UNIVERSIDAD DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

Proteomics of Seminal Plasma and Porcine Spermatozoa

Proteómica del Plasma Seminal y de los Espermatozoides de Porcino

Dª Cristina Pérez Patiño

2018

UNIVERSIDAD DE
MURCIA

D. Jordi Roca Aleu, Catedrático del Departamento de Medicina y Cirugía Animal de la Universidad de Murcia e Inmaculada Parrilla Riera Profesora Titular del Departamento de Medicina y Cirugía Animal de la Universidad de Murcia

AUTORIZAN:

La presentación de la Tesis Doctoral titulada "**Proteomics of seminal plasma and porcine spermatozoa (Proteómica del plasma seminal y de los espermatozoides de porcino)"** realizada por Dª. Cristina Pérez Patiño, bajo nuestra inmediata dirección y supervisión, y que presenta para la obtención del grado de Doctor por la Universidad de Murcia

Para que conste y haga su efecto donde proceda firmamos el presente documento

En Murcia, a 16 de julio de 2018

Fdo: Jordi Roca Aleu

Fdo: Inmaculada Parrilla Riera

PhD Thesis

UNIVERSIDAD DE
MURCIA

Dª. María del Rocío Ruiz de Ybañez carnero, Profesora Titular de Universidad del Área de Sanidad Animal y **Presidente Comisión Académica programa doctorado** * Ciencias Veterinarias, INFORMA:

Que una vez evaluado, de conformidad con el procedimiento establecido en el artículo 21 del Reglamento de doctorado de la Universidad de Murcia, el expediente completo de la tesis doctoral titulada "Proteomics of seminal plasma and porcine spermatozoa (Proteómica del plasma seminal y de los espermatozoides de porcino)", realizada por Dª Cristina Pérez Patiño, bajo la inmediata dirección y supervisión de D. Jordi Roca Aleu y Dª Inmaculada Parrilla Riera, esta Comisión Académica, en sesión celebrada en fecha 23 de julio de 2018, ha dado su autorización para su presentación ante la Comisión General de Doctorado.

Murcia, a 23 de julio de 2018

Doctorando: Dª. Cristina Pérez Patiño

Mod: T-40

Universidad de Concepción
Facultad de Ciencias Veterinarias
Departamento de Ciencia Animal

Chillán, julio de 2018

A quien esto concierna

He recibido para revisión el trabajo de Tesis Doctoral realizado por la Srta. Cristina Pérez Patiño, titulado: **"Proteómica del Plasma Seminal y de los Espermatozoides de Porcino"**

Se trata de un trabajo original muy bien estructurado, que sigue una secuencia de experimentos estrechamente relacionados y que apuntan a responder una hipótesis bien concebida. Este trabajo de tesis está basado en cinco artículos, donde la Srta. Pérez Patiño es la primera autora, y de los cuales tres ya están publicados, en revistas internacionales y con alto factor de impacto (JCR), y dos se encuentran enviados a revistas del mismo orden y todas ellas con sistema de revisión por pares. Los artículos considerados en este trabajo fueron los siguientes:

Artículo 1: "Characterization of the porcine seminal plasma proteome comparing ejaculate portions". **Cristina Pérez-Patiño,** Isabel Barranco, Inmaculada Parrilla, M. Luz Valero, Emilio A. Martínez, Heriberto Rodríguez-Martínez y Jordi Roca. Journal of Proteomics, 2016, 142:15-23. DOI: 10.1016/j.jprot.2016.04.026.
Artículo 2: "Extensive dataset of boar seminal plasma proteome displaying putative reproductive functions of identified proteins". **Cristina Pérez-Patiño,** Isabel Barranco, Inmaculada Parrilla, Emilio A. Martínez, Heriberto Rodríguez-Martínez y Jordi Roca.
Data in Brief, 2016:(8):1370-1373. DOI: 10.1016/j.dib.2016.07.037.
Artículo 3: "New in-depth analytical approach of the porcine seminal plasma proteome reveals potential fertility biomarkers". **Cristina Pérez-Patiño,** Inmaculada Parrilla, Isabel Barranco, María Vergara-Barberán, Ernesto F. Simó-Alfonso, José M. Herrero-Martínez, Heriberto Rodríguez-Martínez, Emilio A. Martínez y Jordi Roca. Journal of Proteome Research 2018:(17):1065-1076. DOI: 10.1021/acs.jproteome.7b00728.
Artículo 4: "iTRAQ-based analyses prove the protein profile of boar spermatozoa is remodelled during ejaculation". **Cristina Pérez-Patiño** y colaboradores. Enviado para publicación a la revista: Mollecular and Celular Proteomics.
Artículo 5: "The proteome of frozen-thawed pig spermatozoa is dependent on ejaculate fraction source". **Cristina Pérez-Patiño** y colaboradores. Enviado para publicación a la revista: Scientific Reports.

Esta tesis, además de los artículos referidos está constituida por una introducción general del tema tratado, un resumen extendido donde se explicitan los objetivos del trabajo, se detallan los materiales y métodos de los experimentos involucrados, además del detalle de los diseños experimentales y resultados de cada uno de los artículos que componen este trabajo. Asimismo, se enumeran las conclusiones obtenidas de cada uno de los estudios realizados. Finalmente se adiciona un apéndice gráfico con los principales instrumentos y procedimientos relacionados con el trabajo realizado.

En su tesis, la Srta. Pérez, ha desarrollado, hasta la fecha, la más completa, específica y actualizada base de datos en relación con las proteínas (proteoma) asociadas al plasma

seminal y espermatozoides del verraco. Complementariamente describió diferencias cuantitativas en la expresión de proteínas según las diferentes fracciones del eyaculado, además con implicancia en la funcionalidad de los espermatozoides constituyentes de cada fracción. Esto se vio reflejado en datos de fertilidad real, mostrando que algunas proteínas podrían ser utilizadas/consideradas como potenciales biomarcadores de fertilidad, asimismo se podrían identificar dentro del grupo de reproductores de similar merito genético, aquellos individuos, que podría ser sub-fértiles. Además de toda la información relacionada con semen fresco/conservado también se estableció que las diferencias en funcionalidad de espermatozoides descongelados podrían deberse a diferencias en la expresión proteica de estas células espermáticas. Concomitantemente, este trabajo continúa complementando y reafirmando la información ya existente acerca de las diferencias que existen en diferentes niveles, ahora sumando la proteómica de plasma seminal y de los espermatozoides, entre las distintas fracciones del eyaculado del verraco.

Los cinco artículos que conforman la estructura central de este trabajo tienen a la Candidata como primera autora lo que indica claramente su involucramiento y protagonismo en el trabajo realizado. El diseño de todos los experimentos es claro y permitió responder la hipótesis planteada. La metodología utilizada fue la más adecuada (actualizada) para los objetivos perseguidos, a su vez los datos obtenidos fueron analizados correctamente y presentados de manera clara y bien ilustrados. El sentido y las implicancias de los resultados obtenidos en los diferentes experimentos fueron contextualizados correctamente y en concordancia con la literatura actualmente disponible, además de ser bien discutidos, además de constituirse plenamente dentro del estado del arte de la temática abordada. En conjunto esto permitió alcanzar conclusiones relevantes que además conducen a nuevos desafíos dentro de este campo de investigación, para los cuales la candidata parece estar muy bien preparada.

Considerando la calidad de este estudio (tesis más los cinco artículos, publicados y/o enviados), estoy plenamente convencido de que este trabajo es muy adecuado para constituirse en una Tesis Doctoral y por lo tanto apoyo a la Srta. Cristina Pérez Patiño como candidata al doctorado de la Facultad de Veterinaria de la Universidad de Murcia.

UNIVERSIDAD DE CONCEPCION
FACULTAD DE CIENCIAS VETERINARIAS
DEPTO. DE CIENCIA ANIMAL

Fernando Saravia Ramos, MV, MSc, PhD.
Profesor Asociado
Laboratorio de Reproducción Animal, Departamento de Ciencia Animal
Facultad de Ciencias Veterinarias, Universidad de Concepción, Chile

Stiftung Tierärztliche Hochschule Hannover
University of Veterinary Medicine Hannover

Stiftung Tierärztliche Hochschule Hannover, Reproduktionsmedizinische Einheit der Kliniken, Bünteweg 15, 30559 Hannover

Unit for Reproductive Medicine of Clinics
University of Veterinary Medicine

Prof. Dr. Dagmar Waberski
Bünteweg 15
D-30559 Hannover
Germany

Tel. +49 511 953-8524
Fax +49 511 953-8504
dagmar.waberski@tiho-hannover.de

Datum
Hannover, 23.07.2018

Review of PhD Thesis entitled "Proteomics of seminal plasma and porcine spermatozoa ", submitted by Cristina Pérez-Patiño

The presented thesis consists of five studies with two major goals: firstly, to perform in depth proteome analysis of boar spermatozoa and seminal plasma and secondly, to elucidate the potential value of proteome data to serve as biomarkers for boar fertility. A major proteome mapping encoded in *Sus scrofa* taxonomy for boar seminal plasma and spermatozoa is provided (Studies 1, 2 and 4), and differential protein expression in relation to the fertility of sires (Study 3) and to the origin of ejaculate fractions in fresh and frozen semen (Studies 4 and 5) is demonstrated. Potential functions of differentially expressed proteins and their relevance for fertility are extensively discussed.

Two studies have been published in high ranked peer-reviewed journals and a data set description in an international open access journal was added. The latest two studies have been submitted to outstanding high impact factor-journals with good chances for being accepted.

Stiftung Tierärztliche Hochschule Hannover
Reproduktionsmedizinische Einheit der Kliniken
Bünteweg 15
30559 Hannover
Steuer-Nr. 25/202/26506
Ust-ID-Nr. DE 233060166

Bankverbindung
Norddeutsche Landesbank Hannover
BLZ 250 500 00
Konto 106 031 206
IBAN DE31 2505 0000 0106 0312 06
SWIFT-BIC: NOLA DE 2H

www.tiho-hannover.de

The results of the presented thesis are of high originality and high scientific value as they give a novel insight in boar semen proteomics and open new doors for the identification of biomarkers for fertility and the improvement of semen preservation methods.

In conclusion, the presented thesis is highly suitable to be accepted as PhD thesis.

Prof. Dagmar Waberski

LINKÖPING UNIVERSITY
Department of Clinical and Experimental Medicine
Professor Heriberto Rodríguez-Martínez

Date
2018-07-04

To whom it may concern

This is to certify that **Mrs Cristina Perez Patiño,** PhD-student at the University of Murcia, carried out two periods of training and research at the Department of Clinical & Experimental Medicine (IKE), under my supervision between May 2 and August 2 of 2015. Mrs Perez-Patiño successfully learned and carried out studies by using the Luminex's xMAP® technology, a multiplex microsphere-based flow cytometric assay for the measurement of the concentration of cytokines present in boar seminal plasma samples.

After obtaining the results, Mrs Perez-Patiño summarized them and wrote the correspondent manuscripts with the undersigned as co-author. Mrs Perez-Patiño was diligent, highly interested in her tasks and successfully adjoined our research group under her research periods in Linköping University.

The present certification is issued in Linköping on July 4th 2018.

Prof Dr Heriberto Rodríguez-Martínez
Senior Professor

Office address	Telephone No	Telefax
IKE—HU Campus US Lasarettsgatan 64-Floor 12	+46(0)10 1030000	+46(0)13 132257
Faculty of Medicine &Health Sciences Linköping University	+46(0)10 1031522 (direct)	
SE-581 85 Linköping SWEDEN	E-mail Heriberto.Rodriguez-martinez@liu.se	

Department of Medicine and Animal Surgery

Faculty of Veterinary Sciences

University of Murcia

My financial support was provided by Seneca Foundation Murcia (Spain).
All the studies performed in the present PhD Thesis were supported by:
The Spanish Ministry of Economy and Competitiveness and FEDER EU funds
(AGL2012-39903; AGL2015-69738-R; CTQ2014-52765-R),
Seneca Foundation, Murcia (Spain, 19892/GERM/15),
Generalitat Valenciana (Spain, PROMETEO/2016/145)
FORMAS-Stockholm (221-2011-512; 2017-00946),
Swedish Research Council VR-Stockholm (521-2011-6353),
and Research Council in Southeast Sweden (FORSS, 378091/312971; 745971).

To my parents,

To my best coincidence Pablo,

To my favorite partner.

ACKNOWLEDGEMENTS

I can hardly believe it, I started working on this project almost six years ago. It has been a long road, full of illusion, and joys but also hard moments. It is precisely for those people responsible that today I am only able to remember the good times, smiles and love, to whom I dedicate these acknowledgements.

Firstly, I would like to thank my director, **Professor Jordi Roca**. For choosing me to carry out the present PhD Thesis under your supervision, when I was wandered around the laboratories without direction. For your patience, your advices and your support. After these years working together I am more demanding of myself, more methodical in the work, more professional and better person.

I would also like to thank my co-director, **Dr. Inmaculada Parrilla**, for teaching me not to give up, for being a model researcher, for your dedication and passion in everything you do. Thank you for your encouraging words and hugs whenever I needed them. I will never forget it.

I would like to give my sincerely thanks to the **Professor Emilio Martinez** for showing me that being an excellent professional is not incompatible with being a great person, for making me always feel that I count on your support, and with your wise advice. I hope to count on it for many more years. Also my sincere thanks to the **Professor Heriberto Rodriguez Martinez**. Thank you for every suggestion, every inspiring word, for your selfless help. I am grateful for all your guidance and assistance in the completion of this PhD Thesis. It has been a pleasure to work with you.

To the **Drs. Cristina Cuello** and **Maria Antonia Gil** for being exemplary women for me, a great inspiration and models to follow and to **Dr. Xiomara Lucas**, for whom I feel a special weakness. Thank you for your wise advice and for always treatating me with so much love.

To my eternal friend **Dr. Isabel Barranco**, much more than the perfect partner, my friend, my sister. Because meeting you has been a reason and not a coincidence. The reason to

make that trip to the worst corner of Spain a "planazo" or to stay for several months in another country, away from our family, and turn it into one of my best memories. Thank you for guiding me on this road, for making these years easier for me and for loving me so much. If this experience has been wonderful is simply because of you. To **Lorena Padilla** thank you for your joy, for your generosity and predisposition. For making the last years so easy and for having completed perfectly what seemed impossible to improve. Now I know that you will be always in my life.

To my girls, thanks for the laughs, the tears, the hugs, the dinners, the walks, ... thanks for EVERYTHING. **Dr. Carmen Rodenas** for making us happy every day even when you are “crispi”, to **Dr. Cristina Martinez** for your great kindness and generosity, I do not remember very well how you reached my heart, but I am sure that you arrived to stay, come what may. **Dr. Alicia Nohalez** my “ñusa”, because since we met life has kept us together, have been an incredible years that have allowed us to know each other perfectly and love us as we are. It seems that now our lifes take different ways but I know that at some point they will intersect again. To **Dr. Carolina** Maside for bringing commun sense to the office, for your advice as an older sister, for your comforting hugs.

I would like to express my appreciation to **Dr. Maria Jose Martinez**, I have always felt that our lives were in a certain way parallel, so I see in you as a great example to follow. **Dr. Junwei Li** for your huge heart and for helping me with a smile whenever I needed it. I hope to see you soon. Here in Murcia you will always have your "other" family. To **Lola Ortega** Thank you for your help in every lab work, but especially in the personal matter. My back to back partner. To **Dr. Diego Vilela** and **Tatiana Tarantini** for treating me so well in my first steps. I still miss both of you. And of course, thanks you to the boys “beyond the wall” **Moisés** and **Miquel** for your predisposition and kindness always.

A very special gratitude goes to **Drs. Alejandro Vicente-Carrillo**, **Manuel Álvarez** and **Marie Rùber** for having turned Linköping into one of my best memories for your help and friendship inside and outside the laboratory.

I would like to thank the team of **Proteomics Unit of the University of Valencia**, especially to Luz Valero and also Raquel Gavidia of the Statistics Section from the same

university, thank you for your infinite patience, for making me discover the fascinating world of proteomics and make my path easier.

I would like to thank the company **AIM Ibérica**, for providing us semen samples for all the experiments in this thesis. I am grateful to Dr. Alfonso Bolarin, Dr Jonatan Sanchez, Manuel Molina, José Ramón and Paco for their generosity and kindness.

I would like to show my gratitude to the **Seneca Foundation** of Murcia for the financial support and for giveing me the opportunity to grow professionally.

To my great friends **Tatiana, Mari** and **Maria Vigueras** because despite the distance and the absences I know that I can always count on you.

To all my **family** for showing me what is truly important in life. The union and your unconditional support. For being present in all the important moments of my life. I love all of you.

To my "other family" especially **Pablo** and **Maria del Mar** for trusting me, for having shared with pride every little achievement, for loving me as a daughter.

To my **parents** because of you I am who I am. For teaching me the importance of being a person with values and principles. For transmitting to me that I could achieve any goal that I would propose if I worked hard. For always believing in me. My favorite place will always be by your side. To **you**, for completing our lives, making us so happy since the first day we met you and for, without being your conscious, became to be my motivation, my hope and my joy.

Finally, I would like to thank you **Pablo** for believing in me as no one else has before, for believing myself capable of everything even when I did not trust myself. For loving me unconditionally in the good days but also in my bad ones. For supporting me in every decision and feeling my triumphs and worries like your own. Knowing that I have my whole lifetime to share with you is my greatest joy.

INDEX

INTRODUCTION

1. Introduction

Consumption demands for pork meat of high quality is continuously increasing. Spain is the main European producer with almost 30 million pigs registered in 2017 (Interporc, 2017). **Artificial insemination** (**AI**) is used worldwide by the swine industry, being considered the best breeding tool to disseminate genetic progress and, in consequence, to efficiently improve swine production and the quality of pork meat (Riesenbeck, 2011). Semen AI-doses, usually at liquid state, are daily produced by specialized AI-centers following the strictest sanitary quality controls to ensure the highest standards of sperm quality. The use of innovative technologies for semen evaluation, such as computer-assisted sperm analysis and flow-cytometry based procedures, are relevant tools during this process (Broekhuijse et al., 2015). Despite this monitoring of quality and sperm function, between 5 to 7% of boars with apparently good sperm quality and thereby regularly used for producing semen AI-doses are sub-fertile, constraining fertility and ultimately leading to serious productive and economic losses to farmers (Roca et al., 2015). Furthermore, the efficient use of **frozen-thawed** (**FT**) **semen** in swine commercial AI-programs remains a challenge (Knox, 2016). The improvements made in protocols for freezing and thawing boar semen in the recent years have led to a noticeable improvement in quality and sperm function after thawing (Yeste et al., 2017), but the fertility outcomes of FT-semen remain lower than those of liquid-stored semen (Knox, 2016). This reality highlights the need of developing new evaluation tests, capable of identifying features, either in spermatozoa or the accompanying **seminal plasma** (**SP**), that allow to better predict the fertility potential of boars used in AI-programs, independently of the semen AI-doses being in a liquid state or cryopreserved.

The boar ejaculate is expelled in fractions easily recognized during ejaculation, being the so-called **sperm-rich ejaculate fraction** (**SRF**) and de post-SRF the two main fractions. The first one emitted, the SRF, contains most spermatozoa and a comparatively little volume of SP while the composition of the post-SRF contains distinctively few spermatozoa and a large SP volume (Rodriguez-Martinez et al., 2011). The SRF is the fraction targeted when semen collection is manually done (gloved-hand method), initially because if facilitates the handling of the major part of the spermatozoa in relatively small volumes (approx. 50-70 mL) during production of AI-doses. Moreover, it is well known

that spermatozoa from the two boar ejaculate fractions response differently to sperm biotechnologies, such as cryopreservation (Saravia et al., 2009; Alkmin et al., 2014). Recent research has proven that spermatozoa from the SRF have a higher cryosurvival than those derived from an entire ejaculate (i.e. SRF plus post-SRF) (Li et al., 2018). Moreover, it has been experimentally demonstrated that the spermatozoa entering the female genital tract, as those the first fraction during natural mating, are those that overwhelmingly colonize the oviductal sperm reservoirs, and thus are those potentially first involved in fertilization (Wallgren et al., 2010). However, the AI-centers are replacing the usual gloved-hand collecting method for semiautomatic methods, for reasons related to sanitary conditions and lower costs of labor; which implies abandoning the current practice of selectively collecting solely the SRF to necessarily end collecting the entire ejaculate in a single vial (Roca et al., 2015). Whether this new practice has biological relevance for sperm function, fertility and the way spermatozoa of the different fractions of the boar ejaculate respond to biotechnologies demands clarification.

Proteins, from both from spermatozoa and the SP, play an important role in the success of fertilization (Rodriguez-Martinez et al., 2011; Caballero et al., 2012). They are involved in main sperm functions, i.e. motility, capacitation, oocyte binding and acrosome reaction (Caballero et al., 2012). They also facilitate the “safe passage” of spermatozoa throughout the uterus and contribute to embryo implantation (Bromfield, 2016). Moreover, in porcine, alike other mammalian species, some specific sperm- or SP-proteins have been identified as potential biomarkers of both fertility and sperm freezability (Yeste, 2016; Rahman et al., 2017). These findings are promising and encourage further study of the proteins of both spermatozoa and SP in pigs. Accordingly, proteomics, a large-scale study of a proteome, is the ideal tool to address these issues. To do this, there are current available accurate and highly sensitive methodologies based on mass spectrometry, which allow the global identification and quantification of the cell proteome (Gallien and Domon, 2015). These methods would allow to decode the entire proteome of both spermatozoa and boar SP and, at the same time, to identify quantitative differences in proteins having direct impact in both functionality and sperm fertility ability.

References

Alkmin DV, Perez-Patiño C, Barranco I, Parrilla I, Vazquez JM, Martinez EA, Rodriguez-Martinez H, Roca J. Boar sperm cryosurvival is better after exposure to seminal plasma from selected fractions than to those from entire ejaculate. Cryobiology. 2014, 69:203-10.

Broekhuijse ML, Gaustad AH, Bolarin Guillén A, Knol EF. Efficient Boar Semen Production and Genetic Contribution: The Impact of Low-Dose Artificial Insemination on Fertility. Reprod Domest Anim. 2015, 50:103-9.

Bromfield JJ. A role for seminal plasma in modulating pregnancy outcomes in domestic species. Reproduction. 2016, 152:R223-R232.

Caballero I, Parrilla I, Almiñana C, del Olmo D, Roca J, Martínez EA, Vázquez JM. Seminal plasma proteins as modulators of the sperm function and their application in sperm biotechnologies. Reprod Domest Anim. 2012, 47:12-21.

Gallien S, Domon B. Advances in high-resolution quantitative proteomics: implications for clinical applications. Expert Rev Proteomics. 2015, 12:489-98.

Knox RV. Artificial insemination in pigs today. Theriogenology. 2016, 85:83-93.

Li J, Roca J, Pérez-Patiño C, Barranco I, Martinez EA, Rodriguez-Martinez H, Parrilla I. Is boar sperm freezability more intrinsically linked to spermatozoa than to the surrounding seminal plasma? Anim Reprod Sci. 2018, pii: S0378-4320(17)30936-3.

Rahman MS, Kwon WS, Pang MG. Prediction of male fertility using capacitation-associated proteins in spermatozoa. Mol Reprod Dev. 2017, 84:749-759.

Riesenbeck A. Review on international trade with boar semen. Reprod Domest Anim. 2011, 46:1-3.

Roca J, Broekhuijse ML, Parrilla I, Rodriguez-Martinez H, Martinez EA, Bolarin A. Boar Differences In Artificial Insemination Outcomes: Can They Be Minimized? Reprod Domest Anim. 2015, 50:48-55.

Rodriguez-Martinez H, Kvist U, Ernerudh J, Sanz L, Calvete JJ. Seminal plasma proteins: what role do they play? Am J Reprod Immunol. 2011, 66:11-22.

Saravia F, Wallgren M, Johannisson A, Calvete JJ, Sanz L, Peña FJ, Roca J, Rodríguez-Martínez H. Exposure to the seminal plasma of different portions of the boar ejaculate modulates the survival of spermatozoa cryopreserved in MiniFlatPacks. Theriogenology. 2009, 71:662-75.

Wallgren M, Saravia F, Rodríguez-Martínez H. The vanguard sperm cohort of the boar ejaculate is overrepresented in the tubal sperm reservoir *in vivo*. J Reprod Dev. 2010, 56: 68-72

Yeste M. Sperm cryopreservation update: Cryodamage, markers, and factors affecting the sperm freezability in pigs. Theriogenology. 2016, 85:47-64.

Yeste M, Rodríguez-Gil JE, Bonet S. Artificial insemination with frozen-thawed boar sperm. Mol Reprod Dev. 2017, 84:802-813.

OBJECTIVES

2. Objectives

The **main objective** of this PhD Thesis was to **update and increase current knowledge of the proteome of the spermatozoa and boar seminal plasma using mass spectrometry-based technologies, looking for potential differences between ejaculate fractions; to later explore which of the identified proteins would explain differences in sperm fertility and freezability.**

To achieve this purpose the following specific objectives were proposed:

Objective 1: Decode the boar seminal plasma proteome and update the boar sperm proteome, improving current knowledge and generating new accessible databases. – **Studies 1, 2, 3 and 4** –

Objective 2: Identify functional proteins in seminal plasma and boar spermatozoa differentially expressed between ejaculate fractions. – **Studies 1 and 4** –

Objective 3: Evaluate putative differences in the seminal plasma proteome between boars with differences in fertility. – **Study 3** –

Objective 4: Identify and characterize boar sperm proteins differentially expressed between frozen-thawed spermatozoa from semen samples showing good or bad sperm freezability. – **Study 5** –

REVIEW OF THE LITERATURE

3. Review of the literature

3.1 Physiology of the pig ejaculate

The boar ejaculate is characterized by a large volume, between 200 and 300 mL, and a relatively low sperm concentration, averaging between 150 and 350 x 10^6 sperm per mL, although showing one of the largest total sperm numbers per ejaculate among mammals (120-200 x10^9 spermatozoa). Ejaculation is, as in all mammalian species, a concerted series of neurological and muscular events that provokes the lengthy (lasting more than 10 min) emission of semen in waves or fractions that characterizes the boar ejaculate. Those semen fractions are easily recognizable during ejaculation by their differences in color and density, in relation to the different proportions of spermatozoa and **seminal plasma** (**SP**). The contribution of the accessory sexual glands differs according to the ejaculate fraction, resulting in a different SP-composition among them. Two main fractions stand out in the boar ejaculate, the first in time, the so-called **sperm-rich ejaculate fraction** (**SRF**), is comparatively smallest in volume (50-70 mL) but cell rich as it contains between 80 and 90% of total ejaculated spermatozoa, which are mainly bathing in secretions from the epididymis and the prostate gland. Of this fraction is worth mentioning the relevance of the first 10 mL, as they are the sperm-peak and the ones least mixed with accessory glands secretions (Rodriguez-Martinez et al., 2009). The second in time is the so-called post-SRF fractions, the largest in volume (usually above 150 mL) and poorest in spermatozoa (between 10 and 20% of total ejaculated spermatozoa). These spermatozoa bathe in secretions mainly from the seminal vesicles and the bulbourethral glands, building a SP-fraction particularly rich in proteins (Mann and Lutwak-Mann, 1981).

Spermatozoa have a unique and very definite goal, to deliver the paternal genome to mature oocytes. To achieve this key goal, the spermatozoa leave the seminiferous tubes at spermiogenesis to be transported through the epididymis where they complete maturation, acquiring motile capacity and fertilization competence, to ultimately remain stored in a quiescent state in the cauda epididymis, awaiting ejaculation (Sullivan and Mieusset, 2016). Mature spermatozoa are highly differentiated cells with a minimal amount of cytosol and organelles. Moreover, they are structurally complex cells having a proportionally large surface of plasma membrane compared to other cells, a membrane

which plays an essential role in sperm functionality (Gadella, 2008). The sperm plasma membranes is structured by a lipid bilayer and polar lipids associated with many proteins. The lipids constitute the basic structure of the membrane, while the inserted proteins are responsible for its function (Singer and Nicolson, 1972). Sperm surface proteins are involved in capacitation, oocyte zona binding, acrosome reaction and fusion to oocyte oolemma (Gadella, 2008). Primary and hyperactivated sperm motility are two essential attributes, by way of which spermatozoa can colonize the utero-tubal junction reservoirs and further complete the fertilization of the oocytes (Freitas et al., 2017). The axoneme, located in the sperm flagellum, is the main motility machinery and its structure includes a highly organized network of proteins (Inaba, 2007). Regarding sperm proteins, it is relevant to point out that mature spermatozoa are considered both transcriptionally and translationally silent cells. Consequently, they are unable for *de novo* protein biosynthesis, beyond a very limited capacity of protein synthesis in the mitochondria (Baker, 2016). Therefore, they only undergo modifications in their existing protein composition either by post-translational modifications (for instance, phosphorylation glycosylation, ubiquitination and acetylation) or by the incorporation of proteins existing in the fluids that surround them during storage in the cauda epididymis and during ejaculation, namely from the SP.

In the pig, the SP is a fluid built by the secretions from the cauda epididymis and accessory sexual glands, specifically, the seminal vesicles, prostate and bulbourethral glands (**Fig. 1**). The SP represents between 95 and 98% of the total volume of the ejaculate (Rodriguez-Martinez et al., 2009). The composition of SP differs, more quantitative than qualitative, between SRF and post-SRF fractions. This is mainly because the plasma of both fractions has a different source. The SP of the SRF mainly comes from the epididymis and prostate, while that from the post-SRF is mainly constituted by secretions of the seminal vesicles and the bulbourethral glands (Rodriguez-Martinez et al., 2009). Remarkably, the bulbourethral glands of boars and stallions secretes a gelatin-bead like secretion that coagulates into a gel once ejaculated (Mann and Lutwak-Mann, 1981). This is mainly present at the end of the ejaculate and generates a cervical plug to the cervix preventing seminal backflow once the large ejaculate volume has in the genital tract of the sow during mating. Possibly it was also evolutionarily intended to prevent that other males could impregnate sows in the wild. Obviously, during the collection of ejaculates

for **artificial insemination** (**AI**) purposes, these gelatin beads must be eliminated since they would otherwise immobilize the spermatozoa.

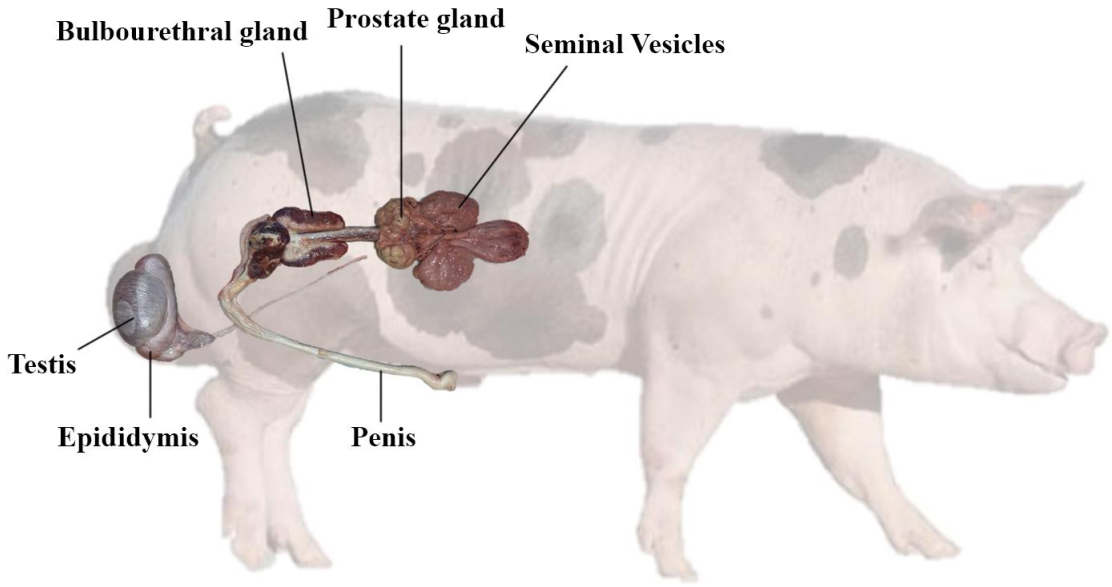


Fig. 1 Boar reproductive system.

The SP has a complex composition including proteins, electrolytes, sugars and lipids. Proteins are a major component in boar SP, ranging 30 to 60 g/L, mainly provided by the seminal vesicles, which implies that the relative concentration of proteins is substantially larger in the post-SRF than in the SRF fraction (**Fig. 2**) (García et al., 2009). Seminal plasma proteins are involved in the most relevant sperm functions, even fertilizing ability. Specifically, many of the SP proteins bind to the surface of the spermatozoa, modulating its functional capacity by either promoting or preventing key events related to specific sperm functions, including the most relevant such as motility, capacitation, binding to oocytes and acrosome reaction (Caballero et al., 2012; Juyena and Stelletta, 2012). Seminal plasma proteins would also protect spermatozoa during its transit in the sow genital tract (Troedsson et al., 2005). Moreover, once delivered in the genital tract of sows, some boar SP proteins would also contribute to regulate the temporal kinetics of ovulation and the subsequent corpus luteum development (Waberski et al., 1997). Some SP components, including proteins, facilitates early pregnancy success (Bromfield,

2018). Finally, SP proteins would contribute to minimize the negative effects of cooling or cryoinjury effects (Caballero et al., 2012; Yeste, 2016).

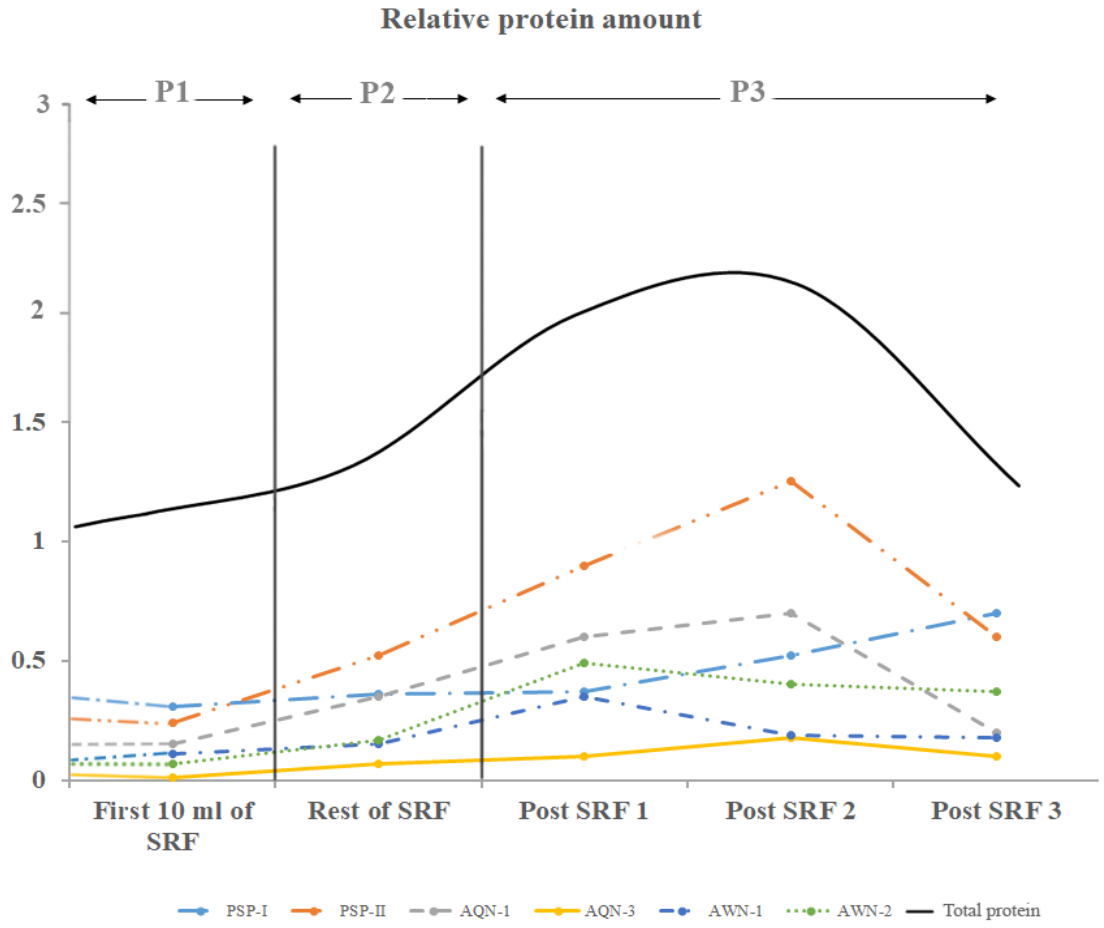


Fig. 2 Relative amount of seminal plasma proteins in consecutive portions of the boar ejaculate (first 10 mL of the sperm rich fraction [SRF], rest of SRF and post-SRF). Modified from Rodriguez-Martinez et al., 2009.

In a natural mating scenario, most spermatozoa are deposited into the sow genital tract surrounded by the secretions of SRF, avoiding any contact with secretions of post-SRF (Foxcroft et al. 2008). This scenario has been imitated for production of semen AI-doses by collecting only the SRF fraction using the routine gloved hand method. In addition, SP is either substantially extended (for liquid semen) or removed (for freezing) when semen is prepared for AI-doses. Currently this picture is changing, shifting from manual to semi-automatic semen collection where the entire ejaculate is collected in a single vial, mixing all fractions (Roca et al., 2016). In consequence, the spermatozoa necessarily

come into contact with the protein-rich SP of the post-SRF. Moreover, the extension rate of semen with commercial extenders for production of liquid-stored semen AI-doses has been reduced substantially as the volume of AI-doses has diminished from 100-80 mL to 50-40 mL when using the post-cervical insemination approach (Roca et al., 2016). This new scenario means that boar spermatozoa meet a mixture of SP of both ejaculate fractions (SRF plus post-SRF) and that a significant proportion of SP is to be inseminated.

3.2 Strategies to study the proteome of seminal plasma and spermatozoa

Proteomics is an emerging tool focused on the qualitative and quantitative study of gene expression at proteomic level. It is intended to improve the knowledge of molecular mechanisms involved in biological processes. Proteomics, defined as a large-scale study of a proteome, provide relevant information of the entire set of proteins expressed by a cell type, tissue, organism or body fluid, including semen (Chandramouli and Qian, 2009).

Mass spectrometry (**MS**)-based technologies are currently the pivotal techniques for proteomic studies (Hedrick et al., 2015). However, the MS requires additional techniques for the complete characterization of proteome (Mirza and Olivier, 2008). Then, the proteomics study of complex samples in a single experiment requires an appropriate workflow design combining analytical techniques for efficient separation or fractionation, identification and protein quantification. In addition, complex fluids, such as semen, require first of a carefully preparation of sample to avoid putative cross reactions.

3.2.1 Handling and preparation of samples

Semen is basically composed by spermatozoa and SP. Thereby, two different proteomic studies ought to be considered, one focused in sperm cells and the other to SP.

For sperm proteomic studies, the first step should be to isolate spermatozoa from the other putative cells present, such as epithelial, somatic or blood cells (Jodar et al., 2017). The most common procedures are based on the centrifugation of semen samples through a colloid gradient, allowing the separation of spermatozoa from SP and other putative cells (Morrell and Rodriguez-Martinez, 2010). However, these procedures are designed to recover the most robust, functional spermatozoa, discarding those non-functional.

Consequently, the sperm sample recovered using colloids does not resemble the spermatozoa in the native semen sample. In contrast to humans, the semen of healthy sires routinely used in livestock AI-programs does not contain too many non-sperm cells (Ford, 2004). Then, gentle washing of the native sample in appropriate buffers can allow the recovery of a cell population mostly constituted by spermatozoa. This approach has been used in a porcine recent sperm proteomic study carried out by Guimaraes et al. (2017).

For SP proteomic studies, the cellular portion of semen can be easily removed by simple, albeit strong, double centrifugation (Barranco et al., 2016). Attention must, however, be directed to avoid proteolytic events because SP is rich in proteases (Pilch and Mann, 2006; Casares-Crespo et al., 2016). The activity of proteases can be dramatically reduced by adding a protease inhibitor cocktail to the SP samples.

To carry out a diligent research in proteomics requires a human team expert in handling the samples, skilled in the use of technologies and competent in the analysis of the data set generated (McDermott et al., 2013). In Spain, like in other European countries, national network have been built with the purpose of supporting Spanish researchers in the field of proteomics, the so called Spanish National Institute for Proteomics (ProteoRed). ProteoRed is composed by seven working expert teams which support the coordination, integration and use of the proteomics facilities and attached laboratories ensuring the diligency of conducting proteomic studies (Martinez-Bartolome et al., 2010).

3.2.2 Proteomics analysis

Proteomics, which involves the identification and relative quantification of a set of proteins, contemplates the use of several alternative steps depending on the complexity of the protein sample and the extent of the study. Such steps include separation or fractionation, fragmentation, sequencing, identification and protein quantification (**Fig. 3**).

Firstly, proteins must be extracted from the sample and then separated, a key step for evaluating complex mixtures of proteins. This separation can be done at either protein or peptide level. At protein level, the traditional procedures used were based on gel electrophoresis. Both **one-dimensional** (**1D**) or **two-dimensional** (**2D**) **sodium dodecyl sulfate polyacrylamide gel electrophoresis** (**SDS-PAGE**) allow protein fractionation, the 2D being more useful for complex samples since it allows the fractionation of the

proteins by both their isoelectric charge and their molecular weight instead of only the molecular weight as the 1D does (du Plessis et al., 2011). Gel electrophoresis is especially useful to visualize post-translational protein modifications (Aitken and Baker, 2008), but it has some relevant technical limitations to be used as conventional method for protein fractionation in proteomic approaches. For instance, it can hardly handle solubilization of protein complexes and provides low resolution for highly charged, too basic acid, hydrophobic or extreme-molecular weight proteins (Brewis and Gadella, 2010). Alternatively, fractionation can be made after protein fragmentation at peptide level using **liquid chromatography** (**LC**). Irrespective of the method used for protein separation, proteins must be fragmented into peptides before MS analysis using proteolytic enzymes such as trypsin. The LC separates peptides according to some specific peptide characteristics such as hydrophobicity, size, charge or the presence of specific molecules, and, therefore, it substantially increases the number of proteins identified with respect to the methods based on gel electrophoresis (Hubner et al., 2010). Consequently, LC is currently the most useful method for peptide fractionation or separation. In samples with complex protein composition, such as SP, abundant proteins usually mask other proteins, hindering their identification. Some of these less abundant proteins are often the most important in biological processes (Brewis and Gadella, 2010). Then, it is imperative to make these proteins emerge so they can be identified, which requires using more than one fractionation procedure before MS analysis, for instance, gel electrophoresis together with LC.

For protein identification, shotgun methods as MS have proven effective, reliable, sensitive and accurate to identify hydrophobic and lowest abundant proteins in samples with complex composition in proteins (Mirza and Olivier, 2008). Protein identification by MS can be carried out by one of two methods: **peptide mass fingerprinting** (**PMF**) or sequence-specific peptide fragmentation (Thiede et al., 2005). The PMF by **matrix assisted laser/desorption ionization-mass spectrometry** (**MALDI-MS**) compares the intact peptide mases from sample problem with an *in silico* digested database (Brewis and Gadella, 2010). This method is currently used but it is not enough for complete protein identification in extent proteomics studies. For this purpose, the sequence-specific peptide fragmentation is more sensitive than MALDI-MS, becoming a key platform for proteomic analysis. Amino acid sequence is achieved by **tandem mass spectrometry** (**MS/MS**) on-line coupled with LC, so peptides eluted from the LC column are sequentially fragmented

during collision-induced process in a collision cell of the MS, and a resulting MS/MS spectra is acquired for each fragmented peptide. This method, LC together with tandem mass spectrometry (**LC-MS/MS**), allows for an accurate and efficient analysis of proteins existing in complex samples (Oliva et al., 2009).

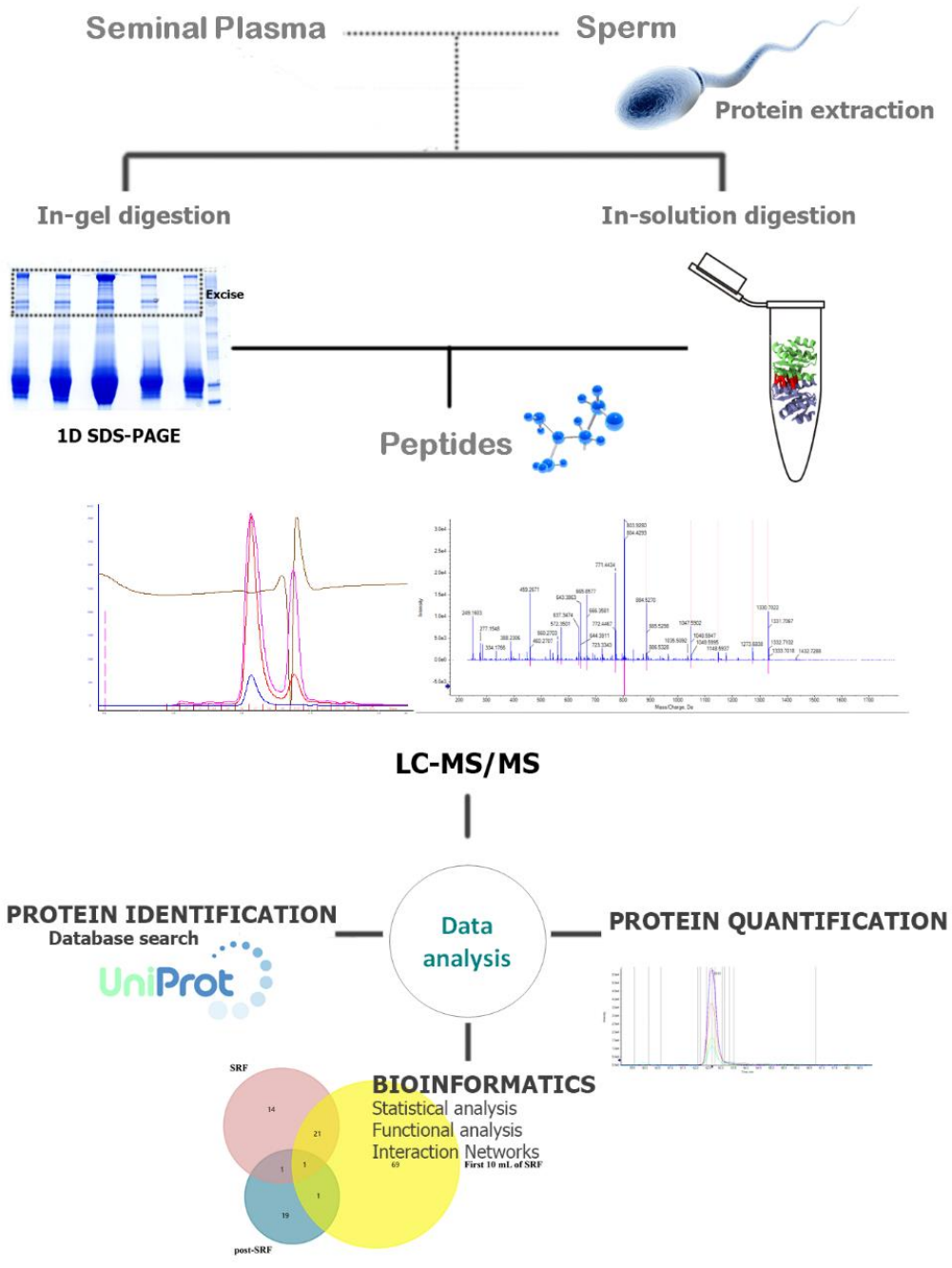


Fig. 3 Diagram of proteomics workflow illustrating the sample collection, protein extraction and purification, sample fractionation, mass spectrometry analysis and data analysis.

The comparative proteomics and the protein quantification allow to search differences in protein composition among individuals and samples, identifying protein markers of biological and pathological processes, such as sperm function and fertility. **Differential proteomics in gel electrophoresis** (**DIGE**) is a form of gel electrophoresis used to detect quantitative differences in protein composition between samples (du Plessis et al., 2011). This method allows, with great resolution, comparisons of the relative amount of proteins between few samples. For this, each sample must be pre-labelled with a differential fluorescent dye (CyDyes), mixed and then separate the using the 2D SDS-PAGE (Lilley and Friedmann, 2004). Approaches like **stable isotope labeling with amino acids in cell culture** (**SILAC**) and **isotope coded affinity tag** (**ICAT**) can compare protein concentrations of two or three different samples in a single LC-MS run (Gygi et al., 1999; Ong et al., 2002). Nowadays, the most used labeling method is the so-called **isobaric tag for relative and absolute quantitation** (**iTRAQ**) because it allows simultaneous comparison of protein quantification of up to 8 samples. The iTRAQ is based on the labeling of the peptides resulting from protein digestions with variable mass labels. Once labelled, the samples are mixed and fractionated by LC and analyzed by MS/MS (Zieske, 2006). The main limitations of this labeling-based method is the number of samples that can be analyzed simultaneously. Recently, a label-free method has been developed to relative protein quantification, the so called **sequential window acquisition of all theoretical mass spectra** (**SWATH**; Gillet et al., 2012). This method allows to quantify relative differences in peptides, and then in the protein expression, by using data-independent acquisition. The SWATH continuously fragments all the peptides with stepped m/z windows and the resulting transition ions are matched to a spectral library, simultaneously generating protein identification and its label-free quantification. Quantitative values are obtained as areas under the peaks and normalized to minimize non-biological variation. The SWATH is less restrictive than iTRAQ to identify proteins, because iTRAQ requires high collision energies leading to a loss of sequence-informative fragments (Wiese et al., 2007).

Bioinformatics is the last key step of proteomics (Persson, 2000). The significant improvements in bioinformatics databases carried out in the last few years facilitate the comprehensive analysis of proteins found in complex samples. The most commonly used databases for livestock species proteomic-studies are the **Dataset for Annotation, Visualization and Integrated Discovery** (**DAVID**), **Protein ANalysis THrough**

Evolutionary Relationships (**PANTHER**), **Gene Set Enrichment Analysis** (**GSEA**), UniProt KB, ENSEMBL and the **National Center for Biotechnology Information-non redundant** (**NCBI-nr**). In addition to the amino acid sequence and related gene of each one of the proteins for each vertebrate species, these databases also provide functional information for comparative proteomics, protein annotation, compute multiple alignments, predict regulatory function and collects data of biological or pathological processes where proteins are involved. These databases are being continuously updated, but information about domestic animals such as *Sus scrofa* is still quite limited. Given this scenario, the combination of several of these databases for the functional study of proteins seems a good option but there is a great limitation to extrapolate identifiers from one platform to another; generating important information losses. Furthermore, results of recent studies of the boar sperm proteome, as the one conducted for Feugang et al. (2018), highlights the lack of information in the above listed proteomics databases available, forcing to convert the obtained identifiers for boar sperm to their human homologs to annotate putative functions for the identified proteins. This scenario highlights the current need to homogenize and update databases to face and manage the proteomics-derived data.

3.3 Current proteome knowledge of sperm and boar seminal plasma

With the advance in analytical techniques, the research in proteins expressed by either spermatozoa or boar SP has substantially increased in the last years. Most of them have focused on studying specific proteins, evaluating their involvement in sperm physiology and thereby their potential value as biomarkers of functionality or even boar sperm fertilizing capacity (Dyck et al., 2011; Caballero et al., 2012; Rahman et al., 2017). However, few are still comprehensive proteomic studies.

Regarding spermatozoa, remarkable is the research carried out in the Chung-Ang University (Republic of Korea), which has provided impressive proteomic studies decrypting proteins involved in boar sperm capacitation (Kwon et al., 2014) and their value as predictive biomarkers of boar fertility (Kwon et al., 2015). More recently Feugang and coworkers (2018) have published the first large-scale study of boar sperm proteome, identifying more than 2,700 proteins. Unfortunately, only a few number of these proteins could be encoded into *Sus scrofa* taxonomy. Future sperm proteomic

studies will probably encode more proteins into *Sus scrofa* taxonomy as the annotation of protein-coding genes has recently been substantially extended, with 22,737 new entries created since November of 2017 of a total of 40,710 proteins registered in the UniProt KD database (Marx et al., 2017). Once boar sperm proteome fully decoded, efforts should be directed toward relating proteome changes with sperm functionality. In this regard, interesting findings have been achieved in other species, mainly in human (Sharma et al., 2013; Pereira et al., 2017; Lamy et al., 2018)

Regarding boar SP, a large-scale study of its proteome is still pending. So far, two proteomic studies have been carried out, identifying a total of 80 (Druart et al., 2013) and 39 proteins (Gonzalez-Calavid et al., 2014). This reality contrasts to other livestock species where the SP proteome is more decoded, for instance bovine (Kelly et al., 2006), equine (Novak et al., 2010), ovine (Soleilhavoup et al., 2014) or poultry (Labas et al., 2014). Fully decrypting the boar SP-proteome is still a challenge because most total protein mass is taken up by a few number of low molecular weight proteins, the so-called most abundant proteins, that mask the less abundant ones, preventing their isolation and identification.

References

Aitken RJ, Baker MA. The role of proteomics in understanding sperm cell biology. Int J Androl. 2008, 31:295-302.

Baker MA. Proteomics of post-translational modifications of mammalian spermatozoa. Cell Tissue Res. 2016, 363:279-87.

Barranco I, Tvarijonaviciute A, Perez-Patiño C, Vicente-Carrillo A, Parrilla I, Ceron JJ, Martinez EA, Rodriguez-Martinez H, Roca J. Glutathione Peroxidase 5 Is Expressed by the Entire Pig Male Genital Tract and Once in the Seminal Plasma Contributes to Sperm Survival and In Vivo Fertility. PLoS One. 2016, 11:e0162958.

Brewis IA, Gadella BM. Sperm surface proteomics: from protein lists to biological function. Mol Hum Reprod. 2010, 16:68-79.

Bromfield JJ. Review: The potential of seminal fluid mediated paternal-maternal communication to optimise pregnancy success. Animal. 2018, 19:1-6.

Caballero I, Parrilla I, Almiñana C, del Olmo D, Roca J, Martínez EA, Vázquez JM. Seminal plasma proteins as modulators of the sperm function and their application in sperm biotechnologies. Reprod Domest Anim. 2012, 47:12-21.

Casares-Crespo L, Vicente JS, Talaván AM, Viudes-de-Castro MP. Does the inclusion of protease inhibitors in the insemination extender affect rabbit reproductive performance? Theriogenology. 2016, 85:928-932.

Chandramouli K, Qian PY. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. Hum Genomics Proteomics. 2009, pii: 239204.

Druart X, Rickard JP, Mactier S, Kohnke PL, Kershaw-Young CM, Bathgate R, Gibb Z, Crossett B, Tsikis G, Labas V, Harichaux G, Grupen CG, de Graaf SP. Proteomic characterization and cross species comparison of mammalian seminal plasma. J Proteomics. 2013, 91:13-22.

du Plessis SS, Kashou AH, Benjamin DJ, Yadav SP, Agarwal A. Proteomics: a subcellular look at spermatozoa. Reprod Biol Endocrinol. 2011, 9:36.

Dyck, MK, Foxcroft GR, Novak S, Ruiz-Sanchez A, Patterson J, Dixon WT. Biological markers of boar fertility. Reprod Domest Anim. 2011, 46:55-8.

Feugang JM, Liao SF, Willard ST, Ryan PL. In-depth proteomic analysis of boar spermatozoa through shotgun and gel-based methods. BMC Genomics. 2018, 19:62.

Ford, WC. Regulation of sperm function by reactive oxygen species. Hum Reprod Update. 2004, 10:387-99.

Foxcroft GR, Dyck MK, Ruiz-Sanchez A, Novak S, Dixon WT. Identifying useable semen. Theriogenology. 2008, 70:1324-36.

Freitas MJ, Vijayaraghavan S, Fardilha M. Signaling mechanisms in mammalian sperm motility. Biol Reprod. 2017, 96:2-12.

Gadella BM. The assembly of a zona pellucida binding protein complex in sperm. Reprod Domest Anim. 2008, 43:12-9.

García EM, Calvete JJ, Sanz L, Roca J, Martínez EA, Vázquez JM. Distinct effects of boar seminal plasma fractions exhibiting different protein profiles on the functionality of highly diluted boar spermatozoa. Reprod Domest Anim. 2009, 44:200-5.

Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics. 2012, 11:O111.016717.

González-Cadavid, V., Martins, J.A., Moreno, F.B., Andrade, T.S., Santos, A.C., Monteiro-Moreira, A.C., Moreira, R.A., and Moura, A.A. Seminal plasma proteins of adult boars and correlations with sperm parameters. Theriogenology. 2014, 82:697-707.

Guimarães DB, Barros TB, van Tilburg MF, Martins JAM, Moura AA, Moreno FB, Monteiro-Moreira AC, Moreira RA, Toniolli R. Sperm membrane proteins associated with the boar semen cryopreservation. Anim Reprod Sci. 2017, 183:27-38.

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol. 1999, 17:994-9.

Hedrick VE, LaLand MN, Nakayasu ES, Paul LN. Digestion, Purification, and Enrichment of Protein Samples for Mass Spectrometry. Curr Protoc Chem Biol. 2015, 7:201-22.

Hubner NC, Ren S, Mann M. Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. Proteomics. 2008, 8:4862-72. Brewis IA, Gadella BM. Sperm surface proteomics: from protein lists to biological function. Mol Hum Reprod. 2010, 16:68-79.

Inaba K. Molecular basis of sperm flagellar axonemes: structural and evolutionary aspects. Ann N Y Acad Sci. 2007, 1101:506-26.

Jodar M, Soler-Ventura A, Oliva R, Molecular Biology of Reproduction and Development Research Group. Semen proteomics and male infertility. J Proteomics. 2017, 162:125-134.

Juyena NS, Stelletta C. Seminal plasma: an essential attribute to spermatozoa. J Androl. 2012, 33:536-51.

Kelly VC, Kuy S, Palmer DJ, Xu Z, Davis SR, Cooper GJ. Characterization of bovine seminal plasma by proteomics. Proteomics. 2006, 6:5826-33.

Kwon WS, Rahman MS, Lee JS, Kim J, Yoon SJ, Park YJ, You YA, Hwang S, Pang MG. A comprehensive proteomic approach to identifying capacitation related proteins in boar spermatozoa. BMC Genomics. 2014, 15:897.

Kwon WS, Rahman MS, Lee JS, Yoon SJ, Park YJ, Pang MG. Discovery of predictive biomarkers for litter size in boar spermatozoa. Mol Cell Proteomics. 2015, 14:1230-40.

Labas V, Grasseau I, Cahier K, Gargaros A, Harichaux G, Teixeira-Gomes AP, Alves S, Bourin M, Gérard N, Blesbois E. Data for chicken semen proteome and label free quantitative analyses displaying sperm quality biomarkers. Data Brief. 2014, 1:37-41.

Lamy J, Nogues P, Combes-Soia L, Tsikis G, Labas V, Mermillod P, Druart X, Saint-Dizier M. Identification by proteomics of oviductal sperm-interacting proteins. Reproduction. 2018, 155:457-466.

Lilley KS, Friedman DB. All about DIGE: quantification technology for differential-display 2D-gel proteomics. Expert Rev Proteomics. 2004, 1:401-9.

Mann T, Lutwak-Mann C. Male reproductive function and semen.Berlin/ Heidelberg/ New York: Springer Verlag, 1981.

Martinez-Bartolome S, Blanco F, Albar JP. Relevance of proteomics standards for the ProteoRed Spanish organization. J Proteomics. 2010, 73:1061-6.

Marx, H., Hahne, H., Ulbrich, S.E., Schnieke, A., Rottmann, O., Frishman, D., and Kuster, B. Annotation of the domestic pig genome by quantitative proteogenomics. J Proteome Res. 2017, 16:2887-2898.

McDermott JE, Wang J, Mitchell H, Webb-Robertson BJ, Hafen R, Ramey J, Rodland KD. Challenges in Biomarker Discovery: Combining Expert Insights with Statistical Analysis of Complex Omics Data. Expert Opin Med Diagn. 2013, 7:37-51.

Mirza SP, Olivier M. Methods and approaches for the comprehensive characterization and quantification of cellular proteomes using mass spectrometry. Physiol Genomics. 2008, 33:3-11.

Morrell JM, Rodriguez-Martinez H. Practical applications of sperm selection techniques as a tool for improving reproductive efficiency. Vet Med Int. 2010, pii: 894767.

Novak S, Smith TA, Paradis F, Burwash L, Dyck MK, Foxcroft GR, Dixon WT. Biomarkers of in vivo fertility in sperm and seminal plasma of fertile stallions. Theriogenology. 2010, 74:956-67.

Oliva R, de Mateo S, Estanyol JM. Sperm cell proteomics. Proteomics. 2009, 9:1004-17.

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002, 1:376-86.

Pereira R, Sá R, Barros A, Sousa M. Major regulatory mechanisms involved in sperm motility. Asian J Androl. 2017, 19:5-14.

Persson B. Bioinformatics in protein analysis. EXS. 2000; 88:215-31.

Pilch B, Mann M. Large-scale and high-confidence proteomic analysis of human seminal plasma. Genome Biol. 2006, 7:R40.

Rahman MS, Kwon WS, Pang MG. Prediction of male fertility using capacitation-associated proteins in spermatozoa. Mol Reprod Dev. 2017, 84:749-759.

Roca J, Parrilla I, Bolarin A, Martinez EA, Rodriguez-Martinez H. Will AI in pigs become more efficient? Theriogenology. 2016, 86:187-93.

Rodríguez-Martínez H, Kvist U, Saravia F, Wallgren M, Johannisson A, Sanz L, Peña FJ, Martínez EA, Roca J, Vázquez JM, Calvete JJ. The physiological roles of the boar ejaculate. Soc Reprod Fertil Suppl. 2009, 66:1-21.

Sharma R, Agarwal A, Mohanty G, Hamada AJ, Gopalan B, Willard B, Yadav S, du Plessis S. Proteomic analysis of human spermatozoa proteins with oxidative stress. Reprod Biol Endocrinol. 2013, 11:48.

Singer SJ, Nicolson GL. The fluid mosaic model of the structure of cell membranes. Science. 1972, 175:720-31.

Soleilhavoup C, Tsikis G, Labas V, Harichaux G, Kohnke PL, Dacheux JL, Guérin Y, Gatti JL, de Graaf SP, Druart X. Ram seminal plasma proteome and its impact on liquid preservation of spermatozoa. J Proteomics. 2014, 109:245-60.

Sullivan R, Mieusset R. The human epididymis: its function in sperm maturation. Hum Reprod Update. 2016, 22:574-87.

Thiede B, Höhenwarter W, Krah A, Mattow J, Schmid M, Schmidt F, Jungblut PR. Peptide mass fingerprinting. Methods. 2005, 35:237-47.

Troedsson MH, Desvousges A, Alghamdi AS, Dahms B, Dow CA, Hayna J, Valesco R, Collahan PT, Macpherson ML, Pozor M, Buhi WC. Components in seminal plasma regulating sperm transport and elimination. Anim Reprod Sci. 2005, 89:171-86.

Waberski D, Claassen R, Hahn T, Jungblut PW, Parvizi N, Kallweit E, Weitze KF. LH profile and advancement of ovulation after transcervical infusion of seminal plasma at different stages of oestrus in gilts. J Reprod Fertil. 1997, 109:29-34.

Wiese S, Reidegeld KA, Meyer HE, Warscheid B. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics. 2007, 7:340-50.

Yeste M. Sperm cryopreservation update: Cryodamage, markers, and factors affecting the sperm freezability in pigs. Theriogenology. 2016, 85:47-64.

Zieske LR. A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. J Exp Bot. 2006, 57:1501-8.

STUDY 1

This study has been published in **Journal of Proteomics** with the title

"**CHARACTERIZATION OF THE PORCINE SEMINAL PLASMA PROTEOME COMPARING EJACULATE PORTIONS**"

Journal of Proteomics 2016, 142:15-23

DOI: 10.1016/j.jprot.2016.04.026

Abstract

Full identification of boar seminal plasma (**SP**)-proteins remains challenging. This study aims to provide an extensive proteomic analysis of boar SP and to generate an accessible database of boar SP-proteome. A SP-pool (33 entire ejaculates/11 boars; 3 ejaculates/boar) was analyzed to characterize the boar SP-proteome. Twenty ejaculates (5 boars, 4 ejaculates/boar) collected in portions (P1: first 10 mL of sperm-rich ejaculate fraction (**SRF**), P2: rest of SRF and P3: post-SRF) were analyzed to evaluate differentially expressed SP-proteins among portions. SP-samples were analyzed using a combination of size exclusion chromatography (**SEC**), one dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (**1D SDS-PAGE**) and Nano liquid chromatography-electrospray ionization-tandem mass spectrometry (**LC-ESI-MS/MS**) followed by functional bioinformatics. The identified proteins were quantified from normalized label free quantification (**LFQ**) intensity data. A total of 536 SP-proteins were identified, 409 of them in *Sus scrofa* taxonomy (374 validated with ≥ 99% Confidence). Barely 20 of the identified SP-proteins were specifically implicated in reproductive processes, albeit other SP-proteins could be indirectly involved in functionality and fertility of boar spermatozoa. Thirty-four proteins (16 identified in *Sus scrofa* taxonomy) were differentially expressed among ejaculate portions, 16 being over-expressed and 18 under-expressed in P1-P2 regarding to P3. This major proteome mapping of the boar SP provides a complex inventory of proteins with potential roles as sperm function- and fertility- biomarkers.

Keywords: porcine, ejaculate, seminal plasma, proteome

IV. Study 1

4.1 Introduction

Artificial insemination (**AI**) is used worldwide in the swine industry (Riesenbeck, 2011). Nowadays, commercial semen AI-doses are used as liquid state, but increasing research is done for the practical incorporation of frozen-thawed semen and even sex-sorted spermatozoa into commercial swine AI-programs (Roca et al., 2015a). Boar ejaculates are selected if most relevant sperm attributes, such as motility, morphology or membrane sperm integrity, are within acceptable limits when evaluated using innovative technologies such as computer-assisted analysis and flow-cytometry based procedures (Broekhuijse et al., 2012). Despite this pre-screening, not all selected ejaculates perform well, neither fertilizing nor overcoming technological handling (Parrilla et al., 2012). Consequently, complementary criteria for boar ejaculate selection are currently being explored, specially focusing on seminal plasma (**SP**) a composite fluid that surrounds and interacts with sperm from ejaculation to deposition into the uterus of the sow.

Seminal plasma is a complex mixture composed of exocrine secretions from the testis, epididymis and the male accessory sexual glands, containing a wide variety of both organic and inorganic components, being proteins the main one and probably the major conditioners of the reproductive success involving the use of processed semen (Jonakova et al., 2007; Rodriguez-Martinez et al., 2011). Certainly, in different mammalian species some SP-proteins have been evidenced to stabilize sperm membrane (Caballero et al., 2009), regulate sperm motility (Luna et al., 2015) and capacitation (Caballero et al., 2009; Vadnais and Roberts 2010), and further facilitate the colonization of the sperm reservoirs in the female reproductive tract (Wagner et al., 2002). Moreover, some SP-proteins, once into the female genital tract after natural mating or AI, modulate the uterine immune response against the foreign semen, facilitating sperm transport (Rodriguez-Martinez et al., 2010), sperm-oocyte fusion (Töpfer-Petersen et al., 2005) and the development of a healthy embryo (O'Leary et al., 2004). These findings provide clear evidence of the relevance of SP-proteins for sperm functional performance and further highlight the possible use of SP-proteins as potential markers for ejaculate selection. As indicated by Park et al. (2013), extensive and comprehensive knowledge of SP-proteome is an essential prerequisite before SP-proteins could be used as biomarkers of sperm functionality, or even fertility. Unfortunately, the boar SP-proteome is still far from being

completely decoded, requiring further research (Rodriguez-Martinez et al., 2011; Druart et al., 2013). Despite the economic relevance, porcine is currently the livestock species with the fewer SP-proteins identified, less than 100 (Druart et al., 2013; Gonzalez-Cadavid et al., 2014), which contrasts with other livestock species, such as poultry or ovine with more than 600 SP-proteins identified (Soleilhavoup et al., 2014; Labas et al., 2014). This shortage in porcine is surprising, considering boar ejaculates yield a large SP-volume, usually above 150 mL. Consequently, the first aim of the current study will be to improve the knowledge of the boar SP-proteome. To achieve this, SP-samples from boars with proven fertility would be examined using one dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (**1D SDS-PAGE**) and liquid chromatography-electrospray ionization-tandem mass spectrometry (**LC-ESI-MS/MS**) followed by bioinformatics of Gene Ontology (**GO**) to evaluate cellular location, molecular function and biological process of the identified proteins. In addition, the boar SP-proteome will be described so that a spectral library can be built for relative 'label free' protein quantification with sequential window acquisition of all theoretical mass spectra (**SWATH**) approach (Gillet et al., 2012).

The boar ejaculate is fractionally emitted, alike other mammalian species such as equine and even human (Rodriguez-Martinez et al., 2011). Then, it is possible to visually differentiate two main fractions during the ejaculation, a first fraction named sperm rich fraction (**SRF**), containing most of the total sperm, and a second named post-SRF, largest in volume but containing very few sperm. As boar ejaculates are traditionally hand-collected, the SRF is the only fraction sampled, discarding the post-SP and thereby most of the total SP. However, the ejaculate collection method is currently shifting from manual to semi-automated, for hygienic and labour reasons, moving from collecting just the SRF to the entire ejaculate (Roca et al., 2015b). The change increases the relevance of SP, as the large SP-volume of post-SRF fraction will now become a part of the collected ejaculate. There are yet no studies evaluating how this procedural change can affect functional sperm performance. However, in this regard, it is well known that sperm from the SRF freezes better than those from the entire ejaculate, suggesting that the SP from the post-SRF, owing to its rich protein composition, impairs sperm freezability (Alkmin et al., 2014a). Furthermore, large proportions of SP in boar semen samples impair sperm sortability and sorting efficiency in ejaculates subjected to flow cytometry for sperm sex-sorting (Alkmin et al., 2014b). In addition, (Saravia et al., 2009)

demonstrated that sperm from the first 10 mL of the SRF sustain cryopreservation better than those from the rest of the ejaculate, which was attributable to differences in SP-proteins between the ejaculate fractions. Accordingly, this study also aims to identify possible qualitative and quantitative differences in SP-proteome among the most relevant boar ejaculate portions: P1: the first 10 mL of SRF, P2: the rest of SRF and P3: the post-SRF.

4.2 Material and methods

4.2.1 Boars and ejaculates

All procedures involving boars were performed according to international guidelines (Directive 2010/63/EU), following the approval of the Bioethics Committee of Murcia University (research code: 639/2012).

Healthy and sexually mature Large White and Landrace boars, with proven fertility and undergoing regular semen collection for commercial AI (AIM Iberica, Spain), were used as ejaculate donors. Boars were subjected to diet and housing conditions characteristics of an AI-center. All the ejaculates used fulfilled the standards of sperm quantity and quality thresholds for the preparation of semen doses for use in AI programs. Thirty-three entire ejaculates collected using the semi-automatic system Collectis® (IMV Technologies, L'Aigle, France) from 11 boars (3 ejaculates per boar), were used to characterize the boar SP-proteome. In order to evaluate differences in SP-protein composition among ejaculate portions, twenty ejaculates were collected fractionated (P1-3) using the gloved hand method, from 5 boars (4 ejaculates per boar). All ejaculates used met the minimum quality requirements for AI-semen doses production.

4.2.2 Preparation of seminal plasma samples for proteome analysis

Immediately after collection, fully filled 15 mL tubes of semen from each ejaculate/portion were centrifuged twice at 1,500 *xg* for 10 min (Rotofix 32A; Hettich Zentrifugen, Tuttlingen, Germany). The second SP-supernatant recovered was microscopically verified as sperm free. A protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO, USA) was added to the SP-samples (1%, v:v) and then they were stored at -80 °C (Ultra Low Freezer; Haier, Schomberg, Ontario, Canada) until proteomic analysis.

The proteomic analyses were carried out in the Proteomics Unit of the University of Valencia, Valencia, Spain (member of the PRB2-ISCIII ProteoRed Proteomics Platform). Seminal plasma samples were thawed at room temperature, centrifuged at 16,100 *xg* at 4 °C for 1 min. The protein concentration of the SP-sample was measured using a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA) according to manufacturer instructions. Before starting the proteome analysis, the 33 SP-samples collected for characterization of boar SP-proteome were mixed in a single pool. Similarly, the 4 SP-samples of each ejaculate portion collected from each boar to identify and quantify differentially expressed proteins among ejaculate portions were also mixed in a single pool, thereby generating a total of 5 SP-pools for each ejaculate portion, one per boar.

A size exclusion chromatography (**SEC**) was carried out in an ETTAN LC system (GE Healthcare Life Sciences, Little Chalfont, United Kingdom) using a Superdex 200 5/150 GL column (GE Healthcare Life Sciences) controlled by an AKTA ETTAN LC system (GE Healthcare Life Sciences), which furthermore would ease a large-scale proteomic study. Fifty µL of the SP-pool were injected into the column, equilibrated with 200 mM ammonium bicarbonate and 1 mM dithiothreitol (**DTT**) (Sigma-Aldrich) at a flow rate of 0.18 mL/min at 4 °C. The eluent was collected in 0.2 mL fractions. It was immediately evident that it was not possible to achieve only one fraction with all proteins except the dominant ones. Therefore, 2.5 to 5 µg of protein (depending of the total amount of the sample) from the fractions were collected after SEC step (A3-A10), dried in a rotatory evaporator and loaded onto different wells of 12% Tris-HCl precast 1D SDS-PAGE (Bio-Rad, Richmond, CA, USA). The gel was run at a constant voltage of 200 mV for 30 min at room temperature including a molecular weight marker (ECL Plex Fluorescent Rainbow Marker, GE Healthcare Life Sciences), and Coomassie Brilliant R250 Blue stain (Bio-Rad) was used to visualize protein bands on the gel. Thereafter, the gel was sliced at 38 kDa, as indicated in **Fig. 4**, and the top of the gel used to analyze less abundant proteins by in-gel digestion processing. The 1D Gel fraction between 38 kDa and 20 kDa was discarded for the analysis because in a previous analysis (data not shown) this gel area was highly contaminated by the more abundant SP-proteins, providing little information to the global study of the boar SP-proteome. The more abundant proteins were identified from an aliquot of the mixed SP-sample analyzed by in-solution processing.

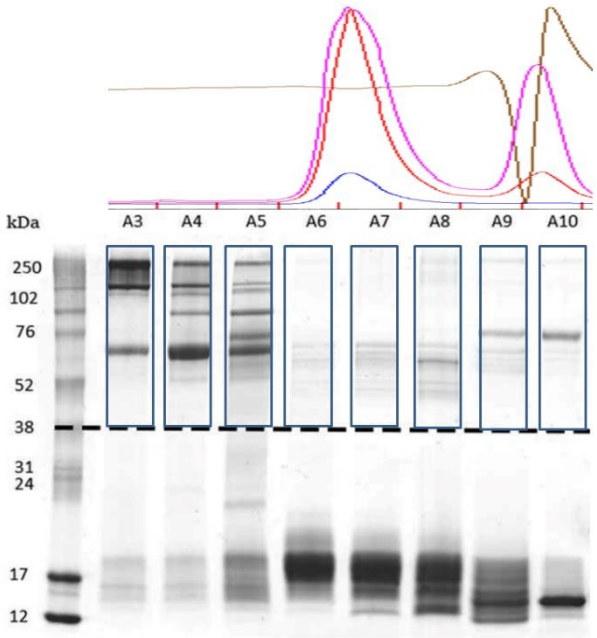


Fig. 4 Representative 1D SDS-PAGE resulting from the electrophoresis of SEC fractions obtained from a pool of boar seminal plasma (SP, 33 ejaculates). The first column corresponds to a molecular weight marker (ECL Plex Fluorescent Rainbow Marker, GE Healthcare Life Sciences) and the rest of columns (A3-A10) correspond to different times eluted fractions obtained by SEC when the pooled boar SP was injected in an ETTAN LC system (GE Healthcare Life Sciences). The gel was sliced at 38 kDa in order to reject dominant proteins (low molecular weight proteins, <38 kDa) and the top of gel used to analyze minor proteins (high molecular weight proteins, >38 kDa) by in-gel digestion processing.

4.2.3 Building a MS/MS library for SWATH analysis of boar seminal plasma

4.2.3.1 Complete proteome. In-solution digestion processing

The more abundant proteins were analyzed using in-solution digestion. Three µL of the pooled SP, containing 10 µg of proteins, were digested with Sequencing Grade Trypsin (Promega Corporation, Madison, USA) to generate peptides of each individual protein according to the following steps: (1) the proteins were reduced using 2 mM DTT reducing reagent in 50 mM NH_4HCO_3 (Sigma-Aldrich) to a final volume of 25 µL and the mixture incubated for 20 min at 60 °C; (2) the proteins were alkylated using 5.5 mM iodoacetamide (**IAM**) (Sigma-Aldrich) alkylating reagent in 50 mM NH_4HCO_3 to a final volume of 30 µL and incubated for 30 min in the dark; (3) 10 mM DTT in 50 mM

NH_4HCO_3 were added to a final volume of 60 µL and sample was vortex and incubated for 30 min at room temperature; (4) trypsin was added in a 1:20 ratio (Trypsin:Protein, w:w) to a final volume of 70 µL, the sample was carefully mixed and digestion was carried out over night at 37 °C. The digestion was stopped on the following morning by acidifying the solution (pH<6) with 7 µL 10% trifluoroacetic (**TFA**) (Fisher Scientific, Madrid, Spain). The final concentration of protein in the digested sample was 0.13 µg/µL.

4.2.3.2 Less abundant proteins. In-gel digestion processing

The portion of the 1D SDS-PAGE containing proteins with a molecular weight higher than 38 kDa was processed as a unique sample. Following by washing in MilliQ water (Merck Millipore, Darmstadt, Germany), the gel was dehydrated in acetonitrile (**ACN**) (Fisher Scientific), reduced with DTT and alkylated with IAM. Each slide was cut and small pieces of approximately 1 mm^2 in size were transferred into 1.5 mL Eppendorf tubes. Sequencing Grade Trypsin digestion of the sliced gel was performed following the protocol used by Shevchenko et al. (1996). Briefly, for protein digestion 200 µL of trypsin digestion buffer were added to each dried gel piece and digestion was set at 37 °C overnight. The trypsin digestion was stopped with 10% TFA and the supernatant, containing the non-extracted digests, was carefully removed, leaving behind the sliced gels in the Eppendorf tube. Then, for peptide extraction, 200 µL of pure ACN, were added to each tube and incubated for 15 min at 37 °C in a shaker. The new supernatant containing the peptide mixture was carefully withdrawn. Both supernatants were mixed in a tube and dried in a speed vacuum (ISS 110 SpeedVac System, Thermo Savant, ThermoScientific, Langenselbold, Germany) for 20 min and re-suspended in 7 µL of 2% ACN and 0.1% TFA prior to LC and MS analysis.

4.2.3.3 LC-MS/MS analysis

The peptides recovered from in-gel and in-solution digestion processing were examined by LC using a NanoLC Ultra 1D plus Eksigent (Eksigent Technologies, Dublin, CA, USA) which was directly connected to an AB SCIEX TripleTOF 5600 mass spectrometer (AB SCIEX, Framingham, MA, USA) in direct injection mode. Briefly, 5 µL from each digested sample were trapped on a NanoLC pre-column (3 µm particles size C18-CL, 350 µm diameter x 0.5 mm long; Eksigent Technologies) and desalted with 0.1% TFA at 3µL/min during 5 min. Then, the digested peptides present in the samples were separated using an analytical LC column (3 µm particles size C18-CL, 75 µm diameter x 12 cm

long, Nikkyo Technos Co®, Tokyo, Japan) equilibrated in 5% ACN 0.1% formic acid (**FA**) (Fisher Scientific). Peptide elution was performed by applying a mixture of solvents A and B; solvent A being 0.1% FA in water and solvent B being 0.1% FA in ACN. The peptides were eluted from the column with a linear gradient from 5% to 35% of solvent B at a constant flow rate of 300 nL/min over 90 min.

The eluted peptides were thereafter direction-ionized using an ESI Nanospray III (AB SCIEX) and then analyzed on an AB SCIEX TripleTOF 5600 mass spectrometer coupled to the NanoLC system. The samples were ionized applying 2.8 kV to the spray emitter and the TripleTOF was operated in data-dependent mode, in which a time of flight (**TOF**) MS scan was made from 350 to 1259 m/z, accumulated for 250 ms TOF followed by 150 ms TOF with the same scan rage for MS, and the 25 most abundant multiply charged (2+, 3+, 4+ or 5+) precursor peptide ions automatically selected. Ions with 1+ and unassigned charge states were rejected from the MS/MS analysis. Collision energy was automatically set by the instrument according to the equation |CE|=(slope)x(m/z)+(intercept) with Charge=2; Slope= 0.0575 and Intercept=9.

The proteomics data and result-files from the analysis have been deposited to the ProteomeXchange Consortium (Vizcaino et al., 2014) via the PRIDE partner repository, with the dataset identifier PXD003579 and DOI: 10.6019/PXD003579.

4.2.3.4 LC-SWATH-MS acquisition

For SWATH-MS-based experiments the instrument (SCIEX TripleTOF 5600) was configured as described by Gillet et al. (2012). Briefly, the mass spectrometer was operated in a looped product ion mode. In this mode, 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 and direct current voltages on the isolation quadrupole in an independent manner. Using an isolation width of 16 Da (15 Da of optimal ion transmission efficiency and 1 Da for the window overlap), a set of 37 overlapping windows was constructed covering the mass range 450–1000 Da. 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, 5 µL of each sample were loaded onto a trap column (NanoLC Column, 3µm C18-CL, 75 µm x 15 cm; Eksigent)

and desalted with 0.1% TFA at 3µL/min during 5 min. The peptides were loaded onto an analytical column (LC Column, 3 µm C18-CL, 75µmx12cm, Nikkyo) equilibrated in 5% ACN 0.1% FA. Peptide elution was carried out with a linear gradient of 5 to 40% B in 90 min (A: 0.1% FA; B: ACN, 0.1% FA) at a flow rate of 300 nL/min. Eluted peptides were infused in the spectrometer nanoESI qQTOF (SCIEX TripleTOF 5600). The TripleTOF was operated in swath mode, in which a 0.050 s TOF MS scan from 350 to 1250 m/z was performed, followed by 0.080 s product ion scans from 230 to 1800 m/z on the 37 defined windows (3.05 sec/cycle). Collision energy was set to optimum energy for a 2+ ion at the center of each SWATH block with a 15 eV collision energy spread. The mass spectrometer was always operated in high sensitivity mode.

4.2.4 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 (**NCBInr**; 70353186 proteins searched) protein sequence database with the following parameters: trypsin specificity, cys-alkylation (IAM), no taxonomy restricted, and the search effort set to through. The same analysis was also made for *Sus scrofa* taxonomy. 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 ≥ 1.3) and a false discovery rate (**FDR**) lower than 1%. It was not set up to a minimum number of peptides for the identification. However, the proteins identified from a single peptide matched were validated manually (see **Supplementary Data 1**). For this, the MS/MS spectra of the assigned peptides were extracted by Protein-Pilot, and only the proteins that fulfilled the following criteria were validated: (1) peptide mass tolerance lower than 10 ppm, (2) 99% of confidence level in peptide identification, and (3) complete b/y ions series found in the MS/MS spectrum. The identified proteins were quantified using PeakView® from normalized label-free quantification (**LFQ**) intensity data.

4.2.5 Gene ontology and bioinformatics analysis

Bioinformatics analysis of identified and validated SP-proteins was manually performed using the comprehensive bioinformatics tool for functional annotation UniProt KB database (www.uniprot.org) in combination with Protein Analysis Through Evolutionary Relationships (**PANTHER**, www.pantherdb.org). Both databases downloaded on 22/03/2016, containing 34,272 and 21,483 entries in UniProt KB and PANTHER, respectively. This analysis allowed elucidation of the different functions and processes in which the identified and validated proteins would be putatively involved. Three independent sets of ontology were used in the annotation: "the molecular function", "the biological processes", in which the proteins participate, and their "cellular component". Proteins without similarity to database entries were not considered for collation.

4.2.6 Statistical analysis

The quantitative data obtained by PeakView® were analyzed using MarkerView® (v1.2, AB SCIEX). First, areas were normalized by total areas summa. Principal Component Analysis (**PCA**) was performed to evaluate the discriminative ability of proteins in ejaculate portions. Multiexperiment Viewer (**MeV**) (http://www.tm4.org/mev.html) was used to identify the protein differentially expressed among ejaculate portions using Student's t-test subjected to Bonferroni correction. Proteins were considered significantly differentially expressed among ejaculate portions with an adjusted P-value < 0.05. The mean quantity of proteins in each ejaculate portion and the fold-change (**FC**) (ratio between the mean of ejaculate portions) were calculated. The results of Hierarchical clustering analysis of proteome profile of different samples were shown by means of heat map after z-score normalization, using Euclidean distances.

4.3 Results

4.3.1 Characterization of boar seminal plasma proteome

The complete spectral library included 33,557 spectral corresponding to 8,189 distinct peptides and 536 SP-proteins with a FDR ≤ 1% is shown in **Supplementary Table 1**. A total of 31,237 spectra corresponding to 7,840 distinct peptides and 409 SP-proteins were identified in *Sus scrofa* taxonomy from the nrNCBI databank with a FDR ≤ 1%, and 374

of them were validated with ≥ 99% Confidence (Unused Score ≥ 1.3). A total of 359 proteins were identified from in-gel digestion and 70 from in-solution digestion processing. Of the latter, just 17 proteins were not identified in the in-gel approach. The complete list of the 374 well-identified boar SP-proteins, their gi Accession Number, Protein Name, Gene Name, Species, Matched Peptides, Unused Score, % of Sequence Coverage (% Cov), protein previously reported indication, Detection Method of each protein, and the band sec to which they belong (if applicable) are provided in **Supplementary Table 2**. The complete list with the Peptide sequence, Confidence and Contribution from all matched peptides for the 374 proteins identified is shown in **Supplementary Table 3**.

4.3.2 Bioinformatics Analysis

Of the 374 proteins identified and validated in *Sus scrofa* taxonomy, a total of 195 proteins were mapped to 201 IDs in UniProt KB and 115 proteins to 119 IDs in PANTHER for protein enrichment. The results are shown in **Fig. 5**. For molecular function (Fig. 5a), a total of 255 hits were found. Most proteins (117, 46%) were engaged in catalytic activity, while 81 (32%) and 32 (12%) were linked to binding- and regulatory activity, respectively. Only two proteins were classified as antioxidants. Of the proteins identified with catalytic activity, 78 belonged to the hydrolases. Most of those with binding activity showed ion binding activity (49). Regarding biological process (Fig. 5b), 126 out of 601 identified hits (21%) were related to metabolic processes and 101 (17%) related to cellular processes. Noticeably, 29 proteins (5%) would be implicated in immune system processes and solely 20 (3%) were specifically implicated in reproductive processes. Finally, a total of 400 hits belonging to cellular components (Fig. 5c). Most of them were related to intracellular processes, predominantly from cell organelles (95, 24%) and cell part (99, 25%). Other proteins belong to extracellular region (104, 26%) and extracellular matrix (13, 3%). Sixty-five proteins (16%) originated from the plasma membrane.

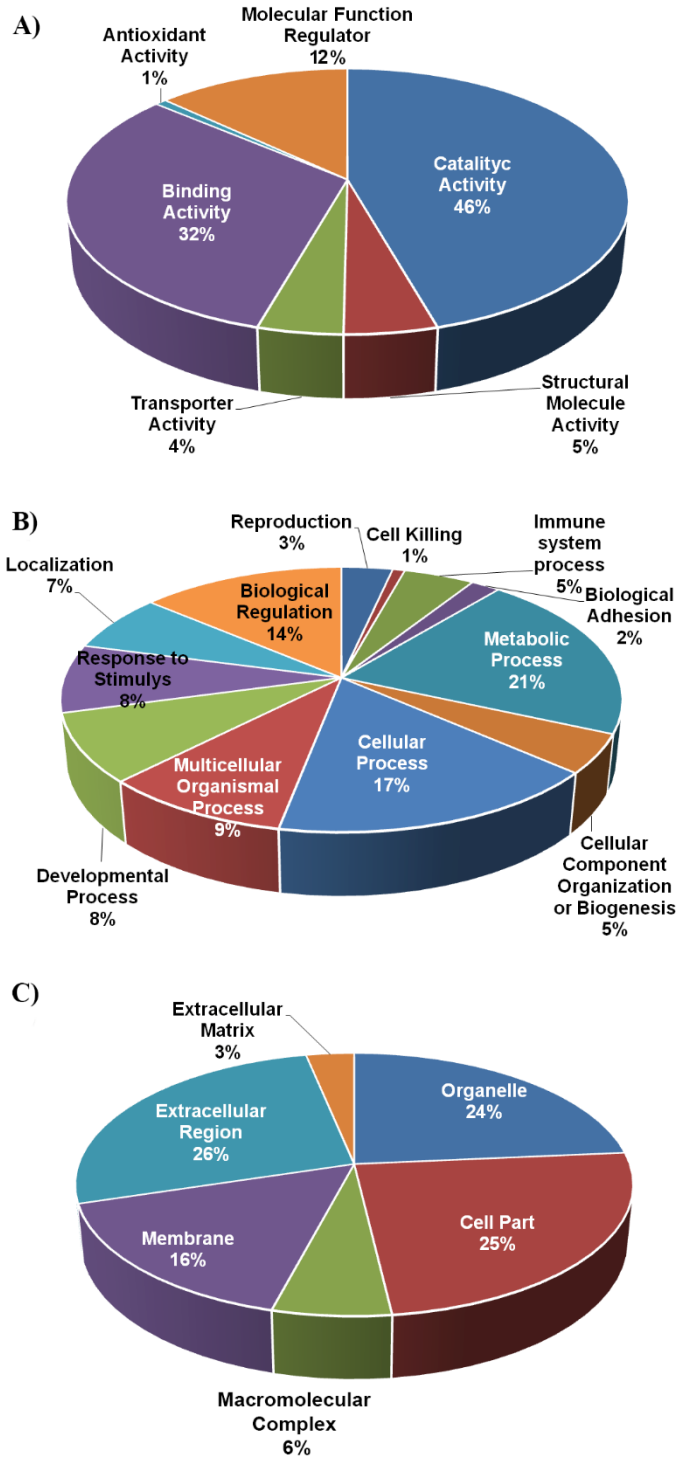


Fig. 5 Pie charts representing the distribution of the proteins identified in boar seminal plasma according to (A) molecular function; (B) biological process and (C) cellular component, using UniProt KB database (www.uniprot.org) in combination with PANTHER (www.pantherdb.org).

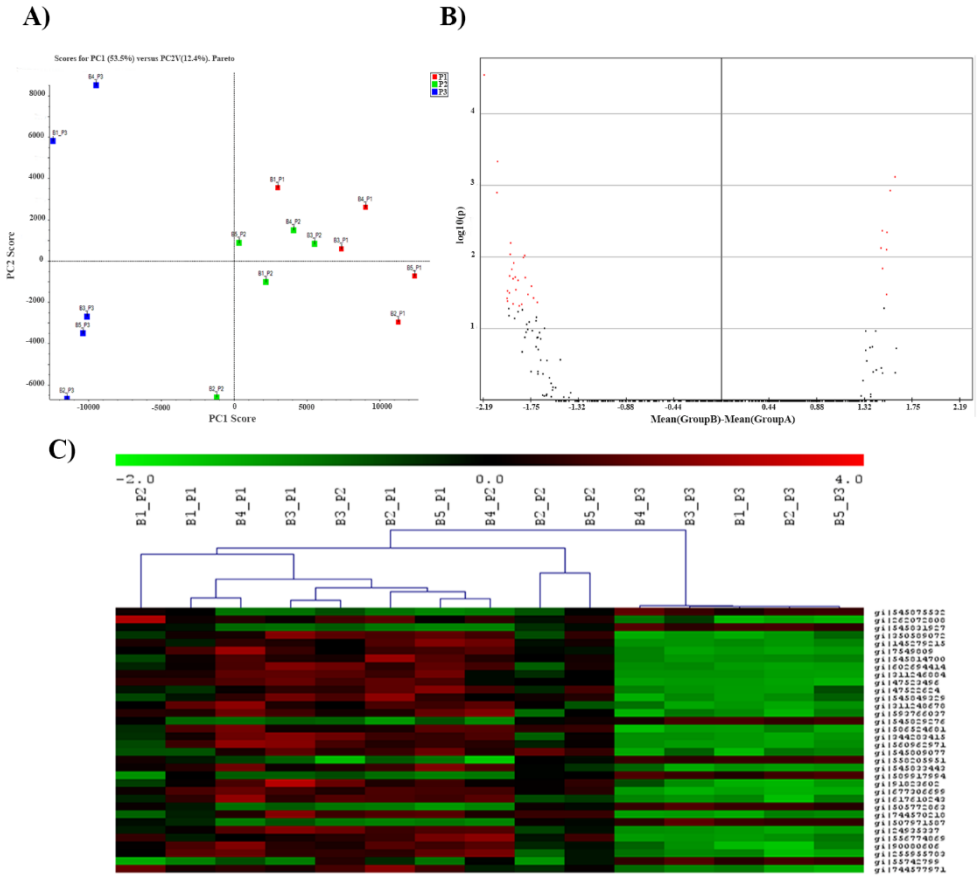


Fig. 6 Bioinformatics analysis of proteins identified and quantified using label-free LC-MS/MS analysis in 15 boar seminal plasma samples of three different ejaculate portions: (P1: first 10 mL of sperm rich ejaculate fraction (SRF), P2: rest of SRF, and P3: post-SRF) from 5 boars. (A) PCA showing the separation between P1-P2 samples and P3 samples in each of the five boars, based on relative protein amount. (B) Volcano plot showing differences in protein expression between P1-P2 and P3 according to magnitude and p-value (t-test). 34 proteins were differentially ($P < 0.05$) expressed between P1-P2 and P3 (indicated as red points). (C) Heat map, showing levels of differentially expressed proteins among ejaculate portions, and hierarchical clustering, showing two main clusters comprising P1-P2 samples and P3 samples, respectively.

4.3.3 Comparison of seminal plasma proteome profiles among ejaculate portions

A total of 447 proteins were identified and all of them were identified and quantified in the three-ejaculate portions. Of them, 223 belong to *Sus scrofa* taxonomy. The full list of these proteins, including relative amount of each protein in each one of the three-ejaculate portions, is shown in **Supplementary Table 4**. PCA was used to determine whether the expression of SP-proteins would differ among ejaculate portions. The biplot, displaying the first and the second components (accounting for 53.5% and 12.4% of variance), did not show a clear separation based in protein expression between P1 and P2 in the five

boars, but P3 samples were clearly separated from the P1-P2 in component 1 of the PCA (Fig. 6a). The most discriminating protein spots between P1-P2 and P3 were identified in Volcano plot. A total of 34 proteins were differentially (P <0.05) expressed between P1-P2 and P3 samples (Fig. 6b). Among them, 16 were identified in *Sus scrofa* taxonomy.

Table 1 List of significant differentially expressed proteins in boar seminal plasma proteins belong the sperm rich ejaculate fraction (SRF) and post-SRF in *Sus scrofa* taxonomy.

gi|Accesion number	Protein Name	Gene Id	FC (log2)*	P value (T-Test)
Over-expressed proteins in SRF				
gi|55742799	Corticosteroid-binding globulin	CBG	0.462	0.008
gi|262072808	Hexosaminidase B	HEXB	0.649	0.044
gi|545814700	Pancreatic secretory granule membrane major glycoprotein GP2	GP2	2.136	0.036
gi|311246884	Epididymal-specific lipocalin-5	LCN5	0.632	0.009
gi|47522624	Arylsulfatase A precursor	ARSA	1.993	0.042
gi|545849329	Galactosidase, beta 1-like 3	GLB1L3	0.826	0.032
gi|311248678	Choline transporter-like protein 2	CTL2	0.242	0.028
gi|545829276	Golgi apparatus protein 1	GLG1	1.025	0.007
Down-expressed proteins in SRF				
gi|545833443	Alpha-enolase	ENO1	-0.057	0.009
gi|24935337	Alkaline phosphatase	ALP	-0.697	0.029
gi|545875532	Fibronectin	FN1	-0.062	0.033
gi|545831927	Nucleobindin-1	NUCB1	-1.226	0.001
gi|350589072	Sulfhydryl oxidase 1	QSOX1	-0.309	0.014
gi|145279215	Angiotensin-converting enzyme isoform 2	ACE	-0.935	0.019
gi|47523496	Epididymal secretory protein E1	NPC2	-0.658	0.001
gi|545809077	Deoxyribonuclease-2-alpha	DNASE2	-0.913	0.024

*FC: Fold change

SRF includes two ejaculate portions: first 10 ml of SRF (P1 in text) and rest of SRF (P2 in text).

Post-SRF is P3 in text.

Of the 34 differentially expressed proteins, 16 proteins were over-expressed and 18 proteins under-expressed in P1-P2 regarding to P3 (**Tables 1 and 2**). The hierarchical

clustering of SP-samples, based on the expression of these 34 differentially expressed proteins, separated the fifteen SP-samples (5 of each ejaculate portion) into two different main clusters, corresponding to P1-P2 samples and to P3 samples, respectively.

Table 2 List of significant differentially expressed proteins in boar seminal plasma proteins belong the sperm rich ejaculate fraction (SRF) and post-SRF without establish taxonomy.

gi|Accesion number	Protein Name	Specie	Gene Id	FC (log2)	P value (T-Test)
Over-expressed proteins in SRF					
gi|677306699	Heat shock cognate 71 kDa protein	Apaloderma vittatum	HSPA8	0.645	< 0.001
gi|507971587	Putative phospholipase B-like 2	Condylura cristata	PLBD2	0.886	0.004
gi|617610243	Guanine nucleotide-binding protein subunit alpha-11	Erinaceus europaeus	GNA11	1.984	0.047
gi|90080606	Unnamed protein product	Macaca fascicularis	PGK1	0.652	0.006
gi|558205951	Polypeptide N-acetylgalactosaminyltransferase	Myotis lucifugus	GALNT2	0.762	0.014
gi|589917994	Fibronectin	Peromyscus maniculatus	FN1	0.449	0.001
gi|593766037	Ezrin	Physeter catodon	EZR	1.873	0.018
gi|505772863	Fibronectin	Sorex araneus	FN1	1.399	0.004
Down-expressed proteins in SRF					
gi|744570218	EGF-like repeat and discoidin I-like domain-containing protein	Camelus dromedarius	EDIL3	-0.582	0.012
gi|744577971	Myelin protein zero-like protein 1	Camelus dromedarius	MPZL1	-1.819	0.009
gi|7549809	Plastin-3 isoform 1	Homo Sapiens	PLS3	-0.834	0.018
gi|91823602	Ectonucleotide pyrophosphatase/phosphodiester	Homo Sapiens	ENPP2	-1.266	0.044
gi|602694414	Alkaline phosphatase	Lipotes vexillifer	ALPL	-1.112	0.031
gi|344283415	Alkaline phosphatase	Loxodonta africana	ALPL	-0.964	0.040
gi|556774869	Beta-galactosidase-1-like protein 2-like	Pantholops hodgsonii	GLB1L2	-0.002	0.020
gi|255955783	Pc21g16370	Penicillium rubens Wisconsin	Pc21g16370	-1.155	0.019
gi|586524681	Syntaxin-binding protein 2	Pteropus alecto	STXBP2	-2.021	< 0.001
gi|560962971	Prominin-2	Vicugna pacos	PROM2	-0.565	0.037

*FC: Fold change

SRF includes two ejaculate portions: first 10 ml of SRF (P1 in text) and rest of SRF (P2 in text).

Post-SRF is P3 in text.

4.4 Discussion

To the best of our knowledge this study generates the largest library of SP-proteins reported to date in boars as more than 250 proteins were first time identified in boar SP. One of the most important contributions of this study would be the identification of a large number of proteins present in small quantities in the boar SP. Probably the approach followed with SEC, 1D SDS-PAGE and post digestion processing in both gel and solution, was decisive for the emergence of these less abundant proteins, which are usually hidden by the most abundant proteins. Three hundred fifty-nine of the 374 identified proteins were spotted in gel analysis due to the chromatographic fractionation and the in gel depletion of the more abundant SP-proteins. In solution digestion of the SP allowed the identification of other 70 proteins, 17 of them were not identified in the gel analysis. These results show that both approaches are complementary for a more deep description of boar SP-proteome. The correct identification of these less abundant proteins could be essential to understand the reproductive response of ejaculates spermatozoa.

A full decrypting of the boar SP-proteome is challenging since most total protein mass is taken up by a few number of low molecular weight proteins, the so-called most abundant proteins. Previous proteomic studies in boar SP using gel electrophoresis allowed identifying simultaneously a limited number of proteins (Druart et al., 2013; Gonzalez-Cadavid et al., 2014). It is well known that the gel-based technique present a limited resolution and most abundant proteins typically mask the identification of less abundant ones (Rabilloud and Lelong, 2011). To overcome these limitations and to identify the largest possible number of SP-proteins, a pre-fractionation step by SEC is needed to fractionate proteins according to molecular weight and thereby improving accuracy of later LC-MS/MS analysis, as SEC is a chromatographic method allowing separating molecules in solution by their size, and in some cases by their molecular weight. The chromatographic fractionation of our SP-samples does not allow achieving SEC fractions without the presence of low molecular weight proteins (most abundant proteins). Thereby, a subsequent analysis by 1D SDS-PAGE proved necessary to achieve an adequate SP fractionation for spectral library construction.

Regarding bioinformatics analysis, it is noticeable that only 20 of the 374 proteins identified in boar SP are -until date- recorded in GO as directly associated with reproductive processes. This is absolutely surprising for a heterogeneous fluid as SP, a secretion composite closely related to spermatozoa from their origin to their transport along the female genitalia. It is also surprising that many of the hereby-identified proteins in boar SP have not yet been attributed a specific role in the GO. Despite of this, we are aware that many the identified SP-proteins, included or not in GO, are indirectly involved in reproductive processes, such as those regulating immune responses by the female genital tract. Certainly, some SP-proteins such as complement factor H, plasma protease C1 inhibitor, cathepsin H, dipeptidyl peptidase 4 and transforming growth factor-ß2 regulate immune responses in the female genital tract (Rodriguez-Martinez et al., 2010; Sakaue et al., 2010), facilitating sperm transport and fertilization. Many of the identified SP-proteins show catalytic activity, specifically hydrolase activity, and some of them would also indirectly be related with reproductive functions, as occurs with acrosin. It is known that acrosin is involved in the proacrosin-acrosin system necessary for boar sperm capacitation (Puigmule et al., 2011). Cathepsins (pro-cathepsin H, cathepsin L1, cathepsin H, cathepsin Z and cathepsin D) are also SP-proteins with hydrolase activity, specifically protease activity, some secreted in the epididymis (Okamura et al., 1995; Moura et al., 2006). This origin suggests that they may rule remodeling sperm membranes during epididymal sperm transit. Moreover, SP-cathepsins could also interact with sperm membrane proteins affecting sperm fertilizing ability, as previously suggested in bull by Moura et al. (2010). A large number of the boar SP-proteins identified in the present study showed binding activity, which would also show important reproductive functions because they are able to bind to the sperm plasma membranes modulating sperm functionality (Jelinkova et al., 2003). Among them, the most abundant in boar SP are spermadhesins, including AWN-1, AWN-3, AQN, PSP-I and PSP-II, with proved implication in the fertilizing ability of boar spermatozoa (Caballero et al., 2008). Heat shock 70 kDa protein and lysosomal type alpha-D-mannosidase are another two SP-proteins encoded in GO as binding activity that have proven impact on sperm functionality (Jin et al., 1999; Spinaci et al., 2005). Boar SP includes a large number of glycoproteins, which also includes the aforementioned spermadhesins. Their facility for glycosylation gives them an important role in sperm functionality (Calvete et al., 1994), particularly in regulating sperm motility, capacitation, acrosome reaction and sperm-oocyte binding (Pang et al., 2011). Some identified SP-proteins show ion-binding

properties, being particularly relevant for sperm performance those showing zinc- and calcium-binding properties. Zinc-binding proteins play an important role to preserve sperm chromatin stability and to regulate sperm motility (Strzezek and Hopfer, 1987). Moreover, they have antibacterial and antioxidant functions and participate in the regulatory immune response in the female reproductive tract (Vivacqua et al., 2004; Edström et al., 2008). Some calcium-binding proteins have also been identified in boar SP. Calcium, especially extracellular calcium, is necessary for sperm exocytosis processes such as the acrosome reaction (Gadella and Harrison, 2000). Therefore, the presence of calcium-binding proteins in boar SP could limit the availability of free calcium by sperm, preventing premature sperm capacitation. Just a few number of boar SP-proteins are encoded in GO as antioxidants. However, identified SP-proteins have proven antioxidant activity. For instance, albumin, a common SP-protein encoded as binding activity, which has proved capacity in protecting sperm against oxidative stress caused by lipid peroxidation (Gonzalez-Cadavid et al., 2014). Likewise, some identified SP-proteins encoded in GO as metabolic function show oxidoreductase activity, for instance epididymal secretory glutathione peroxidase and superoxide dismutase. Boar spermatozoa are particularly vulnerable to oxidative stress because their plasma membrane is particularly rich in polyunsaturated fatty acids and the above-mentioned SP-enzymes have proved effectiveness protecting boar sperm against oxidative stress (Bathgate, 2011). Regarding cellular location, it is also noticeable that a large proportion of SP-proteins had an intracellular localization. One possible explanation would be they come from SP-exosomes, whose content is rich in proteins (Siciliano et al., 2008). Fifty-three proteins identified in boar SP have been encoded in GO as originated from exosomes. Prevailing data demonstrate that proteins in the epididymal fluid associated with post-testicular sperm maturation are transferred to the sperm by exosomes and consequently the protein content of these extracellular vesicles have been shown to be involved in post-testicular sperm maturation, including sperm motility acquisition and reduction of oxidative stress and acrosome reaction (Siciliano et al., 2008; Machtinger et al., 2016).

In addition to the different proteomic analytical approach used, another reason for discrepancy among studies about the number of SP-proteins identified in boar ejaculates could be related to the source of SP. The boar SP samples evaluated by Druart et al. (2013) and Gonzalez-Cadavid et al. (2014) seem to come from SRF as they were collected using

glove-hand method, while SP-samples evaluated in the present study come from the entire ejaculate, including SRF and post-SRF. In this context, it is well known that the total protein composition differs among boar ejaculate fractions (Rodriguez-Martinez et al., 2009). As in human, the boar ejaculate is expelled in fractions, the SRF and the post-SRF being the two main fractions, with relevant differences in SP-composition (Rodriguez-Martinez et al., 2009) including cytokine-contents (Barranco et al., 2015a) and antioxidant enzymes (Barranco et al., 2015b). Further research is required to verify possible divergences in the SP-proteome among ejaculate fractions/portions, particularly in the light of the ongoing discussions as to whether to use only the SRF or the entire ejaculate for both AI and sperm technologies, such as cryopreservation and sex-sorting (Roca et al., 2015a). Noticeable all expressed SP-proteins were present in each of the three ejaculate portions, which would indicate that the possible influence of SP-proteins on boar sperm performance, both AI-fertility and aid overcome sperm technologies, would be more linked to quantitative rather than qualitative differences. In this regard, it was also noticeable that there were no statistically quantitative differences between P1 and P2 in any of the 447 expressed SP-proteins and that only 34 of them showed statistically quantitative differences between P1-P2 and P3. The protein quantitative data were subjected to Bonferroni correction before statistical analysis. The Bonferroni sequential correction is considered conservative as it has the effect of "cleaning" data from false positives resulting from the random nature of sampling process. Cleaning data before statistical analysis deserves special consideration when analyzing biological compounds undergoing to large natural quantitative variations among samples, as in the case of SP-proteins. Consequently, the quantitative differences on SP-proteins among ejaculate portions showed in the present study should be considered statistically robust, but the conservative statistic procedure developed might overlook some other expressed proteins that could also show quantitative differences between ejaculate portions, although to a lesser degree and always open to misinterpretation. The absence of quantitative differences in SP-protein expression between P1 and P2 may be understandable, as P1 and P2 together comprise the first naturally occurring ejaculate fraction (SRF), whose SP has a common origin, albeit having amount differences (Rodriguez-Martinez et al., 2011).

Looking at the 34 differentially expressed SP-proteins between P1-P2 and P3, 16 of them were identified in *Sus scrofa* taxonomy and eight of them were more expressed in P1-P2

than P3. Six of these eight proteins were previously identified in SP of livestock species and all of them have a testicular (pancreatic secretory granule membrane major glycoprotein [GP2], galactosidase beta 1-like 3 [GLB1L3] and golgi apparatus protein 1 [GLG1]) or epididymal (hexosaminidase B [HEXB], epididymal-specific lipocalin-5 [LCN5] and arylsulfatase A [ARSA]) origin. Many of these proteins have been related to some specific sperm functionalities, as membrane stability and permeability (GLB1L3, [Dacheux et al., 1989]; ARSA, [Gadella et al., 1993]), capacitation (HEXB, [Wysocki et al., 2015]), acrosome reaction (GP2, [Guyonnet et al., 2009]), fertilizing ability (LCN-5, [Watanabe et al., 2014]) or sperm-zone pellucid binding (ARSA, [Carmona et al., 2002]).

The other eight SP-proteins identified in *Sus scrofa* taxonomy were less expressed in P1-P2 than in P3. Four of them (fibronectin [FN1], sulfhydryl oxidase 1, angiotensin-converting enzyme and epididymal secretory protein E1) have also a testicular or epididymal origin and all of them have been related to sperm maturation. Other of these SP-proteins more expressed in P3, as alpha-enolase or alkaline phosphatase (ALP), have been related to sperm motility (Nakamura et al., 2013; Zhang et al., 2015). Noticeable, deoxyribonuclease -2-alpha, an acid endonuclease secreted by male accessory glands, is involved in the degradation of exogenous DNA (Carballada and Esponda, 2001) and it also provides a bactericide activity protecting sperm in the transit along female genital tract (Cheng et al., 2009). Nucleobinding-1 belongs to a family of proteins with calcium and DNA binding properties (Gonzalez et al., 2012), but its reproductive function is still unknown. In addition, some of the above related proteins, as HEXB (Wysocki et al., 2015), FN1 (Vilagran et al., 2015) or AP (Bucci et al., 2016), have been linked to the ability of sperm to overcome cryopreservation.

Eighteen of the 34 proteins differently expressed between P1-P2 and P3 were not identified in *Sus scrofa* taxonomy but some of them have been related to spermatogenesis, sperm maturation or mature sperm functionality in other mammalian species. For instance, myelin protein zero-like protein 1 (Nakata et al., 2012), guanine nucleotide-binding protein subunit alpha-11 (Hu et al., 2008), plastin-3 isoform 1 (Li et al., 2015) would be related to spermatogenesis; and polypeptide N-acetylgalactosaminyltransferase 2 and beta-galactosidase-1-like protein 2-like with sperm maturation along epididymal transit (Hall et al., 1987). Another five of these 18 SP-proteins (EGF-like repeat and discoidin I-like domain-containing protein 3, putative phospholipase B-like 2 isoform 1, prominin-2, ezrin and syntaxin-binding protein 2) have been related with sperm

capacitation (Furuya et al., 1993; Douard et al., 2004; Florek et al., 2007; Salvolini et al., 2013; Tomes et al., 2002). Heat shock cognate 71 kDa is a protein related to environmental or physiological stresses (Lowe and Moran, 1984) synthesized by sperm cells (Dun et al., 2012) that has been recently related to boar sperm cryotolerance (Vilagran et al., 2014). Finally, ectonucleotide pyrophosphatase/phosphodiesterase family member is an ATP-hydrolyzing enzyme described in epididymis suggesting that this protein could be involved in the modulation of the epididymal environment (Balleannee et al., 2011).

Based on the foregoing, it is clear that most of the SP-proteins differentially expressed among the ejaculate portions show a clear connection with the development and/or functionality of spermatozoa. However, it does not seem clear that an increased expression of a particular functional SP-protein was linked to a better sperm performance. For instance, some SP-proteins, as FN1, have been positively related to sperm cryoresistence, as sperm from boars with high FN1 in SP showed best freezability (Vilagran et al., 2015). This would indicate that sperm belonging to the ejaculate portion with more FN1 expression (more in P3 than P1+P2) should cryopreserve better. This does not seem be the case, as sperm from the entire ejaculate (P1+P2+P3) cryopreserve worse than those from SRF (P1+P2) (Alkmin et al., 2014a). This example could also be applied to other differentially expressed SP-proteins as AP, which has been related to sperm motility. That a SP-protein carries out its maximal function when expressed within a range, declining it when under- or over-expressed, could explain these apparent inconsistencies. All together, the presented data clearly demonstrate the current limited knowledge of the number and role of boar SP-proteins, which, in turn, provides exciting research opportunities. Clarifying the role of SP-proteins on reproductive processes, including the ability of sperm to successfully overcome sperm technologies, will possibly allow the identification of SP-protein biomarkers but also to promote the design of semen extenders able to maintain sperm fertilizing capacity for longer.

4.5 Conclusion

In summary, this experimental study provided an extensive proteomic analysis of boar SP, which still remained poorly understood, and generated a public accessible database of boar SP-proteome. The boar SP-proteome is described so that a spectral library can be

built for relative 'label free' protein quantification with SWATH approach. The analytical approach used, including SEC, 1D SDS-PAGE and post digestion processing in both gel and solution, has proved to be crucial for the identification of the large number of proteins present in small quantities, in the boar SP. Additionally, the comparative SP-proteome study of the boar ejaculate portions resulted in the identification of a number of differentially expressed proteins, many of which are linked to sperm reproductive performance. Overall, the results of the present study intend to provide an elementary core for further studies focused on understanding the role of SP-proteins in reproductive outcomes of boar semen as well as for identification of biomarkers for sperm quality and fertility.

Supplementary material

All supplementary material is available in the website:

https://cristinaperez61.wixsite.com/phdthesis

Acknowledgements

This experimental study was supported by MINECO (AGL2012-39903) Madrid (Spain), FEDER funds (EU), 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.

References

Alkmin DV, Perez-Patiño C, Barranco I, Parrilla I, Vazquez JM, Martinez EA, Rodriguez-Martinez H, Roca J. Boar sperm cryosurvival is better after exposure to seminal plasma from selected fractions than to those from entire ejaculate. Cryobiology. 2014a, 69:203-10.

Alkmin DV, Parrilla I, Tarantini T, Del Olmo D, Vazquez JM, Martinez EA, Roca J. Seminal plasma affects sperm sex sorting in boars. Reprod Fertil Dev. 2014b, doi: 10.1071/RD14088.

Barranco I, Rubér M, Perez-Patiño C, Atikuzzaman M, Martinez EA, Roca J, Rodriguez-Martinez H. The Seminal Plasma of the Boar is Rich in Cytokines, with Significant Individual and Intra-Ejaculate Variation. Am J Reprod Immunol. 2015a, doi: 10.1111/aji.12432.

Barranco I, Tvarijonaviciute A, Perez-Patiño C, Parrilla I, Ceron JJ, Martinez EA, Rodriguez-Martinez H, Roca J. High total antioxidant capacity of the porcine seminal plasma (SP-TAC) relates to sperm survival and fertility. Sci Rep. 2015b, doi: 10.1038/srep18538.

Bathgate R. Antioxidant mechanism and their benefit on post-thaw boar sperm quality. Reprod Domest Anim. 2011, 46:23-5.

Belleannée C, Labas V, Teixeira-Gomes AP, Gatti JL, Dacheux JL, Dacheux F. Identification of luminal and secreted proteins in bull epididymis. J Proteomics. 2011, 74:59-78.

Broekhuijse ML, Šoštarić E, Feitsma H, Gadella BM. Relationship of flow cytometric sperm integrity assessments with boar fertility performance under optimized field conditions. J Anim Sci. 2012, 90:4327-36.

Bucci D, Giaretta E, Spinaci M, Rizzato G, et al. Characterization of alkaline phosphatase activity in seminal plasma and in fresh and frozen-thawed stallion spermatozoa. Theriogenology. 2016, 85,:288-295.

Caballero I, Vazquez JM, García EM, Parrilla I, Roca J, Calvete JJ, Sanz L, Martínez EA. Major proteins of boar seminal plasma as a tool for biotechnological preservation of spermatozoa. Theriogenology. 2008, 70:1352-5.

Caballero I, Vazquez JM, Mayor GM, Almiñana C, Calvete JJ, Sanz L, Roca J, Martinez EA. PSP-I/PSP-II spermadhesin exert a decapacitation effect on highly extended boar spermatozoa. Int J Androl. 2009, 32:505-13.

Calvete JJ, Solís D, Sanz L, Díaz-Mauriño T, Töpfer-Petersen E. Glycosylated boar spermadhesin AWN-1 isoforms. Biological origin, structural characterization by lectin mapping, localization of O-glycosylation sites, and effect of glycosylation on ligand binding. Biol Chem Hoppe Seyler. 1994, 375:667-73.

Carballada R, Esponda P. Regulation of foreign DNA uptake by mouse spermatozoa. Exp Cell Res. 2001, 262:104-13.

Carmona E, Weerachatyanukul W, Soboloff T, Fluharty AL, White D, Promdee L, Ekker M, Berger T, Buhr M, Tanphaichitr N. Arylsulfatase a is present on the pig sperm surface and is involved in sperm-zona pellucida binding. Dev Biol. 2002, 247:182-96.

Cheng GZ, Li JY, Li F, Wang HY, Shi GX. Human ribonuclease 9, a member of ribonuclease A superfamily, specifically expressed in epididymis, is a novel sperm-binding protein. Asian J Androl. 2009, 11:240-51.

Dacheux JL, Dacheux F, Paquignon M. Changes in sperm surface membrane and luminal protein fluid content during epididymal transit in the boar. Biol Reprod. 1989, 40:635-51.

Douard V, Gassama-Diagne A, Hermier D, Blesbois E. Activity of phospholipases A and lysophospholipase in turkey semen and oviducal fluid. Poult Sci. 2004, 83:1385-93.

Druart X, Rickard JP, Mactier S, Kohnke PL, Kershaw-Young CM, Bathgate R, Gibb Z, Crossett B, Tsikis G, Labas V, Harichaux G, Grupen CG, de Graaf SP. Proteomic characterization and cross species comparison of mammalian seminal plasma. J Proteomics. 2013, 91:13-22.

Dun MD, Aitken RJ, Nixon B. The role of molecular chaperones in spermatogenesis and the post-testicular maturation of mammalian spermatozoa. Hum Reprod Update. 2012, 18:420-35.

Edström AM, Malm J, Frohm B, Martellini JA, Giwercman A, Mörgelin M, Cole AM, Sørensen OE. The major bactericidal activity of human seminal plasma is zinc-dependent and derived from fragmentation of the semenogelins. J Immunol. 2008, 181:3413-21.

Florek M, Bauer N, Janich P, Wilsch-Braeuninger M, Fargeas CA, Marzesco AM, Ehninger G, Thiele C, Huttner WB, Corbeil D. Prominin-2 is a cholesterol-binding protein associated with apical and basolateral plasmalemmal protrusions in polarized epithelial cells and released into urine. Cell Tissue Res. 2007, 328:31-47.

Furuya S, Endo Y, Oba M, Ito T, Suzuki S, Nozawa S. Effect of epidermal growth factor (EGF) on mouse sperm capacitation. Nihon Sanka Fujinka Gakkai Zasshi. 1993, 45:1313-9.

Gadella BM, Colenbrander B, Van Golde LM, Lopes-Cardozo M. Boar seminal vesicles secrete arylsulfatases into seminal plasma: evidence that desulfation of seminolipid occurs only after ejaculation. Biol Reprod. 1993, 48:483-9.

Gadella BM, Harrison RA. The capacitating agent bicarbonate induces protein kinase A-dependent changes in phospholipid transbilayer behavior in the sperm plasma membrane. Development. 2000, 127:2407-20.

Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics. 2012, 11:O111.016717.

Gonzalez R, Mohan H, Unniappan S. Nucleobindins: bioactive precursor proteins encoding putative endocrine factors? Gen Comp Endocrinol. 2012, 176:341-6.

Gonzalez-Cadavid V, Martins JA, Moreno FB, Andrade TS, Santos AC, Monteiro-Moreira AC, Moreira RA, Moura AA. Seminal plasma proteins of adult boars and correlations with sperm parameters. Theriogenology. 2014, 82:697-707.

Guyonnet B, Marot G, Dacheux JL, Mercat MJ, Schwob S, Jaffrézic F, Gatti JL. The adult boar testicular and epididymal transcriptomes. BMC Genomics. 2009, 10:369.

Hall JC, Killian GJ. Changes in rat sperm membrane glycosidase activities and carbohydrate and protein contents associated with epididymal transit. Biol Reprod. 1987, 36:709-18.

Hu Y, Xing J, Chen L, Guo X, Du Y, Zhao C, Zhu Y, Lin M, Zhou Z, Sha J. RGS22, a novel testis-specific regulator of G-protein signaling involved in human and mouse spermiogenesis along with GNA12/13 subunits. Biol Reprod. 2008, 79:1021-9.

Jelínková P, Manásková P, Tichá M, Jonáková V. Proteinase inhibitors in aggregated forms of boar seminal plasma proteins. Int J Biol Macromol. 2003, 32:99-107.

Jin YZ, Dacheux F, Dacheux JL, Bannai S, Sugita Y, Okamura N. Purification and properties of major alpha-D-mannosidase in the luminal fluid of porcine epididymis. Biochim Biophys Acta. 1999, 1432:382-92.

Jonáková V, Manásková P, Tichá M. Separation, characterization and identification of boar seminal plasma proteins. J Chromatogr B Analyt Technol Biomed Life Sci. 2007, 849:307-14.

Labas V, Grasseau I, Cahier K, Gargaros A, Harichaux G, Teixeira-Gomes AP, Alves S, Bourin M, Gérard N, Blesbois E. Data for chicken semen proteome and label free quantitative analyses displaying sperm quality biomarkers. Data Brief. 2014, 1:37-41.

Li N, Mruk DD, Wong CK, Lee WM, Han D, Cheng CY. Actin-bundling protein plastin 3 is a regulator of ectoplasmic specialization dynamics during spermatogenesis in the rat testis. FASEB J. 2015, 29:3788-805.

Lowe DG, Moran LA. Proteins related to the mouse L-cell major heat shock protein are synthesized in the absence of heat shock gene expression. Proc Natl Acad Sci U S A. 1984, 81:2317-21.

Luna C, Yeste M, Rivera Del Alamo MM, Domingo J, Casao A, Rodriguez-Gil JE, Pérez-Pé R, Cebrián-Pérez JA, Muiño-Blanco T. Effect of seminal plasma proteins on the motile sperm subpopulations in ram ejaculates. Reprod Fertil Dev. 2015, doi: 10.1071/RD15231.

Machtinger R, Lauren LC, Baccarelli AA. Extracellular vesicles: roles in gamete maturation, fertilization and embryo implantation. Human Reprod Update. 2016, 2:182-93.

Moura AA, Chapman DA, Koc H, Killian GJ. Proteins of the cauda epididymal fluid associated with fertility of mature dairy bulls. J Androl. 2006, 27:534-41.

Moura AA, Souza CE, Stanley BA, Chapman DA, Killian GJ. Proteomics of cauda epididymal fluid from mature Holstein bulls. J Proteomics. 2010, 73:2006-20.

Nakamura N, Dai Q, Williams J, Goulding EH, Willis WD, Brown PR, Eddy EM. Disruption of a spermatogenic cell-specific mouse enolase 4 (eno4) gene causes sperm structural defects and male infertility. Biol Reprod. 2013, 88:90.

Nakata H, Wakayama T, Adthapanyawanich K, Nishiuchi T, Murakami Y, Takai Y, Iseki S. Compensatory upregulation of myelin protein zero-like 2 expression in spermatogenic cells in cell adhesion molecule-1-deficient mice. Acta Histochem Cytochem. 2012, 45:47-56.

Okamura N, Tamba M, Uchiyama Y, Sugita Y, Dacheux F, Syntin P, Dacheux JL. Direct evidence for the elevated synthesis and secretion of procathepsin L in the distal caput epididymis of boar. Biochim Biophys Acta. 1995, 1245:221-6.

O'Leary S, Jasper MJ, Warnes GM, Armstrong DT, Robertson SA. Seminal plasma regulates endometrial cytokine expression, leukocyte recruitment and embryo development in the pig. Reproduction. 2004, 128:237-47.

Pang PC, Chiu PC, Lee CL, Chang LY, Panico M, Morris HR, Haslam SM, Khoo KH, Clark GF, Yeung WS, Dell A. Human sperm binding is mediated by the sialyl-Lewis(x) oligosaccharide on the zona pellucida. Science. 2011, 333:1761-4.

Park Y-J, Kim J, You Y-A, Pang M-G. Proteomic revolution to improve tools for evaluating male fertility in animals. J Proteome Res. 2013, 12:4738-4747.

Parrilla I, del Olmo D, Sijses L, Martinez-Alborcia MJ, Cuello C, Vazquez JM, Martinez EA, Roca J. Differences in the ability of spermatozoa from individual boar ejaculates to withstand different semen-processing techniques. Anim Reprod Sci. 2012, 132:66-73

Puigmulé M, Fàbrega A, Yeste M, Bonet S, Pinart E. Study of the proacrosin-acrosin system in epididymal, ejaculated and in vitro capacitated boar spermatozoa. Reprod Fertil Dev. 2011, 23:837-45.

Rabilloud T, Lelong C. Two-dimensional gel electrophoresis in proteomics: a tutorial. J Proteomics. 2011, 74:1829-41.

Riesenbeck A. Review on international trade with boar semen. Reprod Dom Anim. 2011, 46:1-3.

Roca J, Parrilla I, Bolarin A, Martinez EA, Rodriguez-Martinez H. Will artificial insemination in pigs become more efficient? Theriogenology. 2015a, S0093-691X(15)00651-2.

Roca J, Broekhuijse ML, Parrilla I, Rodriguez-Martinez H, Martinez EA, Bolarin A. Boar Differences In Artificial Insemination Outcomes: Can They Be Minimized? Reprod Domest Anim. 2015b, 50:48-55.

Rodríguez-Martínez H, Kvist U, Saravia F, Wallgren M, Johannisson A, Sanz L, Peña FJ, Martínez EA, Roca J, Vázquez JM, Calvete JJ. The physiological roles of the boar ejaculate. Soc Reprod Fertil. 2009, 66:1-21.

Rodriguez-Martinez H, Saravia F, Wallgren M, Martinez EA, Sanz L, Roca J, Vazquez JM, Calvete JJ. Spermadhesin PSP-I/PSP-II heterodimer induces migration of polymorphonuclear neutrophils into the uterine cavity of the sow. J Reprod Immunol. 2010, 84:57-65.

Rodríguez-Martínez H, Kvist U, Ernerudh J, Sanz L, Calvete JJ. Seminal plasma proteins: what role do they play? Am J Reprod Immunol. 2011, 66:11-22.

Sakaue T, Takeuchi K, Maeda T, Yamamoto Y, Nishi K, Ohkubo I. Factor H in porcine seminal plasma protects sperm against complement attack in genital tracts. J Biol Chem. 2010, 285:2184-92.

Salvolini E, Buldreghini E, Lucarini G, Vignini A, Lenzi A, Di Primio R, Balercia G. Involvement of sperm plasma membrane and cytoskeletal proteins in human male infertility. Fertil Steril. 2013, 99:697-704.

Saravia F, Wallgren M, Johannisson A, Calvete JJ, Sanz L, Peña FJ, Roca J, Rodríguez-Martínez H. Exposure to the seminal plasma of different portions of the boar ejaculate modulates the survival of spermatozoa cryopreserved in MiniFlatPacks. Theriogenology. 2009, 71:662-75.

Shevchenko A, Jensen ON, Podtelejnikov AV, Sagliocco F, Wilm M, Vorm O, Mortensen P, Shevchenko A, Boucherie H, Mann M. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. Proc Natl Acad Sci. 1996, 93:14440-5.

Siciliano L, Marcianò V, Carpino A. Prostasome-like vesicles stimulate acrosome reaction of pig spermatozoa. Reprod Biol Endocrinol 2008, 6:5.

Soleilhavoup C, Tsikis G, Labas V, Harichaux G, Kohnke PL, Dacheux JL, Guérin Y, Gatti JL, de Graaf SP, Druart X. Ram seminal plasma proteome and its impact on liquid preservation of spermatozoa. J Proteomics. 2014, 109:245-60.

Spinaci M, Volpe S, Bernardini C, De Ambrogi M, Tamanini C, Seren E, Galeati G. Immunolocalization of heat shock protein 70 (Hsp 70) in boar spermatozoa and its role during fertilization. Mol Reprod Dev. 2005, 72:534-41.

Strzezek J and Hopfer E. Zinc ion-dependent protein in boar Semen. I. Egg-yolk precipitating activity and some biochemical properties. Anim Reprod Sci. 1987, 13:117-131.

Tomes CN, Michaut M, De Blas G, Visconti P, Matti U, Mayorga LS. SNARE complex assembly is required for human sperm acrosome reaction. Dev Biol. 2002, 243:326-38.

Töpfer-Petersen E, Ekhlasi-Hundrieser M, Kirchhoff C, Leeb T, Sieme H. The role of stallion seminal proteins in fertilisation. Anim Reprod Sci. 2005, 89:159-70.

Vadnais ML, Roberts KP. Seminal plasma proteins inhibit in vitro- and cooling-induced capacitation in boar spermatozoa. Reprod Fertil Dev. 2010, 22:893-900.

Vilagran I, Yeste M, Sancho S, Casas I, Rivera del Álamo MM, Bonet S. Relationship of sperm small heat-shock protein 10 and voltage-dependent anion channel 2 with semen freezability in boars. Theriogenology. 2014, 82:418-26.

Vilagran I, Yeste M, Sancho S, Castillo J, Oliva R, Bonet S. Comparative analysis of boar seminal plasma proteome from different freezability ejaculates and identification of Fibronectin 1 as sperm freezability marker. Andrology. 2015, 3:345-56.

Vivacqua A, Siciliano L, Sabato M, Palma A, Carpino A. Prostasomes as zinc ligands in human seminal plasma. Int J Androl. 2004, 27:27-31.

Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nature Biotechnol. 2014, 30:223-6.

Wagner A, Ekhlasi-Hundrieser M, Hettel C, Petrunkina A, Waberski D, Nimtz M, Töpfer-Petersen E. Carbohydrate-based interactions of oviductal sperm reservoir formation-studies in the pig. Mol Reprod Dev. 2002, 61:249-57.

Watanabe H, Takeo T, Tojo H, Sakoh K, Berger T, Nakagata N, Mak TW, Kondoh G. Lipocalin 2 binds to membrane phosphatidylethanolamine to induce lipid raft movement in a PKA-dependent manner and modulates sperm maturation. Development. 2014, 141:2157-64.

Wysocki P, Orzołek A, Strzeżek J, Koziorowska-Gilun M, Zasiadczyk Ł, Kordan W. The activity of N-acetyl-β-hexosaminidase in boar seminal plasma is linked with semen quality and its suitability for cryopreservation. Theriogenology. 2015, 83:1194-202.

Zhang HY, Lu JC, Feng RX. Correlations of 24 biochemical markers in seminal plasma with routine semen parameters. Zhonghua Nan Ke Xue. 2015, 21:1087-92.

STUDY 2

This study has been published in **Data in Brief** with the title

"**EXTENSIVE DATASET OF BOAR SEMINAL PLASMA PROTEOME DISPLAYING PUTATIVE REPRODUCTIVE FUNCTIONS OF IDENTIFIED PROTEINS**"

Data in Brief 2016, 8:1370-1373

DOI: 10.1016/j.dib.2016.07.037

Abstract

A complete proteomic profile of seminal plasma (**SP**) remains challenging, particularly in porcine. The present experiment reports the analysis of boar SP-proteins by using a combination of size exclusion chromatography (**SEC**), one dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (**1D SDS-PAGE**) and Nano liquid chromatography-electrospray ionization-tandem mass spectrometry (**LC-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 ≥ 95% (Unused Score > 1.3) and a false discovery rate (**FDR**) ≤ 1%, is provided. In addition, the relative abundance of 432 of them is also showed. Gene ontology annotation of the complete SP-proteome complemented by an extensive review of the putative reproductive role of SP-proteins is also described, 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" published by Perez-Patiño et al. (2016) in the Journal of Proteomics.

Keywords: porcine, ejaculate, seminal plasma, proteome

V. Study 2

5.1 Data

The present study detailed a unique dataset 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. Furthermore, an extensive review 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.

5.2 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®, 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 *xg* 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 size exclusion chromatography (**SEC**), one dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (**1D SDS-PAGE**) followed by Nano liquid chromatography-electrospray ionization-tandem mass spectrometry (**LC-ESI-MS/MS**). The proteomics data and result-files from the analysis have been deposited to the ProteomeXchange Consortium (Vizcaino et al., 2014) via the PRIDE partner repository, with the dataset identifier PXD003579 and DOI: 10.6019/PXD003579. Single SP-pools from each boar were analyzed by LC-sequential window acquisition of all theoretical spectra (**SWATH**)-MS acquisition for determination of protein relative abundance.

5.2.1 Sample preparation

Seminal plasma samples were thawed at room temperature, centrifuged (16,100 *xg*, 4 °C, 1 min) and the protein concentration measured using a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA) according to manufacturer instructions. 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), to quantify the relative abundance of SP-proteins.

A SEC was carried out in an ETTAN LC system (GE Healthcare Life Sciences, Little Chalfont, United Kingdom) using a Superdex 200 5/150 GL column (GE Healthcare Life Sciences) controlled by an AKTA ETTAN LC system (GE Healthcare Life Sciences), which furthermore would ease a large-scale proteomic study. Fifty µL of the SP-pool were injected into the column, equilibrated with 200 mM ammonium bicarbonate and 1 mM dithiothreitol (**DTT**) (Sigma-Aldrich, St. Louis, MO, USA) at a flow rate of 0.18 mL/min at 4 °C. The eluent was collected in 0.2 mL fractions. It was immediately evident that it was not possible to achieve only one fraction with all proteins except the dominant ones. Therefore, 2.5 to 5 µg of protein (depending of the total amount of sample) from the fractions were collected after SEC step, dried in a rotatory evaporator and loaded onto different wells of 12% Tris-HCl precast 1D SDS-PAGE (Bio-Rad, Richmond, CA, USA). The gel was run at a constant voltage of 200 mV for 30 min at room temperature including a molecular weight marker (ECL Plex Fluorescent Rainbow Marker, GE Healthcare Life Sciences), and Coomassie Brilliant R250 Blue stain (Bio-Rad) was used to visualize protein bands on the gel. Thereafter, the gel was sliced at 38 kDa and the top of the gel used to analyze less abundant proteins by in-gel digestion processing. The more abundant proteins were identified from an aliquot of the mixed SP-sample analyzed by in-solution processing.

5.2.2 Proteome analysis

5.2.2.1 Complete proteome. In-solution digestion processing

The more abundant proteins were analyzed using in-solution digestion. Three µL of the pooled SP, containing 10 µg of proteins, were digested with Sequencing Grade Trypsin (Promega Corporation, Madison, USA) to generate peptides of each individual protein. The final concentration of protein in the digested sample was 0.13 µg/µL.

5.2.2.2 Less abundant proteins. In-gel digestion processing

The portion of the 1D SDS-PAGE containing proteins with a molecular weight higher than 38 kDa was processed as a unique sample. Following washing in MilliQ water (Merck Millipore, Darmstadt, Germany), the gel was dehydrated in acetonitrile (**ACN**) (Fisher Scientific, Madrid, Spain), reduced with DTT and alkylated with iodoacetamide

(**IAM**). Each slide was cut and small pieces of approximately 1 mm^2 in size were transferred into 1.5 mL Eppendorf tubes. Sequencing Grade Trypsin digestion of the sliced gel was performed following the protocol used by Shevchenko et al. (1996).

5.2.2.3 LC-MS/MS analysis

The peptides recovered from in-gel and in-solution digestion processing were examined by LC using a NanoLC Ultra 1D plus Eksigent (Eksigent Technologies, Dublin, CA, USA) which was directly connected to an AB SCIEX TripleTOF 5600 mass spectrometer (AB SCIEX, Framingham, MA, USA) in direct injection mode. Briefly, 5 µL from each digested sample were trapped on a NanoLC pre-column (3 µm particles size C18-CL, 350 µm diameter x 0.5 mm long; Eksigent Technologies) and desalted with 0.1% trifluoroacetic acid (**TFA**) at 3 µL/min during 5 min. Then, the digested peptides present in the samples were separated using an analytical LC-column (3 µm particles size C18-CL, 75 µm diameter x 12 cm long, Nikkyo Technos Co®, Tokyo, Japan) equilibrated in 5% ACN 0.1% formic acid (**FA**) (Fisher Scientific). Peptide elution was performed by applying a mixture of solvents A and B; solvent A being 0.1% FA in water and solvent B being 0.1% FA in ACN. The peptides were eluted from the column with a linear gradient from 5% to 35% of solvent B at a constant flow rate of 300 nL/min over 90 min.

The eluted peptides were thereafter direction-ionized using an ESI Nanospray III (AB SCIEX) and then analyzed on an AB SCIEX TripleTOF 5600 mass spectrometer coupled to the NanoLC system. The samples were ionized applying 2.8 kV to the spray emitter and the TripleTOF was operated in data-dependent mode, in which a time of flight (**TOF**) MS scan was made from 350 to 1259 m/z, accumulated for 250 ms TOF followed by 150 ms TOF with the same scan rage for MS, and the 25 most abundant multiply charged (2+, 3+, 4+ or 5+) precursor peptide ions automatically selected. Ions with 1+ and unassigned charge states were rejected from the MS/MS analysis. Collision energy was automatically set by the instrument according to the equation |CE|=(slope)x(m/z)+(intercept) with Charge=2; Slope= 0.0575 and Intercept=9.

5.2.3 LC-SWATH-MS acquisition

The TripleTOF 5600 (SCIEX) was configured as described by Gillet et al. (2012) for SWATH-MS-based experiments. 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 and direct current 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, 5 µL of each single pool per boar was loaded onto a trap column (NanoLC Column, 3µm C18-CL, 75 µm x 15 cm; Eksigent Technologies) and desalted with 0.1% TFA at 3µL/min during 5 min. The peptides were loaded onto an analytical column (LC Column, 3 µm C18-CL, 75µmx12cm, Nikkyo) equilibrated in 5% ACN 0.1% FA. Peptide elution was carried out with a linear gradient of 5 to 40% B in 90 min (A: 0.1% FA; B: ACN, 0.1% FA) at a flow rate of 300 nL/min. Eluted peptides were infused in the spectrometer nanoESI qQTOF (SCIEX TripleTOF 5600). The TripleTOF was operated in swath mode, in which a 0.050 s TOF MS scan from 350 to 1250 m/z was performed, followed by 0.080 s product ion scans from 230 to 1800 m/z on the 37 defined windows (3.05 sec/cycle). Collision energy was set to optimum energy for a 2+ ion at the center of each SWATH block with a 15 eV collision energy spread. The mass spectrometer was always operated in high sensitivity mode.

5.2.4 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 (**NCBInr**; 70353186 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 ≥ 1.3) and a false

discovery rate (**FDR**) less than 1%. The identified proteins were quantified using PeakView® from normalized label-free quantification (**LFQ**) intensity data.

5.3 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) in combination with Protein Analysis THrough Evolutionary Relationships (**PANTHER**, www.pantherdb.org). Both databases downloaded 06/04/2016, containing 63,686,057 and 1,424,953 entries in UniProt KB and PANTHER, respectively.

5.4 Reviewed reproductive role of identified boar seminal plasma proteins

In addition to bioinformatics, the current published literature about the relationship between proteins and reproductive functionality was extensively reviewed in order to determine the evidenced or putative reproductive role of the proteins identified in the boar SP (**Supplementary Table 5**).

Supplementary material

All supplementary material is available in the website:

https://cristinaperez61.wixsite.com/phdthesis

Acknowledgements

This experimental study was supported by MINECO (AGL2012-39903) Madrid (Spain), FEDER funds (EU), 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.

References

Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics. 2012, 11:O111.016717.

Perez-Patiño C, Barranco I, Parrilla I, Valero ML, Martinez EA, Rodriguez- Martinez H, Roca J. Characterization of the porcine seminal plasma proteome comparing ejaculate portions. J. Proteomics. 2016, 142:15-23.

Shevchenko A, Jensen ON, Podtelejnikov AV, Sagliocco F, Wilm M, Vorm O, Mortensen P, Shevchenko A, Boucherie H, Mann M. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. Proc. Natl. Acad. Sci. 1006, 93:14440-5.

Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nature Biotechnol. 2014, 30:223-6.

STUDY 3

This study has been published in **Journal of Proteome Research**

with the title "**NEW IN-DEPTH ANALYTICAL APPROACH OF THE PORCINE SEMINAL PLASMA PROTEOME REVEALS POTENTIAL FERTILITY BIOMARKERS**"

Journal of Proteome Research 2018, 17:1065-1076

DOI: 10.1021/acs.jproteome.7b00728

Abstract

A complete characterization of the proteome of seminal plasma (**SP**) is an essential step to understand how SP influences sperm function and fertility after artificial insemination (**AI**). The purpose of this study was to identify which among characterized proteins in boar-SP were differently expressed among AI-boars with significantly different fertility outcomes. A total of 872 SP-proteins, 390 of them belonging specifically to *Sus scrofa* taxonomy, were identified (Experiment 1) by using a novel proteomic approach that combined size exclusion chromatography and solid phase extraction as pre-fractionation steps prior to Nano liquid chromatography-electrospray ionization-tandem mass spectrometry analysis. The SP-proteomes of 26 boars showing significant differences in farrowing rate (n=13) and litter size (n=13) after the AI of 10,526 sows were further analyzed (Experiment 2). A total of 679 SP-proteins were then quantified by the sequential window acquisition of all theoretical mass spectra (**SWATH**) approach where the penalized linear regression least absolute shrinkage and selection operator (**LASSO**) revealed differentially expressed SP-proteins for farrowing rate (FURIN, AKR1B1, UBA1, PIN1, SPAM1, BLMH, SMPDL3A, KRT17, KRT10, TTC23 and AGT) and litter size (PN-1, THBS1, DSC1 and CAT). This study extended our knowledge of the SP-proteome and revealed some SP-proteins as potential biomarkers of fertility in AI-boars.

Keywords: seminal plasma, proteome, fertility, pig

VI. Study 3

6.1 Introduction

Artificial insemination (**AI**) is worldwide used by the porcine industry as it is considered the best breeding tool to disseminate genetic progress and, in consequence, to efficiently improve the production of high-quality pork meat (Riesenbeck, 2011). Currently, large numbers of semen AI-doses, usually at liquid state, are daily produced by specialized AI-centers following rigorous sanitary and semen quality controls. Despite this, fertility drops are routinely recorded among swine farms and attributed to males; most often linked to individual boars than to breeds or genetic lines (Roca et al., 2015). Between 5 to 7% of boars in AI-centers showing normal ejaculates and thereby used to produce semen AI-doses are sub-fertile, leading to serious productive and economic losses to farmers (Roca et al., 2015). This reality highlights current semen assessments performed by AI-centers are unfortunately unable to identify sub-fertile boars, despite being carried out using innovative technologies such as computer assisted sperm analysis (**CASA**) or flow-cytometry-based procedures (Broekhuijse et al., 2015). Consequently, pig AI-centers are calling for new semen evaluation tests capable to early identify sub-fertile boars, ideally before their semen is incorporated into commercial AI-programs (Roca et al., 2016). An area of relevant analysis is the seminal plasma (**SP**), since it is the fluid accompanying the spermatozoa. Analyses of the SP-composition might explain the evident individual variability in fertility among selected AI-boars.

Boar SP consists of a mixture of secretions from the testis, epididymis and, mainly, from the male accessory sexual glands which thus yields a complex composition (Mann and Lutwak-Mann, 1982), with potential relevance for sperm functionality and even fertility (Dyck et al., 2011; Bromfield, 2014). Recent findings in pigs support this relevance, with the findings that some SP-proteins could contribute to fertility outcomes of liquid stored boar semen AI-doses by either promoting it, as Glutathione Peroxidase 5, Paraoxonase-I, Osteopontin or Heat Shock Protein 70 (Barranco et al., 2005; Novak et al., 2010; Barranco et al., 2016; Flowers et al., 2016) or hindering it, as PSP-I and AQN-3 (Novak et al., 2010). These findings, albeit preliminary in nature as they were focused exclusively on specific proteins, strongly indicate that SP-proteins are involved in regulating sperm fertility in boars. Moreover, the above results also highlighted that an extensive knowledge of SP-proteome is a prerequisite in the search of SP-proteins with prediction

value for fertility of AI-boars, at the earliest possible stage. Recently, some large-scale studies have been performed to decode the boar SP-proteome and more than 500 proteins have been identified (Druart et al., 2013; Gonzalez-Cadavid et al., 2014; Perez-Patiño et al., 2016). Although worthy, this number of proteins is far below the identified numbers in i.e. human SP, with more than 2,000 SP-proteins described (Batruch et al., 2011; Batruch et al., 2012; Rolland et al., 2013; Pilch and Mann, 2006), thus calling for further research to completely decode the boar SP-proteome. Thus, the present study firstly aimed to increase our knowledge of the boar SP-proteome, to later look for qualitative and/or quantitative differences between fertile and subfertile boars. To achieve the first aim, the effectiveness of a fractionation approach based on novel solid-phase extraction (**SPE**) sorbent instead of the traditional used size exclusion chromatography (**SEC**) was tested (Experiment 1). This sorbent has been successfully applied for selective retention of some high-abundant proteins, such as bovine serum albumin (Vergara-Barberan et al., 2016), as well as for the isolation/pre-concentration of low-abundant proteins in complex matrices, e.g. lectins and viscotoxins in mistletoe extracts (Vergara-Barberan et al., 2016, 2017). This scenario of highly-abundant proteins masking low-abundant ones is a recurring trouble when aiming the complete decoding of the proteome of biological fluids containing complex mixtures of proteins, such as the SP (Perez-Patiño et al., 2016). Once the new spectral library was generated, the second objective was addressed (Experiment 2), disclosing the SP-proteome profile of AI-boars (n=64) with significant differences in fertility, measured in terms of farrowing rate (**FR**) and litter size (**LS**) of a very large number of inseminated females (>20,000). This profiling focused on the quantification of relative 'label free' proteins using the sequential window acquisition for all theoretical mass spectra (**SWATH**) approach.

6.2 Material and methods

6.2.1 Animals and fertility records

All procedures involving animals were performed according to international guidelines (Directive 2010/63/EU), following the approval of the Bioethics Committee of Murcia University (research code: 639/2012).

An initial population of 64 healthy and sexually mature boars of Landrace, Large White or the Pietrain breed delivered semen for the study. Preliminary fertility records of these boars suggested that they could show deviation in fertility outcomes from the average boar population within each breed. All boars belonged to the Topigs Norsvin España (Madrid, Spain) and were housed in four AI-centers located in Murcia, Soria, Lérida and León (Spain). The boars had high genetic merit and were included in AI-programs for genetic improvement. The boars were subjected to the same management conditions, specifically housed in individual pens in buildings with controlled light-regime (16 h per day) and temperature (15-25 °C) and with free access to water. The boars were fed with commercial feedstuff according to the nutritional requirements for adult boars subjected to regular ejaculate collection.

The inseminated sows, a total of 25,069 animals, were multiparous (1–7 farrowings) belonging to Large White, Landrace or Pietrain breed, and housed in different commercial farms located in Spain. The sows were cervically inseminated twice using 24-72 h liquid-stored semen AI-doses (2,500 x 10^6 sperm extended to 80 mL of a commercial extender). Semen AI-doses were prepared from entire ejaculates that fulfilled the standard of quantity and sperm quality thresholds, specifically > 200 x 10^6 sperm/mL, 75% of them motile and 85% depicting normal morphology. Fertility data was recorded over a 12-month period in terms of FR (number of farrowing sows respect to the number of inseminated sows) and LS (total number of piglets born per litter).

6.2.2 Seminal plasma sampling and storage

A total of 256 entire ejaculates were collected in a single container, using a semi-automatic collection procedure, during the 12 months that the boars were used in the AI-program. Immediately after ejaculate collection, two fully filled 15-mL tubes from each ejaculate were centrifuged twice at 1,500 *xg* at room temperature for 10 min (Rotofix 32A; Hettich Zentrifugen, Tuttlingen, Germany). The second supernatant was microscopically examined to ensure that it was sperm-free and thereafter split in 2-mL cryotubes, stored at -20 °C. The cryostored SP-samples were shipped in insulated containers with dry ice (-79 °C) to the Andrology Laboratory of the University of Murcia (Spain), where they were stored at -80 °C (Ultra Low Freezer; Hier, Schomberg, Ontorio, Canada) until proteome analysis.

6.2.3 Proteome analysis

6.2.3.1 Preparation of seminal plasma samples

The proteome analyses were carried out in the Proteomics Unit of the University of Valencia, Valencia, Spain (member of the PRB2-ISCIII ProteoRed Proteomics Platform). The SP-samples were thawed at room temperature and a protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO, USA) was added (1%, v:v), before centrifugation to 16,100 *xg* at 4 °C for 1 min. Total protein concentration was measured using a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA) following the manufacturer instructions. Aliquots of each of all SP-samples (64 SP-samples from 16 boars, 4 samples per boar) were mixed in a single pool for full characterization of the proteome (Experiment 1). Similarly, SP-samples (from 4 ejaculates) of 26 boars were used to identify and quantify differentially expressed proteins between boars showing highest and lowest fertility records (Experiment 2).

Two pre-fractionation approaches, SEC and SPE, were used to reduce the complexity of protein composition of the SP-samples. The SEC approach was carried out in an ETTAN LC system (GE Healthcare Life Science, Little Chalfont, United Kingdom) using a Superdex 200 5/150 GL column (GE Healthcare Life Science) following our previously described protocol (Perez-Patiño et al., 2016). Briefly, 50 µL of the pooled SP-sample (75 µg of protein) were injected into the column at a flow rate of 0.18 mL/min at 4 °C. The eluent was collected into different 0.2 mL-fractions which were dried in a rotatory evaporator and loaded in a 12% Tris-HCl precast one dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (**1D SDS-PAGE**) (Bio-Rad, Richmond, CA, USA). After this run, the gel was sliced at 38 kDa and the top of the gel was used in order to analyze the masked proteins by in gel-digestion processing. Secondly, a SPE with glycidyl methacrylate (**GMA**)-based polymer modified with cysteamine and treated with gold nanoparticles (**AuNPs**) was carried out to retain and separate proteins according to their pI value as described recently by Vergara-Barberán et al. (2017). Briefly, a GMA-co-ethylene dimethacrylate (**EDMA**) polymer was prepared by mixing GMA (20 wt%; Sigma-Aldrich), EDMA (5 wt%; Sigma-Aldrich), cyclohexanol (70 wt%), 1-dodecanol (5 wt%) and azobisisobutyronitrile (1 wt%; Fluka, Buchs, Switzerland) followed by thermal polymerization at 60 °C for 24 h. The resulting material was grounded and sieved (with pore size ≤ 100 µm), treated with an aqueous solution of cysteamine 2.5 M during

2 h, and then washed with deionized water until reaching a neutral pH. Finally, 400 mg of thiol-modified GMA were saturated with 100 mL of AuNPs metallic suspension (Alfa Aesar, Landcashire, United Kingdom). To perform the SPE approach, the cartridges were prepared as described by Vergara-Barberán et al. (2017) placing 50 mg of the modified polymer onto 1-mL propylene SPE cartridge (Análisis Vínicos, Tomelloso, Spain) using two frits (1/16', 20 µm). The SPE sorbents were activated with 200 µL of acetonitrile (**ACN**; Scharlab, Barcelona, Spain) and equilibrated with 500 µL of deionized water. Then, 200 µL of the SP-pool sample in 25 mM phosphate buffer solution (**PBS**) adjusted at either pH 8.2 or pH 9, were passed through the SPE material at a flow rate of 0.1 mL/min. The washing step was carried out with PBS (under the same pH conditions as the loading step) at 0.7 mL/min. Finally, the elution of the retained proteins was carried out using 200 µL of 25 mM PBS (pH 12) at 0.1 mL/min. A total of 75 µg of protein from all steps fractions (loading, washing and elution fractions) was collected (**Supplementary Table 6**). Each collected fraction was loaded onto different wells of 12% Tris-HCl precast 1D SDS-PAGE (Bio-Rad). The gel was run at a constant voltage of 200 mV for 30 min at room temperature including a molecular weight marker (ECL Plex Fluorescent Rainbow Marker, GE Healthcare Life Sciences). Coomassie Brilliant R250 Blue stain (Bio-Rad) was used to visualize protein bands on the gel. The eluted gel band was sliced in 10 fragments and used to analyze masked proteins by in-gel digestion processing.

6.2.3.2 Building a MS/MS library for SWATH analysis of boar SP

6.2.3.2.1 Complete proteome: In-solution digestion processing

The complete proteome (Experiment 1) was analyzed from an aliquot of the mixed SP-sample treated by in-solution processing following the steps described previously by Perez-Patiño et al. (2016). Briefly, three µL of the pooled SP for proteome analysis, containing 10 µg of proteins, were digested with Sequencing Grade Trypsin (Promega Corporation, Madison, USA) to generate peptides of each individual protein. The final concentration of protein in the digested sample was 0.13 µg/µL.

6.2.3.2.2 Low-abundant proteins: In-gel digestion processing

The portion of the 1D SDS-PAGE from SEC containing proteins with a molecular weight higher than 38 kDa and the portion of the 1D SDS-PAGE from SPE containing proteins from the eluted fraction were processed by in-gel digestion processing. After washing

with deionized water, the gel was dehydrated in ACN, reduced with dithiothreitol (**DTT**) and alkylated with iodoacetamide (**IAM**). The slide was cut into 10 small pieces of approximately 1 mm^2 in size and then transferred into 1.5-mL Eppendorf tubes. Sequencing Grade Trypsin digestion of the sliced gel was performed following the protocol used by Shevchenko et al. (1996).

6.2.3.2.3 LC-MS/MS analysis

The peptides recovered from in-gel and in-solution digestion processing were examined by liquid chromatography (**LC**) using a NanoLC Ultra 1D plus Eksigent (Eksigent Technologies, Dublin, CA, USA) which was directly connected to an AB SCIEX TripleTOF 5600 mass spectrometer (AB SCIEX, Framingham, MA, USA) in direct injection mode. Briefly, 5 µL from each digested sample were trapped on a NanoLC pre-column (3 µm particles size C18-CL, 350 µm diameter x 0.5 mm long; Eksigent Technologies) and desalted with 0.1% trifluoroacetic acid (**TFA**) at 3 µL/min during 5 min. Then, the digested peptides present in the samples were separated using an analytical LC-column (3 µm particles size C18-CL, 75 µm diameter x 12 cm long, Nikkyo Technos Co®, Tokyo, Japan) equilibrated in 5% ACN 0.1% formic acid (**FA**; Fisher Scientific). Peptide elution was performed by applying a mixture of solvents A and B; solvent A was 0.1% FA in water and solvent B was 0.1% FA in ACN. The peptides were eluted from the column with a linear gradient from 5% to 35% of solvent B at a constant flow rate of 300 nL/min over 90 min.

The eluted peptides were thereafter direction-ionized using an electrosparay ionization (**ESI**) Nanospray III (AB SCIEX) and then analyzed on an AB SCIEX TripleTOF 5600 mass spectrometer coupled to the NanoLC system. The samples were ionized applying 2.8 kV to the spray emitter and the TripleTOF was operated in data-dependent mode, in which a time of flight (**TOF**) mass sperctrometry (**MS**) scan was made from 350 to 1,259 m/z, accumulated for 250 ms TOF followed by 150 ms TOF with the same scan rage for MS, and the 25 most abundant multiply charged (2+, 3+, 4+ or 5+) precursor peptide ions automatically selected. Ions with 1+ and unassigned charge states were rejected from the MS/MS analysis.

6.2.3.3 LC-SWATH-MS acquisition

To determine quantitative differences in SP-protein composition among boars (Experiment 2), the SWATH analysis of individual SP-pool samples followed the same procedure previously described by Perez-Patiño et al. (2016) tuning the TripleTOF 5600 (AB SCIEX) as described by Gillet et al. (2012) for SWATH-MS-based experiments. 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 kept by the operation of the Radio Frequency and Direct Current 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, 5 µL of each single pool was loaded onto a trap column (NanoLC Column, 3 µm C18-CL, 75 µm x 15 cm; Eksigent Technologies) and desalted with 0.1% TFA at 3µL/min during 5 min. The peptides were loaded onto an analytical column (LC Column, 3 µm C18-CL, 75 µm x 12 cm, Nikkyo Technos Co®) equilibrated in 5% ACN 0.1% FA. Peptide elution was carried out with a linear gradient of 5 to 40% B in 90 min (A: 0.1% FA; B: ACN, 0.1% FA) at a flow rate of 300 nL/min. Eluted peptides were infused in the spectrometer nanoESI qQTOF (SCIEX TripleTOF 5600). The TripleTOF was operated in SWATH mode, in which a 0.050 s TOF MS scan from 350 to 1250 m/z was performed, followed by 0.080 s product ion scans from 230 to 1800 m/z on the 37 defined windows (3.05 sec/cycle). Collision energy was set to optimum energy for a 2+ ion at the center of each SWATH block with a 15 eV collision energy spread. The mass spectrometer was always operated in high sensitivity mode.

6.2.4 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 protein sequence database (**NCBInr**; 70353186 proteins searched) 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 97% confidence and a false discovery rate (**FDR**) less than 1%. The peptide confidence threshold was not set for a minimum number of peptides quantified but 6 transitions per peptide were necessary for quantify one peptide. The extracted ions chromatograms were integrated and the areas used to calculate total protein. A normalization of the calculated areas was done by total sum and the sum of all areas was equalized for all the samples.

6.2.5 Gene ontology and bioinformatics analysis

The bioinformatics of identified and validated SP-proteins was manually performed using the comprehensive bioinformatics tool for functional annotation UniProt KB database (www.uniprot.org) downloaded 15/05/2017, containing 553,941 reviewed entries of them 1,419 in *Sus scrofa* taxonomy. This analysis allowed elucidation of the different functions and processes in which the identified and validated proteins would be putatively involved. Three independent sets of ontology were used in the annotation: “the molecular function”, “the biological processes”, with a special mention to the reproductive process, and their “cellular component”. Proteins without similarity to database entries were not considered for collation.

6.2.6 Statistical analysis

Data were statistically analyzed using IBM SPSS (v19.0, IBM Spain, Madrid) and R software (R Foundation Members Supporters. www.r-project.org/, June 2014). Fertility data of each boar were recorded as direct boar effect (**DBE**). To do this, the raw fertility dataset was corrected for parameters related to farm and sow by using the multivariate statistical model previously described by Broekhuijse et al. (2012). The quantitative data obtained by PeakView® were analyzed using MarkerView® (v1.2, AB SCIEX). Firstly, peak areas were normalized by the sum of peak areas of all identified peptides and then, a penalized linear regression model using **LASSO** (least absolute shrinkage and selection

operator) (Tibshirani et al., 1996) was used to identify quantitative differently expressed SP-proteins among boars exhibiting different reproductive outcomes, specifically in FR and LS. Two different shrinkage factors (λ) were used for running the LASSO regression analyses, specifically λ_1 and λ_2, which were the median value obtained after replicating 100 times the cross validation and the minimum value obtained for the model, respectively. The explanatory ability of resulting selected proteins was showed using heatmaps after z-score normalization.

6.3 Results

6.3.1 Fertility records

The initial fertility data set included 64 boars with a total of 25,069 inseminated sows (**Supplementary Table 7**). Of the 64 boars, those showing largest FR and LS deviations regarding to average values of its genetic line averaged were finally selected, totaling 26 boars with 10,526 inseminated sows (**Fig. 7**). **Table 3** shows the fertility outcomes of the 26 boars, including the number of sows inseminated per boar. Thirteen boars were selected for showing highest deviations in FR and other 13 for showing largest deviation in LS (**Fig. 8**) with 5,449 and 5,077 inseminated sows, respectively. The selected boars showed deviations from genetic line average by at least 1.5% in FR or 0.3 litters in LS.

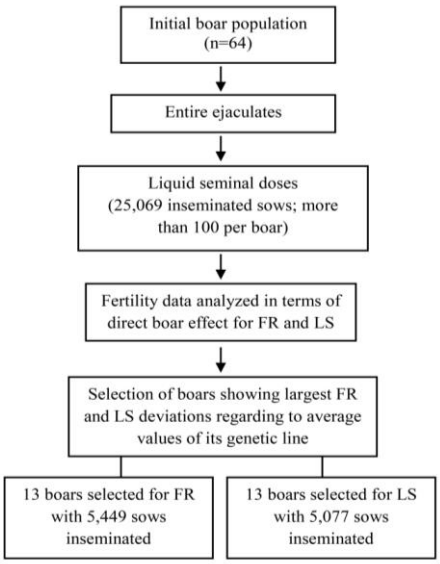


Fig. 7 Boar selection criteria according to fertility records of farrowing rate (**FR**) and litter size (**LS**).

Table 3 Fertility outcomes of 26 selected boars showing the largest deviations for farrowing rate (**FR**, n=13) and litter size (**LS**, n=13), including the inseminated sows per boar and the mean for each genetic line. Fertility data are in terms of direct boar effect.

Boars with larger deviations in Farrowing Rate				Boars with larger deviations in Litter Size			
Boar number	Inseminated sows (n)	FR (%)		Boar number	Inseminated sows (n)	LS (n)	
		Boar	Genetic Line			Boar	Genetic Line
1	397	85.59	87.10	**1**	273	13.79	13.20
2	679	89.18	87.50	**2**	628	12.85	13.60
3	605	90.63	88.10	**3**	516	13.97	13.50
4	334	89.69	88.10	**4**	254	14.01	13.50
5	211	85.55	87.50	**5**	135	13.30	13.60
6	567	83.49	88.10	**6**	592	12.57	13.50
7	215	89.62	88.10	**7**	187	12.44	13.50
8	586	88.63	87.10	**8**	527	12.89	13.20
9	556	86.43	88.10	**9**	213	13.85	13.50
10	258	80.84	88.10	**10**	598	13.10	13.50
11	174	84.48	87.50	**11**	415	14.23	13.60
12	475	91.14	87.50	**12**	512	14.04	13.60
13	392	91.05	88.10	**13**	227	13.99	13.50

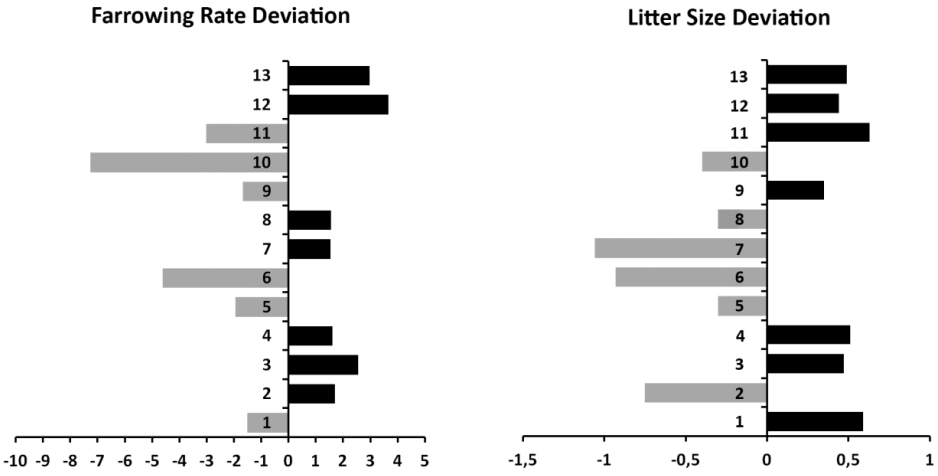


Fig. 8 Bar chart showing the deviation in farrowing rate and litter size of selected boars. The 0 represent the average for each genetic line in terms of direct boar effect. (See Table 3).

6.3.2 Characterization of the boar seminal plasma proteome

In Experiment 1, a single pooled SP sample from 64 ejaculates of 16 boars (4 ejaculates per boar) was analyzed. The use of SEC as pre-fractionation step allowed identifying a total of 35,093 spectra corresponding to 8,118 distinct peptides and 524 SP-proteins with a FDR ≤ 1% at protein level (**Supplementary Table 8**). The use of SPE as pre-fractionation step allowed to increase the number of spectra identified to 94,585, corresponding to 9,849 distinct peptides and 810 SP-proteins with a FDR ≤ 1% at protein level (**Supplementary Table 9**). These 810 SP-proteins resulted of the sum of SP-proteins identified in the two SPE performed at different pHs. At pH 8.4 618 SP-proteins were identified, whereas at pH 9.2 678 SP-proteins were identified. In sum, the combination of the two pre-fractionations steps (SEC and SPE), revealed a total of 134,605 spectra corresponding to 13,975 distinct peptides and 872 SP-proteins identified with a FDR ≤ 1% at protein level. The complete list of the 872 SP-proteins identified, including their Unused Score, UniProt Accession number, Protein Name, Species, % of Sequence Coverage and Matched Peptides is provided in **Supplementary Table 10**. A total of 390 SP-proteins was characterized as belonging to *Sus scrofa* taxonomy. The SWATH approach allowed the quantification of 679 SP-proteins present in all the SP-samples analyzed (**Supplementary Table 11**).

6.3.3 Bioinformatics Analysis

A total of 842 of the 872 SP-proteins identified were successfully mapped to UniProt KB for protein enrichment. The results are shown in **Fig. 9**. A total of 854 hits where framed into molecular function (Fig. 9A), showing many of them catalytic (349, 41%) and binding (239, 28%) activities. Some others appeared showing regulatory (89, 10%) and structural molecule (83, 10%) activities. Only 13 (2%) of them showed antioxidant activity. A total of 1,730 hits were enclosed into biological process (Fig.9B) and more than half of them were included into four fundamental biological issues: cellular (317, 18%), single-organism (312, 18%), biological regulation (256, 15%) and metabolic (246, 14%) processes. Noticeably, a total of 44 (2%) hits corresponding to 37 SP-proteins were classified as specifically implicated in reproductive processes (Fig. 9D). In particular, fourteen of these 37 SP-proteins (32%) were involved in fertilization process, 8 (18%) in modulating spermatogenesis and sperm capacitation and 5 (11%) in placental development. The rest of SP-proteins implicated in reproductive process were involved

in reproductive structure development (6, 14%), male accessory sexual gland development (4, 9%), oocyte and ovarian follicle development (3, 7%), oestrus cycle regulation and oogenesis (2, 5%) and finally, embryo implantation (2, 4%). At the end, a total of 1,359 hits were enriched for cellular components (Fig. 9C). Most of them belonged to the cell part group (421, 31%), predominantly in cell organelles (190, 14%) and membranes (167, 12%). A total of 366 (27%) SP-proteins belonged specifically to the extracellular region.

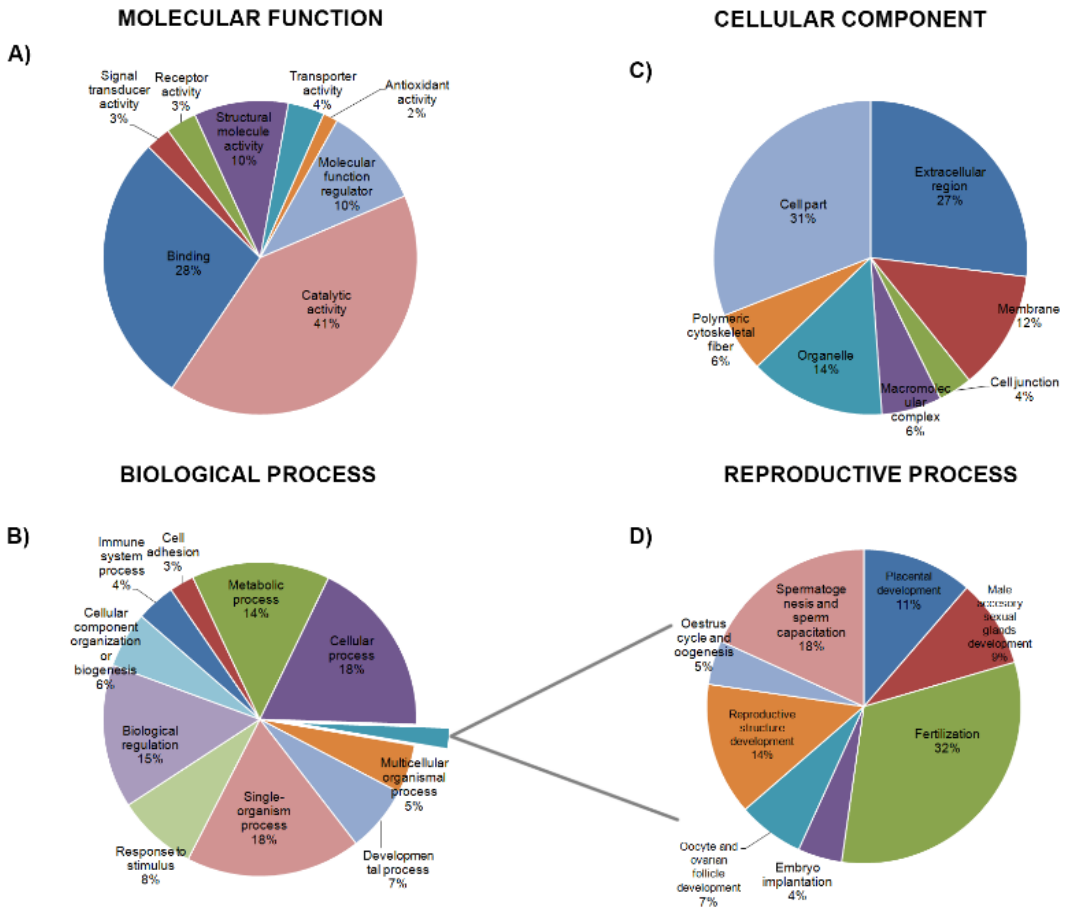


Fig. 9 Pie charts representing the distribution of the proteins identified in boar seminal plasma according to (A) molecular function; (B) biological process; (C) cellular component and those specifically involved in reproductive process (D) using UniProt KB database (www.uniprot.org).

6.3.4 Differences in SP-proteome profile among boars with different fertility

In Experiment 2, 26 SP samples (each one constituted as a pool of 4 ejaculates from each single boar) from 26 boars of different breeds were separately analyzed. The LASSO penalized regression analyses were used to identify quantitative differentially expressed

VI. Study 3

SP-proteins among boar populations with significant differences in fertility outcomes. For FR, 11 SP-proteins were quantitative differentially expressed between high- and low-FR boars using the penalty parameter λ_2, as illustrated in the heatmap of **Fig. 10**. Relative amount of these 11 differentially expressed SP-proteins in each boar is showed in **Supplementary Table 12**. Eight and three of these 11 SP-proteins were over- respectively under-expressed in boars showing high-FR (**Table 4**). To LS, the LASSO analysis using λ_1 identified four SP-proteins quantitative differentially expressed between boars showing large- and small-LS, as illustrated in the heatmap of **Fig. 11**. Two of these SP-proteins were over-expressed and the other two under-expressed in boars showing large-LS (**Table 5**). The relative amounts of these four differentially expressed SP-proteins in each boar are given in **Supplementary Table 13**.

Table 4 List of quantitative differentially expressed seminal plasma proteins between boars showing high- and low-farrowing rate from Lasso regression using λ_2.

UniProtKB IDs	Protein Name	Gene Name	Correlation
H0YNB5_HUMAN	Furin	FURIN	0.44
A0A140TAK7_PIG	Aldose reductase	AKR1B1	0.29
K7GRY0_PIG	Ubiquitin-like modifier-	UBA1	0.22
Q307R2_RABIT	Peptidyl-prolyl cis-trans	PIN1	0.18
Q8MI02_PIG	Sperm adhesion molecule	SPAM1	0.70
L5JSI4_PTEAL	Bleomycin hydrolase	BLMH	0.16
I3LV23_PIG	Sphingomyelin	SMPDL3A	0.09
H2QCZ8_PANTR	Keratin type I cytoskeletal	KRT17	1.21
F7BV15_ORNAN	Keratin type I cytoskeletal	KRT10	-0.33
E9QKU9_MOUSE	Tetratricopeptide repeat	TTC23	-0.95
U5L198_DELLE	Angiotensinogen	AGT	-0.43

λ_2: the minimum value obtained for the model

Table 5 List of quantitative differentially expressed seminal plasma proteins between boars showing large- and small-litter size from Lasso regression using λ_1.

UniProtKB IDs	Protein Name	Gene Name	Correlation
Q8WNW8_PIG	Nexin-1	PN-1	-0.02
F1SS26_PIG	Trombospondin-1	THBS1	-0.03×10^{-3}
Q9HB00_HUMAN	Desmocollin-1	DSC1	0.30
H2Q3E5_PANTR	Catalase	CAT	0.05

λ_1: the median value obtained after replicating 100 times the cross validation

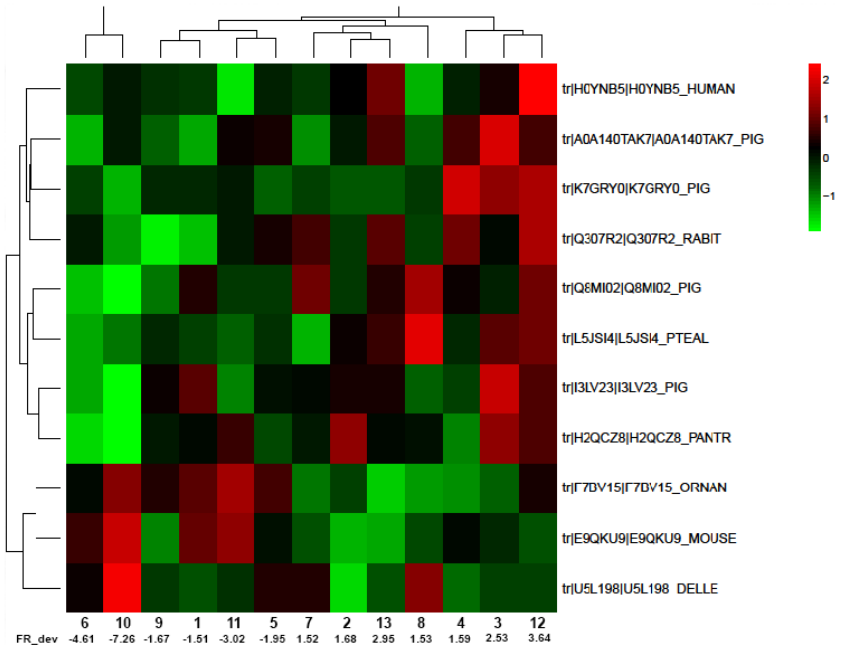


Fig. 10 Heatmap depicting the 11 boar seminal plasma proteins differentially expressed between sires with high- or low-farrowing rate after applying LASSO regression shrinkage parameter λ_2 (minimum value obtained for the model). Expression shown is mean centered. The red indicates increased and green down-regulated expression while black indicates mean value. **FR_dev**: farrowing rate deviation from average in terms of direct boar effect of each boar respecting to its genetic line.

6.4 Discussion

To the best of our knowledge, this would be the first large-scale SP-proteome study carried out in a livestock polytocous species highlighting a quantitative profile of differentially expressed SP-proteins among AI-sires showing differences in fertility outcome. Moreover, the findings revealed the presence of SP-proteins that would be specifically related to either FR or LS, indicating their value for the earliest possible prognosis of fertility and prolificacy by targeted analyses of the SP. This work should be considered as a first step that should be continued with others focused on evaluating the specific relevance to fertility of each of the SP-proteins disclosed herein.

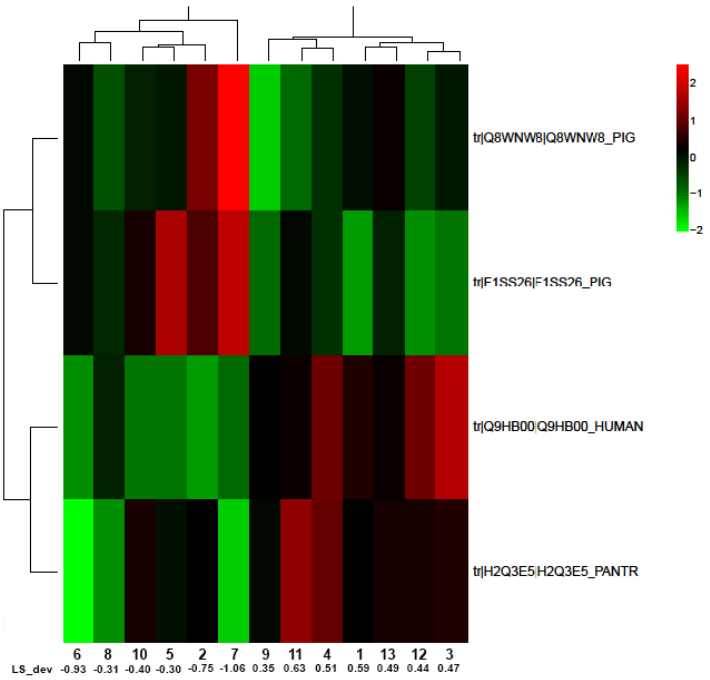


Fig. 11 Heatmap depicting the four boar seminal plasma proteins differentially expressed between sires yielding large- or small-litters after applying LASSO regression shrinkage parameter λ_1 (median value obtained after replicating 100 times the cross validation). Expression shown is mean centered. The red indicates increased and green down-regulated expression while black indicates mean value. **LS_dev**: litter size deviation from average in terms of direct boar effect of each boar respecting to its genetic line.

The presence of sub-fertile sires in AI-centers is a serious problem for the livestock industry because it compromises the reproductive performance of production farms and, thereby, the economic profit of farmers. In swine, where currently a single AI-boar can yearly produce semen AI-doses for inseminating more than 5,000 sows (Roca et al., 2016), a single AI-boar producing 0.5 piglet less per litter than expected could cause economic losses above 30,000 euros/year (Roca et al., 2015). Therefore, AI-boars showing downward deviation of 0.5 piglets per litter or 2-3% in farrowing rates are considered sub-fertile (Roca et al., 2015). Currently, it is estimated that between 5 to 7 % of boars housed in AI-centers could be considered as sub-fertile (Roca et al., 2015). The present study has found some AI-boars that fulfill the above requirements for sub-fertility thus accurately reflecting the situation in the field, and proving the selected animals are a properly selected cohort sample.

To identify the largest possible number of SP-proteins related to fertility outcomes, the first challenge of the present study was to try to enhance the number of existing boar SP-

proteins identified. To date, the most complete proteome of boar SP gathered a total of 536 proteins, 14 which is below what was found in other species. Those results highlighted that many low-abundant SP-proteins cannot be identified because they were masked by high-abundant SP-proteins, a recurrent problem in complex biological fluids as SP (Anderson and Anderson, 2002; Pilch and Mann, 2006) and clearly visible in pigs (Perez-Patiño et al., 2016), where the analytical method used was a limiting factor for full identification of boar SP-proteins. To solve this methodological trouble, a protein pre-fractionation step based on SPE using sorbents with proper selectivity was used instead of the traditional SEC. The SPE allowed to remove the high-abundant SP-proteins and, consequently, to reveal the low-abundant ones. The SPE support is based on a polymeric solvent modified with cysteamine as ligand for a posterior assembly of AuNPs (Vergara-Barberan et al., 2017). The polymer provides a porous structure that leads to a low-back pressure thus improving the extraction efficiency, allowing larger flow rates and shortening the operation time (Zhang et al., 2011). Besides, the incorporation of AuNPs to these materials constitutes a promising way of increasing the surface areas as well as to serve as new platforms for further tailoring its selective properties (Cao et al., 2010; Xu et al., 2010). In the present study, SPE was carried out at two different pHs, following the recommendation of Vergara-Barberán et al. (2016, 2017) that demonstrated a key role for pH to efficiently retain proteins onto the surface of AuNPs. The performed SPE allowed the recovery of a total of 13,975 peptides characterizing a total of 810 proteins. The combination of the two spectral libraries generated by using SEC and SPE resulted in a SP-proteome of 872 proteins, 336 more than the earlier largest SP-proteome known (Perez-Patiño et al., 2016). Other previous studies focused on the description of the porcine SP-proteome (Druart et al., 2013; Gonzalez-Cadavid et al., 2014) have identified only four SP-proteins which were not identified in the present study, specifically the serine protease inhibitor Kazal-type 13-like, seminal vesicles 14 kDa protein, ras-related protein Rab-22A and CUE domain-containing protein-1 like. These differences could be attributed to the search engine used in the present study. In fact, the two above quoted papers used MASCOT, whereas in the present study the Paragon algorithm of ProteinPilot was selected (Shilov et al., 2007; Shteynberg et al., 2013).

Thirty-seven of the 872 SP-proteins recorded in the present boar SP-samples were assorted in databases as involved in reproductive functions, highlighting the highly abundant low-molecular weight glycoproteins spermadhesins (Calvete and Sanz, 2007).

VI. Study 3

The spermadhesins identified in the current SP-samples were the five previously identified as characteristics of boar SP, namely PSP-I, PSP-II, AWN, AQN-1 and AQN-3, which could represent more than 90% of the total protein load in boar SP (Calvete and Sanz, 2007; Caballero et al., 2012). Spermadhesins play relevant roles in sperm capacitation, sperm-oocyte interaction and in modulating the uterine immune environment for a later successful embryo development (Rodriguez-Martinez et al., 2011). None of these spermadhesins were quantitative differentially expressed between SP-samples of boars showing differences in fertility in the present study. The relationship between some of these SP-spermadhesins and boar fertility has been previously evaluated and the PSP-I was the only one showing a negative relation with fertility records (LS) (Novak et al., 2010). These authors also noted that the concentration of PSP-I was higher in the SP from the sperm-poor ejaculate fraction than in the sperm-rich fraction, which is consistent with previous studies demonstrating that the sperm-poor ejaculate fraction, not only contains the largest SP volume in boar ejaculate, but also the highest amount of proteins per mL (Caballero et al., 2012; Perez-Patiño et al., 2016). In addition, spermadhesins are considered carrier proteins with a wide range of ligand-binding abilities (Töpfer-Petersen et al., 1998), leading they are over-expressed in SP, minimizing putative differential expression between SP-samples, particularly in those from entire ejaculates. These facts could explain the apparent discrepancy between our results and those reported by Novak and coworkers (Novak et al., 2010). They used only the sperm-rich fraction to prepare the AI-doses while we used the entire ejaculate (both ejaculate fractions collected in a single container) for both analyses and preparation the AI-doses used to render information on fertility. A tempting possibility is that differences in fertility may arise from differences in the relative volume of the post-SRF (the sperm poorest fraction) among boars, a matter that has -to the best of our knowledge- not been carefully studied/documented. In relation to this question, it is important to mention that insemination centers are currently moving, for hygiene and labour reasons, from collecting only the sperm-rich ejaculate fraction to collect the entire ejaculate (Roca et al., 2016), as performed in the present study, thus requiring further studies on SP-relative volume and fertility. Only one of these 37 SP-proteins classified with reproductive function in databases was quantitatively differentially expressed among the SP-samples of high and less fertile boars. Specifically, the hyaluronidase sperm adhesion molecule 1 (SPAM1), whose significance will be discussed later.

The data set of quantified SP-proteins was statistically analyzed using LASSO approaches for a consistent identification of the SP-proteins more related to boar fertility traits. The LASSO approaches allow both to select highly explanatory variables, in our case for differences in fertility traits, and to discard those variables that are not very significant (Li and Sillampaa, 2007). The final number of variables selected in LASSO approaches depends of the level of shrinkage, defined by factor λ. The factors λ used in the present study were particularly restrictive, leading to the selection of a small number of SP-proteins, albeit with high explanatory power. The fertility traits evaluated were FR and LS, where their relationship with SP-proteins was independently evaluated. The LASSO approaches identified 11 SP-proteins quantitatively differentially expressed between boars with high- or low-FR. Eight of the proteins were over-expressed and the other three under-expressed in the boars showing high-FR. Four of the eight SP-proteins over-expressed in boars showing high-FR, specifically Keratin I type cytoskeletal 17 (KRT17), a Peptidyl-prolyl cis-trans isomerase (PIN1), Sphingomyelin phosphodiesterase acid like 3A (SMPDL3A) and Bleomycin Hydrolase (BLMH), were so far not directly related to reproductive function in either male or female mammals. Keratin proteins protect epithelial cells against mechanical and non-mechanical stresses (Gu and Coulombe, 2007) and the present study would be the first highlighting its relevance in SP. There are no previous studies relating keratin proteins with reproductive functions. The KRT17 is expressed in cervical mucosa and it is over-expressed in cervical carcinoma (Escobar-Hoyos et al., 2014). With this background, it is difficult to elucidate how this SP-protein can influence the sperm fertilizing capacity and/or embryo development in the porcine species, but it could reflect the power of SP as a signal for the female, either initiating a transient inflammation or the long-lasting attainment of immune tolerance to paternal antigens by the female (Robertson, 2007). Interestingly, keratin proteins also participate in regulating cell motility (Chiang et al., 2016), maybe also sperm motility. The peptidyl-prolyl cis-trans isomerases (PPlases) are cellular enzymes widely distributed in the organism playing a relevant role as regulators of immune cell response (Nath et al., 2015). SMPDL3A is a protein carried by lysosomes, still poorly characterized and with putative anti-inflammatory and lipid-related roles (Gorelik et al., 2016). In the latter role, extracellular SMPDL3A would act hydrolyzing nucleotide substrates rather than membrane lipids (Gorelik et al., 2016), which in mammalian sperm would lead to capacitation (Cisneros-Mejorado and Sanchez Herrera, 2012). The BLMH is a cysteine aminopeptidase that protects cells against homocysteine, a metabolite whose excess is

very toxic (Zimny et al., 2006). In reproductive functionality, evidence suggests that homocysteine and related thiols are associated with male infertility, since the accumulation of homocysteine affects sperm functionality by inducing the generation of high intracellular reactive oxygen species (**ROS**) and subsequently the alteration of sperm proteins (Ebisch et al., 2006; Aitken et al., 2016). The other four SP-proteins, Ubiquitin-like modifier-activating enzyme 1 (UBA1), Sperm adhesion molecule 1 (SPAM1), Furin and Aldose reductase (AKR1B1), showed a clear implication in male reproductive success. The UBA1 is responsible of ubiquitin activation leading to protein ubiquitination. Yi et al. (2012) described UBA1 in boar sperm acrosome and they suggested that it would play a pivotal role in sperm capacitation and zona pellucida (**ZP**) penetration. In boar, cytometric values of sperm ubiquitin were positively correlated with FR, but negatively with LS (Lovercamp et al., 2007). Hyaluronidases are a family of proteins with well-known roles in mammalian sperm fertilization (Martin-DeLeon, 2006). The SPAM1, also known as PH-20, is the main hyase protein described in the boar reproductive tract which, secreted by the seminal vesicle is present in SP (Cibulkova et al., 2007). SPAM1 is essential for the fertilizing capacity of boar spermatozoa as it disperses the cumulus cell mass of oocyte and facilitates sperm-ZP binding (Yoon et al., 2014). Furin is a calcium-dependent serine endoprotease ubiquitously expressed in mammals, including the epididymal fluid of boars, where it would play an important role in sperm maturation, specifically promoting the acquisition of motility and fertilization capability (Thimon et al., 2006). The enzyme AKR1B1 is essential for the synthesis of prostaglandin (**PG**) F2α in the uterine endometrium (Bresson et al., 2011). In pregnant sows, AKR1B1 would play a decisive role regulating the endometrial synthesis of PGF2α during the first pregnancy stage facilitating the maintenance of pregnancy (Seo et al., 2014). The AKR1B1 may also play important roles in sperm functionality. Katoh et al. (2014) suggested that AKR1B1 would contribute to the acquisition of fertilizing capacity during the epididymal maturation of spermatozoa. Moreover, once in the female genital tract, AKR1B1 could play pivotal role in boar sperm capacitation mainly by regulating the sperm ability to generate ROS (Katoh et al., 2014). The three under-expressed SP-proteins were Angiotensinogen (AGT), Keratin type I cytoskeletal 10 (KRT10) and Tetratricopeptide repeat protein 23 (TTC23). The first one, AGT, belongs to the serpin family and it is the substrate precursor of all bioactive angiotensin (AT) peptides implicated in the renin-angiotensin system that regulates blood pressure (Lu et al., 2016). In the male reproductive tract, AGT was identified in the rat epididymis (Leung et al.,

1999) and the resulting AT peptides in human SP, with a presumed involvement in male fertility (O'Mahony et al., 2005) although their specific roles remaining unclear. One of the resulting AT, specifically AT II, was found at high levels in azoospermic men (Zheng et al., 2003). As mentioned above, the main function of keratin protein is to protect epithelial cells against mechanical stresses (Gu and Coulombe, 2007). Related to reproductive outcomes, a recent study found that serum concentration of KRT10 was particularly high at birth in intrauterine growth restricted infants (Ruis-Gonzalez et al., 2015). Tetratricopeptide repeat proteins (TPRs) are amino acid sequences involved in both protein-protein interaction and functionality of some chaperones, such as heat-shock proteins (Hsp), more specifically Hsp70 and Hsp90, that may influence on the action of reproductive hormones (Roberts et al., 2005). It is not currently known whether TPRs are involved in male reproductive functions. However, some of them are over-expressed in canine prostate cancer (Azakami et al., 2017) and involved in rat testis FSH receptor signaling (Ohta et al., 2012).

Focusing on LS, the LASSO approach identified 4 SP-proteins quantitatively expressed differently between boar exhibiting largest or smallest LS. Two of them were over-expressed (Desmocollin-1 and Catalase) and the other two under-expressed (Protease Nexin-1 and Trombospondin-1) in the boars showing the largest litter sizes. Desmocollin-1 (DSC1) is a cell surface serine protease expressed in male genital tract playing an essential role for tissue maintenance (Hobson et al., 2004) and it is involved in the spermatogenesis success (Mruk et al., 2014). Catalase (CAT) is an enzyme present in SP of several species including porcine (Marti et al., 2007; Koziorowska-Gilun et al., 2011), showing well characterized functionality as ROS-scavenger. Specifically, CAT decomposes hydrogen peroxide, the most dangerous ROS for boar spermatozoa (Awda et al., 2009), to water and oxygen, thereby protecting sperm against oxidative stress and consequently preventing the cells from suffering lipid peroxidation. Protease Nexin-1 (PN-1) is a serine protease inhibitor synthesized in the seminal vesicles that would contribute to regulate proteolytic activity of SP. The PN-1, like many other proteins found in SP, would have a bipolar effect. The lack of this protein would lead to infertility (verified in mice with deletion of the PN-1 gene) and secreted in excess to seminal vesicles dysfunctionality, as described in men (Murer et al., 2001). Trombospondin-1 (THBS1) is a potent anti-angiogenic protein (Rusnati et al., 2010) that would alter the

maternal-fetal interface during early pregnancy when it was present in the SP deposited during insemination in the female genital tract (Edwards et al., 2011).

6.5 Conclusion

In conclusion, a new pre-fractionation step combining the already used SEC and an approach based on SPE with a novel sorbent for protein isolation has been used for in-depth proteomic SP-characterization in boars with differential fertility (FR) and prolificacy (LS). This is evidently the most extensive annotation of the boar SP-proteome published to date. In addition, the quantification of identified SP-proteins with a SWATH approach among boars showing real fertility differences in terms of FR and LS, resulted in the identification of a panel of differentially expressed SP-proteins, some of them playing a pivotal role in reproductive processes. These results provide novel insight about the role of SP-proteins on boar fertility, acting as inducers of a series of processes at the sperm level, but also at the interaction with the female genital tract, including paradoxal immune responses to foreign proteins, which might eventually serve as biomarkers of fertility, at the earliest possible stage. Prediction of fertility and/or prolificacy using biomarkers in semen is a long-wish by academics and by the production and breeding industries. However, further validation of these proteins as bio-markers is necessary, probably with double-blind prospective testing for their suitability as boar fertility biomarkers.

Supplementary material

All supplementary material is available in the website:

https://cristinaperez61.wixsite.com/phdthesis

Acknowledgements

This experimental study was supported by MINECO and FEDER (AGL2015-69738-R and CTQ2014-52765-R) Madrid (Spain), Seneca Foundation (19892/GERM/15) Murcia

(Spain), PROMETEO/2016/145 (Generalitat Valenciana, 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 M. Vergara-Barberan were financially supported by the Seneca Foundation (Murcia, Spain) and MINECO (Madrid, Spain), respectively. The authors are grateful to AIM Iberica (Topigs Norsvin Iberica) for supplying the boar ejaculates.

References

Aitken RJ, Flanagan HM, Connaughton H, Whiting S, Hedges A, Baker MA. Involvement of homocysteine, homocysteine thiolactone, and paraoxonase type 1 (PON-1) in the etiology of defective human sperm function. Andrology. 2016, 4:345-60.

Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics. 2002, 1:845-67.

Awda BJ, Mackenzie-Bell M, Buhr MM. Reactive oxygen species and boar sperm function. Biol Reprod. 2009, 81:553-61.

Azakami D, Nakahira R, Kato Y, Michishita M, Kobayashi M, Onozawa E, Bonkobara M, Kobayashi M, Takahashi K, Watanabe M, Ishioka K, Sako T, Ochiai K, Omi T. The canine prostate cancer cell line CHP-1 shows over-expression of the co-chaperone small glutamine-rich tetratricopeptide repeat-containing protein α. Vet Comp Oncol. 2017, 15:557-562.

Barranco I, Tvarijonaviciute A, Perez-Patiño C, Alkmin DV, Ceron JJ, Martinez EA, Rodriguez-Martinez H, Roca J. The activity of paraoxonase type 1 (PON-1) in boar seminal plasma and its relationship with sperm quality, functionality, and in vivo fertility. Andrology. 2015, 3:315-20.

Barranco I, Tvarijonaviciute A, Perez-Patiño C, Vicente-Carrillo A, Parrilla I, Ceron J.J, Martinez E.A, Rodriguez-Martinez H, Roca J. Glutathione Peroxidase 5 Is Expressed by the Entire Pig Male Genital Tract and Once in the Seminal Plasma Contributes to Sperm Survival and In Vivo Fertility. PLoS One. 2016, 11:e0162958.

Batruch I, Lecker I, Kagedan D, Smith CR, Mullen BJ, Grober E, Lo KC, Diamandis EP, Jarvi KA. 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, 10:941-53.

Batruch I, Smith CR, Mullen BJ, Grober E, Lo KC, Diamandis EP, Jarvi KA. Analysis of seminal plasma from patients with non-obstructive azoospermia and identification of candidate biomarkers of male infertility. J Proteome Res. 2012, 11:1503-11.

Bresson E, Boucher-Kovalik S, Chapdelaine P, Madore E, Harvey N, Laberge PY, Leboeuf M, Fortier MA. The human aldose reductase AKR1B1 qualifies as the primary prostaglandin F synthase in the endometrium. J Clin Endocrinol Metab. 2011, 96:210-9.

Broekhuijse ML, Soštarić E, Feitsma H, Gadella BM. The value of microscopic semen motility assessment at collection for a commercial artificial insemination center, a retrospective study on factors explaining variation in pig fertility. Theriogenology. 2012, 77, 1466-1479.

Broekhuijse ML, Gaustad AH, Bolarin A, Knol EF. Efficient Boar Semen Production and Genetic Contribution: The Impact of Low-Dose Artificial Insemination on Fertility. Reprod Domest Anim. 2015, 50:103-9.

Bromfield JJ. Seminal fluid and reproduction: much more than previously thought. J Assist Reprod Genet. 2014, 31:627-36.

Caballero I, Parrilla I, Almiñana C, del Olmo D, Roca J, Martínez EA, Vázquez JM. Seminal plasma proteins as modulators of the sperm function and their application in sperm biotechnologies. Reprod Domest Anim. 2012, 47:12-21.

Calvete JJ, Sanz L. Insights into structure-function correlations of ungulate seminal plasma proteins. Soc Reprod Fertil Suppl. 2007, 65:201-15.

Cao Q, Xu Y, Liu F, Svec F, Fréchet JM. Polymer monoliths with exchangeable chemistries: use of gold nanoparticles as intermediate ligands for capillary columns with varying surface functionalities. Anal Chem. 2010, 82:7416-21.

Chiang CH, Wu CC, Lee LY, Li YC, Liu HP, Hsu CW, Lu YC, Chang JT, Cheng AJ. Proteomics Analysis Reveals Involvement of Krt17 in Areca Nut-Induced Oral Carcinogenesis. J Proteome Res. 2016, 15:2981-97.

Cibulková E, Manásková P, Jonáková V, Tichá M. Preliminary characterization of multiple hyaluronidase forms in boar reproductive tract. Theriogenology. 2007, 68, 1047-54.

Cisneros-Mejorado A, Sánchez Herrera DP. cGMP and cyclic nucleotide-gated channels participate in mouse sperm capacitation. FEBS Lett. 2012, 586:149-53.

Druart X, Rickard JP, Mactier S, Kohnke PL, Kershaw-Young CM, Bathgate R, Gibb Z, Crossett B, Tsikis G, Labas V, Harichaux G, Grupen CG, de Graaf SP. Proteomic characterization and cross species comparison of mammalian seminal plasma. J Proteomics. 2013, 91:13-22.

Dyck MK, Foxcroft GR, Novak S, Ruiz-Sanchez A, Patterson J, Dixon WT. Biological markers of boar fertility. Reprod Domest Anim. 2011, 46:55-8.

Ebisch IM, Peters WH, Thomas CM, Wetzels AM, Peer PG, Steegers-Theunissen RP. Homocysteine, glutathione and related thiols affect fertility parameters in the (sub)fertile couple. Hum Reprod. 2006, 21:1725-33.

Edwards AK, van den Heuvel MJ, Wessels JM, Lamarre J, Croy BA, Tayade C. Expression of angiogenic basic fibroblast growth factor, platelet derived growth factor, thrombospondin-1 and their receptors at the porcine maternal-fetal interface. Reprod Biol Endocrinol. 2011, 9:5.

Escobar-Hoyos LF, Yang J, Zhu J, Cavallo JA, Zhai H, Burke S, Koller A, Chen EI, Shroyer KR. Keratin 17 in premalignant and malignant squamous lesions of the cervix: proteomic discovery and immunohistochemical validation as a diagnostic and prognostic biomarker. Mod Pathol. 2014, 27:621-30.

Flowers WL, Deller F, Stewart KR. Use of heterospermic inseminations and paternity testing to evaluate the relative contributions of common sperm traits and seminal plasma proteins in boar fertility. Anim Reprod Sci. 2016, 174:123-131.

Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics. 2012, 11:O111.016717.

González-Cadavid V, Martins JA, Moreno FB, Andrade TS, Santos AC, Monteiro-Moreira AC, Moreira RA, Moura AA. Seminal plasma proteins of adult boars and correlations with sperm parameters.Theriogenology. 2014, 82:697-707.

Gorelik A, Illes K, Superti-Furga G, Nagar B. Structural Basis for Nucleotide Hydrolysis by the Acid Sphingomyelinase-like Phosphodiesterase SMPDL3A. J Biol Chem. 2016, 291:6376-85.

Gu LH, Coulombe PA. Keratin function in skin epithelia: a broadening palette with surprising shades. Curr Opin Cell Biol. 2007, 19:13-23.

Hobson JP, Netzel-Arnett S, Szabo R, Réhault SM, Church FC, Strickland DK, Lawrence DA, Antalis TM, Bugge TH. Mouse DESC1 is located within a cluster of seven DESC1-like genes and encodes a type II transmembrane serine protease that forms serpin inhibitory complexes. J Biol Chem. 2004, 279:46981-94.

Katoh Y, Takebayashi K, Kikuchi A, Iki A, Kikuchi K, Tamba M, Kawashima A, Matsuda M, Okamura N. Porcine sperm capacitation involves tyrosine phosphorylation and activation of aldose reductase. Reproduction. 2014, 148:389-401.

Koziorowska-Gilun M, Koziorowski M, Fraser L, Strzeżek J. Antioxidant defence system of boar cauda epididymidal spermatozoa and reproductive tract fluids. Reprod Domest Anim. 2011, 46:527-33.

Leung PS, Wong TP, Sernia C. Angiotensinogen expression by rat epididymis: evidence for an intrinsic, angiotensin-generating system. Mol Cell Endocrinol. 1999, 155:115-22.

Li Z, Sillanpää MJ. Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection. Theor Appl Genet. 2012, 125:419-35.

Lovercamp KW, Safranski TJ, Fischer KA, Manandhar G, Sutovsky M, Herring W, Sutovsky P. Arachidonate 15-lipoxygenase and ubiquitin as fertility markers in boars. Theriogenology. 2007, 67:704-18.

Lu H, Cassis LA, Kooi CW, Daugherty A. Structure and functions of angiotensinogen. Hypertens Res. 2016, 39:492-500.

Mann T, Lutwak-Mann C. Passage of chemicals into human and animal semen: mechanisms and significance. Crit Rev Toxicol. 1982, 11:1-14.

Marti E, Mara L, Marti JI, Muiño-Blanco T, Cebrián-Pérez JA. Seasonal variations in antioxidant enzyme activity in ram seminal plasma. Theriogenology. 2007, 67:1446-54.

Martin-DeLeon PA. Epididymal SPAM1 and its impact on sperm function. Mol Cell Endocrinol. 2006, 250:114-21.

Mruk DD, Xiao X, Lydka M, Li MW, Bilinska B, Cheng CY. Intercellular adhesion molecule 1: recent findings and new concepts involved in mammalian spermatogenesis. Semin Cell Dev Biol. 2014, 29:43-54.

Murer V, Spetz JF, Hengst U, Altrogge LM, de Agostini A, Monard D. Male fertility defects in mice lacking the serine protease inhibitor protease nexin-1. Proc Natl Acad Sci U S A. 2001, 98:3029-33.

Nath PR, Isakov N. Insights into peptidyl-prolyl cis-trans isomerase structure and function in immunocytes. Immunol Lett. 2015, 163:120-31.

Novak S, Ruiz-Sánchez A, Dixon WT, Foxcroft GR, Dyck MK. Seminal plasma proteins as potential markers of relative fertility in boars. J Androl. 2010, 31:188-200.

Ohta M, Ohyama K, Asano A, Yokota S, Khalid AM, Yamano Y. Regulation of rat tetratricopeptide repeat domain 29 gene expression by follicle-stimulating hormone. Biosci Biotechnol Biochem. 2012, 76:1540-3.

O'Mahony, OA, Barke S, Puddefoot JR, Vinson GP. Synthesis and secretion of angiotensin II by the prostate gland in vitro. Endocrinology. 2005, 146:392-8.

Perez-Patiño C, Barranco I, Parrilla I, Valero ML, Martinez EA, Rodriguez-Martinez H, Roca J. Characterization of the porcine seminal plasma proteome comparing ejaculate portions. J Proteomics. 2016, 142:15-23.

Pilch B, Mann M. Large-scale and high-confidence proteomic analysis of human seminal plasma. Genome Biol. 2006, 7:R40.

Riesenbeck A. Review on international trade with boar semen. Reprod Domest Anim. 2011, 46:1-3.

Roberts JD, Thapaliya A, Martínez-Lumbreras S, Krysztofinska EM, Isaacson RL. Structural and Functional Insights into Small, Glutamine-Rich, Tetratricopeptide Repeat Protein Alpha. Front Mol Biosci. 2015, 2:71.

Robertson SA. Seminal fluid signaling in the female reproductive tract: lessons from rodents and pigs. J Anim Sci. 2007, 85:36-44.

Roca J, Broekhuijse ML, Parrilla I, Rodriguez-Martinez H, Martinez EA, Bolarin A. Boar Differences In Artificial Insemination Outcomes: Can They Be Minimized? Reprod Domest Anim. 2015, 50:48-55.

Roca J, Parrilla I, Bolarin A, Martinez EA, Rodriguez-Martinez H. Will AI in pigs become more efficient? Theriogenology. 2016, 86:187-93.

Rodríguez-Martínez H, Kvist U, Ernerudh J, Sanz L, Calvete JJ. Seminal plasma proteins: what role do they play? Am J Reprod Immunol. 2011, 66:11-22.

Rolland AD, Lavigne R, Dauly, C, Calvel P, Kervarrec C, Freour T, Evrard B, Rioux-Leclercq N, Auger J, Pineau C. Identification of genital tract markers in the human seminal plasma using an integrative genomics approach. Hum Reprod. 2013, 28:199-209.

Ruis-González MD, Cañete MD, Gómez-Chaparro JL, Abril N, Cañete R, López-Barea J. Alterations of protein expression in serum of infants with intrauterine growth restriction and different gestational ages. J Proteomics. 2015, 119:169-82.

Rusnati M, Urbinati C, Bonifacio S, Presta M, Taraboletti G. Thrombospondin-1 as a Paradigm for the Development of Antiangiogenic Agents Endowed with Multiple Mechanisms of Action. Pharmaceuticals (Basel). 2010, 3:1241-1278.

Seo H, Choi Y, Shim J, Yoo I, Ka H. Comprehensive analysis of prostaglandin metabolic enzyme expression during pregnancy and the characterization of AKR1B1 as a prostaglandin F synthase at the maternal-conceptus interface in pigs. Biol Reprod. 2014:90, 99.

Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem. 1996, 68:850-8.

Shilov IV, Seymour SL, Patel AA, Loboda A, Tang WH, Keating SP, Hunter CL, Nuwaysir LM, Schaeffer DA. The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. Mol Cell Proteomics. 2007, 6:1638-55.

Shteynberg D, Nesvizhskii AI, Moritz RL, Deutsch EW. Combining results of multiple search engines in proteomics. Mol Cell Proteomics. 2013, 12:2383-93.

Thimon V, Belghazi M, Dacheux JL, Gatti JL. Analysis of furin ectodomain shedding in epididymal fluid of mammals: demonstration that shedding of furin occurs in vivo. Reproduction. 2006, 132:899-908.

Tibshirani R. Regression shrinkage and selection via the lasso. J. R. Statist. Soc. 1996, 58:267–288.

Töpfer-Petersen E, Romero A, Varela PF, Ekhlasi-Hundrieser M, Dostàlovà Z, Sanz L, Calvete JJ. Spermadhesins: a new protein family. Facts, hypotheses and perspectives. Andrologia. 1998, 30:217-24.

Vergara-Barberán M, Lerma-García MJ, Simó-Alfonso EF, Herrero-Martínez JM. Solid-phase extraction based on ground methacrylate monolith modified with gold nanoparticles for isolation of proteins. Anal Chim Acta. 2016, 917:37-43.

Vergara-Barberán M, Lerma-García MJ, Simó-Alfonso EF, Herrero-Martínez JM. Polymeric sorbents modified with gold and silver nanoparticles for solid-phase extraction of proteins followed by MALDI-TOF analysis. Microchimica Acta. 2017, 184:1683–1690

Xu Y, Cao Q, Svec F, Fréchet JM. Porous polymer monolithic column with surface-bound gold nanoparticles for the capture and separation of cysteine-containing peptides. Anal Chem. 2010, 82:3352-8.

Yi YJ, Zimmerman SW, Manandhar G, Odhiambo JF, Kennedy C, Jonáková V, Maňásková-Postlerová P, Sutovsky M, Park CS, Sutovsky P. Ubiquitin-activating enzyme (UBA1) is required for sperm capacitation, acrosomal exocytosis and sperm-egg coat penetration during porcine fertilization. Int J Androl. 2012, 35:196-210.

Yoon S, Chang KT, Cho H, Moon J, Kim JS, Min SH, Koo DB, Lee SR, Kim SH, Park KE, Park YI, Kim E. Characterization of pig sperm hyaluronidase and improvement of the digestibility of cumulus cell mass by recombinant pSPAM1 hyaluronidase in an in vitro fertilization assay. Anim Reprod Sci. 2014, 150:107-14.

Zhang A, YE F, Lu J, Wei Z, Peng Y, Zhao S. Preparation and Characterization of Polymer Solid Phase Microextraction Monolith Modified with Gold Nanoparticle. Chinese J Anal chem. 2011, 39:1247-1250.

Zheng S, Li Z, Wang YX, Xiang ZQ. Seminal plasma angiotensin II detection and its clinical implication. Zhonghua Nan Ke Xue. 2003, 9:669-72.

Zimny J, Sikora M, Guranowski A, Jakubowski H. Protective mechanisms against homocysteine toxicity: the role of bleomycin hydrolase. J Biol Chem. 2006, 281:22485-92.

STUDY 4

This manuscript has been submitted to for its publication to the **Mollecular and Cellular Proteomics** journal with the title "**ITRAQ-BASED ANALYSES PROVE THE PROTEIN PROFILE OF BOAR SPERMATOZOA IS REMODELLED DURING EJACULATION**"

Abstract

Proteins are essential for sperm function, including fertilizing capacity. Pig spermatozoa, emitted in well-defined ejaculate fractions, clearly vary in their functionality, which would be related to differences in protein composition. This study aimed firstly to update the porcine sperm proteome and secondly to identify proteins differentially expressed among mature spermatozoa from cauda epididymis and those fortuitously delivered in different ejaculate portions. Nine ejaculates from nine mature and fertile boars were manually collected in three separate ejaculate portions: the first 10 mL of sperm-rich ejaculate fraction (**SRF**), the rest of SRF and the post-SRF. The contents of cauda epididymides of the same boars was collected post-mortem by perfusion. All samples were centrifuged, pooling the resulting sperm pellets within the respective source-sample, which were later split to generate two technical replicates per source. The final eight sperm samples were subjected to isobaric tags for relative and absolute quantitation (**iTRAQ**)-based liquid chromatography tandem mass sperctrometry (**LC-MS/MS**) for protein identification and quantification. A total of 1,723 proteins were identified (974 of *Sus scrofa* taxonomy) and 1,602 of them were also quantified (960 of *Sus scrofa* taxonomy). After an ANOVA test, 32 *Sus scrofa* proteins showed quantitative differences ($P < 0.01$) among the sperm samples, which were particularly relevant for the functionality of spermatozoa fortuitously ejaculated in the post-SRF. The present study is the first showing quantitative differences in the protein profile of mature spermatozoa, involving proteins clearly implicated in sperm function, that prove the protein profile of boar spermatozoa is remodelled during ejaculation. These findings provide a valuable groundwork for further studies focused on identifying protein biomarkers of sperm fertility.

Keywords: proteomics, spermatozoa, ejaculate, porcine, iTRAQ

VII. Study 4

7.1 Introduction

Spermatozoa are highly differentiated, structurally complex and dynamic cells with a single but very specific function, to deliver the paternal genome to mature oocytes. To achieve this goal, the spermatozoon must possess full fertilization potential, which implies to follow a series of molecular and/or functional changes that include complete maturation, display an active forward movement and ability to undergo capacitation, zona binding and the acrosome reaction. Most of these essential functional processes are dependent on proteins, many of them of structural character (Park et al., 2013).

Moreover, spermatozoa undergo continuous molecular and structural remodelling through their interaction with the surrounding environment during epididymal maturation and storage and subsequent ejaculation (Gadella, 2017); interactive phenomena essential to attain fertilizing capacity (Mohanty et al., 2015). Since the ejaculate contains a heterogeneous suspension of spermatozoa, sperm sub-populations of different quality and functionality are present (Roca et al., 2016). D'Amours et al. (2018) suggested that qualitative and/or quantitative differences in protein expression among these sperm sub-populations resulted -probably- of an unbalanced interaction between spermatozoa and the surrounding environment during ejaculation. This noticeable generalist theory, derived from empirical studies in bulls based on sperm sub-populations generated post-ejaculation using cell separation methods, is calling for confirmation. The porcine ejaculate is an excellent model for clarifying the above theory as it is (as in human) emitted in fractions, with different contents of surrounding fluids ranging from cauda epididymal fluid to the differential secretion of the sexual accessory glands (Rodriguez-Martinez et al., 2009). These different environments built by ejaculate fractions result in remarkable differences in the capability of spermatozoa to withstand technologies such as cryopreservation (Saravia et al., 2009; Alkmin et al., 2014; Li et al., 2018a). A recent study (Li et al., 2018b) suggested differences in protein profile among spermatozoa in the different boar ejaculate fractions. Therefore, besides updating the porcine sperm proteome (Feugang et al., 2018), particularly considering the recent improvement of the annotation of protein-coding genes (Marx et al., 2017); the main objective of the present study was to identify and measure, by using isobaric Tags for Relative and Absolute Quantification (**iTRAQ**), eventual putative differences in protein composition between

mature spermatozoa from the cauda epididymis and those retrieved from the most representative ejaculate fractions, specifically the sperm-peak (first 10 mL of the sperm-rich ejaculate fraction, **SRF**), the rest of the SRF and the post-SRF. This approach would allow to measure the exchange in proteins experienced by the spermatozoa during ejaculation as well as evidencing the relevance of the different sexual accessory glands in such exchange.

7.2 Experimental procedures

7.2.1 Boars and collection of ejaculate and cauda epididymal contents

All procedures involving animals were performed according to international guidelines (Directive 2010/63/EU), following the approval of the Bioethics Committee of Murcia University (research code: 639/2012) and of the Local Ethical Committee for Experimentation with Animals at Linköping, Sweden (permit nr ID-1400).

Semen and cauda epididymal contents were retrieved from mature and fertile boars (*Sus scrofa*) of different breeds and crossbreds. The boars were housed in an artificial insemination (**AI**) centre (Topigs Norsvin España) located in Murcia (Spain) and regularly used in conventional AI-programmes (two ejaculates collected per week). The boars were kept under the usual management conditions of an AI-centre, being housed in individual pens under controlled regimens of both temperature (15-25 °C) and light (16 h per day), with free access to water and being fed with a commercial feedstuff for mature boars.

Boars, still healthy and fit to deliver semen, were removed from the AI-centre for genetic replacement reasons and slaughtered (Slaughterhouse Agroalimentaria de Teruel, Teruel, Spain). During the last month of their productive life, one ejaculate per boar was collected in separate portions by using the gloved-hand method. Specifically, the ejaculate portions collected were the first 10 mL of the SRF (P1), the rest of SRF (P2) and the post-SRF (P3). A proportional mixture of each ejaculate fraction was used to verify that all collected ejaculates fulfilled the standard thresholds of sperm quantity and quality needed for the preparation of AI-semen doses, specifically more than 200 x 10^6 sperm/mL (SP-100 NucleoCounter; ChemoMetec A/S, Allerød, Denmark) with more than 70% and 75% of them motile (objectively evaluated using ISASV1® CASA; Proiser R+D, Valencia,

Spain) and viable (cytometrically evaluated after staining using Hoechst 33342 and propidium iodide; BD FACSCanto II cytometer; Becton Dickinson Co, Franklin Lakes, NJ, USA), respectively. Once commercially slaughtered, the scrotal content of each boar were collected and transported in insulated containers (5 °C) to the Andrology Laboratory of Veterinary Teaching Hospital of University of Murcia (**VTH**), arriving within 4 h post-mortem.

7.2.2 Sperm sample preparation

Immediately after ejaculation in the AI-centre, the ejaculate portions were centrifuged twice (1,500 x*g* at room temperature for 10 min, Rotofix 32 A, Hettich Zentrifuge, Tuttlingen, Germany) to separate the seminal plasma (**SP**) and built a sperm pellet. The sperm pellets were washed twice with phosphate buffer solution (**PBS**, 1,500 x*g*, room temperature, 10 min) and the resulting pellets extended with PBS (1:3, vol:vol). The extended sperm pellets were transported in thermal containers to the VTH, arriving within 2 h of ejaculate collection. Once in the VTH-laboratory, the PBS-extended sperm pellets were again centrifuged (2,400 x*g*, room temperature, 3 min; Megafuge 1.0 R, Heraeus, Hanau, Germany). Although porcine semen barely contains other cells in addition to spermatozoa, as epithelial or blood cells (Ford, 2004); the resulting extended sperm pellets were cytometrically (stained with Hoechst 33342) evaluated to ensure that they were composed mainly of sperm cells (above 97% in all samples). Thereafter, the resulting sperm pellets were re-extended in PBS to a final concentration of 1,000 x 10^6 spermatozoa/mL, aliquoted in 1 mL volume and stored at -80 °C (Ultra Low Freezer; Haier, Schomberg, Ontorio, Canada) until use.

Contents of the epididymal caudae were collected following the procedure described by Alkmin et al. (2014) with slight modifications. Briefly, the cauda epididymides were carefully dissected, the ductus sectioned at the corpus-cauda limit, and retrograde infusion of air from the ductus deferens used to collect cauda luminal fluid from the sectioned ducts. The harvested fluid of the two caudae of each boar were pooled and microscopically evaluated to confirm that above 75% of viable, mature spermatozoa were present in all collected epididymal samples. The cauda epididymal sperm pellets were processed and stored following the same above protocol.

7.2.3 Proteome analysis

7.2.3.1 Protein extraction

The proteomic analyses were carried out in the Proteomics Unit of the University of Valencia, Valencia, Spain (member of the PRB2-ISCIII ProteoRed Proteomics Platform). Sperm pellets were thawed at room temperature and centrifuged at 14,000 x*g* at 10 ºC for 10 min (Eppendorf 5424R, Eppendorf AG, Hamburg, Germany) to obtain the protein-enriched fractions. Then, the sperm pellets were subjected to two extraction cycles, which involve dilution in 200 µL of UTC lysis buffer (7 M Urea, 2 M thiourea and 4% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate, **CHAPS**) and constant rotation at 5 ºC during 1 h. Then, the samples of the two extraction cycles were combined and treated with 10% (final concentration) of trichloroacetic acid (**TCA**) (Fisher Scientific, Madrid, Spain) and stored overnight at 5 ºC to achieve complete sperm lysis. Thereafter, samples were diluted in 200 µL of miliQ water (Merck Millipore, Darmstadt, Germany) and centrifuged at 14,000 x*g* during 2 h.

The TCA/Acetone protocol was used for protein precipitation. Briefly, the treated samples were diluted (4 x volume) in TCA-cold acetone solution, stirred and stored in ice during 15 min. Then, the samples were centrifuged (14,000 x*g* during 20 min at 4 °C; Eppendorf 5424R) and the resulting pellets were twice washed with 1 mL ice-cold acetone and centrifuged (14,000 x*g* during 20 min at 4 °C; Eppendorf 5424R). The pellets were incubated overnight at room temperature to volatilize the residual acetone. The dried pellets were lysed by dilution in 200 µL of protein extraction reagent (8 M urea and 0.5 M triethylammonium bicarbonate buffer [**TEAB**]), sonication and constant rotation at room temperature during 1 h. The protein concentration was measured by Lowry modified RC DC™ Protein Assay Kit (Bio-Rad, Richmond, CA, USA) following manufacturer instructions. A total of 100 µg of final protein extract per sample was used for iTRAQ analysis.

7.2.3.2 iTRAQ labelling

The iTRAQ labelling was performed following the protocol provided in the kit (AB SCIEX, Framingham, MA, USA). Cysteine residues were blocked by incubation in 4 µL of 50 mM tris (2-carboxyethyl) phosphine (**TCEP**) at 37 °C for 180 min. Sulfhydryl groups were alkylated with 1 µL of 200 mM S-methyl methanethiosulphonate (**MMTS**) at room temperature for 10 min. Urea was diluted to 2 M with 0.5 M of TEAB buffer to a final volume of 100 µL.

The protein samples were digested with 10 µg of sequencing-grade modified trypsin (Promega Corporation, Madison, WI, USA) diluted in 0.5 M of TEAB buffer, and incubated at 37 °C, overnight The digested protein samples were dried in a centrifugal vacuum concentrator (ISS 110 SpeedVac System, Thermo Savant, ThermoScientific, Langenselbold, Germany), dissolved in 80 µL of TEAB buffer in ethanol solution (3:7, vol:vol) and sonicated during 10 min. Then, the resulting peptides mixtures were labelled with the appropriate iTRAQ reagents following the protocol of 8-plex iTRAQ labelling kit (AB SCIEX). Thereafter, the iTRAQ-labelled peptides were incubated at room temperature for 3 h, mixed, aliquoted in 250 µg and dried by vacuum centrifugation.

7.2.3.3 Peptide fractionation by isoelectric focusing

The dried iTRAQ labelled peptides were subjected to fractionation by isoelectric focusing (**IEF**) separation following the protocol of Krijgsveld et al. (2006) with minor modifications. Briefly, 250 µg of the peptide mixture were brought up to 8 M urea in the presence of immobilized pH gradient (**IPG**) buffer, 3-11 NL (GE Healthcare Life Sciences, Little Chalfont, United Kingdom) and applied to 13 cm IPG dry strips, 3-11 NL (GE Healthcare), which were isoelectrofocused with 5000 V up to 25000 Vh. Thereafter, strips were washed with miliQ grade water and cut into 11 equal pieces. The isoelectrofocused peptides were extracted from the strips with 120 µL of the following sequential five extracted solutions (Fisher Scientific): (1) 5% aqueous acetonitrile (**ACN**) 0.1% trifluoroacetic acid (**TFA**), (2) 20% ACN 0.1% TFA, (3) 50% ACN 0.1% TFA, (4) 70% ACN 0.1% TFA and (5) 99.9% ACN 0.1% TFA. All the peptide fractions were combined, dried by vacuum centrifugation and re-dissolved with 40 µL of 2% ACN 0.1% TFA. The samples were cleaned and concentrated by C18 silica homemade “in tip” columns, dried by speed vacuum and re-suspended to a concentration of ca 0.30 µg/µL in 2% ACN 0.1% TFA.

7.2.3.4 Liquid chromatography and tandem mass spectrometry analysis (LC-MS/MS)

The labelled peptides were analysed by liquid chromatograohy (**LC**) using a NanoLC Ultra 1D plus Eksigent (Eksigent Technologies, Dublin, CA, USA), which was directly connected to a TripleTOF 5600 mass spectrometer (AB SCIEX). Briefly, 5 µL from each sample were loaded onto a trap column (NanoLC column, Chrom XP C18-3µm, 350 µm x 0.5 mm; Eksigent Technologies) and desalted with 0.1% TFA at 3 µL/min during 5 min. Then, the peptides were eluted from the trap column and separated using an

analytical LC-column (3 µm particles size C18-CL, 75 µm x 12 cm, Nikkyo Technos Co®, Tokyo, Japan) equilibrated in 5% ACN 0.1% formic acid (**FA**) (Fisher Scientific). Peptide elution was performed by applying a linear gradient of solvents A (0.1% FA in water) and B (0.1% FA in ACN) from 5% to 35% of solvent B in A at a constant flow rate of 300 nL/min over 90 min.

The eluted peptides were thereafter ionized using an ESI Nanospray III ion source (AB SCIEX) and then analyzed with the TripleTOF 5600 mass spectrometer coupled to the NanoLC system. The samples were ionized applying 2.8 kV to the spray emitter and the TripleTOF was operated in information-dependent acquisition mode, in which a time of flight (**TOF**) mass spectrometry (**MS**) scan was made from 350 to 1250 m/z, accumulated for 250 ms TOF followed by 75 product ion scans from 100-1500 m/z, on the 25 most abundant multiply charged (2+, 3+, 4+ or 5+) precursor peptide ions were automatically selected. Ions with 1+ and unassigned charge states were rejected from the MS/MS analysis. Collision energy was automatically set by the instrument rolling collision energies for iTRAQ labelled peptides.

7.2.4 Data processing: protein identification, validation and quantification

The generated SCIEX.wiff data-files were processed using ProteinPilot v5.0 search engine (AB SCIEX) for protein identification and quantification with a peptide confidence threshold of 95% and a false discovery rate (**FDR**) less than 1% at protein level. The Paragon algorithm (4.0.0.0, 4767) of ProteinPilot was used to search against the Uniprot_mammalias database with the following parameters: iTRAQ quantitation, trypsin specificity, cys-alkylation (MMTS), no taxonomy restricted, and the search effort set to throughout. The identified proteins were grouped by the Pro Group™ algorithm (ProteinPilot™ Solfware) following the Pro Group Report recommendation (http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_040586.pdf). Protein groups were exclusively made from observed peptides and the grouping was guided by usage of spectra. Consequently, unobserved regions of protein sequence were not considered for explaining the data.

7.2.5 Gene ontology analysis and bioinformatics

Bioinformatics of all identified and differentially expressed sperm-proteins was manually performed using the comprehensive bioinformatics tool for functional annotation UniProt

KB database (www.uniprot.org) downloaded 28/03/2018, containing 111,425,245 total entries with 40,710 of them encoded in *Sus scrofa* taxonomy. This analysis allowed elucidation of the different functions and processes in which the differentially expressed sperm proteins would be putatively involved. Three independent sets of ontology were used in the annotation: "the molecular function", "the biological processes", in which the proteins participate, and their "cellular component". Proteins without similarity to database entries were not considered for collation.

7.2.6 Experimental Design and Statistical Rationale

Sperm samples of nine mature and fertile boars were used. Four different sperm sources were contemplated per boar; (P0) cauda epididymis, (P1) the first 10 mL of the SRF, (P2) the rest of the SRF and (P3) the post-SRF. Before proteomic analysis, the nine sperm pellets derived from the same source (either cauda epididymis or fractions of the ejaculate) were pooled together to diminish individual effects. Consequently, four single sperm pools were built: P0-P4 to investigate the proteome of mature boar spermatozoa, pre- and post-ejaculation. In order to validate analytical reliability, each sperm pool was in turn split into two aliquots to generate two technical replicates. Labelled samples with the appropriate iTRAQ reagents were as follows: spermatozoa from P0 113 and 117; spermatozoa from P1 114 and 118; spermatozoa from P2 115 and 119; spermatozoa from P3 116 and 121; for the first and the second technical replicate, respectively.

The relative quantification of sperm proteins was achieved by comparison of the relative intensities of reporter ions of different sperm samples (P0, P1, P2 and P3), dividing the iTRAQ reporter groups (114, 115, 116, 117, 118, 119 and 121) by the peak intensity of 113. Principal Component Analysis (**PCA**) was performed to evaluate the discriminative ability of sperm-proteins in the four sperm samples. Thereafter, the Multiexperiment Viewer (**MeV**) software (version 4.8) (http://www.tm4.org/mev.html) was used for statistical normalization following instructions in the software manual. An ANOVA test was used to identify the sperm protein differentially expressed among the four sperm samples. Proteins were considered differentially expressed with an adjusted P-value < 0.01, and those with a fold change (**FC**) ≥ 1.5 after log2 transformation were then highlighted. The results of hierarchical clustering analysis of proteome profile of the different sperm samples were shown by means of heat-mapping after z-score normalization, using Euclidean distances. Quantitative analysis was done only in proteins

identified in all sperm samples. Differences in the gene ontology (**GO**)-distribution between total and differentially expressed proteins were analyzed using Chi-square analysis.

7.3 Results

7.3.1 Sperm proteome profile

A total of 1,723 proteins were identified when assuming a cut-off of unused prot score > 1.3 (corresponds to a confidence limit of 95% and a FDR less than 1%), belonging 974 of them to *Sus scrofa* taxonomy. Among the identified proteins, 1,602 were successfully quantified, 960 of them belonging to *Sus scrofa* taxonomy. All the proteins identified were present in the spermatozoa of the 4 sperm source samples, each examined as pools of 9 boars. The complete list of identified and quantified proteins and the ratio of the peak area of the iTRAQ reporter ion displaying the relative abundance of each protein is shown in **Supplementary Table 14**.

7.3.2 Differentially expressed proteins

The PCA analysis of total identified proteins counted 39.3% and 13.4% of variance in PC1 and PC2, respectively (**Supplementary Fig. 1**). The PCA of identified proteins encoded in *Sus scrofa* taxonomy showed similar results, explaining 36.2% and 13.9% of variance in PC1 and PC2, respectively (**Fig. 12**). The PC1 clearly discriminated sperm source samples. Specifically, the sperm samples were grouped in three separate branches: the first one includes the P0 and P1 source samples and the second and third the P2 and P3, respectively. Each branch was clearly separate from each other. The ANOVA test revealed a total of 43 proteins differentially expressed ($P < 0.01$) among the sperm samples, 32 of them belonging to *Sus scrofa* taxonomy. The quantitative value of the 43 differentially expressed proteins, after data normalization for each one of the sperm samples and the FC estimation of the groups after log2 transformation is shown in **Supplementary Table 15**. The expression pattern of the differentially expressed sperm proteins of the two technical replicates of each sperm source sample is graphically presented as heat-map in **Fig. 13** (proteins encoded in *Sus scrofa* taxonomy) and **Supplementary Fig. 2** (total differentially expressed proteins). The dendrograms of the two heat-maps made evident that the technical replicates of each sperm source sample

were grouped into a same cluster merged at a short distance; highlighting the robustness of the analysis carried out. The dendrograms also showed that the distance between the cluster grouping P0-P2 source samples and P3 source samples is large, making evident that the greatest differences in protein expression are between these two clusters. **Table 6** shows the FC between two sperm source samples for each one of the 32 differentially expressed protein encoded in *Sus scrofa* taxonomy. Only 3 of these proteins were differentially expressed, with a FC ≥ 1.5 between epididymis (P0) and ejaculate (P1-3) sperm sources. In contrast, a larger number of proteins, specifically 20, were differentially expressed with a FC ≥ 1.5 between post-SRF (P3) and the other sperm source samples (epididymis -P0-, first 10 mL of SRF -P1- and rest of SRF -P2-). Noticeable, these 20 differentially expressed proteins were overexpressed in the spermatozoa of the post-SRF fraction.

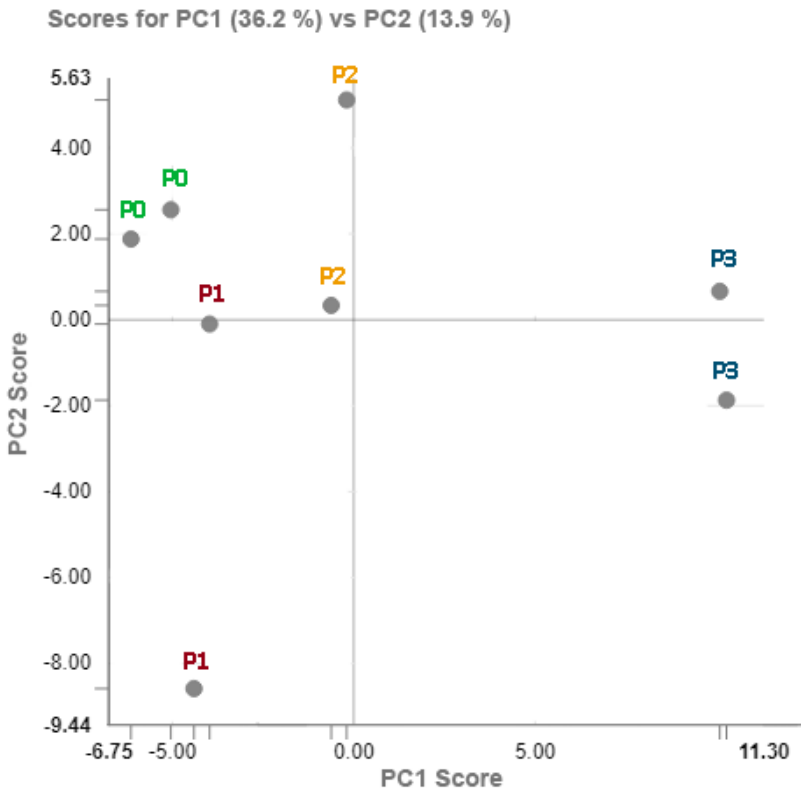


Fig. 12 Principal Component Analysis (PCA) showing the separation among the four sperm source samples for the sperm proteins identified in *Sus scrofa* taxonomy: (P0) mature spermatozoa from cauda epididymis, (P1) spermatozoa from first 10 mL of the sperm-rich ejaculate fraction (SRF), (P2) the rest of SRF and (P3) the post-SRF. The points represent two technical replicates for each sperm source sample and are based on the relative amounts quantified in each one of them.

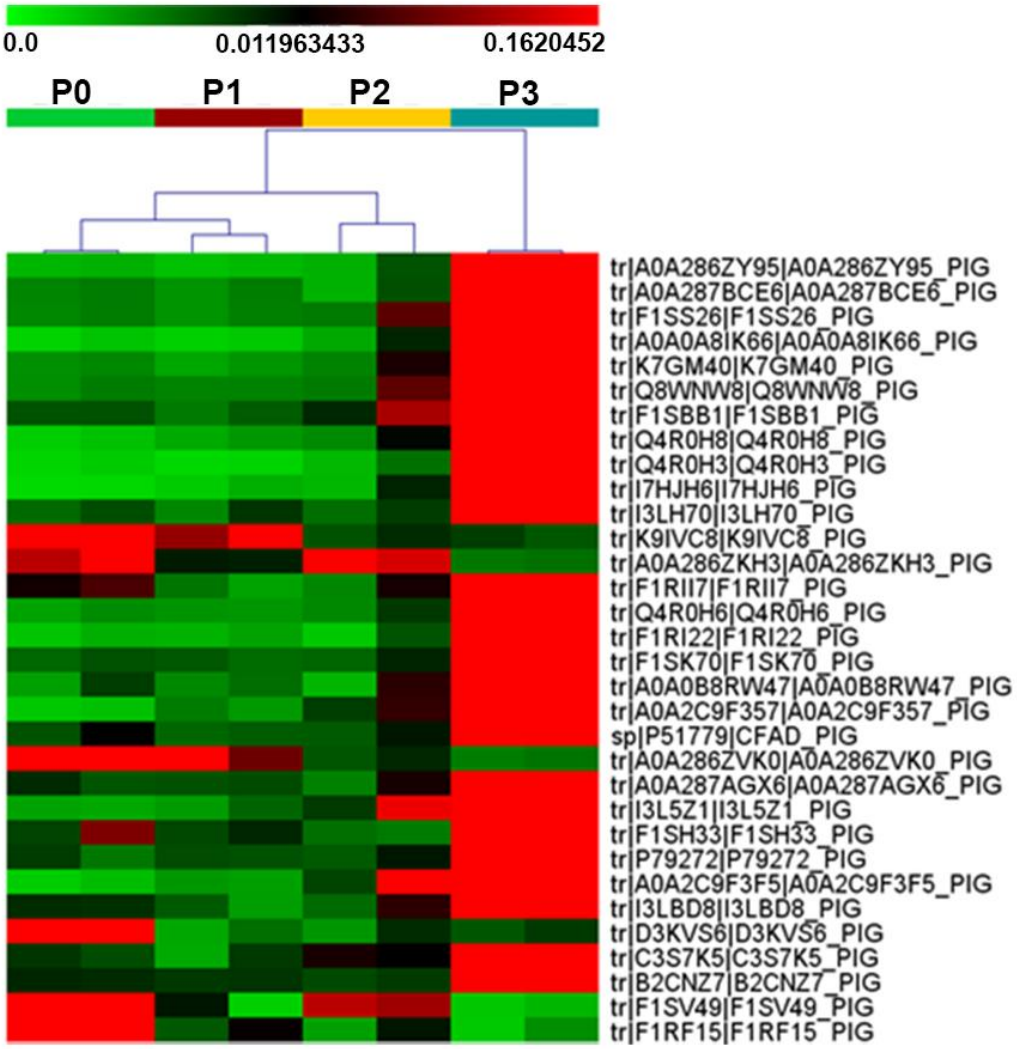


Fig. 13 Heat-map with dendrograms representing the differentially expressed proteins belonging to *Sus scrofa* among the four sperm samples: P0: mature spermatozoa from cauda epididymis, P1: spermatozoa from the first 10 mL of sperm-rich ejaculate fraction (SRF), P2: the rest of SRF, and P3: the post-SRF. Data retrieved from two technical replicates of each sperm source pooled sample. The hierarchical clustering tree of sperm samples is shown at the top. The relative expression level of each differentially expressed protein is shown on a colour scale from red (highest level) to green (lowest level).

Table 6 *Sus scrofa* proteins differentially expressed among boar sperm samples from cauda epididymis (P0), 10 first mL of sperm-rich ejaculate fraction (SRF) (P1), rest of SRF (P2) and post-SRF (P3). The proteins showing a fold change ≥ 1.5 among sperm samples are highlighted in bold.

PROTEIN NAME	ACCESSION	SEMINAL PLASMA[a]	FOLD CHANGE					
			P0 vs P1	P0 vs P2	P0 vs P3	P1 vs P2	P1 vs P3	P2 vs P3
Fibronectin 1	A0A286ZY95	Yes	0.20	-0.74	**-3.43**	-0.94	**-3.63**	**-2.69**
Fc fragment IgG binding protein	A0A287BCE6	Yes	0.11	0.00	**-2.47**	-0.10	**-2.58**	**-2.48**
Thrombospondin 1	F1SS26	Yes	0.09	-0.75	**-2.31**	-0.85	**-2.40**	**-1.55**
Heparin-binding protein WGA16	A0A0A8IK66	Yes	0.04	-1.28	**-4.00**	-1.32	**-4.04**	**-2.42**
Apolipoprotein A-I	K7GM40	No	0.25	-0.76	**-2.53**	-1.01	**-2.78**	**-1.77**
Nexin-1	Q8WNW8	Yes	-0.03	-0.80	**-2.34**	-0.77	**-2.32**	**-1.54**
Ca-binding phosphorylation	F1SBB1	No	0.23	-0.62	**-1.48**	-0.85	**-1.71**	-0.86
Spermadhesin AWN	Q4R0H8	Yes	-0.76	-1.14	**-3.75**	-0.97	**-2.99**	**-2.01**
Spermadhesin AQN1	Q4R0H3	Yes	0.22	-1.17	**-4.21**	-1.39	**-4.43**	**-3.05**
Spermadhesin AQN3	I7HJH6	Yes	-0.79	-1.35	**-4.45**	-1.16	**-3.66**	**-2.50**
Collagen type XVIII alpha 1 chain	I3LH70	Yes	0.02	-0.06	**-1.81**	-0.08	**-1.84**	**-1.76**
Nardilysin convertase	K9IVC8	No	0.09	0.90	0.99	0.80	0.90	0.10
Sepiapterin reductase	A0A286ZKH3	No	0.63	-0.02	1.33	-0.65	0.70	1.35
Hemoglobin subunit beta	F1RII7	Yes	1.24	0.50	-0.84	-0.74	**-2.08**	-1.35
Spermadhesin PSPI	Q4R0H6	Yes	-0.02	-0.61	**-2.77**	-0.60	**-2.75**	**-2.15**
Spermadhesin PSPII	F1RI22	Yes	-0.30	-0.73	**-3.55**	-0.43	**-3.26**	**-2.83**
Protein S	F1SK70	No	0.05	-0.19	**-1.80**	-0.25	**-1.85**	**-1.60**
Adenosine deaminase	A0A0B8RW47	No	0.14	-0.25	**-2.10**	-0.39	**-2.24**	**-1.86**
Seminal pB1	A0A2C9F357	Yes	-0.99	-1.00	**-3.50**	-1.01	**-2.51**	**-1.50**
Complement factor D	P51779	Yes	0.44	0.10	-1.22	-0.34	**-1.66**	-1.32
Transmembrane protein 38B	A0A286ZVK0	No	0.31	1.09	1.31	0.78	1.30	0.52
Glutathione S-transferase alpha 1	A0A287AGX6	No	0.11	-0.03	**-1.58**	-0.14	**-1.52**	**-1.48**
Cadherin 1	I3L5Z1	Yes	-0.51	-1.29	**-2.71**	-1.08	**-2.21**	**-1.52**
Serpin family I member 1	F1SH33	Yes	0.28	0.82	-1.00	0.53	-1.28	**-1.82**
Vitronectin	P79272	Yes	-0.10	-0.27	**-1.68**	-0.17	**-1.58**	**-1.51**
Acrosin inhibitor	A0A2C9F3F5	Yes	-0.77	-1.24	**-3.47**	-1.27	**-2.70**	**-1.63**
Semaphorin-3C	I3LBD8	Yes	0.69	0.00	**-1.51**	-0.69	**-1.99**	**-1.51**
16kDa secretory protein	D3KVS6	Yes	**2.40**	**1.94**	**1.74**	-0.46	-0.66	-0.20
Protein S100 A8	C3S7K5	No	0.44	-0.45	-1.32	-0.89	**-1.76**	-0.87
Cathepsin B	B2CNZ7	No	0.10	0.19	-1.11	0.08	-1.21	-1.29
DNA damage inducible 1	F1SV49	No	**1.97**	0.79	**3.11**	-1.18	1.13	1.32
Carboxylic ester hydrolase	F1RF15	No	**1.52**	**1.90**	**2.88**	0.38	1.36	0.98

[a]Perez-Patiño et al., 2016

7.3.3 Bioinformatics

The identified sperm proteins as a whole and those encoded into *Sus scrofa* taxonomy followed a similar pattern of distribution among the functional GO categories. The GO-distribution of total identified proteins and those differentially expressed is showed in **Supplementary Fig. 3**. The results shown here focus on in the proteins encoded into *Sus scrofa* taxonomy whose GO-distribution of total and differentially expressed proteins is shown in **Fig. 14A-C**. Regarding total identified *Sus scrofa* proteins (outer circles in **Fig. 14**), a total of 808 hits were found as linked to molecular function (Fig. 14A) with most of them involved in catalytic (420 proteins, 52%) and binding (354 proteins, 44%) activities. A total of 751 hits were related to biological process (Fig. 14B), with most of them involved in cellular (381, 34%) and metabolic (293, 27%) processes. Remarkably, only 58 proteins (5%) found were linked to reproductive process, 39 of them belonging to sperm functionality (**Table 7**). Finally, a total of 947 hits were associated to cellular component (Fig. 14C) with the related proteins mainly distributed in cell part (461, 49%), organelle (229, 24%) and membrane (207, 22%) regions.

The functional distribution of the differentially expressed *Sus scrofa* proteins is shown in the inner circles of **Fig. 14**. The 19 hits associated to molecular function followed a similar pattern of distribution to that of total identified *Sus scrofa* proteins, since 10 (53%) and 8 (42%) proteins were involved in binding and catalytic activities, respectively (Fig. 14A). However, the distribution pattern was different ($P < 0.001$) to biological process and cellular component (**Fig. 14B-C**). Most of the 23 hits associated with biological process were involved in metabolic (7, 19%), reproductive (6, 16%) and cellular (6, 16%) processes and biological regulation (6, 16%). Regarding to cellular component (Fig. 3C), the 21 hits were mainly attributed to extracellular region (8, 38%) and cell part (7, 34%).

VII. Study 4

Table 7 Proteins belonging to *Sus scrofa* taxonomy identified in boar spermatozoa linked to sperm functionality.

SPERM FUNCTIONALITY RELATED PROTEINS		
	ACCESION	PROTEIN NAME
ACROSOME REACTION	P08001	Acrosin
	A0A287B423	Sperm equatorial segment protein 1
	F1RFQ6	Syntaxin 2
	F1SRY6	Phosphoinositide phospholipase C
	I3LT46	Proprotein convertase subtilisin/kexin type 4
BINDING TO ZONA PELLUCIDA	F1SDH8	Phospholipid-transporting ATPase
	D0G0C8	Chaperonin containing TCP1, subunit 2
	F1SQN1	T-complex protein 1 subunit delta
	I3LR32	Chaperonin containing TCP1 subunit 5
	I3LCA2	Chaperonin containing TCP1 subunit 8
	I3LQ01	Calmegin
	A5A8V7	Heat shock 70 kDa protein 1-like
	I3LT46	Proprotein convertase subtilisin/kexin type 4
	F1S7B1	Sperm surface protein Sp17
	F1SB63	t-complex 1
	F1SF28	Zona pellucida-binding protein 1
	C8C4M8	Zona pellucida binding protein 2 transcript variant 1
CAPACITATION	I3LT05	Rhophilin associated tail protein 1
	F1SAF0	Dihydrolipoyl dehydrogenase
	A0A2C9F357	Seminal plasma protein pB1
	I3LT46	Proprotein convertase subtilisin/kexin type 4
SPERMATOGENESIS	F1S2A8	B-cell receptor associated protein 31
	P36968	Phospholipid hydroperoxide glutathione peroxidase
	F1RRW5	Angiotensin-converting enzyme
	F1RZG8	Meiosis specific nuclear structural 1
	I3LTK6	Ornithine decarboxylase antizyme 3
	D5K8A6	Sperm associated antigen 4-like protein
	F1RR24	Spermatogenesis associated 32
	F1S5Q9	Theg spermatid protein
	F1STE2	Testis specific 10
	F1SND1	Thioredoxin domain containing 8
SPERM MOTILITY	F1SFX0	L-lactate dehydrogenase
	F1SGC2	Coiled-coil domain containing 39
	I3LAH7	Coiled-coil domain containing 40
	F1SLL4	DEAD (Asp-Glu-Ala-Asp) box polypeptide 4
	F1S6P4	Growth arrest specific 8
	X4ZHK0	Phosphoglycerate kinase
	I3LT05	rhophilin associated tail protein 1
	F1RGN9	Tektin 1
	F1SV60	Tektin 2
	F1SDE8	Tektin 3
	F1RL08	Tektin 5

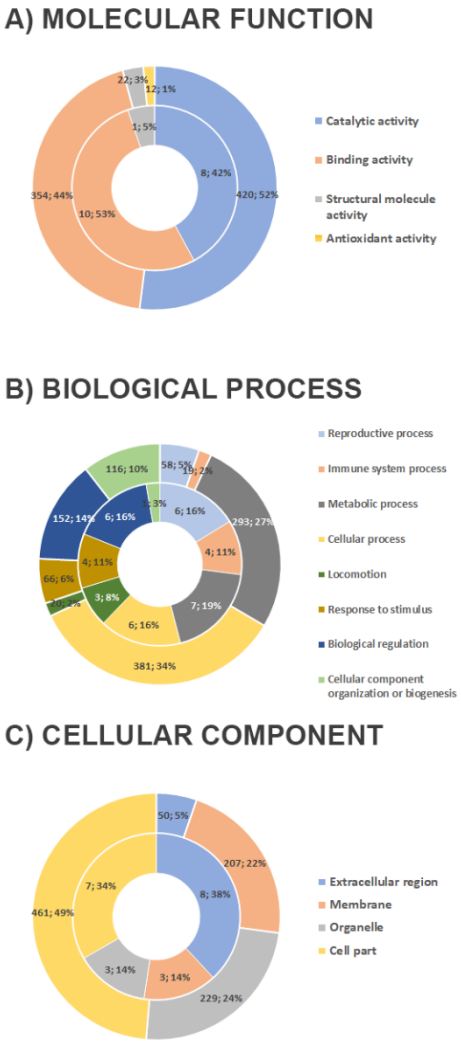


Fig. 14 Functional distribution of *Sus scrofa* sperm proteins into the three categories available in the UniProtKB/Swiss-Prot Web site (www.uniprot.org) for total identified (outer circles) and those differentially expressed (inner circles) sperm proteins belonging to *Sus scrofa* taxonomy.

7.4 Discussion

The first contribution of the present study was the generation, using iTRAQ, of a dataset of mature porcine sperm proteins that includes 1,723 proteins identified and 1,602 of them quantified, with about 60% of them encoded into *Sus scrofa* taxonomy. The number of identified proteins was substantially lower than the 2,728 proteins identified recently by Feugang et al. (2018) using shotgun approach. The difference in the number of identified proteins between both studies would be related to the differences in the methodological approach used for peptide detection and data processing for protein identification and

validation, including cut-off data acquisition, algorithm and search database. For instance, the iTRAQ is more restrictive than the shotgun approach to identify proteins, as iTRAQ requires high collision energies that cause a loss of sequence-informative fragments (Wiese et al., 2007). In addition, the FDR differed substantially between the studies, as Feugang et al. (2018) set it under 10% while we used a FDR below 1%, which substantially improved the confidence for the identified proteins. For the sake of comparing results from either study, if we had used an FDR lower than 10%, the number of identified proteins in our study would be 2,203, which approaches the number found by Feugang et al. (2018). The quantitative approach used, together with the opportunity to access the latest updates of porcine proteome, allowed us to identify and quantify a proportionally highest number of proteins encoded into *Sus scrofa* taxonomy. Specifically, 256 of the 974 proteins identified and encoded in *Sus scrofa* taxonomy have been incorporated into the UniProt KB database since November 2017, date in which more than 20,000 new proteins encoded in *Sus scrofa* taxonomy were include in the dataset.

Recently, D'Amours et al. (2018), in an experiment using frozen-thawed bovine sperm samples subjected to Percoll-based gradient centrifugation enrichment suggested the possibility that the mammalian ejaculated sperm population could be heterogeneous in protein composition, a bold consideration since they had used bull ejaculates, a species where the ejaculate is delivered as a single volume. The main purpose of the present study was to try to demonstrate this possibility using different sub-populations of mature porcine spermatozoa, specifically from cauda epididymis and the most recognizable portions of the porcine ejaculate. Since the porcine ejaculate (or human for that matter) is delivered in clearly identifiable fractions which can be manually collected separately the findings might prove of comparative value. In our case, moreover, samples could be obtained from boars that were later slaughtered still healthy and fertile, allowing collecting their cauda epididymis contents for comparison of relevant proteomic changes before and after ejaculation. Interestingly, there were no qualitative differences in the proteome profile between cauda epididymal spermatozoa and those of the three ejaculate portions studied, specifically the sperm-peak first 10 mL of the SRF, the rest of SRF and the post-SRF. Pini et al. (2016) reported similar results in ovine spermatozoa where only 0.5% of the proteins identified in epididymal spermatozoa were absent in ejaculated spermatozoa. These findings were somewhat expected knowing the limited capacity of

mature spermatozoa to generate new proteins (Baker, 2011; Gadella, 2017). However, there were quantitative differences in the protein profile between cauda epididymal spermatozoa and those spermatozoa fortuitously derived from three specific ejaculate portions, as iTRAQ revealed 43 proteins differentially expressed, 32 of them encoded in *Sus scrofa* taxonomy. These findings clearly demonstrate that the protein profile of boar spermatozoa is remodelled during ejaculation, as differences between cauda epididymal and ejaculated spermatozoa were evident, and they also demonstrated that the remodelling of protein profile during ejaculation was not equal in all spermatozoa, as quantitative differences among spermatozoa of the three ejaculate portions were found. This latter finding is especially relevant as it showed for the first time that mature ejaculate spermatozoa are heterogeneous in protein composition as they are in other parameters as sperm motility, morphology or fertilizing capacity, variables previously used to define sperm sub-populations (Roca et al., 2016).

The distribution of total identified proteins into the GO categories followed a similar pattern to that observed in the sperm proteome of other mammalian species, including human (Martinez-Heredia et al., 2006; Swegen et al., 2015) and porcine (Feugang et al., 2018). Focusing on the proteins encoded into *Sus scrofa* taxonomy, as they are the only ones validated in porcine it was noticeable that the GO-distribution of those differentially expressed proteins differed to that of total identified, particularly for biological process and cellular component. Compared to the total identified proteins, the proportion of differentially expressed proteins related to reproductive process and those located in extracellular region were significantly higher. It is also remarkable that the differentially expressed proteins related to reproductive process are involved in sperm functionality, including fertilizing capacity; and those located in extracellular region have also been identified in boar SP (Perez-Patiño et al., 2016). This reality suggests that the differences in protein expression among mature boar spermatozoa would have repercussions on sperm functionality and that the interaction between sperm and surrounding media during ejaculation could explain such differences. In other words, either the SP-protein coating was kept, or the SP causes relevant differences at the extracellular region of the spermatozoa.

During ejaculation in pigs, all mature spermatozoa stored in the caudal portion of the epididymis duct are propelled through the vas deferens and urethra where they are resuspended with secretions from the seminal vesicles, prostate and bulbourethral glands

(Rodriguez-Martinez et al., 2009). The spermatozoa present in the SRF (10 first mL of SRF and rest of SRF portions) are mainly exposed to epididymal fluids and prostate secretion, either one not especially rich in proteins (Rodriguez-Martinez, 2007). This reality would support the similarity in the protein profile between the mature spermatozoa collected from the caudal region of the epididymis (P0) and those of the SRF portions (P1: 10 first mL of SRF and P2: the rest of SRF), as they only shown differences in 3 proteins with a FC ≥ 1.5, which -moreover- were found overexpressed in epididymal mature spermatozoa. Specifically, the 16-kDa secretory protein, DNA damage inducible 1 homolog 1 (DDI1) and Carboxyl ester hydrolase (CES1), which are involved in sperm maturation. So, the 16-kDa secretory protein would be involved in the regulation of membrane lipids during epididymal sperm maturation (Okamura et al., 1999). The DDI proteins would attenuate apoptosis during sperm maturation (Bhat et al., 2006; Gong et al., 2017) and the CES proteins, present in the epididymal fluid (Ecroyd et al., 2006), are considered a safeguarding mechanism for sperm maturation when overexpressed in male reproductive tract (Mikhailov and Torrado, 1999).

In contrast to "SRF-spermatozoa", those spermatozoa fortuitously present in the post-SRF are mainly resuspended in secretions from the vesicular and bulbourethral glands during ejaculation (Einarsson, 1971; Mann and Lutwak-Mann, 1981). The post-SRF is proportionally the ejaculate fraction containing the largest amount of SP, and in turn being the protein-richest (Strzezek, 2002). Therefore, the spermatozoa of the post-SRF fraction are proportionally more exposed to absorbable SP-proteins. This fact would explain that the spermatozoa of the post-SRF contained more differentially expressed proteins with a FC ≥ 1.5, specifically 25 out of 28, all of them overexpressed which mainly provided by the SP, as 19 of them have also been identified in boar SP by a previous study of our own (Perez-Patiño et al., 2016). The SP-proteins bind to the sperm surface at ejaculation and many of them remodel protein domains of sperm membranes and, consequently, influence sperm functionality (Caballero et al., 2012). Jonakova et al. (2000) suggested that boar SP-proteins bind to the sperm surface mostly in aggregates than individually. These authors highlighted an aggregate that included the five spermadhesins identified in boar SP (PSPI, PSPII, AWN, AQN1 and AQN3) together with porcine B1 (pB1), a protein of bovine seminal plasma (**BSP**) family. All these six SP-proteins are among those overexpressed in spermatozoa of the post-SRF portion. Spermadhesins are low-molecular-weight glycoproteins with multifunctional properties that are mainly secreted

by the seminal vesicles (Calvete et al., 1995; Töpfer-Petersen et al., 1998; Kwok et al., 1993; Manaskova and Jonakova, 2008). Once bound to the sperm plasma membrane, they are involved in regulating some of the most relevant sperm functions, as motility, capacitation, acrosome reaction or sperm-zona pellucida binding (Dostalova et al., 1995; Boerke et al., 2008; Kwon et al., 2014). However, if overexpressed they may also be detrimental to sperm functionality. For instance, the PSPI/PSPII heterodimer showed decapacitating properties in highly-extended boar spermatozoa (Caballero et al., 2009); which mimics the high extension spermatozoa would experience in the post-SRF fraction where a relatively low number of spermatozoa are surrounded by a large volume of protein-rich SP. Similarly, Dyck et al. (2011) found that the SP-PSPI was negatively correlated with sperm motility in liquid-stored boar semen AI-doses. Spermadhesins can even impair boar sperm fertility if overexpressed in spermatozoa or when spermatozoa are confronted to an excess in SP; Novak et al. (2010) showed that boar ejaculates with high SP-concentration of AWN1 or PSPI had lower fertility outcomes in artificial inseminated sows; and Kwon et al. (2015) showed that boar spermatozoa overexpressing AQN3 had lower litter size. The other protein of the aggregate, the pB1, is a heparin-binding glycoprotein also secreted by the seminal vesicles (Calvete et al., 1997). The role of SP-pB1, once bound to the sperm membrane, is yet unclear. More is known about the DQH protein, a homologue of pB1 (Bezouska et al., 1999), which once attached to the sperm surface would play a crucial role in the ability of boar spermatozoa for binding to the oocyte zona pellucida, an ability that would be substantially reduced if the protein is overexpressed (Manaskova and Jonakova, 2008). It is interesting also to note that long-time exposure of bull spermatozoa to BSP proteins increase their sensitivity to cooling stress (Manjunath and Therien, 2002; Manjunath et al., 2007). Unfortunately, since the relationship with sperm function of all other overexpressed proteins is yet largely unknown; we now focus on those proteins showing evidence of their involvement in sperm functionality. One of them is the Fibronectin 1 (FN1), a cell adhesion glycoprotein overexpressed in SP of the post-SRF (Perez-Patiño et al., 2016) positively related with the percentage of spermatozoa with abnormal morphology (Gonzalez-Cadavid et al., 2014) but also with good sperm freezability (Vilagran et al., 2015). Regarding to FN1 bound to the sperm membrane, Pinke et al. (1997) showed that human spermatozoa rich in FN1 were less functional and more cryosensitive. Furthermore, Martinez-Leon et al. (2015) demonstrated that human spermatozoa incubated with FN1 are more likely to undergo capacitation and acrosome reaction, and are less prone to bind oocyte zona

pellucida. The SP IgG-binding protein was also related positively with the percentage of boar spermatozoa with abnormal morphology (Gonzalez-Cadavid et al., 2014) and with low sperm motility in human ejaculates (Chan et al., 2009). Thrombospondin-1 (TSP1) is a matricellular protein that binds to cell surface receptors modulating cellular function (Murphy-Ullrich and Iozzo, 2012). Although specific roles for sperm function are not known, the TSP1 contributes to the production of reactive oxygen species (**ROS**) that can lead to cellular dysfunction when in excess (LeBlanc and Kelm, 2017). Moreover, SP-concentration of TSP1 was recently related with boar ejaculates generating the smallest litter sizes in artificially inseminated sows (Perez-Patiño et al., 2018). The wheat germ agglutinin (WGA)-reactive protein on lipid rafts (WGA16) is a heparin-binding protein secreted by the prostate gland that binds to the sperm surface; a protein which must be removed so that spermatozoa can undergo the capacitation process (Garenaux et al., 2015). Accordingly, the WGA16 could hinder the capacitation process when overexpressed in the sperm membrane. Nexin-1, a seminal vesicle-derived protein, was also recently related with small litter size in artificially inseminated sows (Perez-Patiño et al., 2018). The type XVIII Collagen α1 chain (COL18A1) is an extracellular matrix protein that has been related with spermatogenic dysfunction in humans when overexpressed (Dobashi et al., 2003). The COL18A1 is elevated in the SP of men with non-obstructive azoospermia and therefore considered a potential biomarker for infertility (Batruch et al., 2011). Overall, the overexpression of the above-mentioned proteins exerts deleterious effects in sperm functionality, which would explain why post-SRF spermatozoa are proportionally less functional and more sensitive to cryopreservation than those of the other boar ejaculate portions and even of those of the cauda epididymis (Alkmin et al., 2014; Li et al., 2018a).

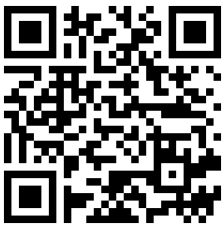
7.5 Conclussion

In summary, the present study offers an updated proteome of the porcine spermatozoa, proving for the first time that the protein profile of boar spermatozoa is remodelled during ejaculation, with ejaculated pig spermatozoa showing quantitative differences in proteins clearly implicated in sperm functionality. The interaction between spermatozoa with specific secretions they bathe in during ejaculation of well-delineated fractions could cause such differences in protein composition. This in turn, could explain the documented differences between spermatozoa of the same ejaculate in their response to sustain certain sperm biotechnologies, such as cryopreservation, or to achieve successful fertility. These findings warranted further studies to disclose the inner details of these differentially expressed proteins.

Supplementary material

All supplementary material is available in the website:

https://cristinaperez61.wixsite.com/phdthesis



Acknowledgements

The authors are grateful to AIM Iberica (Topigs Norsvin Iberica) for supplying the boar ejaculates. This experimental study was supported by MINECO (Spain) and FEDER funds (EU) (AGL2015-69738-R), and Seneca Foundation Murcia, Spain (19892/GERM-15); FORSS (745971) and The Swedish Research Council FORMAS (2017-00946), Stockholm, Sweden. C. Perez-Patiño, I. Barranco and J. Li were financially supported by Seneca Foundation (Murcia, Spain), MECD (Madrid, Spain) and the China Scholarship Council, respectively.

References

Alkmin DV, Perez-Patiño C, Barranco I, Parrilla I, Vazquez JM, Martinez EA, Rodriguez-Martinez H, Roca J. Boar sperm cryosurvival is better after exposure to seminal plasma from selected fractions than to those from entire ejaculate. Cryobiology. 2014, 69:203-10.

Baker MA. The 'omics revolution and our understanding of sperm cell biology. Asian J Androl. 2011, 13:6-10.

Batruch I, Lecker I, Kagedan D, Smith CR, Mullen BJ, Grober E, Lo KC, Diamandis EP, Jarvi KA. 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, 10:941-53.

Bezouska K, Sklenár J, Novák P, Halada P, Havlícek V, Kraus M, Tichá M, Jonáková V. Determination of the complete covalent structure of the major glycoform of DQH sperm surface protein, a novel trypsin-resistant boar seminal plasma O-glycoprotein related to pB1 protein. Protein Sci. 1999, 8:1551-6.

Bhat GK, Sea TL, Olatinwo MO, Simorangkir D, Ford GD, Ford BD, Mann DR. Influence of a leptin deficiency on testicular morphology, germ cell apoptosis, and expression levels of apoptosis-related genes in the mouse. J Androl. 2006, 27:302-10.

Boerke A, Tsai PS, Garcia-Gil N, Brewis IA, Gadella BM. Capacitation-dependent reorganization of microdomains in the apical sperm head plasma membrane: functional relationship with zona binding and the zona-induced acrosome reaction. Theriogenology. 2008, 70:1188-96.

Caballero I, Vazquez JM, Mayor GM, Almiñana C, Calvete JJ, Sanz L, Roca J, Martinez EA. PSP-I/PSP-II spermadhesin exert a decapacitation effect on highly extended boar spermatozoa. Int J Androl. 2009, 32:505-13.

Caballero I, Parrilla I, Almiñana C, del Olmo D, Roca J, Martínez EA, Vázquez JM. Seminal plasma proteins as modulators of the sperm function and their application in sperm biotechnologies. Reprod Domest Anim. 2012, 47:12-21.

Calvete JJ, Mann K, Schäfer W, Raida M, Sanz L, Töpfer-Petersen E. Boar spermadhesin PSP-II: location of posttranslational modifications, heterodimer formation with PSP-I glycoforms and effect of dimerization on the ligand-binding capabilities of the subunits. FEBS Lett. 1995, 365:179-82.

Calvete JJ. Raida M, Gentzel M, Urbanke C, Sanz L, Töpfer-Petersen E. Isolation and characterization of heparin- and phosphorylcholine-binding proteins of boar and stallion seminal plasma. Primary structure of porcine pB1. FEBS Lett. 1997, 407:201-6.

Chan CC, Shui, HA, Wu CH, Wang CY, Sun GH, Chen HM, Wu GJ. Motility and protein phosphorylation in healthy and asthenozoospermic sperm. J Proteome Res. 2009, 8:5382-6.

D'Amours O, Frenette G, Bourassa S, Calvo É, Blondin P, Sullivan R. Proteomic markers of functional sperm population in bovines: Comparison of low- and high-density spermatozoa following cryopreservation. J Proteome Res. 2018, 17:177-188.

Dobashi M, Fujisawa M, Naito I, Yamazaki T, Okada H, Kamidono S. Distribution of type IV collagen subtypes in human testes and their association with spermatogenesis. Fertil Steril. 2003, 80:755-60.

Dostálová Z, Calvete JJ, Sanz L, Töpfer-Petersen E. Boar spermadhesin AWN-1. Oligosaccharide and zona pellucida binding characteristics. Eur J Biochem. 1995, 230:329-36.

Dyck MK. Foxcroft GR, Novak S, Ruiz-Sanchez A, Patterson J, Dixon WT. Biological markers of boar fertility. Reprod Domest Anim. 2011, 46:55-8.

Ecroyd H, Belghazi M, Dacheux JL, Miyazaki M, Yamashita T, Gatti JL. An epididymal form of cauxin, a carboxylesterase-like enzyme, is present and active in mammalian male reproductive fluids. Biol Reprod. 2006, 74:439-47.

Einarsson S. Studies on the composition of epididymal content and semen in the boar. Acta Vet Scand Suppl. 1971, 36:1-80.

Feugang JM, Liao SF, Willard ST, Ryan PL. In-depth proteomic analysis of boar spermatozoa through shotgun and gel-based methods. BMC Genomics. 2018, 19:62.
Ford WC. Regulation of sperm function by reactive oxygen species. Hum Reprod Update. 2004, 10:387-99.
Gadella BM. Reproductive tract modifications of the boar sperm surface. Mol Reprod Dev. 2017, 84:822-831.
Garénaux E, Kanagawa M, Tsuchiyama T, Hori K, Kanazawa T, Goshima A, Chiba M, Yasue H, Ikeda A, Yamaguchi Y, Sato C, Kitajima K. Discovery, primary, and crystal structures and capacitation-related properties of a prostate-derived heparin-binding protein WGA16 from boar sperm. J Biol Chem. 2015, 290:5484-501.
Gong X, Xie H, Li X, Wu J, Lin Y. Bisphenol A induced apoptosis and transcriptome differences of spermatogonial stem cells in vitro. Acta Biochim Biophys Sin (Shanghai). 2017, 49:780-791.
González-Cadavid V, Martins JA, Moreno FB, Andrade TS, Santos AC, Monteiro-Moreira AC, Moreira RA, Moura AA. Seminal plasma proteins of adult boars and correlations with sperm parameters. Theriogenology. 2014, 82:697-707.
Jonáková V, Manásková P, Kraus M, Liberda J, Tichá M. Sperm surface proteins in mammalian fertilization. Mol Reprod Dev. 2000, 56:275-7.
Krijgsveld J, Gauci S, Dormeyer W, Heck AJ. In-gel isoelectric focusing of peptides as a tool for improved protein identification. J Proteome Res. 2006, 5:1721-30.
Kwok SC, Yang D, Dai G, Soares MJ, Chen S, McMurtry JP. Molecular cloning and sequence analysis of two porcine seminal proteins, PSP-I and PSP-II: new members of the spermadhesin family. DNA Cell Biol. 1993, 12:605-10.
Kwon WS, Rahman MS, Lee JS, Kim J, Yoon SJ, Park YJ, You YA, Hwang S, Pang MG. A comprehensive proteomic approach to identifying capacitation related proteins in boar spermatozoa. BMC Genomics. 2014, 15:897.
Kwon WS, Rahman MS, Lee JS, Yoon SJ, Park YJ, Pang MG. Discovery of predictive biomarkers for litter size in boar spermatozoa. Mol Cell Proteomics. 2015, 14:1230-40.
LeBlanc AJ, Kelm NQ. Thrombospondin-1, free radicals, and the coronary microcirculation: The aging conundrum. Antioxid Redox Signal. 2017, 27:785-801.
Li J, Barranco I, Tvarijonaviciute A, Molina MF, Martinez EA, Rodriguez-Martinez H, Parrilla I, Roca J. Seminal plasma antioxidants are directly involved in boar sperm cryotolerance. Theriogenology. 2018a, 107:27-35.
Li J, Roca J, Perez-Patiño C, Barranco I, Martinez EA, Rodriguez-Martinez H, Parrilla I. Is boar sperm freezability more intrinsically linked to spermatozoa than to the surrounding seminal plasma? Anim Rep Sci. 2018b, pii: S0378-4320(17)30936-3.
Manásková P, Jonáková V. Localization of porcine seminal plasma (PSP) proteins in the boar reproductive tract and spermatozoa. J Reprod Immunol. 2008, 78:40-8.
Manjunath P, Bergeron A, Lefebvre J, Fan J. Seminal plasma proteins: functions and interaction with protective agents during semen preservation. Soc Reprod Fertil Suppl. 2007, 65:217-28.
Manjunath P, Thérien I. Role of seminal plasma phospholipid-binding proteins in sperm membrane lipid modification that occurs during capacitation. J Reprod Immunol. 2002, 53:109-19.
Mann T, Lutwak-Mann C. Male reproductive function and semen. Berlin/Heidelberg/New York: Springer Verlag. 1981
Martínez-Heredia J, Estanyol JM, Ballescà JL, Oliva R. Proteomic identification of human sperm proteins. Proteomics. 2006, 6:4356-69.
Martínez-León E, Osycka-Salut C, Signorelli J, Pozo P, Pérez B, Kong M, Morales P, Pérez-Martínez S, Díaz ES. Fibronectin stimulates human sperm capacitation through the cyclic AMP/protein kinase A pathway. Hum Reprod. 2015, 30:2138-51.
Marx H, Hahne H, Ulbrich SE, Schnieke A, Rottmann O, Frishman D, Kuster B. Annotation of the domestic pig genome by quantitative proteogenomics. J Proteome Res. 2017, 16:2887-2898.

Mikhailov AT, Torrado M. Carboxylesterase overexpression in the male reproductive tract: a universal safeguarding mechanism? Reprod Fertil Dev. 1999, 11:133-45.

Mohanty G, Swain N, Samanta L Sperm Proteome: What is on the horizon? Reprod Sci. 2015, 22:638-53.

Murphy-Ullrich JE, Iozzo RV. Thrombospondins in physiology and disease: new tricks for old dogs. Matrix Biol. 2012, 31:152-4.

Novak S, Ruiz-Sánchez A, Dixon WT, Foxcroft GR, Dyck MK. Seminal plasma proteins as potential markers of relative fertility in boars. J Androl. 2010, 31:188-200.

Okamura N, Kiuchi S, Tamba M, Kashima T, Hiramoto S, Baba T, Dacheux F, Dacheux JL, Sugita Y, Jin YZ. A porcine homolog of the major secretory protein of human epididymis, HE1, specifically binds cholesterol. Biochim Biophys Acta. 1999, 1438:377-87.

Park YJ, Kim J, You YA, Pang MG. Proteomic revolution to improve tools for evaluating male fertility in animals. J Proteome Res. 2013, 12, 4738-47.

Perez-Patiño C, Barranco I, Parrilla I, Valero ML, Martinez EA, Rodriguez-Martinez H, Roca J. Characterization of the porcine seminal plasma proteome comparing ejaculate portions. J Proteomics. 2016, 142:15-23.

Pérez-Patiño C, Parrilla I, Barranco I, Vergara-Barberán M, Simó-Alfonso EF, Herrero-Martínez JM, Rodriguez-Martínez H, Martínez EA, Roca J. New in-depth analytical approach of the porcine seminal plasma proteome reveals potential fertility biomarkers. J Proteome Res. 2018, 17:1065-1076.

Pini T, Leahy T, Soleilhavoup C, Tsikis G, Labas V, Combes-Soia L, Harichaux G, Rickard JP, Druart X, de Graaf SP. Proteomic investigation of ram spermatozoa and the proteins conferred by seminal plasma. J Proteome Res. 2016, 15:3700-3711.

Pinke LA, Swanlund DJ, Hensleigh HC, McCarthy JB, Roberts KP, Pryor JL. Analysis of fibronectin on human sperm. J Urol. 1997, 158:936-41.

Roca J, Parrilla I, Gil MA, Cuello C, Martinez EA, Rodriguez-Martinez H. Non-viable sperm in the ejaculate: Lethal escorts for contemporary viable sperm. Anim Reprod Sci. 2016, 169:24-31.

Rodríguez-Martínez H. State of the art in farm animal sperm evaluation. Reprod Fertil. 2007, 19:91-101.

Rodríguez-Martínez H, Kvist U, Saravia F, Wallgren M, Johannisson A, Sanz L, Peña FJ, Martínez EA, Roca J, Vázquez JM, Calvete JJ. The physiological roles of the boar ejaculate. Soc Reprod Fertil Suppl. 2009, 66:1-21.

Saravia P, Wallgren M, Johannisson A, Calvete JJ, Sanz L, Peña FJ, Roca J, Rodríguez-Martínez H. Exposure to the seminal plasma of different portions of the boar ejaculate modulates the survival of spermatozoa cryopreserved in MiniFlatPacks. Theriogenology. 2009, 71:662-75.

Strzezek, J. Secretory activity of boar seminal vesicle glands. Reprod Biol. 2002, 2:243-66.

Swegen A, Curry BJ, Gibb Z, Lambourne SR, Smith ND, Aitken RJ. Investigation of the stallion sperm proteome by mass spectrometry. Reproduction. 2015, 149:235−244.

Töpfer-Petersen E, Romero A, Varela PF, Ekhlasi-Hundrieser M, Dostàlovà Z, Sanz L, Calvete JJ. Spermadhesins: a new protein family. Facts, hypotheses and perspectives. Andrologia. 1998, 30:217-24.

Vilagran I, Yeste M, Sancho S, Castillo J, Oliva R, Bonet S. Comparative analysis of boar seminal plasma proteome from different freezability ejaculates and identification of Fibronectin 1 as sperm freezability marker. Andrology. 2015, 3:345-56.

Wiese S, Reidegeld KA, Meyer HE, Warscheid B. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics. 2007, 7:340-50.

STUDY 5

This manuscript has been submitted to for its publication to the **Scientific Reports** journal with the title "**THE PROTEOME OF FROZEN-THAWED PIG SPERMATOZOA IS DEPENDENT ON EJACULATE FRACTION SOURCE**"

Abstract

The quality, including functional parameters and fertility, of frozen-thawed (**FT**) spermatozoa is variable in porcine, a species with fractionated ejaculates. Although intrinsic individual differences have primarily been linked to this variation; differences in protein expression among FT-sperm ought to be far more relevant. This study looked for differences in protein expression between FT-sperm samples differing in quality (post-thaw viability, total and progressive motility, early signs of apoptosis, membrane lipid peroxidation and nuclear DNA fragmentation) and derived from the sperm rich ejaculate fraction (**SRF**) vs the entire ejaculate (Experiment 1) or from the first 10 mL of SRF, the rest of SRF and the post-SRF (Experiment 2). Quantitative sperm proteome differences were analyzed using a liquid chromatography-electrospray ionization-tandem mass spectrometry (**LC-ESI-MS/MS**)-based sequential window acquisition of all theoretical mass spectra (**SWATH**) approach. In Experiment 1, FT-sperm from the SRF showed better quality than those from the entire ejaculate and 26 *Sus scrofa* proteins with functional sperm relevance were differentially expressed (FC ≥ 1.5) between FT-sperm of both sources. In Experiment 2, FT-sperm from the first 10 mL of SRF and the rest of SRF were qualitatively better than those from the post-SRF, with 187 proteins differentially expressed among the three ejaculate sources. The results clearly indicate that proteome expression is linked to FT-sperm quality.

Keywords: proteomics, frozen-thawed, spermatozoa, ejaculate, porcine

VIII. Study 5

8.1 Introduction

Improving fertility outcomes of frozen-thawed (**FT**) spermatozoa remains a pending challenge for some livestock species, including porcine (Yeste et al., 2017). In spite of the valuable progress achieved when cryopreserving boar spermatozoa in the past recent years (Yeste, 2015), parameters defining post-thaw sperm quality attributes are still variable thus affecting fertility outcomes of FT-spermatozoa, which remain considerably lower than those achieved by liquid stored semen (Knox, 2015; Yeste, 2016). This status of variable cryosurvival, which impairs the efficient inclusion of FT-spermatozoa in swine commercial artificial insemination (**AI**)-programs (Roca et al., 2016), is not exclusive of porcine since it also occurs in other livestock species, such as ovine (Pini et al., 2018), and even in humans (Marlea et al., 2012). However, it is especially relevant in porcine commercial husbandry, where the magnitude of such variability leads to classify the AI-boars as good or bad sperm freezers (Thurston et al., 2002), impairing the efficient AI-usage of some genetically superior boars.

The usual variability in sperm cryosurvival among boars and ejaculates within boar are compensated in the production of AI doses by matching the numbers of cryosurviving FT-spermatozoa to those of the viable sperm used in AI-doses of liquid stored semen. However, this implies until a 4-fold increase of the total number of FT-sperm per AI-dose (Roca et al., 2016). Moreover, this accommodating and inefficient practice does not match the fertility of FT-spermatozoa to that of the liquid stored semen, but even does not minimize the variability between boars/ejaculates in the AI-fertility outcomes of FT-spermatozoa (Bolarin et al., 2006). Roca et al. (2013) noted that *in vitro* fertility outcomes of FT-boar spermatozoa from semen samples showing bad sperm freezability were lower than those showing good sperm freezability, even after inseminations with a similar number of cryosurvived spermatozoa. This background suggests that there may be putative differences in molecular arrangement affecting fertilizing capacity between the FT-spermatozoa from semen samples differing in sperm freezability. Since proteins are involved in the most critical sperm functions, including fertilizing ability (Gadella and Boerke, 2016), the present study tries to clarify this issue by analyzing the proteome of FT-spermatozoa with documented freezability. It is currently known that cryopreservation remodels the proteome of boar spermatozoa (Chen et al., 2014), but it

is unknown that there may be proteome differences between FT-spermatozoa which could influence their fertility.

The boar ejaculate is emitted in fractions, being the so-called sperm-rich fraction (**SRF**) and the post-SRF the two main fractions (Rodriguez-Martinez et al., 2009). Currently, swine AI-centers to retrieve spermatozoa for AI are moving from selectively collecting the SRF, by using the gloved-hand method, to collecting the entire ejaculate (**EE**), using semi-automatic methods (Roca et al., 2016). Surprisingly, this change in the methodology for ejaculate collection, mainly motivated by sanitary and labor cost reasons, not only does not advance animal welfare, but particularly negatively affects cryosurvival, since post-thawing functional attributes of spermatozoa from the EE are worse than those from the SRF (Alkmin et al., 2014; Li et al., 2018a). Therefore, the current study was split in two experiments to compare the proteome of FT-spermatozoa derived from the SRF and the EE (Experiment 1) or between ejaculate fractions showing clear differences in sperm freezability (Saravia et al., 2009; Li et al., 2018a), specifically, the first 10 mL of the SRF, the rest of SRF and the post-SRF (Experiment 2).

8.2 Material and methods

8.2.1 Boars, ejaculates and sperm sources

All procedures involving boars and semen samples were performed following international guidelines (Directive 2010/63/EU) and were approved by the Bioethics Committee of the University of Murcia (research code: 639/2012). The chemicals used for elaborate semen extenders and the fluorescent probes used for sperm evaluation were purchased from Sigma-Aldrich Co. (St. Louis, MO, USA).

Five healthy, sexually mature and fertility proven boars (two of Large White, two of Duroc and one of Landrace breeds) were used as semen donors. The boars were housed in a commercial AI-center (Topigs Norsvin España, Madrid, Spain) and subjected to regular ejaculate collection (two ejaculates per week). A total of 15 ejaculates (three per boar) were manually collected using the gloved-hand method. All the ejaculates meet the minimum requirements for the preparation of insemination doses (200 x 10^6 spermatozoa/mL, 75% of motility and 80% of normal morphology). The ejaculates were collected in three separate fractions, specifically the first 10 mL of SRF, the rest of SRF

and the post-SRF. Volumes of the first 10 mL of SRF and the rest of SRF were proportionally mixed to generate semen samples simulating entire SRFs. Similarly, volumes of three ejaculate fractions were mixed for generating semen samples simulating EE. The five resulting semen samples of each ejaculate were extended (1:1, vol:vol) in Beltsville Thawing Solution (**BTS**), cooled at 17 ºC, placed into a styrofoam box and transported to the Andrology Laboratory of the Veterinary Teaching Hospital (**VTH**) of the University of Murcia (trip < 2 h). Once in the VHT, the BTS-extended semen samples were stored overnight at 17 ºC to be frozen the next morning.

8.2.2 Semen cryopreservation and post-thaw sperm evaluation

Semen samples were frozen using the straw freezing protocol described by Alkmin et al. (2014). Briefly, semen samples were centrifuged (2,400 *xg* during 3 min; Megafuge 1.0 R, Heraeus, Hanau, Germany) and the resulting sperm pellets extended to 1.5 x 10^9 spermatozoa/mL in a Tris-citric acid-glucose extender supplemented with egg yolk (80:20, vol:vol). Then, the extended spermatozoa were cooled to 5 ºC and re-extended to 1.0 x 10^9 spermatozoa/mL in the same extender supplemented with 9% of glycerol and 1.5% of Equex STM (v:v) (Nova Chemical Sales, Scituate, MA, USA). The spermatozoa were then packed into 0.5 mL polyvinyl chloride French straws (Minitüb, Tiefenbach, Germany) and frozen by placing them on a metal rack 3 cm above liquid nitrogen (**LN_2**) surface for 20 min (Freezing unit, Minitüb). After being stored into a LN_2-tank (GT40, Air Liquide, Paris, France) for at least a week, the straws (two per sperm sample/source) were thawed in a circulating water bath at 37 ºC for 20 s, the contents extended in BTS (1:1, vol:vol) and incubated at 37 ºC during 30 min before sperm analyses.

Post-thaw total and progressive sperm motility was evaluated using a computer-assisted sperm analyzer (ISASV1® CASA, Proiser R+D, Paterna, Spain) and sperm viability using a BD FACS Canto II flow cytometer (Becton Dickinson & Company, Franklin Lakes, NJ, USA) after labeling the spermatozoa (100-µL with 3 x10^6 spermatozoa) with 3 µL Hoechst 33342 (**H-42**; 0.05 mg/mL in PBS), 2 µL propidium iodide (**PI**, 0.5 mg/mL in PBS), and 2 µL fluorescein-conjugated peanut agglutinin (**PNA-FITC**, 100 µg/mL in PBS). Viable spermatozoa were those H-42 positive, PI negative and PNA-FITC negative. Protocols are described at length in Li et al. (2018a).

Further functional attributes of cryosurvived spermatozoa were cytometrically evaluated; as early events of apoptosis and membrane lipid peroxidation (100-µL with 2 x10^6

spermatozoa labelled with 2 µL H-42 and 10 µL of PI) using the fluorescent probes Annexin V-FITC (3 µL) and BODIPY 581/591 C11 (2.5 µL), respectively (Li et al., 2018b). Viable spermatozoa (H-42 positive and PI negative) showing either early signs of apoptosis (Annexin V-FITC positive) or membrane lipid peroxidation (BODIPY positive) were recorded. Lastly, nuclear DNA fragmentation was assessed using the Sperm-Sus-Halomax® kit (Halotech DNA SL, Madrid, Spain) following the method reported by Alkmin et al. (2013).

8.2.3 Sperm proteomics

The proteomics analyses were carried out in the Proteomics Unit of the University of Valencia, Valencia, Spain (member of the PRB2-ISCIII ProteoRed Proteomics Platform).

8.2.3.1 Preparation of sperm samples

Three straws per each one of the 75 sperm sample/source (five semen samples of each one of the three ejaculates from each of the five boars) were thawed a 37 ºC and their content centrifuged at 600 xg during 20 min (Megafuge 1.0 R, Heraeus) on a Percoll monolayer gradient (45% in PBS [v:v], Percoll® P4937; Sigma-Aldrich Co.) to separate the spermatozoa from other putative cells and debris, including egg-yolk remnants. The sperm pellets were extended in PBS (1:10, v:v) and centrifuged again (300 *xg*, 10 min). Then, the resulting sperm pellets were extended in PBS to a concentration of 1,000 x10^6 sperm/mL, stored into cryotubes (2 mL Cryogenic vial, Fisher Scientific, Madrid, Spain) that were preserved at -80 ºC (Ultra Low Freezer; Haier, Schomberg, Ontorio, Canada) until protein analysis.

At time of proteomics analysis, sperm samples were thawed at room temperature and the fifteen sperm samples (three from each of the five boars) of each one of the five semen sources (first 10 mL of SRF, rest of SRF, entire SRF, post-SRF and EE) were pooled; generating five single sperm pools (one pool per source). Each sperm pool was in turn split into three aliquots generating three technical replicates per sperm sample. Thus, a total of 15 sperm samples were analyzed.

8.2.3.2 Protein extraction

Sperm samples were centrifuged at 14,000 *xg* for 10 min (Eppendorf 5424R, Eppendorf AG, Hamburg, Germany) and the supernatant was discarded. The total protein from the

resulting sperm pellets were extracted using 200 µL of UTC buffer [7 M Urea, 2 M thiourea, 4% 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate, (**CHAPS**)], supplemented with protease inhibitor cocktail (1%, v:v), and a vigorous stirring (one hour at 5 °C). The concentration of extracted protein was quantified by RC_DC Lowry (Bio-Rad, Richmond, CA, USA) following the manufacturer instructions. Then, a total of 30 µg of final protein extract was taken for one dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (**1D SDS-PAGE**) loading, to remove the remaining UTC and other interferences for tandem mass spectrometry (**MS/MS**) analysis.

8.2.3.3 SWATH analysis

8.2.3.3.1 In-gel digestion processing

The 1D SDS-PAGE portion containing proteins was digested at 37 °C using 500 ng of sequencing Grade Trypsin (V511, Promega Co., Madison, WI, USA) following the protocol used by Shevchenko et al. (1996). The trypsin digestion was stopped with 10% trifluoroacetic (**TFA**) and the supernatant, containing the non-extracted digests, was carefully removed, leaving behind the sliced gels in the Eppendorf tube. For peptide extraction, 200 µL of pure acetonitrile (**ACN**) were then added, followed by incubation at 37 °C in a shaker for 15 min. The new supernatant containing the peptide mixture was carefully withdrawn and dried in a speed vacuum (ISS 110 SpeedVac System, Thermo Savant, ThermoScientific, Langenselbold, Germany) for 20 min, and then re-suspended in 25 µL of 2% ACN and 0.1% TFA.

8.2.3.3.2 Liquid chromatography and tandem mass spectrometry (LC-MS/MS) analysis

For the spectral library acquisition, a mixture of all digested samples (2 µL of each sample) were examined by liquid chromatography (**LC**) using a NanoLC Ultra 1D plus Eksigent (Eksigent Technologies, Dublin, CA, USA), which was directly connected to an AB SCIEX TripleTOF 5600 mass spectrometer (AB SCIEX, Framingham, MA, USA) in direct injection mode. Five µL of the digested sample were loaded on a trap NanoLC pre-column (3 µm particles size C18-CL, 350 µm diameter x 0.5 mm long; Eksigent Technologies) and desalted with 0.1% TFA at 3µL/min during 5 min. Then, the peptides were separated using an analytical LC column (3 µm particles size C18-CL, 75 µm diameter x 12 cm long, Nikkyo Technos Co®, Tokyo, Japan) equilibrated in 5% ACN

0.1% formic acid (**FA**) (Fisher Scientific). Peptide elution was performed by applying a linear gradient from 5% to 35% of ACN containing 0.1% FA at a constant flow rate of 300 nL/min over 180 min.

The TripleTOF was operated in data-dependent mode, in which a time of flight (**TOF**) MS scan was made from 350 to 1250 m/z, accumulated for 250 ms TOF followed by 150 ms TOF with the same scan range for MS, and the 25 most abundant multiply charged (2+, 3+, 4+ or 5+) precursor peptide ions automatically selected. Ions with 1+ and unassigned charge states were rejected from the MS/MS analysis. The rolling collision energies equations were automatically set by the instrument according to the equation |CE|=(slope)x(m/z)+(intercept) with Charge=2; Slope= 0.0575 and Intercept=9.

8.2.3.4 LC-SWATH-MS acquisition

To determine quantitative differences between SRF and EE and among the first 10 mL of SRF, the rest of SRF and the post-SRF, the sequential window acquisition of all theoretical spectra (**SWATH**) analysis of individual sperm samples was performed following the procedure described by Perez-Patiño et al. (2016) tuning the TripleTOF 5600 (AB SCIEX) as described by Gillet et al. (2012) for SWATH-MS-based experiments. In this way, 5 µL of one of the three technical replicates from each sample were randomly loaded onto a trap column (NanoLC Column, 3 µm C18-CL, 75 µm x 15 cm; Eksigent Technologies) and flushed for 5 min with 0.1% TFA at 3µL/min to remove salts. Peptide separation was achieved using an analytical column (LC Column, 3 µm C18-CL, 75 µm x 12 cm, Nikkyo Technos Co®) equilibrated in 5% ACN 0.1% FA, eluted with a linear gradient from 5 to 40% ACN during 90 min at a flow rate of 300 nL/min. The analysis of eluted peptides was carried out in the spectrometer nanoESI qQTOF (SCIEX TripleTOF 5600) and the TripleTOF operated in SWATH mode, in which a 0.050 s TOF MS scan from 350 to 1250 m/z was performed, followed by 0.080 s product ion scans from 230 to 1800 m/z on the 37 defined windows (3.05 sec/cycle). The collision energy for each window was calculated for 2+ charged ion at the center of each SWATH block with a collision energy spread of 15 eV. The MS was always operated in high sensitivity mode.

8.2.4 Protein identification, validation and quantification

After the LC-MS/MS, the resulting SCIEX.wiff data-files were processed by ProteinPilot v5.0 search engine (AB SCIEX) and the Paragon algorithm (4.0.0.0, 4767) was used to search against the Uniprot_mammals database with the following parameters: trypsin specificity, cys-alkylation and the search effort set to through. To avoid using the same spectral evidence for 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 wiff files obtained from SWATH experiment were analyzed by PeakView® (v2.1, AB SCIEX) and peaks from SWATH were extracted with a peptide confidence threshold of 95% confidence. A false discovery rate (**FDR**) less than 1% and 6 transitions per peptide were required to quantify one peptide. The extracted ions chromatograms were integrated and the areas under the curve (**AUC**) used to calculate total protein. A normalization of the calculated AUCs was done by total sum, and the sum of all areas was equalized for all the samples.

8.2.5 Gene ontology and bioinformatics analysis

The bioinformatics of differentially expressed sperm-proteins was manually annotated by using two comprehensive bioinformatic tools, specifically UniProt KB and DAVID. The UniProt KB database (www.uniprot.org), downloaded 19/04/2018, contains 115,678,811 total entries with 47,499 of them encoded in *Sus scrofa* taxonomy. The DAVID Database for Annotation, Visualization and Integrated Discovery (DAVID Bioinformatics Resources 6.8; https://david.ncifcrf.gov/; [Huang et al., 2009]) integrates numerous public sources of protein annotation and, consequently, contains information of more than 1.5 million genes from more than 65,000 species.

8.2.6 Statistical analysis

A mixed ANOVA test (IBM SPSS v24.0 software, IBM Spain, Madrid), using ejaculate as random effect, was performed to evaluate the influence of the semen source (entire SRF vs EE and among the first 10 mL of the SRF, rest of the SRF and the post-SRF) and the boar (n=5) on the post-thaw sperm attributes. In proteomics, the quantitative data obtained by PeakView® were analyzed using MarkerView® (v1.2, AB SCIEX) and peak areas were normalized by the sum of peak areas of all identified peptides. Principal Component Analysis (**PCA**) was performed to evaluate the discriminative ability of FT-sperm proteins between semen sources examined in Experiment 1 and 2. Then, the

Multiexperiment Viewer (**MeV**) software (version 4.8) (http://www.tm4.org/mev.html) was used to identify quantitative differently expressed FT-sperm proteins. A Student's t-test and ANOVA test were used to identify the sperm protein differentially expressed between the FT-sperm sources of Experiment 1 and 2, respectively. Proteins were considered differentially expressed with an adjusted P-value < 0.01, and those with a fold change (**FC**) ≥ 1.50 after log2 transformation were highlighted. The explanatory ability of resulting selected proteins in both experiments was illustrated by mean of heat maps after z-score normalization, using Euclidean distances.

8.3 Results

8.3.1 Sperm proteome profile

A total of 93,457 spectra corresponding to 16,777 distinct peptides and 1,157 proteins were identified when assuming a FDR ≤ 1% at protein level. Of the latter, 673 belonged to *Sus scrofa* taxonomy. The complete list of the 1,157 sperm-proteins identified, including their Unused Score, UniProt Accession number, Protein Name, Species, % of Sequence Coverage and Matched Peptides is provided in **Supplementary Table 16**. The SWATH approach allowed the quantification of 1,094 sperm-proteins, 670 of them belonging to *Sus scrofa* taxonomy (**Supplementary Table 17**). All the quantified proteins were present in FT-spermatozoa of the different sources evaluated, specifically EE, entire SRF, first 10 mL of SRF, rest of SRF and post-SRF.

8.3.2 Differentially expressed proteins between FT-sperm of the different semen sources

The results of the comparison between entire SRF and EE and of those comparing the different portions of the ejaculate (first 10 mL of the SRF, rest of the SRF and post-SRF) are shown separately for better clarity. In addition, the results of the post-thaw assessment of sperm attributes are also included.

8.3.2.1 Differences between FT-spermatozoa retrieved from SRF or EE

Concerning post-thaw sperm attributes, FT-sperm source and boar influenced (P<0.01) all sperm parameters evaluated 30 min after thawing, except for proportion of fragmented nuclear DNA. The interaction between FT-sperm source and boar was not significant for any of the sperm attributes evaluated. Therefore, the data of the five boars were averaged

for each FT-sperm source. The sum of sperm attributes defining FT-sperm quality was better (P<0.001) among spermatozoa from the SRF source than those from the EE (**Table 8**).

Table 8 Attributes (as percentage mean±SEM) of frozen-thawed boar spermatozoa derived from the sperm rich ejaculate fraction (SRF) or the entire ejaculate (EE) (15 ejaculates from 5 boars).

Sperm attributes (%)	**Sperm source**	
	SRF	EE
Total motility	49.02±1.79[a]	36.13±1.16[b]
Progressive motility	40.93±1.55[a]	31.87±1.83[b]
Viability	54.05±1.65[a]	41.53±1.38[b]
Viable sperm with early apoptosis signs	4.17±0.35[a]	10.74±0.84[b]
Viable sperm with lipid peroxidation	4.76±0.27[a]	12.00±0.80[b]
Sperm with fragmented nuclear DNA	1.64±0.24	2.07±0.25

[a,b]P≤0.001

Concerning the total of FT-sperm proteins quantified, the two first components of the PCA explained 94.4% of total variance and the PC1 (explaining 85.2% of total variance) clearly discriminated between the three replicates of the SRF from those of the EE (**Fig. 15**). A total of 34 proteins belonging to *Sus scrofa* taxonomy were differentially expressed (P<0.01) between the FT-sperm from the SRF and those from the EE. The quantitative value of the differentially expressed proteins, after data normalization for each one of the two FT-sperm sources and the FC estimation of the groups after log2 transformation, is shown in **Supplementary Table 18**. Twenty-six of these 34 differentially expressed proteins showed a FC ≥ 1.50 and 11 of them were over-expressed in the FT-spermatozoa retrieved from the SRF, whereas the other 15 were in those derived from the EE (**Table 9**). The heat-map of the **Fig. 16** clearly evidenced that the three technical replicates of each FT-sperm source were grouped in a same cluster merged at a short distance; evidenced further by the dendrogram with a large distance between the cluster of SRF and that of EE.

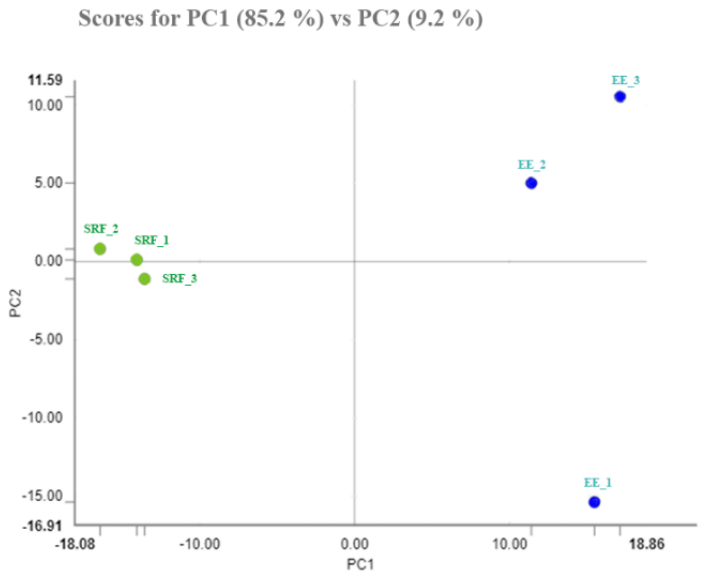


Fig. 15 Principal Component Analysis (PCA) relative to the proteins quantified in frozen-thawed spermatozoa retrieved from the sperm rich fraction (SRF) of the ejaculate or from the entire ejaculate (EE). The points represent the three technical replicates for each sperm source and are based on the relative amounts quantified in each one of them.

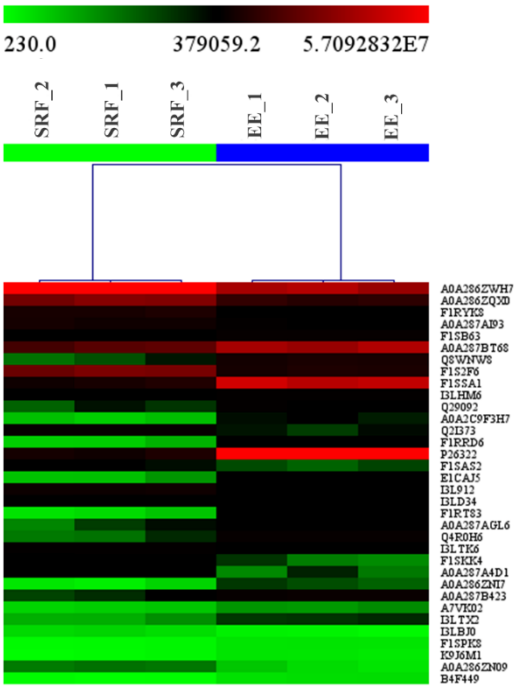


Fig. 16 Heat-map with dendrograms representing the *Sus scrofa* proteins differentially expressed by frozen-thawed spermatozoa derived from either the sperm-rich ejaculate fraction (SRF) or the entire ejaculate (EE). The hierarchical clustering tree of sperm samples is shown at the top. The relative expression level of each differentially expressed protein is shown on a colour scale from green (lowest level) to red (highest level).

Table 9 *Sus scrofa* proteins differentially expressed with a fold change (FC) ≥ 1.50 between frozen-thawed boar spermatozoa retrieved from the sperm-rich ejaculate fraction (SRF) or from the entire ejaculate (EE).

PROTEIN NAME	ACCESSION	FC
A-kinase anchoring protein 4	A0A286ZWH7	**3.46**
Fibrous sheath interacting protein 2	F1RYK8	**2.43**
Uncharacterized protein	A0A287AI93	**1.83**
Nexin-1	Q8WNW8	**-4.49**
Voltage-dependent anion-selective channel protein 2	F1S2F6	**2.01**
Chromosome 1 open reading frame 56	F1SSA1	**-2.84**
Endoplasmin	Q29092	**-2.38**
Dipeptidyl peptidase 4	A0A2C9F3H7	**-2.07**
Fascin	Q2I373	**2.90**
Programmed cell death 6 interacting protein	F1RRD6	**-2.93**
Carbohydrate-binding protein AQN-1	P26322	**-3.80**
Protein disulfide-isomerase	E1CAJ5	**-2.58**
Family with sequence similarity 205 member A	I3L912	**1.83**
Uncharacterized protein	F1RT83	**-4.20**
Radial spoke head 6 homolog A	A0A287AGL6	**-1.54**
Spermadhesin PSP-I	Q4R0H6	**-3.34**
Uncharacterized protein	I3LTK6	**1.68**
Transmembrane protein 89	F1SKK4	**1.91**
IQ motif containing F5	A0A287A4D1	**1.73**
Protease, serine 8	A0A286ZNI7	**-2.69**
Uncharacterized protein	A0A287B423	**-2.99**
Leucine rich repeat and coiled-coil centrosomal protein 1	I3LBJ0	**1.83**
Tetraspanin CD63	F1SPK8	**-1.88**
Ras GTPase-activating-like protein IQGAP2	K9J6M1	**-2.16**
Uncharacterized protein	A0A286ZN09	**1.82**
Complement factor D preproprotein	B4F449	**-3.18**

With respect to bioinformatics, specifically focused to categories related to sperm and reproductive functions, most differentially expressed proteins were closely related to cell membrane (20%, 10 proteins), energy metabolism (14%, 7), cell motility (12%, 6), and reproduction (12%, 6). The six differentially expressed proteins framed into reproductive function were equally distributed between spermatogenesis, sperm capacitation and fertilization (**Fig. 17**). Noteworthy, some of the differentially expressed proteins were related to immune response (10%, 5), apoptosis (6%, 3), stress response (6%, 3) and ADN (2%, 1).

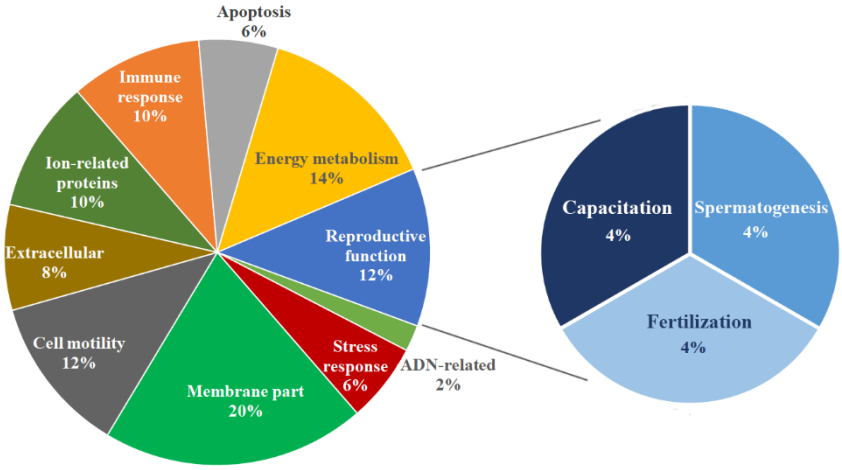


Fig. 17 Functional distribution of *Sus scrofa* sperm proteins into the categories sperm and reproductive functions available in the UniProtKB/Swiss-Prot Web site (www.uniprot.org) and DAVID Bioinformatics Resources 6.8 (https://david.ncifcrf.gov).

8.3.2.2 Differences among FT-spermatozoa retrieved from three specific ejaculate portions

Concerning post-thaw sperm attributes, FT-sperm source and boar influenced ($P<0.01$) all of them, except for fragmented nuclear DNA. The interaction between FT-sperm source and boar was not significant for any of the sperm attributes evaluated. Therefore, the data of the five boars were averaged for each FT-sperm source. The sum of sperm attributes defining sperm quality was better ($P<0.01$) among FT-spermatozoa from the first 10 mL of SRF and the rest of SRF than those from the post-SRF (**Table 10**).

VIII. Study 5

Table 10 Attributes (percentage mean±SEM) of frozen-thawed boar spermatozoa retrieved from three clearly identifiable ejaculate fractions (the first 10 mL of sperm-rich ejaculate fraction [SRF], the rest of SRF and the post-SRF; 15 ejaculates from 5 boars).

Sperm attributes (%)	Sperm source		
	First 10 ml of SRF	Rest of SRF	Post SRF
Total motility	53.67±1.98[a]	47.40±2.35[a]	20.87±1.50[b]
Progressive motility	44.33±1,74[a]	40.87±2.27[a]	15.73±0.84[b]
Viability	58.50±1.99[a]	52.55±2.54[a]	26.09±1.88[b]
Viable sperm with early apoptosis signs	4.26±0.47[a]	4.69±0.44[a]	19.32±1.36[b]
Viable sperm with lipid peroxidation	3.73±0.36[a]	4.69±0.70[a]	16.63±1.27[b]
Sperm with fragmented nuclear DNA	1.62±0.24	1.82±0.25	1.78±0.22

[a,b]P≤0.01

Regarding the total of the FT-sperm proteins quantified, the two first components of the PCA explained 92.1% of total variance and the PC1 (explaining 58.3% of total variance) is the one discriminating best among the three ejaculate portions/sources (**Fig. 18**). Accordingly, the FT-sperm samples were grouped in two separate branches: the first one includes those from the first 10 mL of SRF and the rest of SRF and the second one those from the post-SRF. A total of 257 proteins belonging to *Sus scrofa* taxonomy were differentially expressed (P<0.01) between the FT-spermatozoa derived from the three ejaculate portions/sources. The quantitative value of the differentially expressed proteins, following data normalization for each one of the three FT-sperm sources and the FC estimation of the groups after log2 transformation, is shown in **Supplementary Table 19**. A total of 187 of the differentially expressed proteins showed a FC ≥ 1.50 (**Supplementary Table 20**), most of them differing between FT-sperm from the post-SRF and those from the two-other ejaculate portion/sources. Specifically, 173 and 165 proteins differed between the FT-sperm from the post-SRF and those from the first 10 mL of SRF and rest of SRF, respectively. The FT-spermatozoa from the post-SRF source overexpressed 90 and 103 proteins with those from the first 10 mL of the SRF and the rest of SRF, respectively. The heat-map in **Fig. 19** evidenced that the three technical replicates of FT-spermatozoa from the post-SRF were grouped in a same cluster merged close together being clearly separated than those of the FT-spermatozoa derived from both the first 10 mL of SRF or the rest of SRF, which were grouped in a same cluster.

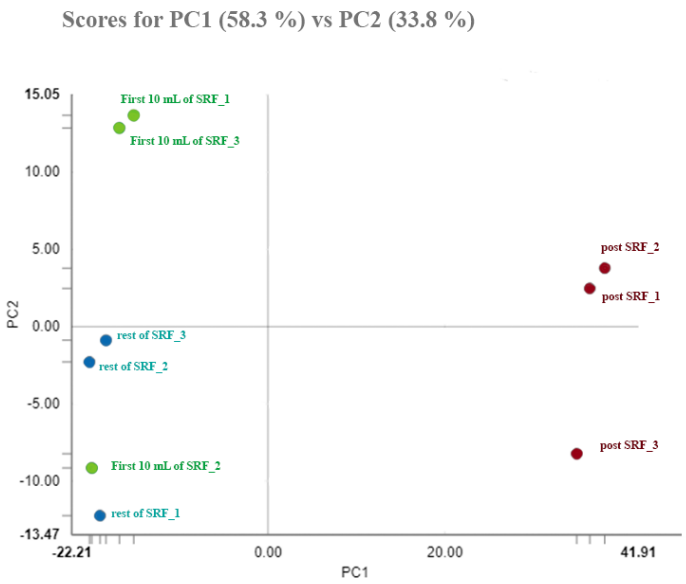


Fig. 18 Principal Component Analysis (PCA) of proteins quantified in frozen-thawed spermatozoa retrieved from three well identifiable portions of the pig ejaculate (the 10 first mL of the sperm-rich ejaculate fraction (SRF), the rest of SRF or the post-SRF). The points represent the three technical replicates for each sperm source and are based on the relative amounts quantified in each one of them.

With respect to bioinformatics, specifically focused to categories related to sperm and reproductive functions, most differentially expressed proteins were related to energy metabolism (22%, 66 proteins), ions (14%, 43), stress response (11%, 35), and reproduction (11%, 33). Most of the differentially expressed proteins framed into reproductive function were related to fertilization (4%, 12), spermatogenesis (3%, 11) and sperm capacitation (2%, 6) (**Fig. 20**). It is remarkable that some of the differentially expressed proteins were also related to cell membrane (10%, 31), lipid metabolism (9%, 27), cell motility (5%, 16) and ADN (5%, 15).

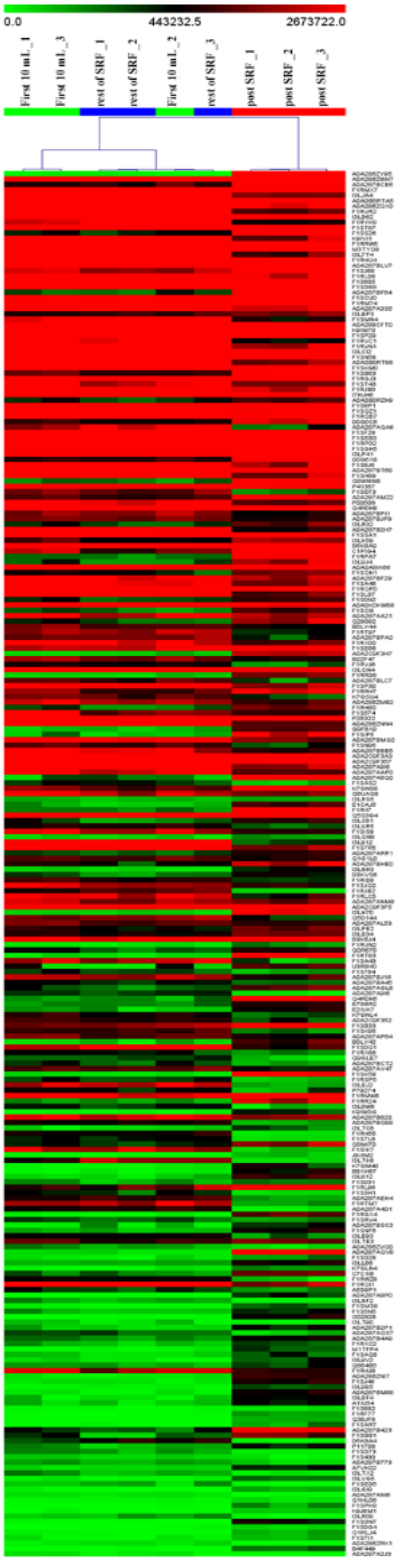


Fig. 19 Heat-map with dendrograms representing the *Sus scrofa* proteins differentially expressed among frozen-thawed spermatozoa from three well identifiable portions of the pig ejaculate (the 10 first mL of the sperm-rich ejaculate fraction (SRF), the rest of SRF or the post-SRF). The hierarchical clustering tree of sperm samples is shown at the top. The relative expression level of each differentially expressed protein is shown on a colour scale from green (lowest level) to red (highest level).

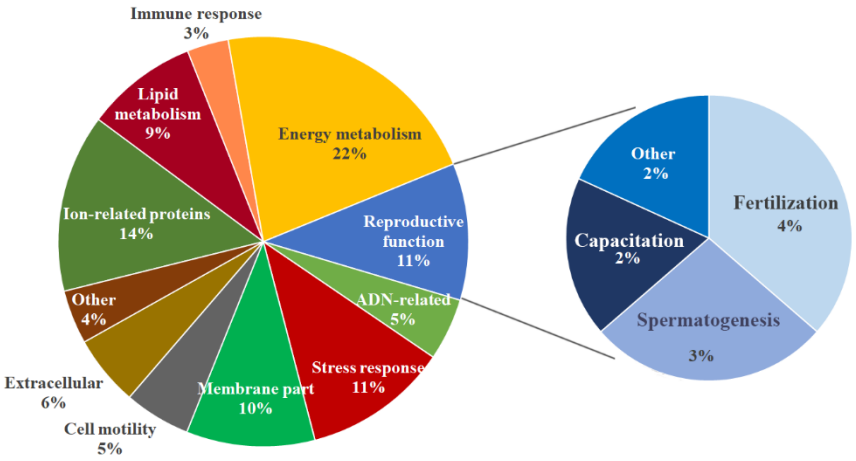


Fig. 20 Functional distribution of *Sus scrofa* sperm proteins into the categories sperm and reproductive functions available in the UniProtKB/Swiss-Prot Web site (www.uniprot.org) and DAVID Bioinformatics Resources 6.8 (https://david.ncifcrf.gov).

8.4 Discussion

The results of the present study show quantitative differences of proteins with relevance for sperm function in FT-spermatozoa showing differences in freezability, retrieved from different sources, namely various portions of the pig ejaculate. The results highlighted that the spermatozoa derived from the EE would be less functional post-thaw than those retrieved from the SRF because of its differential protein expression, a matter which could seriously impair its fertilizing capacity as we will try to argument for next in this discussion.

The results of post-thaw sperm quality confirm previous reports, demonstrating that boar spermatozoa from the SRF cryosurvive better than those retrieved from the EE (Alkmin et al., 2014; Li et al. 2018a). Remarkably, cryosurviving spermatozoa derived from the EE showed more evident dysfunctional signs than those derived from the SRF, as they shown worse quality of sperm movement and higher rates of apoptosis and membrane lipid peroxidation. Molecular derangements, involving proteins, are probably underlying these existing differences in freezability. Subtle changes in protein expression may

compromise the reproductive performance of spermatozoa since they are involved in membrane remodeling, capacitation, oocyte zona binding, acrosome reaction and fusion to the oolemma (Gadella, 2008). Moreover, sperm proteins, together with those of the seminal plasma, also play a critical role in the immunological reaction of internal sow genital tract tissues after insemination (Rozeboom et al., 1999).

A total of 26 proteins encoded into *Sus scrofa* taxonomy were differentially expressed with a FC ≥1.50 between FT-spermatozoa retrieved from the SRF and those derived from the EE; and proteins belonging to functional groups encoded in the proteome databases as related to sperm function. Eleven of these proteins, including two still uncharacterized, were overexpressed in FT-spermatozoa from the SRF, the ones showing best sperm attributes post-thaw. The proteins were directly involved in sperm motility and their overexpression in the FT-spermatozoa from the SRF would explain why they have better quality of movement than those retrieved from the EE. Specifically, A-kinase anchoring protein (AKAP) -4, the most abundant cytoskeletal glycoproteins of the sperm fibrous sheath, Fibrous sheath interacting protein 2, Fascin, Ornithine decarboxylase antizyme 3 and Leucine rich repeat and coiled-coil centrosomal protein 1. These proteins are structural components of the sperm tail playing essential roles in the activation of the flagellum and are, therefore, positively involved in sperm motility (Ruan et al., 2011; Hashemitabar et al., 2015; Inaba and Mizuno, 2015). Similarly, the overexpression of the other four characterized proteins could also clarify the best functionality showed by FT-spermatozoa from the SRF. In this sense, Voltage-dependent anion-selective channel protein 2 (VDAC2) is a mitochondrial outer membrane porin protein, considered both a positive biomarker of boar sperm freezability (Vilagran et al., 2014), and actively participating in the regulation of sperm mitochondrial function (Menzel et al., 2009). Family with sequence similarity 205-mamber A (FAM205A) is a protein found overexpressed in subpopulations of ejaculated human viable spermatozoa with low nuclear DNA fragmentation (Istasqui et al., 2013). Finally, the other two overexpressed proteins in FT-spermatozoa from the SRF, specifically the Transmembrane protein 89 and IQ motif containing F5, are proteins of sperm plasma membrane with involvement in membrane stabilization and permeability regulation (Daskalova et al., 2014).

A larger number of proteins were overexpressed in FT-spermatozoa retrieved from the EE. Specifically, a total of 15 proteins, all of them linked to deterioration of sperm function, were over-expressed. Such an overexpression could explain why FT-

spermatozoa from the EE were less functional than those retrieved from the SRF. For instance, one of them, the Dipeptidyl peptidase 4 (DPP-4 or CD26), a mitochondrial associated protein, would be involved in controlling sperm motility, as demonstrated in mice (Shaw and Nath, 2015; An et al., 2018), or overexpressed in hyperactivated human spermatozoa (Agrawal and Vanha-Perttula, 1987). Hyperactivation is a motility pattern displayed by some cryosurviving boar spermatozoa that experience the so-called cryocapacitation, a premature capacitation-like change linked to suboptimal cryopreservation process (Vadnais and Althause, 2011) and related to low fertility outcomes of FT-spermatozoa (Watson, 2000). Needless to say, presence of a large number of capacitated spermatozoa in an AI-dose eventually leads to sperm death, long before they could encounter the oocyte/s. Endoplasmin, more known as tumor rejection antigen gp96, is a chaperone involved in the rebuilding of sperm surface during capacitation (Mitchell et al., 2007), which could also be overexpressed in cryocapacitated spermatozoa. Similarly, cryocapacitation could lead to the overexpression of the spermadhesin AQN1, an acrosomal-surface bound protein involved in sperm interaction with oviduct cells (Sanz et al., 1992; Yi et al., 2012). The incidence of apoptosis and lipid peroxidation, associated an excessive generation of reactive oxygen species (**ROS**), was larger in viable FT-spermatozoa retrieved from the EE-source than those from the SRF source and both dysfunctions could be linked to cryocapacitation. Recently, Aitken (2017) highlighted the relationship between capacitation and sperm apoptosis, due to a continuous generation of ROS. Moreover, some of the overexpressed proteins would entail that the fertilizing capacity of FT-spermatozoa from the EE source could be compromised. For instance, Nexin 1, Spermadhesins PSPI, Tetraspanin CD63 (CD63), Complement Factor D (CFD) and Ras GTPase-activating-like protein IQGAP2 (IQGAP2) have been related to fertility losses when overexpressed in either spermatozoa or seminal plasma. Overexpression of PSPI and Nexin 1 was related with low fertility outcomes in artificial inseminated sows (Novak et al., 2010; Perez-Patiño et al., 2018). Similarly, CD63, CFD and IQGAP2 were overexpressed in fresh human spermatozoa with poor embryo development capacity after ICSI (Garcia-Herrero et al., 2011; McReynolds et al., 2014). Other of the overexpressed proteins are related with the immune response either through the complement and T-cell activation or cytokine modulation, specifically Syndecan binding protein (SDCBP), DPP-4 and Programmed cell death 6 interacting protein. Their overexpression could make FT-spermatozoa from the EE source particularly susceptible to phagocytosis once inseminated in the internal

genital tract of the sow (Matthijs et al., 2000; Rodriguez-Martinez et al., 2005). The leading consequences on FT-sperm performance by the five remaining overexpressed proteins would be less clear. Overexpression of Protein disulphide-isomerase could be again related to cryocapacitation (Wong et al., 2017). Sperm equatorial segment protein 1 is involved in sperm-oocyte binding (Fujihara et al., 2010), so that it would be overexpressed in capacitated spermatozoa. Radial spoke head 6 homolog A (RSPH6A), a structural proteins of the sperm fibrous sheath and thereby involved in sperm motility (Pereira et al., 2017), was find overexpressed in ovine FT-sperm (Pini et al., 2018). Protease, serine 8 is an enzyme involved in semen coagulation (Yu et al., 1995) and Chromosome 1 open reading frame 56 was recently identified in boar sperm chromatin (Mendonça et al., 2017).

Interesting was to note the peculiar relationship of the overexpressed proteins with seminal plasma. Whereas none of the proteins overexpressed in the FT-spermatozoa from the SRF source were identified in the seminal plasma of such fraction of the ejaculate, eleven of the fifteen overexpressed proteins in EE-derived FT-spermatozoa were also present in the seminal plasma (Perez-Patiño et al., 2018). These findings would indicate that most of the proteins overexpressed in the FT-spermatozoa from the EE source would come from the seminal plasma, binding to the sperm surface either during the ejaculation or in the subsequent overnight storage before freezing. Moreover, as these proteins were found overexpressed in the seminal plasma of the post-SRF source (Perez-Patiño et al., 2016), it is reasonable to think that they primarily bind to the spermatozoa fortuitously present in the post-SRF fraction, which would explain why they were not found overexpressed in the FT-spermatozoa retrieved from the SRF (neither the first 10 mL of SRF not the rest of SRF). It would also suppose that the binding of the proteins to the spermatozoa had probably already occurred during the ejaculation time than during the subsequent overnight storage in contact with seminal plasma before freezing. Spermatozoa of the three main, identifiable, ejaculate fractions, specifically the first 10 mL of SRF, rest of SRF and post-SRF, were separately cryopreserved and the proteome of FT-spermatozoa was evaluated to clarify the above assumptions.

The results of post-thaw sperm attributes agree with those previously reported by Alkmin et al. (2014) and Li et al. (2018a) confirming that spermatozoa derived from the post-SRF are those depicting the worst freezability. The results prove that distinguishable sperm populations with clear differences in freezability are present in a single porcine ejaculate,

suggesting biological reasons for the presence of a fractionated ejaculate in this species. Taken together, the results indicate that the worse freezability of spermatozoa from the EE with respect to those from the SRF source would be mainly due to the negative concerted contribution of spermatozoa and seminal plasma of the post-SRF in building the EE. The clear quantitative differences in the proteome between the FT-spermatozoa from the post-SRF fraction and those from the other two ejaculate fractions could explain the poor quality and functionality of FT-spermatozoa from the post-SRF source, as they overexpressed many proteins directly or indirectly involved in sperm functionality. As expected, the 15 proteins overexpressed in the FT-spermatozoa from the EE with respect to those from the SRF source were among the proteins overexpressed in the FT-spermatozoa from the post-SRF. Remarkable was the finding that most of the overexpressed proteins in FT-spermatozoa from the post-SRF source were previously identified as overexpressed in its seminal plasma (Perez-Patiño et al., 2016). This scenario, with more differentially expressed proteins in FT-spermatozoa from the post-SRF compared to the other two ejaculate fractions, would also support the notion that seminal plasma proteins bind most likely to the spermatozaa during ejaculation than during the overnight sperm storage its own seminal plasma before freezing.

8.5 Conclusion

In conclusion, the present study showed that differences in proteome expression could explain that spermatozoa retrieved from the EE were less functional post-thaw than those derived from the SRF source, a matter with clear biological and practical implications considering that most of the differentially expressed proteins are directly involved in sperm function, including fertilizing capacity. The worst functionality of the FT-spermatozoa derived from the EE source would be caused by the negative contribution of the FT-spermatozoa from the post-SRF, contributing with many seminal plasma proteins which, if overexpressed could negatively influence freezability and cryosurvival. These results should be taken into careful consideration by those responsible of collecting and cryopreserving EE-spermatozoa instead of continue using spermatozoa present in the SRF.

Supplementary material

All supplementary material is available in the website:

https://cristinaperez61.wixsite.com/phdthesis

Acknowledgements

The authors are grateful to AIM Iberica (Topigs Norsvin Iberica) for supplying the boar ejaculates. This experimental study was supported by MINECO (Spain) and FEDER funds (EU) (AGL2015-69738-R), and Seneca Foundation Murcia, Spain (19892/GERM-15); FORSS (745971) and The Swedish Research Council FORMAS (2017-00946), Stockholm, Sweden. C. Perez-Patiño, I. Barranco and J. Li were financially supported by Seneca Foundation (Murcia, Spain), MECD (Madrid, Spain) and the China Scholarship Council, respectively.

References

Agrawal Y, Vanha-Perttula T. Effect of secretory particles in bovine seminal vesicle secretion on sperm motility and acrosome reaction. J Reprod Fertil. 1987, 79:409-19.

Aitken RJ. Reactive oxygen species as mediators of sperm capacitation and pathological damage. Mol Reprod Dev. 2017, 84:1039-1052.

Alkmin DV, Martinez-Alborcia MJ, Parrilla I, Vazquez JM, Martinez EA, Roca J. The nuclear DNA longevity in cryopreserved boar spermatozoa assessed using the Sperm-Sus-Halomax. Theriogenology. 2013, 79:1294-300.

Alkmin DV, Perez-Patiño C, Barranco I, Parrilla I, Vazquez JM, Martinez EA, Rodriguez-Martinez H, and Roca J. Boar sperm cryosurvival is better after exposure to seminal plasma from selected fractions than to those from entire ejaculate. Cryobiology. 2014, 69:203-10.

An T, Wang YF, Liu JX, Pan YY, Liu YF, He ZC, Mo FF, Li J, Kang LH, Gu YJ, Lv BH, Gao SH, Jiang GJ. Comparative analysis of proteomes between diabetic and normal human sperm: Insights into the effects of diabetes on male reproduction based on the regulation of mitochondria-related proteins. Mol Reprod Dev. 2018, 85:7-16.

Bolarín A, Roca J, Rodríguez-Martínez H, Hernández M, Vázquez JM, Martínez EA. Dissimilarities in sows' ovarian status at the insemination time could explain differences in fertility between farms when frozen-thawed semen is used. Theriogenology. 2006, 65:669-80.

Chen X, Zhu H, Hu C, Hao H, Zhang J, Li K, Zhao X, Qin T, Zhao K, Zhu H, Wang D. Identification of differentially expressed proteins in fresh and frozen-thawed boar spermatozoa by iTRAQ-coupled 2D LC-MS/MS. Reproduction. 2014, 147:321-30.

Daskalova D, Kukov A, Kirilova I, Ivanova-Kicheva M. Protein analysis of boar seminal plasma proteins with protective effect during low-temperature storage of spermatozoa. Biotechnol Biotechnol Equip. 2014, 28:716-720.

Fujihara Y, Murakami M, Inoue N, Satouh Y, Kaseda K, Ikawa M, Okabe M. Sperm equatorial segment protein 1, SPESP1, is required for fully fertile sperm in mouse. J Cell Sci. 2010, 123:1531-6.

Gadella BM, Boerke A. An update on post-ejaculatory remodeling of the sperm surface before mammalian fertilization. Theriogenology. 2016, 85:113-24.

Gadella BM. Sperm membrane physiology and relevance for fertilization. Anim Reprod Sci. 2008, 107:229-36.

García-Herrero S, Garrido N, Martínez-Conejero JA, Remohí J, Pellicer A, Meseguer M. Differential transcriptomic profile in spermatozoa achieving pregnancy or not via ICSI. Reprod Biomed Online. 2011, 22:25-36.

Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics. 2012, 11:O111.016717.

Hashemitabar M, Sabbagh S, Orazizadeh M, Ghadiri A, Bahmanzadeh M. A proteomic analysis on human sperm tail: comparison between normozoospermia and asthenozoospermia. J Assist Reprod Genet. 2015, 32:853-63.

Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc. 2009, 4:44-57.

Inaba K, Mizuno K. Sperm dysfunction and ciliopathy. Reprod Med Biol. 2015, 15:77-94.

Intasqui P, Camargo M, Del Giudice PT, Spaine DM, Carvalho VM, Cardozo KH, Cedenho AP, Bertolla RP. Unraveling the sperm proteome and post-genomic pathways associated with sperm nuclear DNA fragmentation. J Assist Reprod Genet. 2013, 30:1187-202.

Knox RV. The Fertility of Frozen Boar Sperm When used for Artificial Insemination. Reprod Domest Anim. 2015, 50:90-7.

Li J, Barranco I, Tvarijonaviciute A, Molina MF, Martinez EA, Rodriguez-Martinez H, Parrilla I, and Roca J. Seminal plasma antioxidants are directly involved in boar sperm cryotolerance. Theriogenology. 2018a, 107:27-35.

Li J, Roca J, Pérez-Patiño C, Barranco I, Martinez EA, Rodriguez-Martinez H, Parrilla I. Is boar sperm freezability more intrinsically linked to spermatozoa than to the surrounding seminal plasma? Anim Reprod Sci. 2018b, pii: S0378-4320(17)30936-3.

Marlea DS, Nicoletta T, Marco N, Andrea B. Human Sperm Cryopreservation: Update on Techniques, Effect on DNA Integrity, and Implications for ART. Adv Urol. 2012, 2012: 854837.

Matthijs A, Harkema W, Engel B, Woelders H. In vitro phagocytosis of boar spermatozoa by neutrophils from peripheral blood of sows. J Reprod Fertil. 2000, 120:265-73.

McReynolds S, Dzieciatkowska M, Stevens J, Hansen KC, Schoolcraft WB, Katz-Jaffe MG. Toward the identification of a subset of unexplained infertility: a sperm proteomic approach. Fertil Steril. 2014, 102:692-9.

Mendonça AG, Filho RM, Souza ET, Gaggini TS, Cruvinel M, Silva-Mendonça M, Robson CA, Beletti ME. Isolation and identification of proteins from swine sperm chromatin and nuclear matrix. Animal Reproduction. 2017, 14:418-428.

Menzel VA, Cassará MC, Benz R, de Pinto V, Messina A, Cunsolo V, Saletti R, Hinsch KD, Hinsch E. Molecular and functional characterization of VDAC2 purified from mammal spermatozoa. Biosci Rep. 2009, 29:351-62.

Mitchell LA, Nixon B, Aitken RJ. Analysis of chaperone proteins associated with human spermatozoa during capacitation. Mol Hum Reprod. 2007, 13:605-13.

Novak S, Ruiz-Sánchez A, Dixon WT, Foxcroft GR, Dyck MK. Seminal plasma proteins as potential markers of relative fertility in boars. J Androl. 2010, 31:188-200.

Pereira R, Sá R, Barros A, Sousa M. Major regulatory mechanisms involved in sperm motility. Asian J Androl. 2017, 19:5-14.

Perez-Patiño C, Barranco I, Parrilla I, Valero ML, Martinez EA, Rodriguez- Martinez H, Roca J. Characterization of the porcine seminal plasma proteome comparing ejaculate portions. J. Proteomics. 2016, 142:15-23

Pérez-Patiño C, Parrilla I, Barranco I, Vergara-Barberán M, Simó-Alfonso EF, Herrero-Martínez JM, Rodriguez-Martínez H, Martínez EA, Roca J. New In-Depth Analytical Approach of the Porcine Seminal Plasma Proteome Reveals Potential Fertility Biomarkers. J Proteome Res. 2018, 17:1065-1076.

Pini T, Rickard JP, Leahy T, Crossett B, Druart X, de Graaf SP. Cryopreservation and egg yolk medium alter the proteome of ram spermatozoa. J Proteomics. 2018, 181:73-82.

Roca J, Martinez-Alborcia MJ, Gil MA, Parrilla I, Martinez EA. Dead spermatozoa in raw semen samples impair in vitro fertilization outcomes of frozen-thawed spermatozoa. Fertil Steril. 2013, 100:875-81.

Roca J, Parrilla I, Bolarin A, Martinez EA, Rodriguez-Martinez H. Will AI in pigs become more efficient? Theriogenology. 2016, 86:187-93.

Rodríguez-Martínez H, Saravia F, Wallgren M, Tienthai P, Johannisson A, Vázquez JM, Martínez E, Roca J, Sanz L, Calvete JJ. Boar spermatozoa in the oviduct. Theriogenology. 2005, 63:514-35.

Rodríguez-Martínez H, Kvist U, Saravia F, Wallgren M, Johannisson A, Sanz L, Peña FJ, Martínez EA, Roca J, Vázquez JM, Calvete JJ. The physiological roles of the boar ejaculate. Soc Reprod Fertil. 2009, 66:1-21.

Rozeboom KJ, Troedsson MH, Molitor TW, Crabo BG. The effect of spermatozoa and seminal plasma on leukocyte migration into the uterus of gilts. J Anim Sci. 1999, 77:2201-6.

Ruan Y, Cheng M, Ou Y, Oko R, van der Hoorn FA. Ornithine decarboxylase antizyme Oaz3 modulates protein phosphatase activity. J Biol Chem. 2011, 286:29417-27.

Sanz L, Calvete JJ, Mann K, Schäfer W, Schmid ER, Töpfer-Petersen E. The complete primary structure of the boar spermadhesin AQN-1, a carbohydrate-binding protein involved in fertilization. Eur J Biochem. 1992, 205:645-52.

Saravia P, Wallgren M, Johannisson A, Calvete JJ, Sanz L, Peña FJ, Roca J, and Rodríguez-Martínez H. Exposure to the seminal plasma of different portions of the boar ejaculate

modulates the survival of spermatozoa cryopreserved in MiniFlatPacks. Theriogenology. 2009, 71:662-75.

Shaw M, Nath D. Characterization of a Dipeptidyl Peptidase and its Role in Motility of Rat Epididymal Maturing Spermatozoa. Reproductive Biology Insights. 2015, 8:9-16

Shevchenko A, Jensen ON, Podtelejnikov AV, Sagliocco F, Wilm M, Vorm O, Mortensen P, Shevchenko A, Boucherie H, Mann M. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. Proc Natl Acad Sci. 1996, 93:14440-5.

Thurston LM, Siggins K, Mileham AJ, Watson PF, Holt WV. Identification of amplified restriction fragment length polymorphism markers linked to genes controlling boar sperm viability following cryopreservation. Biol Reprod. 2002, 66:545-54.

Vadnais ML, Althouse GC. Characterization of capacitation, cryoinjury, and the role of seminal plasma in porcine sperm. Theriogenology. 2011, 76:1508-16.

Vilagran I, Yeste M, Sancho S, Casas I, Rivera del Álamo MM, Bonet S. Relationship of sperm small heat-shock protein 10 and voltage-dependent anion channel 2 with semen freezability in boars. Theriogenology. 2014, 82:418-26.

Watson PF. The causes of reduced fertility with cryopreserved semen. Anim Reprod Sci. 2000, 60-61:481-92.

Wong CW, Lam KKW, Lee CL, Yeung WSB, Zhao WE, Ho PC, Ou JP, Chiu PCN. The roles of protein disulphide isomerase family A, member 3 (ERp57) and surface thiol/disulphide exchange in human spermatozoa-zona pellucida binding. Hum Reprod. 2017, 32:733-742.

Yeste M. Recent Advances in Boar Sperm Cryopreservation: State of the Art and Current Perspectives. Reprod Domest Anim. 2015, 50:71-9.

Yeste M. Sperm cryopreservation update: Cryodamage, markers, and factors affecting the sperm freezability in pigs. Theriogenology. 2016, 85:47-64.

Yeste M, Rodríguez-Gil JE, Bonet S. Artificial insemination with frozen-thawed boar sperm. Mol Reprod Dev. 2017, 84:802-813.

Yi YJ, Zimmerman SW, Manandhar G, Odhiambo JF, Kennedy C, Jonáková V, Maňásková-Postlerová P, Sutovsky M, Park CS, Sutovsky P. Ubiquitin-activating enzyme (UBA1) is required for sperm capacitation, acrosomal exocytosis and sperm-egg coat penetration during porcine fertilization. Int J Androl. 2012, 35:196-210.

Yu JX, Chao L, Chao J. Molecular cloning, tissue-specific expression, and cellular localization of human prostasin mRNA. J Biol Chem. 1995, 270:13483-9.

GENERAL CONCLUSIONS

9. General Conclusions

1 The present study provides the largest and most updated proteome, encoded in *Sus scrofa* taxonomy, of the seminal plasma and pig spermatozoa.

2 The proteome expression of seminal plasma and pig spermatozoa differs among the different portions/fractions of the ejaculate and the differences involve to proteins clearly implicated in sperm functionality.

3 The proteomic profile of seminal plasma differs among boars showing real differences in fertility, and the comparative proteome identify a panel of sperm proteins potentially suitable for using as fertility biomarkers.

4 Differences in quality and functionality between frozen-thawed pig sperm samples would be linked to differences in sperm proteome expression.

RESUMEN

Introducción

Las **proteínas** de los espermatozoides y del plasma seminal (**SP**) son fundamentales para la capacidad fecundante de los espermatozoides (Rodríguez-Martínez y cols., 2011; Caballero y cols., 2012), ya que participan en la regulación de las más importantes funciones espermáticas, como la motilidad, capacitación, fusión con el ovocito y reacción acrosómica (Caballero y cols., 2012). Además, estas proteínas facilitan el viaje de los espermatozoides en el útero y el oviducto, favoreciendo también la implantación embrionaria (Bromfield, 2016). En la especie porcina, como ocurre en otros mamíferos, algunas proteínas específicas de los espermatozoides y del SP han sido identificadas como potenciales biomarcadores tanto de la fertilidad como de la congelabilidad espermática (Yeste, 2016; Rahman y cols., 2017). Estos hallazgos son prometedores y motivan futuros estudios que profundicen en el conocimiento del proteoma de los espermatozoides y del SP. Para ello, la **proteómica**, definida como el estudio a gran escala del proteoma, es el área de investigación a utilizar, y la espectrometría de masas (**MS**) la herramienta básica para desarrollarla (Hedrick y cols., 2015). En este contexto, la presente **Tesis Doctoral** pretende mejorar el conocimiento actual del proteoma de los espermatozoides y del SP de porcino para, después, identificar potenciales diferencias cuantitativas en proteínas que puedan tener un impacto directo sobre la funcionalidad y capacidad fecundante de los espermatozoides.

La presente Tesis Doctoral está compuesta por cinco estudios. El primero de ellos, publicado en la revista **Journal of Proteomics (2016, 142:15-23),** fue diseñado para caracterizar el proteoma del SP de porcino y determinar diferencias en expresión proteica entre el SP de las principales fracciones del eyaculado, específicamente, de los primeros 10 mL de la fracción rica del eyaculado (**SRF**), del resto de la SRF y de la post-SRF. Con parte de los resultados generados en este estudio se diseñó el segundo estudio, publicado en la revista **Data in Brief (2016, 8:1370-1373).** Estudio que mostró una amplia base de datos de acceso público del proteoma del SP de porcino, incluyendo en ella una revisión extensa de la posible implicación reproductiva de cada una de las proteínas identificadas. Esta base de datos supone una valiosa fuente de información para ayudar a conocer la función que las proteínas del SP ejercen sobre la funcionalidad espermática. El tercer estudio, además de actualizar el conocimiento del proteoma del SP de porcino, fue

diseñado para identificar potenciales biomarcadores de fertilidad entre las proteínas identificadas (**Journal of Proteome Research 2018, 17:1065-1076**). El cuarto y el quinto estudio se centraron en el estudio del proteoma del espermatozoide de porcino. En primer lugar, diseñamos un estudio para comparar el proteoma de espermatozoides procedentes de las principales fracciones del eyaculado, ya mencionadas anteriormente (manuscrito enviado para su publicación a la revista **Molecular and Cellular Proteomics**). Finalmente, el último estudio se centró en el análisis proteómico de espermatozoides de porcino congelados-descongelados (**FT**), con el fin de identificar un patrón diferente de expresión proteica entre muestras de espermatozoides FT con diferencias de calidad a la descongelación (manuscrito enviado para su publicación a la revista **Scientific Reports**).

Materiales y métodos

Animales, obtención de muestras de espermatozoides y de plasma seminal y registros de fertilidad

Todos los eyaculados y los epidídimos fueron obtenidos de verracos de diferentes razas todos ellos sanos, maduros y sexualmente activos que estaban incluidos en programas de inseminación artificial (**AI**) y que fueron sacrificados debido a razones de reemplazo genético. Los eyaculados fueron recogidos de manera fraccionada, empleando el método de la mano enguantada, o como eyaculación completa, utilizando un método semiautomático. Solamente aquellos eyaculados que cumplían los estándares de cantidad y calidad espermática fueron utilizados. Las fracciones del eyaculado recogidas fueron los primeros 10 mL de la SRF, la SRF completa o sin los primeros 10 mL y la post-SRF. Los espermatozoides procedentes de la cola del epidídimo fueron obtenidos mediante infusión retrógrada de aire en el conducto deferente.

Para el análisis proteico, las muestras de SP y de espermatozoides frescos (procedentes del eyaculado o del epidídimo) fueron obtenidas mediante doble centrifugación (1.500 *xg* 10 min) y almacenadas a -80 °C hasta el análisis. Para la congelación espermática se utilizó un procedimiento estándar para pajuelas de 0,5 mL con un ratio de congelación y de descongelación de -40 °C/min y +1.200 °C/min, respectivamente.

Para generar los registros de fertilidad, cerdas multíparas (1-7 partos) fueron inseminadas cervicalmente, de 2 a 3 veces por estro, utilizando dosis de AI (2.500 x 10^6 espermatozoides diluidos hasta un volumen de 80 mL en un diluyente comercial). La fertilidad de los verracos fue registrada durante un periodo de 12 meses en términos de tasa de partos (**FR**, porcentaje de cerdas que paren con respecto al número de cerdas inseminadas) y tamaño de la camada (**LS**, número de lechones nacidos por camada). Estos datos se registraron como efecto directo del verraco, anulando estadísticamente los efectos relacionados con la granja o la cerda.

Análisis de la calidad espermática tras la descongelación

La calidad y funcionalidad espermática fue evaluada a los 30 min tras la descongelación en términos de motilidad (total y progresiva), viabilidad, fragmentación del ADN nuclear, peroxidación lipídica (**LPO**) y apoptosis celular temprana. La motilidad espermática fue evaluada usando un sistema objetivo de análisis espermático computarizado (**CASA**) y la fragmentación del ADN nuclear fue evaluada mediante el kit Sperm-Sus-Halomax®. El resto de parámetros espermáticos fueron evaluados mediante citometría de flujo, tiñendo los espermatozoides con los fluorocromos apropiados para cada análisis.

Análisis proteómico

Los análisis de proteínas fueron llevados a cabo en la Unidad de Proteómica de la Universidad de Valencia (Valencia, España) que es miembro de la plataforma PRB2-ISCIII ProteoRed Proteomics Platform.

Análisis proteómico de las muestras de plasma seminal y de los espermatozoides

Las muestras de SP fueron descongeladas a temperatura ambiente y durante el proceso de descongelación a las muestras se les añadió un cóctel inhibidor de proteasas para prevenir la degradación proteica. La concentración total de proteínas se midió usando el fluorómetro Qubit. Una alícuota de cada muestra fue directamente digerida con tripsina y una segunda alícuota fue sometida a electroforesis de una dimensión en gel de poliacrilamida con dodecil sulfato de sodio (**1D SDS-PAGE**) y las proteínas separadas, fueron sometidas a una digestión en gel con tripsina. Los péptidos resultantes fueron

separados mediante cromatografía de exclusión por tamaño (**SEC**) y extracción en fase sólida (**SPE**) con el fin de reducir la complejidad proteica de las muestras.

Las muestras de espermatozoides, una vez descongeladas también a temperatura ambiente, fueron lavadas en un gradiente de densidad de Percoll en mono capa para separar las células espermáticas de la yema de huevo existente y otros restos. Después, las proteínas de los espermatozoides frescos y descongelados fueron extraídas con un tampón de lisis UTC, se cuantificaron mediante RCDC Lowry y se sometieron a 1D SDS-PAGE para eliminar restos de UTC antes de la digestión en gel con tripsina.

Los péptidos extraídos de las muestras de SP y espermatozoides fueron analizados mediante cromatografía líquida acoplada a la espectrometría de masas en tándem (**LC-MS/MS**) usando un NanoLC Ultra 1D plus Eksigent conectado a un espectrómetro de masas AB SCIEX TripleTOF 5600. Los archivos SCIEX.wiff se procesaron usando el programa ProteinPilot v5.0. El algoritmo Paragon de ProteinPilot se usó para cotejar los espectros de MS/MS con las secuencias de péptidos de las bases de datos del Centro Nacional de Información Biotecnológica (**NCBI**) o de UniProt_mammals. Las proteínas se identificaron con un nivel de confidencia peptídica de al menos un 95% y una tasa de falso descubrimiento (**FDR**) menor del 1% a nivel proteico.

Cuantificación de proteínas procedentes de plasma seminal y de los espermatozoides

Las proteínas fueron cuantificadas mediante las tecnologías de adquisición secuencial de ventanas de todos los espectros teóricos (**SWATH**) y por la de isótopos isobáricos para la cuantificación relativa (**iTRAQ**). Para el análisis SWATH, las muestras digeridas fueron cargadas, aleatoriamente y sin marcaje previo en la columna de LC. Los péptidos eluidos fueron analizados con el espectrómetro de masas AB SCIEX TripleTOF 5600 configurado en modo swath. Los documentos de Protein-Pilot derivados del análisis, se cargaron en PeakView® (v2.1) y los picos de los iones fragmentados extraídos del análisis SWATH fueron normalizados dividiendo el área de cada pico por la suma de las áreas de todos los picos, y la suma total de todas las áreas fue igualada para todas las muestras. Para el análisis iTRAQ seguimos el protocolo del kit 8-plex iTRAQ. Brevemente, los residuos de cisteína de todas las proteínas desnaturalizadas fueron bloqueados y las proteínas fueron sometidas a una digestión con tripsina. La mezcla de péptidos resultantes se incubó durante 3 h a temperatura ambiente con el marcador correspondiente de iTRAQ y estos fueron separados mediante isoelectroenfoque (**IEF**)

seguido del análisis mediante LC-MS/MS. La cuantificación relativa de las proteínas espermáticas, fue obtenida, en este caso, mediante la comparación de las intensidades relativas de los iones reporteros de las diferentes muestras.

Análisis bioinformático

El análisis bioinformático de las proteínas identificadas y validadas de SP y de los espermatozoides de porcino fue realizado manualmente usando la base de datos de acceso público UniProtKB (www.uniprot.org) en combinación con PANTHER (Protein Analysis THrough Evolutionary Relationships, www.pantherdb.org) y DAVID (Database for Annotation, Visualization and Integrated Discovery, www.david.ncifcrf.gov).

Análisis estadístico

Los análisis estadísticos se realizaron usando los paquetes estadísticos IBM SPSS v19, R y Multiexperiment Viewer (**MeV**). Se calculó la cantidad media de cada proteína en cada muestra y su diferencia de expresión (fold change, **FC**) entre las muestras. Se realizó análisis de componentes principales (**PCA**) para determinar el poder discriminativo de los perfiles proteicos entre muestras. Se realizaron análisis de T de Student o ANOVA para identificar el nivel de significancia de las proteínas diferencialmente expresadas entre muestras y para evaluar la influencia de la muestra en la calidad espermática a la descongelación. Empleando la contracción absoluta mínima y la regresión del operador de la selección (**LASSO)** se identificaron las proteínas del SP que explicarían diferencias entre verracos en los parámetros de fertilidad. Las diferencias cuantitativas en las proteínas diferencialmente expresadas entre las diferentes muestras se representaron gráficamente en mapas de calor tras la normalización de la puntuación estándar (z-score). Un valor de $p<0.05$ se consideró como el nivel mínimo de significancia estadística.

Diseño experimental y resultados

El **Primer Estudio** (**Journal of Proteomics 2016, 142:15–23**) se diseñó para caracterizar el proteoma del SP de porcino y para identificar diferencias en la expresión proteica del SP procedente de diferentes fracciones del eyaculado, concretamente, de los primeros 10 mL de la SRF, del resto de la SRF y de la post-SRF. Para caracterizar el proteoma del SP, se generó un pool de SP procedente de 33 eyaculados completos (11 verracos, 3 eyaculados/verraco). Para caracterizar diferencias en la expresión proteica del SP de las fracciones del eyaculado, se utilizaron muestras de SP procedentes de 20 eyaculados (5 verracos, 4 eyaculados/verraco). Las muestras de SP se analizaron usando una combinación de SEC, 1D SDS-PAGE y NanoLC-ESI-MS/MS seguido del análisis bioinformático y las proteínas identificadas fueron cuantificadas mediante la tecnología SWATH. Un total de 536 proteínas fueron identificadas, 409 de ellas de *Sus scrofa* (374 validadas con un nivel de confidencia ≥99%). Treinta y cuatro proteínas (16 de ellas de *Sus scrofa*) fueron identificadas como diferencialmente expresadas (16 sobre-expresadas y 18 infra-expresadas) entre la SRF (primeros 10 mL de la SRF y resto de la SRF) y la post-SRF. Este estudio proteómico aporta la mayor caracterización del proteoma del SP de porcino disponible hasta el momento, con más de 250 proteínas descritas por primera vez en el SP. Además, el estudio reveló un conjunto de proteínas diferencialmente expresadas entre las fracciones del eyaculado, muchas de ellas asociadas con la capacidad funcional de los espermatozoides, incluida la fecundante.

El **Segundo Estudio** (**Data in Brief 2016, 8:1370-1373**) recogió los resultados derivados del primer estudio generando una base de datos de acceso público complementada con una detallada revisión de las funciones reproductivas en las que las proteínas del SP porcino podrían estar implicadas.

El **Tercer Estudio** (**Journal of Proteome Research 2018, 17:1065-1076**) se diseñó para identificar proteínas del SP de porcino diferencialmente expresadas entre verracos con diferencias significativas en los parámetros de fertilidad (tasa de partos -FR- y lechones nacidos por parto -LS-). Previamente, se actualizó el proteoma del SP de porcino, utilizando para ello un nuevo procedimiento de fraccionamiento de las proteínas que incluía SEC y SPE seguido de NanoLC-ESI-MS/MS. Este novedoso procedimiento permitió identificar un total de 872 proteínas, 390 de ellas de *Sus scrofa*. A continuación, se analizó el proteoma del SP de 26 verracos (pool de 4 muestras de SP por verraco) con

diferencias significativas en FR (13 verracos) y LS (13 verracos) en cerdas artificialmente inseminadas (10.526 cerdas). Un total de 679 proteínas del SP fueron cuantificadas mediante SWATH y el análisis de regresión LASSO identificó un total de 11 proteínas del SP expresadas diferencialmente entre verracos con una FR alta y baja (FURIN, AKR1B1, UBA1, PIN1, SPAM1, BLMH, SMPDL3A, KRT17, KRT10, TTC23 y AGT) y 4 proteínas expresadas diferencialmente entre verracos con un LS alto y bajo (PN-1, THBS1, DSC1 and CAT). Este estudio contribuye a incrementar el conocimiento del proteoma del SP así como a identificar algunas proteínas que pudieran ser utilizadas como potenciales biomarcadores de fertilidad en verracos incluidos en programas de AI.

El **Cuarto Estudio** (manuscrito enviado para su publicación a la revista **Molecular and Cellular Proteomics**) se diseñó para, en primer lugar, actualizar el proteoma del espermatozoide de porcino y, en segundo lugar, identificar proteínas expresadas diferencialmente entre espermatozoides almacenados en la cola del epidídimo (no eyaculados) y aquellos procedentes de las fracciones del eyaculado. Para lograr estos objetivos, un total de 9 eyaculados procedentes de otros tantos verracos fueron recogidos de forma fraccionada (primeros 10 mL de la SRF, resto de la SRF y post-SRF). Las proteínas extraídas de estas muestras espermáticas fueron sometidas a un análisis de iTRAQ basado en LC-MS/MS permitiendo la identificación y cuantificación simultánea de las proteínas. Un total de 1.723 proteínas fueron identificadas y 1.602 de ellas fueron además cuantificadas. Treinta y dos proteínas *Sus scrofa* mostraron diferencias cuantitativas ($P<0,01$) entre las muestras de espermatozoides. Estas diferencias fueron especialmente relevantes para determinadas proteínas relacionadas con la funcionalidad de los espermatozoides procedentes de la post-SRF. Este estudio caracteriza, por primera vez, diferencias cuantitativas en el perfil proteico de espermatozoides maduros (eyaculados y no eyaculados), las cuales afectan a proteínas claramente involucradas en la funcionalidad espermática. El estudio también evidencia que el perfil proteico de los espermatozoides es remodelado durante la eyaculación, probablemente por la interacción entre espermatozoides y SP. Estos hallazgos suponen, además, una base preliminar para futuros estudios centrados en identificar proteínas espermáticas como marcadores de fertilidad.

El **Quinto Estudio** (manuscrito enviado para su publicación a la revista **Scientific Reports**) se estructuró en dos experimentos diferentes con el fin de determinar diferencias en la expresión proteica entre espermatozoides criopreservados que pudieran explicar

diferencias en calidad y funcionalidad espermática tras la descongelación. Las diferencias en la expresión proteica fueron evaluadas usando un análisis SWATH basado en LC-MS/MS. El primer experimento se diseñó para comparar el proteoma de espermatozoides descongelados procedentes de la SRF con el de aquellos procedentes del eyaculado completo. El segundo experimento, comparó el proteoma de espermatozoides descongelados procedentes de tres fracciones del eyaculado con claras diferencias de congelabilidad espermática (primeros 10 mL de la SRF, resto de la SRF y post-SRF, siendo ésta última la de peor congelabilidad). En el experimento 1, los espermatozoides descongelados procedentes de la SRF mostraron mejor calidad que aquellos procedentes del eyaculado completo y 26 proteínas S*us scrofa* se expresaban diferencialmente (FC ≥ 1.5) entre ambas muestras. En el experimento 2, los espermatozoides descongelados procedentes de los primeros 10 mL de la SRF y del resto de la SRF mostraron mejor calidad a la descongelación que aquellos procedentes de la post-SRF con 187 proteínas expresadas diferencialmente entre los espermatozoides de las tres fracciones. Estos hallazgos evidencian que diferencias en la expresión proteica podrían explicar diferencias en la congelabilidad espermática. Estos resultados poseen una clara implicación biológica y práctica ya que, actualmente, los centros de AI están dejando de recoger únicamente la SRF para recoger el eyaculado completo, un cambio que puede afectar notablemente a la calidad de las dosis de semen preparadas, particularmente las de semen criopreservado.

Conclusiones

1.- Este estudio aporta la descripción más extensa y actualizada del proteoma *Sus scrofa* del plasma seminal y de los espermatozoides de porcino.

2.- La expresión de las proteínas del plasma seminal y de los espermatozoides de porcino difiere entre las fracciones del eyaculado. Estas diferencias cuantitativas implican a proteínas relacionadas con la funcionalidad espermática.

3.- El perfil proteico del plasma seminal difiere entre verracos con diferencias en fertilidad. El estudio comparativo del proteoma señala un conjunto de proteínas del plasma seminal que pueden ser consideradas como potenciales biomarcadores de fertilidad.

4.- Las diferencias en calidad y funcionalidad de los espermatozoides congelados-descongelados de porcino podrían estar relacionadas con diferencias en la expresión proteica entre dichos espermatozoides.

Bibliografía

Bromfield JJ. A role for seminal plasma in modulating pregnancy outcomes in domestic species. Reproduction. 2016, 152:R223-R232.

Caballero I, Parrilla I, Almiñana C, del Olmo D, Roca J, Martínez EA, Vázquez JM. Seminal plasma proteins as modulators of the sperm function and their application in sperm biotechnologies. Reprod Domest Anim. 2012, 47:12-21.

Hedrick VE, LaLand MN, Nakayasu ES, Paul LN. Digestion, Purification, and Enrichment of Protein Samples for Mass Spectrometry. Curr Protoc Chem Biol. 2015, 7:201-22.

Rahman MS, Kwon WS, Pang MG. Prediction of male fertility using capacitation-associated proteins in spermatozoa. Mol Reprod Dev. 2017, 84:749-759.

Rodriguez-Martinez H, Kvist U, Ernerudh J, Sanz L, Calvete JJ. Seminal plasma proteins: what role do they play? Am J Reprod Immunol. 2011, 66:11-22.

Yeste M. Sperm cryopreservation update: Cryodamage, markers, and factors affecting the sperm freezability in pigs. Theriogenology. 2016, 85:47-64.

SUMMARY

Introduction

Proteins, from both spermatozoa and the seminal plasma (**SP**), play important roles in the success of fertilization (Rodriguez-Martinez et al., 2011; Caballero et al., 2012). They are involved in major sperm functions, i.e. motility, capacitation, oocyte binding and acrosome reaction (Caballero et al., 2012). They also facilitate the "safe passage" of spermatozoa throughout the uterus and oviduct and contribute to embryo implantation (Bromfield, 2016). Moreover, in porcine, alike other mammalian species, some specific sperm- or SP-proteins have been identified as potential **biomarkers** of both fertility and sperm freezability (Yeste, 2016; Rahman et al., 2017). These findings are promising and encourage further large-scale studies of the proteome of both spermatozoa and SP in pigs using proteomics. Mass spectrometry (**MS**)-based technologies are currently the pivotal techniques for proteomic analyses (Hedrick et al., 2015). Accordingly, this **PhD Thesis** aimed to increase current knowledge of the boar SP and sperm proteome and to identify quantitative differences of proteins with direct impact in both sperm function and fertility.

The PhD Thesis is built by five separate studies. The first one, already published in **Journal of Proteomics** (**2016, 142:15-23**)**,** was designed to decode the proteome of pig SP, with attention to possible proteomic differences among the most recognized ejaculate fractions, specifically the first 10 mL of the sperm-rich ejaculate fraction (**SRF**), the rest of the SRF and the post-SRF. From these results, a second study (**Data in Brief 2016, 8:1370-1373**) was designed to generate an accessible database of pig SP-proteome, including an extensive review of the putative reproductive roles of the identified proteins, and thus to provide a valuable source for improving our understanding of the role of SP in sperm reproductive performance. The third study, in addition to updating the pig SP-proteome, was designed to identify fertility biomarkers among the decoded pig SP-proteins (**Journal of Proteome Research 2018, 17:1065-1076**). The fourth and fifth studies were focused on pig sperm proteome starting with the comparison of the proteome of spermatozoa retrieved from of the above mentioned relevant fractions of the pig ejaculate (fourth study, manuscript submitted to **Molecular and Cellular Proteomics**). The fifth study focused on frozen-thawed (**FT**) pig spermatozoa and it was designed to identify proteins differentially expressed between FT-sperm from semen samples showing clear differences in freezability (manuscript submitted to **Scientific Reports**).

Material and methods

Animals, sperm and seminal plasma sources and fertility records

Ejaculates and epididymes were obtained from healthy and sexually mature boars of different breeds, routinely used in artificial insemination (**AI**)-programs and slaughtered due to genetic replacement. Ejaculates were collected either as separate fractions (gloved-hand method) or complete in a single vial (semi-automatic collection method) and only those fulfilling current standards of sperm quality thresholds were used. The ejaculate fractions collected were the first 10 mL of the SRF, the SRF (entire or without the first 10 mL) and the post-SRF. The cauda epididymal contents was collected by retrograde infusion of air in the ductus.

Seminal plasma and fresh spermatozoa (from ejaculate and epididymis) used for proteomics were obtained after centrifugation (1,500 *xg*, 10 min, twice) and they were stored at -80 °C until analysis. Standard cryopreservation in 0.5-mL straw, which included a freezing and thawing rates of -40 ºC/min and 1,200 ºC/min, respectively, was used for generating FT-sperm samples.

For fertility records, multiparous (1–7 farrowings) sows were cervically inseminated twice using 24-72 h liquid-stored semen AI-doses (2,500 x 10^6 sperm extended to 80 mL of a commercial extender). Boar fertility was recorded over a 12-month period in terms of farrowing rate (**FR,** number of farrowing sows respect to the number of inseminated sows) and litter size (**LS,** total number of piglets born per litter). Fertility data were recorded as direct boar effect after data correction for parameters related to farm and sow.

Assessment of post-thaw sperm quality

Post-thaw sperm quality, including functional attributes, was evaluated 30 min after thawing in terms of total and progressive motility, cell viability, nuclear DNA fragmentation, lipid peroxidation (**LPO**) and early apoptosis-like changes. Motility was evaluated using a computer assisted sperm analyzer (**CASA**) system and DNA nuclear fragmentation using the Sperm-Sus-Halomax® kit. The remaining sperm parameters were cytometrically assessed after labelling sperm with appropriate fluorochromes.

Proteomic analysis

The proteome analyses were carried out in the Proteomics Unit of the University of Valencia, Valencia, Spain (member of the PRB2-ISCIII ProteoRed Proteomics Platform).

Proteomic analysis of seminal plasma and pig spermatozoa

The SP-samples were thawed at room temperature and a protease inhibitor cocktail was added to avoid protein degradation. Total protein concentration was measured using a Qubit fluorometer and an aliquot was subjected to trypsin digestion. Another aliquot was loaded in a one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (**1D SDS-PAGE**) before in-gel trypsin digestion. Resulting peptides were fractionated by size exclusion chromatography (**SEC**) and solid phase extraction (**SPE**), to reduce the protein complexity of the SP-samples.

Once thawed, the FT-sperm samples were washed through a Percoll monolayer gradient to separate the sperm cells from debris and egg-yolk remnants. Then, total protein of both fresh and FT-sperm pellets was extracted using UTC lysis buffer, quantified by RCDC Lowry and subjected to 1D SDS-PAGE to remove UTC interferences. Finally, resulting proteins were in-gel digested by trypsin. The generated peptides from SP and sperm-samples were examined by liquid chromatography-tandem mass spectrometry (**LC-MS/MS**) using a NanoLC Ultra 1D plus Eksigent connected to an AB SCIEX TripleTOF 5600 mass spectrometer. The resulting SCIEX.wiff data-files were processed using ProteinPilot v5.0 search engine. Then, the Paragon algorithm was used to collate the mass spectrometry data with the National Center for Biotechnology Information non-redundant peptide sequence database (**NCBInr**) or in UniProt_mammals. The identified proteins were those assuming having a peptide confidence threshold of 95% and a false discovery rate (**FDR**) lower than 1% at protein level.

Protein quantification in seminal plasma and spermatozoa

Sequential window acquisition of all theoretical mass spectra (**SWATH**) and isobaric tags for relative and absolute quantitation (**iTRAQ**) were used for protein quantification. For SWATH, digested samples were randomly loaded without labeling onto the LC column and the eluted peptides were analyzed with the TripleTOF 5600 operating in swath mode.

The resulting Protein-Pilot group file was loaded into PeakView® (v2.1) and peaks from SWATH runs were extracted and normalized by total sum, and the sum of all areas under the peaks was equalized for all samples. The protocol of 8-plex iTRAQ labeling kit was used for running the iTRAQ procedure. Briefly, cysteine residues of all the denatured protein samples were blocked and the proteins were digested with trypsin. The resulting peptide mixtures were labeled with appropriate iTRAQ reagents by incubating the samples at room temperature for 3 h. The labeled peptides were subjected to isoelectrofocusing (**IEF**) separation followed by LC-MS/MS analysis. The relative quantification of the sperm proteins was achieved by comparison of the relative intensities of reporter ions of different sperm samples.

Bioinformatics analysis

The bioinformatics of identified and validated SP- and sperm-proteins were manually performed using the comprehensive bioinformatics tool for functional annotation UniProtKB database (www.uniprot.org) in combination with PANTHER (Protein Analysis THrough Evolutionary Relationships, www.pantherdb.org) and DAVID (Database for Annotation, Visualization and Integrated Discovery, www.david.ncifcrf.gov).

Statistical analysis

The statistical software packages IBM SPSS v19, R and Multiexperiment Viewer (**MeV**) were used. The mean quantity of each protein in every sample and the fold-change (**FC**, ratio between the mean of samples) among samples were calculated. Principal Component Analysis (**PCA**) was used to discriminate SP or sperm proteome among samples. Student's t-test and ANOVA test were used in each sample type to identify the protein differentially expressed among samples and to evaluate the influence of sample on post-thaw sperm quality. A penalized linear regression model **LASSO** (least absolute shrinkage and selection operator) was used to quantitatively identify those differently expressed SP-proteins among boars exhibiting differences in fertility. The explanatory ability of those differentially expressed proteins was depicted using heatmaps after z-score normalization. A value of $P < 0.05$ was accepted as the minimal level of significance.

Experimental designs and results

The **First Study** (**Journal of Proteomics 2016, 142:15–23**) aimed to decode the boar SP-proteome and to identify differences in protein expression among the SP of specific ejaculate fractions/portions, namely the first 10 mL of the SRF, the rest of the SRF and the post-SRF. To characterize the boar SP-proteome, a SP-pool from 33 entire ejaculates (11 boars, 3 ejaculates/boar) was used. The SP of twenty split-collected ejaculates from five boars (4 ejaculates/boar) was used to characterize differences between the ejaculate fractions in the protein expression. The SP-samples were analyzed using a combination of SEC, 1D SDS-PAGE and NanoLC-ESI-MS/MS followed by functional bioinformatics. The identified proteins were quantified by the SWATH approach. A total of 536 SP-proteins were identified, 409 of them encoded in *Sus scrofa* taxonomy (374 validated with ≥99% confidence). Thirty-four proteins (16 identified in *Sus scrofa* taxonomy) were differentially expressed among ejaculate fractions/portions. Sixteen of them were over-expressed and 18 under-expressed in the SRF split (first 10 mL of SRF and the rest of SRF) compared to the post-SRF. This proteomic study provides the major characterization of the boar SP-proteome with more than 250 proteins firstly reported. Additionally, the comparative SP-proteome study resulted in the identification of several differentially expressed proteins among ejaculate portions/fractions, many of which are linked to sperm reproductive performance.

The **Second Study** (**Data in Brief 2016, 8:1370-1373**) displayed selected results generated in the first study to provide an accessible database showing comprehensive information and references of putative reproductive functionality of the proteins identified in the boar SP.

The **Third Study** (**Journal of Proteome Research 2018, 17:1065-1076**) aimed to identify boar-SP proteins differently expressed between AI-boars having significant differences in fertility outcome. Previously, the boar SP-proteome was decoded by using a novel proteomic approach that combined SEC and SPE (to reduce the complexity of protein composition of the SP-samples) as pre-fractionation steps prior to Nano LC-ESI-MS/MS analysis. This novel pre-fractionation approach revealed a total of 872 SP-proteins, 390 of them specifically belonging to *Sus scrofa* taxonomy. Then, the SP-

proteomes of 26 boars showing significant differences in FR (n=13) and LS (n=13) after the AI of 10,526 sows were analyzed. A total of 679 SP-proteins were then quantified by the SWATH approach where the penalized linear regression LASSO revealed 11 differentially expressed SP-proteins for FR (FURIN, AKR1B1, UBA1, PIN1, SPAM1, BLMH, SMPDL3A, KRT17, KRT10, TTC23 and AGT) and four for LS (PN-1, THBS1, DSC1 and CAT). This study extended our knowledge of the SP-proteome and revealed some SP-proteins as potential biomarkers of fertility in AI-boars.

The **Fourth Study** (manuscript submitted to **Molecular and Cellular Proteomics**) aimed firstly to update the porcine sperm proteome and secondly to identify proteins differentially expressed among mature spermatozoa stored in the cauda epididymis and those delivered in selected portions of split-ejaculates. To achieve these objectives, nine ejaculates from nine mature and fertile boars were manually collected in three separate ejaculate portions: first 10 mL of the SRF, the rest of the SRF and the post-SRF. The spermatozoa from cauda epididymis were collected after slaughter of the same boars by retrograde infusing the duct with air. All sperm samples were subjected to iTRAQ-based LC-MS/MS for protein identification and quantification. A total of 1,723 proteins were identified and 1,602 of them were also quantified. After an ANOVA test, 32 *Sus scrofa* proteins showed quantitative differences ($P<0.01$) among the sperm samples, differences particularly relevant for the functionality of spermatozoa retrieved from the post-SRF. The present study is the first showing quantitative differences in the protein profile of mature spermatozoa, involving proteins clearly implicated in sperm function; proving the protein profile of boar spermatozoa is remodelled during ejaculation. These findings provide a valuable groundwork for further studies focused on identifying protein biomarkers of sperm fertility.

The **Fifth Study** (manuscript submitted to **Scientific Reports**) carried out two different experiments aimed to unveil differences in protein expression between FT-semen samples differing in sperm quality and function (assessed as indicated above). Differences in protein expression were measured using a LC-MS/MS-based SWATH approach. The first experiment compared the proteome of FT-spermatozoa from the SRF with that of the entire ejaculate. The second experiment compared the proteome of FT-spermatozoa from each one of the three ejaculate fractions/portions explored, each showing clear differences in sperm freezability. In Experiment 1, FT-spermatozoa from the SRF showed better quality than those from the entire ejaculate and 26 *Sus scrofa* proteins with functional

sperm relevance were differentially expressed (FC ≥ 1.5) between FT-spermatozoa of either source. In Experiment 2, FT-spermatozoa from the first 10 mL of the SRF and the rest of the SRF were qualitatively better than those from the post-SRF, with 187 proteins differentially expressed among the three ejaculate split sources. These results clearly showed that differences in proteome expression could explain differences in sperm attributes between FT-sperm samples. These results have clear biological and practical implications considering that swine AI-centers are moving from selectively collecting the SRF to collecting the entire ejaculate, a change that can have consequences for the quality of the AI-doses prepared.

Conclusions

1.- The present study provides the largest and most updated proteome, encoded in Sus scrofa taxonomy, of the pig seminal plasma and spermatozoa.

2.- The proteome expression of seminal plasma and pig spermatozoa differs among the different portions/fractions of the ejaculate and the differences involve proteins clearly implicated in sperm function.

3.- The proteomic profile of seminal plasma differs among boars with consistent differences in fertility; the comparative proteome identifying a panel of seminal plasma proteins potentially defined as biomarkers of fertility.

4.- Differences in quality and functionality between frozen-thawed sperm samples would be linked to differences in sperm proteome expression.

References

Bromfield JJ. A role for seminal plasma in modulating pregnancy outcomes in domestic species. Reproduction. 2016, 152:R223-R232.

Caballero I, Parrilla I, Almiñana C, del Olmo D, Roca J, Martínez EA, Vázquez JM. Seminal plasma proteins as modulators of the sperm function and their application in sperm biotechnologies. Reprod Domest Anim. 2012, 47:12-21.

Hedrick VE, LaLand MN, Nakayasu ES, Paul LN. Digestion, Purification, and Enrichment of Protein Samples for Mass Spectrometry. Curr Protoc Chem Biol. 2015, 7:201-22.

Rahman MS, Kwon WS, Pang MG. Prediction of male fertility using capacitation-associated proteins in spermatozoa. Mol Reprod Dev. 2017, 84:749-759.

Rodriguez-Martinez H, Kvist U, Ernerudh J, Sanz L, Calvete JJ. Seminal plasma proteins: what role do they play? Am J Reprod Immunol. 2011, 66:11-22.

Yeste M. Sperm cryopreservation update: Cryodamage, markers, and factors affecting the sperm freezability in pigs. Theriogenology. 2016, 85:47-64.

ABBREVIATIONS

ABBREVIATIONS

1D	One dimensional
2D	Two dimensional
ACN	Acetonitrile
AI	Artificial insemination
AUC	Area under curve
AuNPs	Gold nanoparticles
BSP	Bovine seminal plasma proteins
BTS	Beltsville Thawing Solution
CASA	Computer assisted sperm analysis
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
DAVID	Database for Annotation, Visualization and Integrated Discovery
DBE	Direct boar effect
DIGE	Differential proteomics in gel electrophoresis
DTT	Dithiothreitol
EDMA	Ethylene dimethacrylate
EE	Entire ejaculate
ESI	Electrospray ionization
FA	Formic acid
FC	Fold change
FDR	False discovery rate
FITC-PNA	Fluorescein isothiocyanate-conjugated peanut agglutinin
FR	Farrowing rate
FT	Frozen-thawed
GMA	Glycidyl methacrylate
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
H42	Hoechst 33342
IAM	Iodoacetamide
ICAT	Isotope coded affinity tag
IEF	Isoelectric focusing
IPG	Immobilized pH gradient
iTRAQ	Isobaric tags for relative and absolute quantitation
LASSO	Least absolute shrinkage and selection operator
LC	Liquid chromatography

XII. Abbreviations

LFQ	Label free quantification
LN_2	Liquid nitrogen
LPO	Lipid peroxidation
LS	Litter size
MALDI-MS	Matrix-assisted laser/desorption ionization-mass spectrometry
MeV	Multiexperiment viewer
MMTS	S-methyl methanethiosulphonate
MS	Mass pectrometry
MS/MS	Tandem mass spectrometry
NCBInr	National Center for Biotechnology Information-non redundant
PANTHER	Protein Analysis THrough Evolutionary Relationships
PBS	Phosphate buffer solution
PCA	Principal component analysis
PG	Prostaglandin
PI	Propidium Iodide
PMF	Peptide mass fingerprinting
PNA-FITC	Fluorescein-conjugated peanut agglutinin
ROS	Reactive oxygen species
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SILAC	Stable isotope labeling with amino acids in cell culture
SP	Seminal plasma
SPE	Solid phase extraction
SRF	Sperm-rich ejaculate fraction
SWATH	Sequential window acquisition of all theoretical mass spectra
TCA	Tricloroacetic acid
TCEP	Tris (2-carboxyethyl) phosphine
TEAB	Triethylammonium bicarbonate buffer
TFA	Trifluoroacetic acid
TOF	Time of flight
UTC	7M Urea, 2M thiourea, 4% CHAPS
VTH	Veterinary teaching hospital
ZP	Zona pellucida

GRAPHIC APPENDIX

XIII. Graphic appendix

Facilities of the artificial insemination center where the studies were run. Exterior (**A**) and inner (**B, C**) views of the artificial insemination center located in Calasparra (Murcia) and belonging to AIM Iberica (Topigs Norsvin España, Madrid, Spain).


A

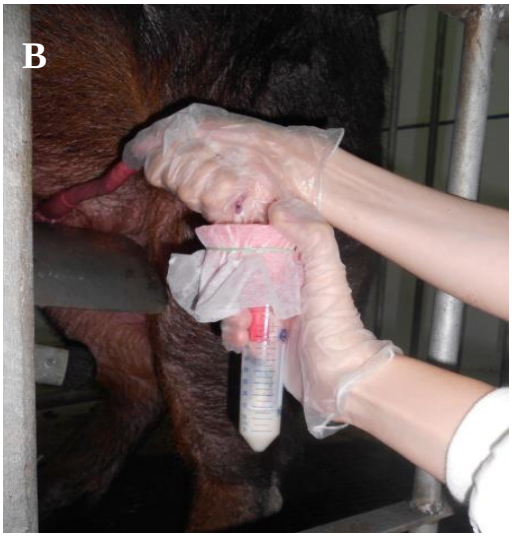

B


C

XIII. Graphic appendix

Procedures used for the collection of ejaculates. The gloved-hand method (**A, B**) and the semi-automatic method Collectis® (**C, D**).


A


B

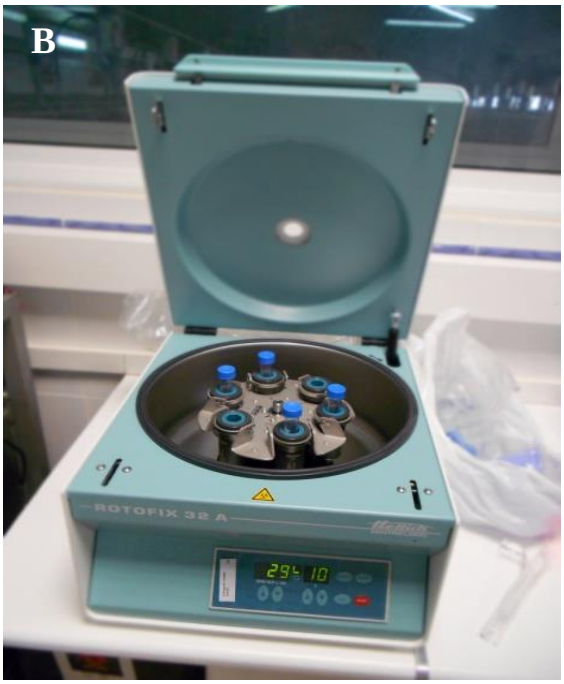

C

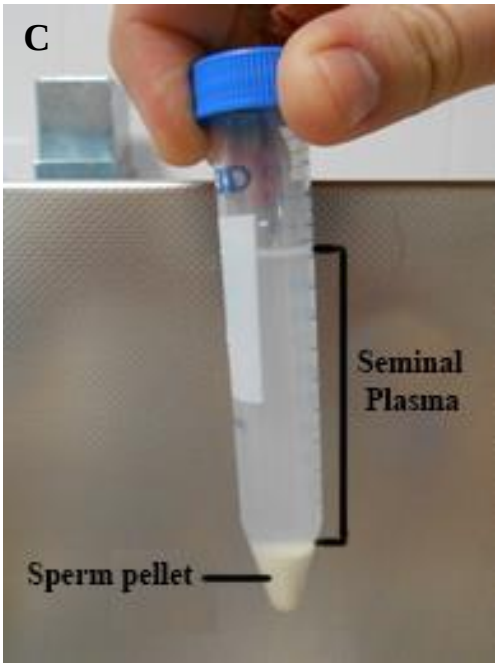

D

XIII. Graphic appendix

Harvest and storage of spermatozoa and seminal plasma samples. Semen samples deposited into 15-mL centrifuge tube (**A**), centrifugation (Rotofix 32A, Hettich Zentrifugen, Tuttlingen, Germany) (**B**), separation of the sperm pellet from seminal plasma (**C**), storage at -80 °C (**D**).


A
14
13
12
11
10
9
8
7
6
5
4
3
2

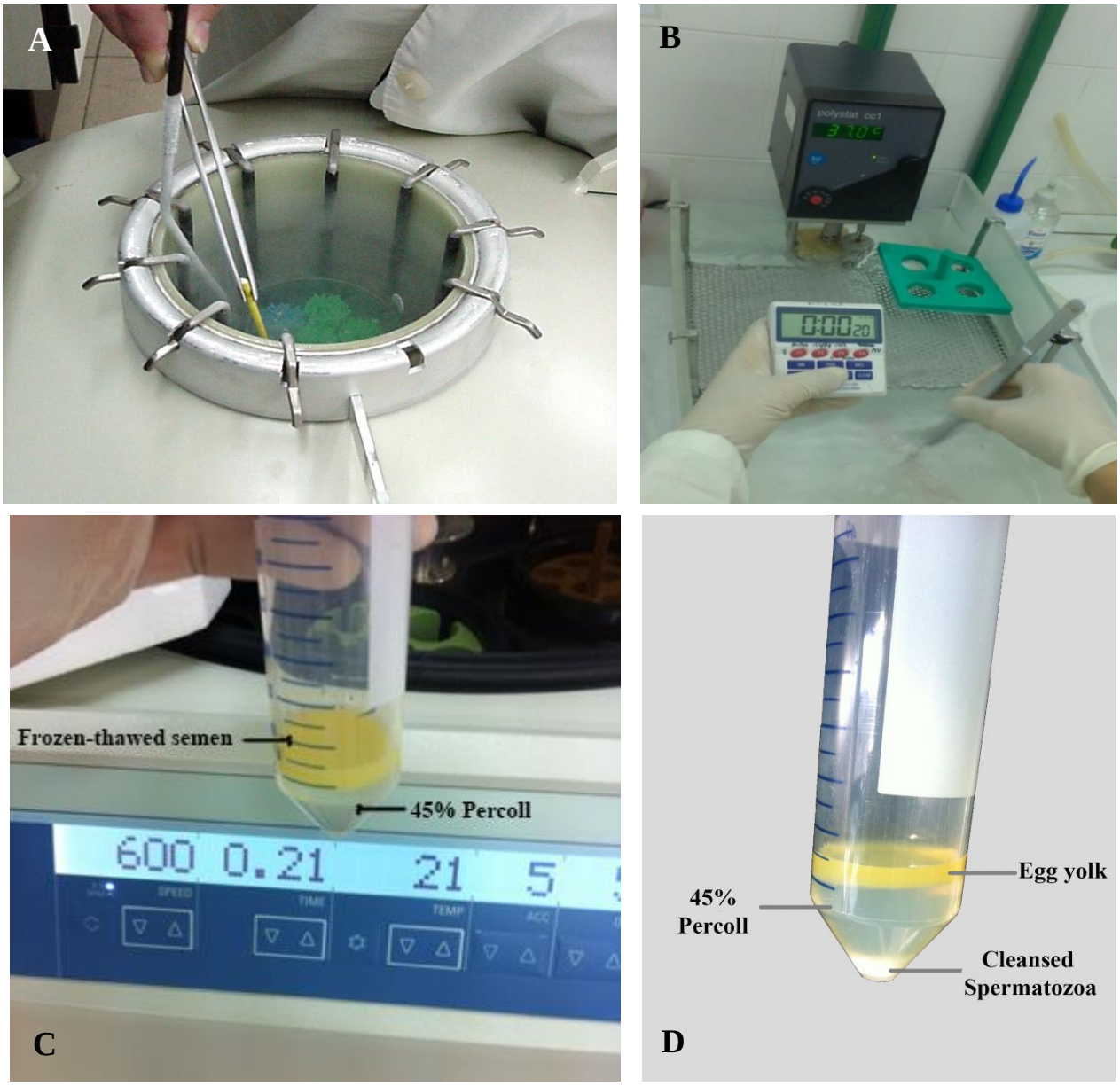

B
ROTOFIX 32 A


C
Seminal Plasma
Sperm pellet


D
-79 °C

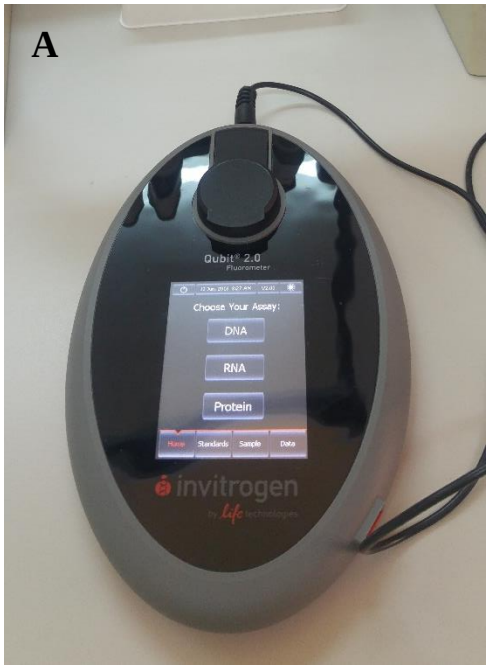
XIII. Graphic appendix

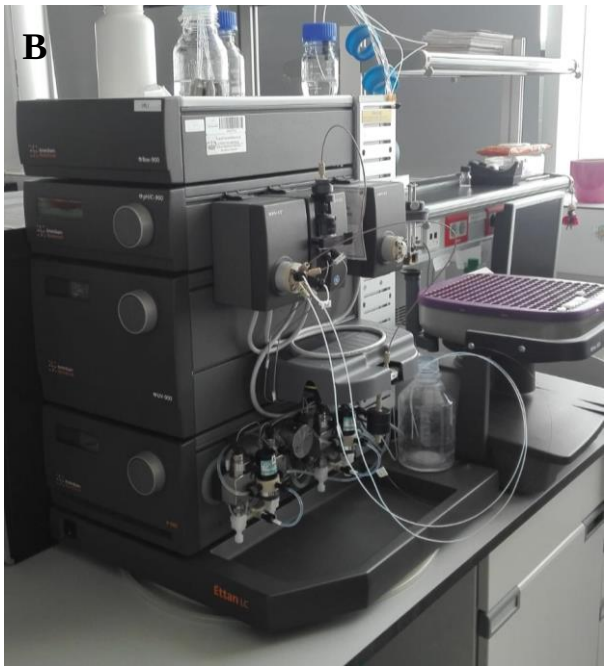
Handling of frozen-thawed boar spermatozoa for proteomic analysis. Storage of the straws of frozen semen in liquid nitrogen (GT40®, Air Liquide, Paris, France) (**A**), thawing of straws in a circulating water bath at 37 °C for 20 s (**B**), placement of the thawed semen on top of the Percoll® monolayer gradient (**C**), the separated layers of egg-yolk and cleansed spermatozoa after centrifugation through the density gradient (**D**).


A
B
Frozen-thawed semen
45% Percoll
600 0.21 21 5
C
Egg yolk
45%
Percoll
Cleansed
Spermatozoa
D

XIII. Graphic appendix

Instruments used for total protein quantification and fractionation of boar seminal plasma proteins. Qubit fluorometer (Invitrogen, Carlsbad, CA, USA) for total protein quantification (**A**). ETTAN LC system (GE Healthcare Life Sciences, Little Chalfont, United Kingdom) for protein fractionation by size exclusion chromatography (**B, C**).

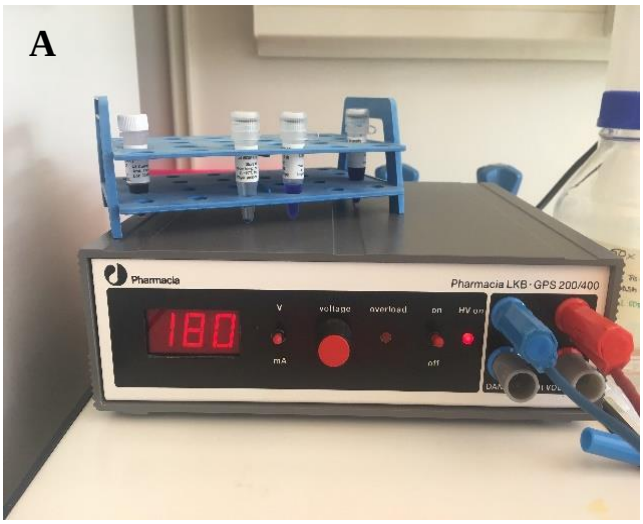

A
Qubit® 2.0
Fluorometer
Choose Your Assay:
DNA
RNA
Protein
Home
Standards
Sample
Data
invitrogen
by life technologies

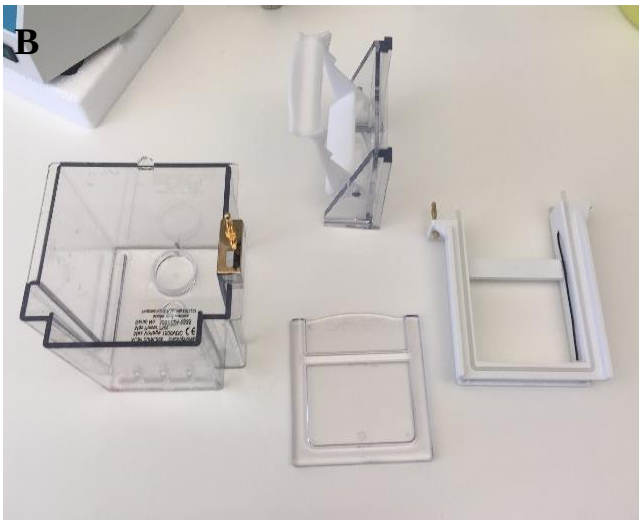

B


C
Warning!
Frac-950
Amersham Biosciences

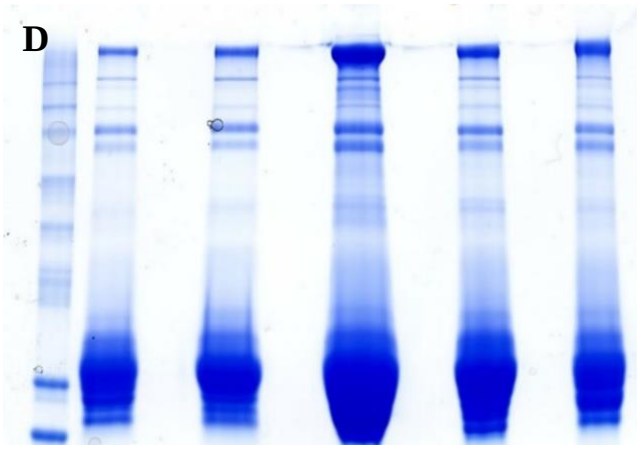
XIII. Graphic appendix

Gel electrophoresis instrumentation. Power supply (**A**), gel electrophoresis chamber (**B**), electrophoresis reagents (**C**) and polyacrylamide gel after electrophoresis step of seminal plasma samples (**D**).


A
Pharmacia
Pharmacia LKB · GPS 200/400
180
V
mA
voltage
overload
on
off

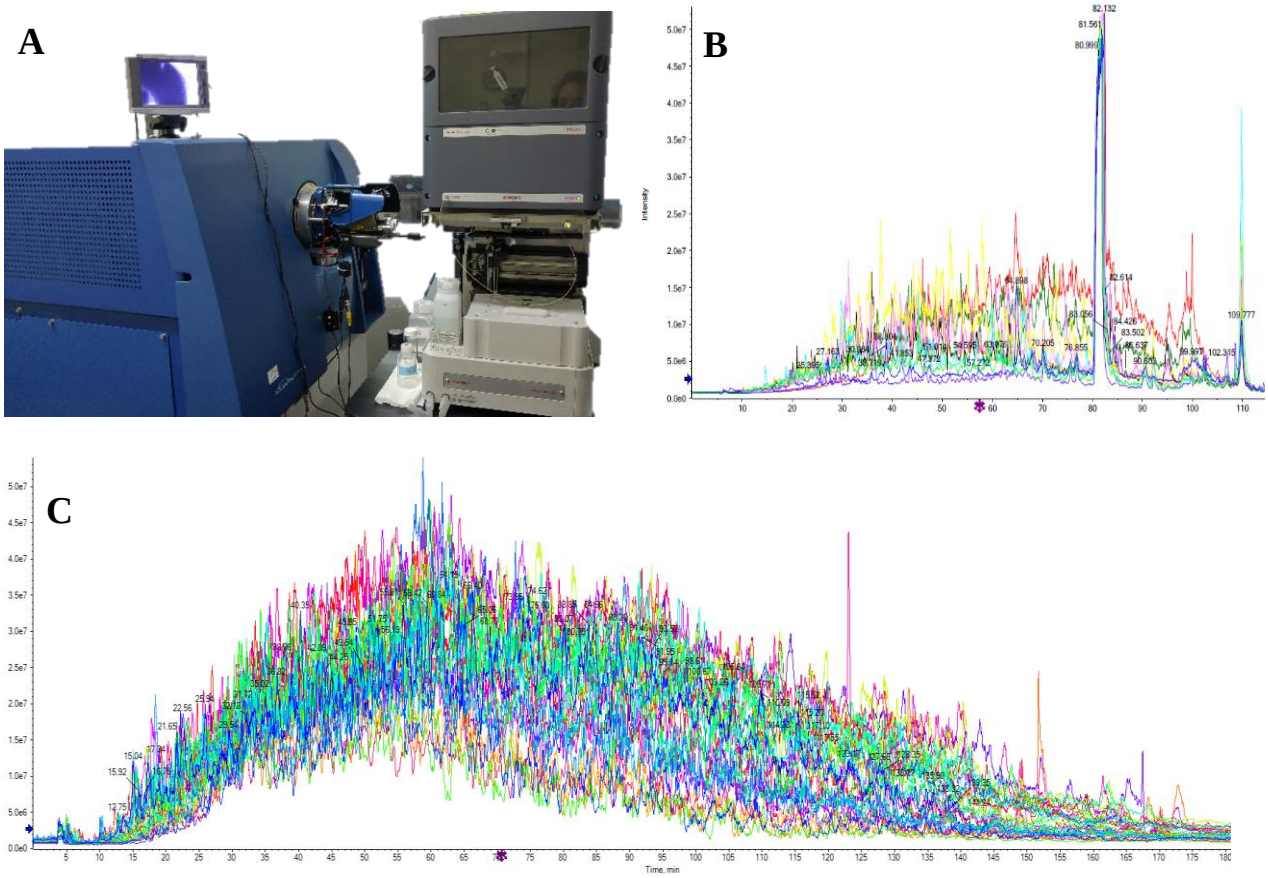

B


C
invitrogen
SeeBlue® Plus2 Prestained
Standard
Qty: 500 µL
1891868
2018-01-31
For Research Use Only.
Not for use in diagnostic procedures.
24-hour Medical or Spill Emergency Response (Chemtrec)
North America +1 866 536 0631 / Other +1 301 431 8586
protein
invitrogen
Antioxidant


D

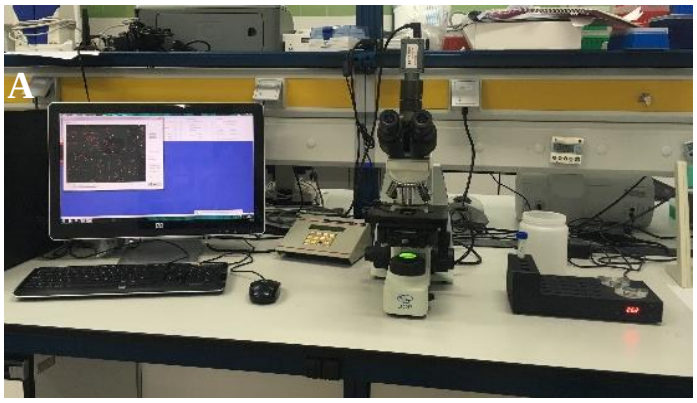
XIII. Graphic appendix

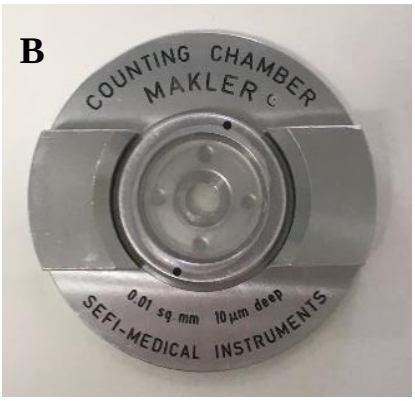
Protein quantification by liquid chromatography tandem mass spectrometry (LC-MS/MS). Mass spectrometer nanoESI qQTOF (5600 TripleTOF, ABSCIEX) (**A**), total ion current chromatogram from both isobaric tags for relative and absolute quantitation (iTRAQ) (**B**) and sequential window acquisition of all theoretical mass spectra (SWATH) (**C**).


A
B
C
Intensity
Time, min

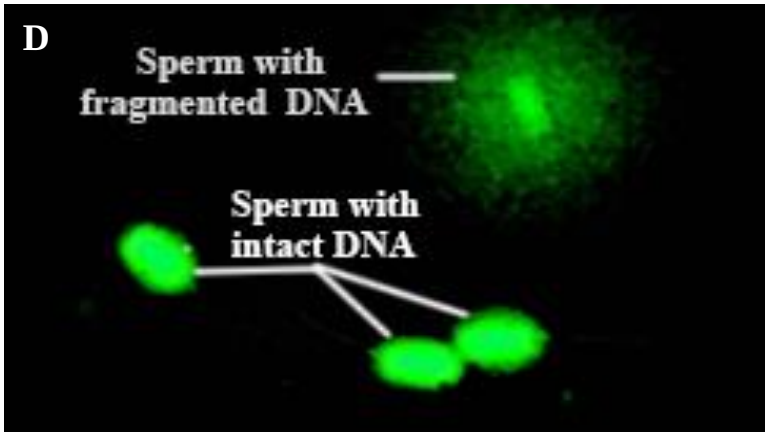
XIII. Graphic appendix

Assessment of sperm attributes. Sperm motility was objectively evaluated using a computer-assisted sperm analyzer (CASA) system (ISASV1®, Proiser R+D S.L, Paterna, Spain) (**A**) using a Makler chamber (**B**). Sperm viability and functionality was assessed using the flow cytometer BD FACSCanto II (Becton Dickinson Co, Franklin Lakes, NJ, EUA) (**C**). Sperm DNA fragmentation was labeled using the Sperm-Sus-Halomax® kit (Halotech DNA, Madrid, Spain) (**D**) and visualized using a fluorescent microscopy (Nikon Eclipse E800, Nikon Instruments Corporation, Melville, New York, USA) (**E**).


A


B
COUNTING CHAMBER
MAKLER
0.01 sq. mm 10 µm deep
SEFI-MEDICAL INSTRUMENTS


C
BD FACSCanto II


D
Sperm with fragmented DNA
Sperm with intact DNA


E