ) were mixed, generating a total of 11 single pools (1 pool per boar).

### 2.2. Proteome analysis

The proteome analysis was performed as described in Ref. [1]. The more abundant proteins were identified from an aliquot of the mixed SP-sample analyzed by in-solution processing. The less abundant proteins were analyzed in-gel digestion processing using the portion of the 1-D SDS PAGE containing proteins with a molecular weight higher than 38 kDa obtained from the fractions collected after a SEC step. The digestion of the sliced gel was performed following the protocol used by Shevchenko et al. [3].

#### 2.2.1. LC-MS/MS analysis

The peptides recovered from in-solution and in-gel digestion processing were analysed as described in Ref. [1]. Briefly, peptides were examined by LC using a NanoLC Ultra 1-D plus Eksigent (Eksigent Technologies, Dublin, CA, USA) and the eluted peptides were direction-ionized using an ESI Nanospray III (AB SCIEX, Framingham, MA, USA) and then analyzed on an AB SCIEX TripleTOF 5600 mass spectrometer (AB SCIEX) in direct injection mode.

### 2.3. LC-SWATH-MS acquisition

For LC-SWATH-MS acquisition the TripleTOF 5600 (SCIEX) was configured as described by Gillet et al. [4] and later adapted to porcine SP by Perez-Patiño et al. [1]. Briefly, the mass spectrometer was operated in a looped product ion mode where the instrument was specifically tuned to allow a quadrupole resolution of Da/mass selection. The stability of the mass selection was maintained by the operation of the Radio Frequency (RF) and Direct Current (DC) voltages on the isolation quadrupole in an independent manner. A set of 37 overlapping windows, covering the mass range 450–1000 Da, was constructed using an isolation width of 16 Da (15 Da of optimal ion transmission efficiency and 1 Da for the window overlap). Consecutive swaths need to be acquired with some precursor isolation window overlap to ensure the transfer of the complete isotopic pattern of any given precursor ion in at least one isolation window and, thereby, to maintain optimal correlation between parent and fragment isotopes peaks at any LC time point. In this way, each single pool was loaded onto a trap column followed by an analytical column and eluted peptides were infused in the spectrometer nanoESI qTOF (SCIEX TripleTOF 5600) operating in swath and in high sensitivity mode.

### 2.4. Data processing: protein identification, validation and quantification

After LC-MS/MS, The SCIEX.wiff data-files were processed using ProteinPilot v5.0 search engine (AB SCIEX). The Paragon algorithm (4.0.0.0, 4767) of ProteinPilot was used to search against the National Center for Biotechnology Information non-redundant (NCBIInr; 70,353,186 proteins searched) protein sequence database with the following parameters: trypsin specificity, cys-alkylation (IAM), no taxonomy restricted, and the search effort set to through. To avoid using the same spectral evidence in more than one protein, the identified proteins were grouped based on MS/MS spectra by the Protein-Pilot Pro Group™ Algorithm, regardless of the peptide sequence assigned. The protein within each group that could explain more spectral data with confidence was depicted as the primary protein of the group. The resulting Protein-Pilot group file was loaded into PeakView® (v2.1, AB

SCIEX) and peaks from SWATH runs were extracted with a peptide confidence threshold of 99% confidence (Unused Score  $\geq$  1.3) and a false discovery rate (FDR) less than 1%. The identified proteins were quantified using PeakView<sup>®</sup> from normalized label-free quantification (LFQ) intensity data.

## Gene ontology

Bioinformatic analysis of identified and validated SP-proteins was manually performed using the comprehensive bioinformatics tool for functional annotation UniProt KB database ([www.uniprot.org](http://www.uniprot.org)) in combination with PANTHER ([www.pantherdb.org](http://www.pantherdb.org)). Both databases downloaded 06/04/2016, containing 63,686,057 and 1,424,953 entries in UniProt KB and PANTHER, respectively.

## Acknowledgments

This experimental study was supported by MINECO Madrid (Spain) and FEDER funds (EU) (AGL2012-39903), Seneca Foundation (19892/GERM/15) Murcia (Spain), The Swedish Research Council (VR, 521-2011-6353), the Swedish Research Council Formas (221-2011-512) and the Research Council in Southeast Sweden (FORSS, 378091/312971), Sweden. C. Perez-Patiño and I. Barranco and were financially supported by the Seneca Foundation (Murcia, Spain) and MECD (Madrid, Spain), respectively. The authors are grateful to AIM Iberica (Topigs Norsvin Iberica) for supplying the boar ejaculates.

## Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at: <http://dx.doi.org/10.1016/j.dib.2016.07.037>.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at: <http://dx.doi.org/10.1016/j.dib.2016.07.037>.

## References

- [1] C. Perez-Patiño, I. Barranco, I. Parrilla, M.L. Valero, E.A. Martinez, H. Rodriguez- Martinez, J. Roca, Characterization of the porcine seminal plasma proteome comparing ejaculate portions, *J. Proteom.* 142 (2016) 15–23.
- [2] J.A. Vizcaíno, E.W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Ríos, et al., ProteomeXchange provides globally coordinated proteomics data submission and dissemination, *Nat. Biotechnol.* 30 (2014) 223–226.
- [3] A. Shevchenko, O.N. Jensen, A.V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie, M. Mann, Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels, *Proc. Natl. Acad. Sci.* 93 (1996) 14440–14445.
- [4] L.C. Gillet, P. Navarro, S. Tate, H. Röst, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis, *Mol. Cell. Proteom.* 11 (2012), O111.016717.