***Post print version corresponding to:*** *"Castillo J, Jodar M, Oliva R. The contribution of human sperm proteins to the development and epigenome of the preimplantation embryo.Hum Reprod Update. 2018 Sep 1;24(5):535-555. doi: 10.1093/humupd/dmy017."*

# The contribution of human sperm proteins to the development and epigenome of the pre-implantational embryo

Running title: Sperm proteome, fertilization and the early embryo

Judit Castillo[1], Meritxell Jodar[1*] and Rafael Oliva[1,2*]

[1]Molecular Biology of Reproduction and Development Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Fundació Clínic per a la Recerca Biomèdica, Faculty of Medicine, University of Barcelona, Casanova 143, 08036 Barcelona, Spain

[2]Biochemistry and Molecular Genetics Service, Hospital Clínic, Villarroel 170, 08036 Barcelona, Spain

*Corresponding authors: roliva@ub.edu ORCID iD 0000-0003-4876-2410 (R Oliva) and meritxell.jodar@ub.edu ORCID iD 0000-0002-3272-0163 (M Jodar)

**Table of contents:**

**Abstract**

Background: The knowledge of the proteomic composition of the gametes is essential to understand the reproductive function. Most of the sperm proteins are related to spermatogenesis and sperm function, but less abundant protein groups with potential post-fertilization roles are also detected. The current data is challenging our understanding of the sperm biology and functionality, demanding an integrated analysis of the proteomic and RNA-seq datasets available for the spermatozoa, oocyte and the early embryo, in order to unravel the impact of the male gamete on the generation of the new individual.

Objective and rationale: The aim of this review is to identify, infer the origin, and discuss the relevance of human sperm proteins during oocyte fecundation, pre-implantational embryogenesis, and epigenetic inheritance.

Search methods: The scientific literature was comprehensively searched for proteomic studies on the human sperm, oocyte, embryo, and additional reproductive cells and fluids. Proteins were compiled and functionally classified according to Gene Ontology annotations, and the mouse phenotypes integrated into the Mouse Genome Informatics database. High-throughput RNA datasets were used to decipher the origin of embryo proteins. The tissue origin of the sperm proteins was inferred on the basis of RNA-seq and protein data available in the Human Protein Atlas database.

Outcomes: So far, 6871 proteins have been identified and reported in the sperm, 1376 in the oocyte, and 1300 in the blastocyst. With a deeper analysis of the sperm proteome, 103 proteins with known roles in the processes of fertilization and 93 in early embryo development have been identified. Additionally, 560 sperm proteins have been found to be involved in modulating gene expression by the regulation of transcription,

DNA methylation, histone post-translational modifications, and non-coding RNA biogenesis. Some of these proteins may be critical for gene expression regulation after embryo genome activation and, therefore, be potentially involved in the epigenetic inheritance of acquired traits. Furthermore, the integrative analysis of the sperm, oocyte and embryo proteomes and transcriptomes revealed a set of embryo proteins with an exclusive paternal origin, being some of them crucial for correct embryogenesis, and most likely for the modulation of the offspring phenotype. The analysis of the expression of sperm proteins, at both RNA and protein levels, in tissues not only from the male reproductive tract but also from peripheral organs, has predicted a putative extra-testicular origin for some sperm proteins. This suggests that these proteins might be imported into sperm from the accessory sex glands and other tissues, most likely through exosomes.

Wider implications: The integrative proteome and transcriptome analyses focused on specific groups of proteins, rather than in enriched pathways, identified important sperm proteins involved in early embryogenesis and provides further evidence to support the hypothesis of the epigenetic inheritance of specific acquired traits. The putative extra-testicular origin of some sperm proteins suggests not only the involvement of accessory sex glands in fertilization, and epigenetic information transmission, but also that some proteins from additional organs may contribute information to the offspring phenotype. These findings should stimulate further research in the field.

*Keywords:* sperm, proteomics, embryo development, fertilization, epigenetic inheritance, RNA, extra-testicular.

## Introduction

Classically, the role and the relative contribution of the sperm cell to the embryo has not been fully appreciated, and nearly minimized only to the contribution of the paternal DNA (Miescher, 1874; Sutton, 1903; Baccetti and Afzelius, 1976; Dahm, 2005; Maruyama and Singson, 2006). However, taking the advantage of the current high-throughput "omic" technologies, growing evidence supports the idea that the male gamete is much more than a vehicle to deliver half of the DNA to the new individual. In fact, the sperm cell provides DNA containing different epigenetic marks, such as the DNA methylation (Hammoud *et al.*, 2010; Siklenka *et al.*, 2015), the post-translational modifications (PTMs) of histones and protamines (Carrell *et al.*, 2008; Brykczynska *et al.*, 2010; Brunner *et al.*, 2014; Castillo *et al.*, 2015; Siklenka *et al.*, 2015), and a differential distribution of genes within the nucleohistone and the nucleoprotamine chromatin domains (Arpanahi *et al.*, 2009; Hammoud *et al.*, 2009, 2011; Erkek *et al.*, 2013; Castillo *et al.*, 2014a) to the zygote. All those epigenetic marks are also combined with a complex population of sperm RNAs and proteins (de Mateo *et al.*, 2011; Jodar *et al.*, 2013, 2015; Amaral *et al.*, 2014a; Castillo *et al.*, 2014a, 2014b), some of which seem to be crucial for early embryogenesis and the future health of the offspring and further generations (Krawetz, 2005; Carrell, 2012; Rando, 2012; Castillo *et al.*, 2015; Chen *et al.*, 2016b).

Focusing on the sperm proteomic contribution to the zygote, the application of strategies based on liquid chromatography (LC) coupled to tandem mass spectrometry (MS/MS) has resulted in the high-confidence identification of thousands of sperm proteins (Baker *et al.*, 2013; Wang *et al.*, 2013; Amaral *et al.*, 2014a; Vandenbrouck *et al.*, 2016; Jodar *et al.*, 2017). However, in order to fully understand the impact of the

sperm proteome to the developing embryo, it is important to also consider the different steps taking place in the biogenesis of both male and female gametes, as well as the initial proteomic changes on the pre-implantational embryo development. The sperm cell is the end cellular product of spermatogenesis, a highly complex process of differentiation that involves very marked genetic, chromatin, biochemical, structural, and cellular changes (Figure 1) (Davies and Mann, 1947; Fawcett and Chemes, 1979; Mezquita, 1985; Poccia, 1986; Oliva and Dixon, 1991; Green *et al.*, 1994; de Kretser *et al.*, 1998; Kimmins and Sassone-Corsi, 2005; Sutovsky and Manandhar, 2006; Oliva, 2006; Carrell *et al.*, 2016). Once released by the testis, the spermatozoa travel through the epididymis, and all along the male reproductive tract, coming into intimate contact with secretions from the different accessory sex glands. This contributes to the acquisition of the potential for sperm motility, and the ability to fertilize the oocyte (Saez *et al.*, 2003; Sullivan *et al.*, 2005; Dacheux and Dacheux, 2014; Sullivan and Mieusset, 2016). In contrast, the oogenesis results in one oocyte accumulating most of the cytoplasm and small polar bodies, which receive little more than a haploid nucleus (Mattson and Albertini, 1990; Bukovsky *et al.*, 2005; Zuccotti *et al.*, 2011; Coticchio *et al.*, 2015; Conti and Franciosi, 2018) (Figure 1). Interestingly, and similarly to the sperm development, the interactions among the oocyte and its surroundings (granulosa cells and the follicular fluid) are also crucial for the correct oocyte maturation (Gilula *et al.*, 1978; Eppig, 1985; Driancourt and Thuel, 1998; Salustri *et al.*, 2003; Chang *et al.*, 2016; Richani and Gilchrist, 2018) (Figure 1). Although most of the proteins present in the mature sperm and oocyte are synthesized during spermatogenesis and oogenesis, respectively, it is important to note that some proteins may be imported from the surroundings and supporting reproductive fluids and cells to the male and female gametes (Salustri *et al.*, 2003; Gilchrist *et al.*, 2008; Baker *et al.*, 2012; Intasqui *et al.*,

2013; Ambekar *et al.*, 2015; Zamah *et al.*, 2015; Johnson *et al.*, 2015; Martin-DeLeon, 2015; Saez and Sullivan, 2016; Sullivan and Mieusset, 2016; Chang *et al.*, 2016; Machtinger *et al.*, 2016; Hamad, 2017; Jodar *et al.*, 2017). In fact, emerging evidences show the existence of an active communication and transference of proteins and other components between gametes and fluids, most probably through the exosome pathway (Johnson *et al.*, 2015; Jodar *et al.*, 2016; Machtinger *et al.*, 2016; Hamad *et al.*, 2017).

The protein profile of the human male gamete has been extensively studied during the past years, and up to 6238 different proteins had been previously compiled (Amaral *et al.*, 2014a; Jodar *et al.*, 2017). However, it remains to be elucidated whether some sperm proteins are required for early embryogenesis, or only represent spermatogenic leftovers required for spermatogenesis, sperm maturation, or sperm functions such as sperm motility. In the present review, the large amount of information and datasets currently available on the proteomic knowledge of gametes, pre-implantational embryo, and related reproductive fluids and supporting cells have been compiled, curated, integrated, and analyzed. This information has been combined with the corresponding RNA-seq datasets, in order to identify, infer the origin, discuss, and provide further evidences of the proteomic paternal contribution to the pre-implantational embryogenesis, and add additional support to the hypothesis of the epigenetic inheritance of some acquired traits through the paternal line.

Because of the large amount of data available, this review is mainly focused on the human model. However, and since very relevant data in different animal model systems are also available, the reader is referred to original articles and excellent reviews which are complementary to the subject (Wasbrough *et al.*, 2010; Skerget *et al.*,

2013, 2015; Kwon *et al.*, 2015; Zhou *et al.*, 2015; Baker, 2016; Holt and Fazeli, 2016; McDonough *et al.*, 2016; Ntostis *et al.*, 2017; Swegen *et al.*, 2017).

## Methods

The cellular and the fluid proteomes, corresponding to human sperm, oocyte, blastocyst, cumulus cells, mid-secretory endometrium, and follicular fluid, were compiled after an exhaustive literature search using the Web of Science database. Only proteomic articles published in English and available online until the end of September 2017 were included in the search. Subsequently, and in order to reduce as much as possible potential protein false-positive identifications, we selected only those proteomic studies whose protein detection relied on mass spectrometry (MS) approaches, and whose protein identification criteria included at least two peptides per protein with a false discovery rate (FDR) $\leq$ 5% for each peptide, which corresponds to an estimated FDR $\leq$ 0.25% per protein (Table 1). Only proteins with a known gene name were included.

For the generation of the human sperm proteome, proteomic profiles were acquired from: 1) original articles (de Mateo *et al.*, 2011; Amaral *et al.*, 2013; Baker *et al.*, 2013; Wang *et al.*, 2013; among others, see the complete list of references in Table 1) already compiled in the comprehensive reviews previously published by our group (Amaral *et al.*, 2014a; Jodar *et al.*, 2017), and 2) additional recently published sperm proteomic catalogues (Zhu *et al.*, 2013; Intasqui *et al.*, 2013; Azpiazu *et al.*, 2014; Frapsauce *et al.*, 2014; Sun *et al.*, 2014; Wang *et al.*, 2015, 2016; Yu *et al.*, 2015; Jumeau *et al.*, 2015; Liu *et al.*, 2015; Hetherington *et al.*, 2016; Vandenbrouck *et al.*,

2016; Carapito *et al.*, 2017; Saraswat *et al.*, 2017). Only those original articles in which specific methods were referenced to ensure the proper elimination of potentially contaminating cells were considered. In total, 46 published proteomic studies performed using purified ejaculated human sperm were included in the analysis reported herein, which are enlisted in Table 1. However, it must be noted that some specific characteristics of the males from which the sperm samples were obtained, such as their fertility, age, body mass index (BMI), life style, or exposure to toxins, were not always specified in the sperm proteomic studies included in this analysis. Thus, it is possible that some of the proteins detailed in the current dataset could reflect an altered phenotypic state. Additionally, in order to further reduce the chance of any false-positive identifications, just those proteins with identified unique peptides were integrated in the final list. This compilation resulted in high-confidence identification of 6871 non-redundant proteins in human spermatozoa (reported in Supplementary Table 1).

Two different strategies were used to infer the functional involvement of the sperm proteins at oocyte fecundation and consecutive processes of human reproduction. The first one was according to their association to the Gene Ontology (GO) Biological Process annotations enclosed in the Gene Ontology Consortium Database (http://www.geneontology.org/; (Ashburner *et al.*, 2000; The Gene Ontology Consortium, 2017)). The second one was according to their association to the Phenotypes & Mutant Alleles data from the Mouse Genome Informatics (MGI) database (http://www.informatics.jax.org/phenotypes.shtml). In particular, gene lists associated to GO terms related to fertilization, formation of the zygote and the blastocyst, and implantation were retrieved from de Gene Ontology Consortium

database (Release date 2016-12-07). The GO-derived list of genes was subsequently compared to the human sperm proteome dataset. Similarly, those sperm proteins functionally involved in epigenetic processes were predicted by retrieving gene lists associated to the GO terms related to the processes of epigenetics, transcription, DNA methylation, histone PTMs, and small non-coding RNA biogenesis from the Gene Ontology Consortium database (Release date 2016-12-07). The complete list of GO terms included in these analyses is described in Supplementary Table 2. However, it cannot be assumed that all the known protein-related functions revealed from the GO data analyses are executed in all the tissues where the proteins are expressed. For that reason, additional tools were complementary used, such as the analysis of published data obtained after the targeted protein depletion in mice. In fact, the animal model data become especially valuable to predict protein functions, due to the ethical limitations associated to the conduction of functional studies in humans. Therefore, sperm proteins whose corresponding gene-disrupted mice resulted in "embryonic lethality before implantation" phenotype, according to the MGI database, were identified and classified according to the developmental stage in which the impairment takes place (first divisions, morula or blastocyst; Figure 2). Additionally, knock-out data related to the "male infertility" phenotype from the MGI database, and GO terms associated to different stages of spermatogenesis and sperm maturation were also retrieved, in order to discriminate those proteins that might represent spermatogenesis leftovers with no further function in subsequent stages (Supplementary Table 2).

In order to allow a high-confident identification of pre-implantational embryo proteins with a potential sperm-specific origin, the proteome of the human oocyte and pre-implantational embryo were also compiled and compared to the human sperm

proteome. The human oocyte proteomic profile was obtained from a single study on mature oocytes without granulosa cells (Virant-Klun *et al.*, 2016), and the human pre-implantational blastocyst proteome profile was compiled from proteomic reports using 6-days embryos (blastocyst stage; Table 1) (Jensen *et al.*, 2013; Kaihola *et al.*, 2016). Those proteins identified exclusively in treated groups or culture media, and those that did not fit the same MS identification criteria indicated above for the sperm proteome were discarded for the current analysis. In total, 1376 and 1300 non-redundant proteins were compiled for the oocyte and the blastocyst, respectively (Supplementary Table 1).

Since the human oocyte and blastocyst proteomes are less covered than the sperm proteome, and due to the fact that these cells are in close contact with their surroundings, the protein catalogues of additional oocyte and pre-implantational embryo interacting cells, tissues and fluids were also compiled (Table 1, Supplementary Table 1). Specifically, the literature was searched for proteins identified in human cumulus cells (179 proteins) (McReynolds *et al.*, 2012; Braga *et al.*, 2016), human follicular fluid (1394 proteins) (Lee *et al.*, 2005; Angelucci *et al.*, 2006; Kim *et al.*, 2006; Liu *et al.*, 2007; Hanrieder *et al.*, 2008, 2009; Estes *et al.*, 2009; Lo Turco *et al.*, 2010, 2013; Jarkovska *et al.*, 2010; Kushnir *et al.*, 2012; Twigt *et al.*, 2012; Ambekar *et al.*, 2013, 2015; Severino *et al.*, 2013; Bianchi *et al.*, 2013; Hashemitabar *et al.*, 2014; Regiani *et al.*, 2015; Wu *et al.*, 2015; Zamah *et al.*, 2015; Lewandowska *et al.*, 2017; Lim *et al.*, 2017; Oh *et al.*, 2017), and human endometrium (tissue and fluid) on mid-secretory and decidualization phases (2082 proteins) (Zhang *et al.*, 2006; Fowler *et al.*, 2007; Chen *et al.*, 2009, 2015; Domínguez *et al.*, 2009; Parmar *et al.*, 2009; Scotchie *et al.*, 2009; Hannan *et al.*, 2010; Paule *et al.*, 2010; Stephens *et al.*, 2010; Garrido-Gomez *et al.*, 2011). Although individual characteristics such as BMI, life style, and exposition to

toxins were not possible to control, only those proteins identified in healthy, fertile or donor women, younger than 40 years old were included in the respective protein profiles (Supplementary Table 1). Furthermore, published RNA data from human sperm (Johnson *et al.*, 2015), oocyte, and pre-implantational embryo (Dang *et al.*, 2016) have been used as an additional tool to predict the gamete-origin of the blastocyst proteins.

Lastly, the tissue expression data enclosed in The Human Protein Atlas database (Uhlén *et al.*, 2005; Uhlen *et al.*, 2015) was retrieved (Release date 2017-06-08) and compared to the comprehensive human sperm proteome, in order to identify the potential tissue origin of each sperm gene product. In particular, sperm proteins were predicted to be acquired in extra-testicular stages of sperm maturation, when fitting the following criteria: RNA levels in testis < 25 transcripts per million (TPM), and no protein expression in testicular seminiferous tubules. Antibody detection on testis sections enclosed in Human Protein Atlas database were individually checked for all those sperm proteins fitting the criteria mentioned above, in order to ensure the lack of antibody staining and discard false-positive matches. The functional involvement of this subset of sperm proteins was predicted by enrichment analyses on GO annotations related to biological processes, by using the tools from the Gene Ontology Consortium database. The significance of the enrichment analyses was calculated by a Fisher Exact Test. P-values <0.05 adjusted for multiple-comparisons with the Bonferroni correction were considered significant.

## Involvement of the sperm proteome at oocyte fecundation and beyond

The first generation of a compiled proteome profile of the sperm cell was published in 2014, and included 6198 non-redundant proteins from 30 different studies (Amaral *et al.*, 2014a). Since then, many additional proteomic reports focused on the male gamete have been published (Table 1) (Amaral *et al.*, 2014b; Azpiazu *et al.*, 2014; Jumeau *et al.*, 2015; Wang *et al.*, 2015, 2016; Hetherington *et al.*, 2016; Vandenbrouck *et al.*, 2016; Carapito *et al.*, 2017; Jodar *et al.*, 2017). In this review we have updated the sperm protein profile while following strict inclusion criteria, in order to minimize as much as possible potential false-positive protein identifications (see Methods). By doing this, the sperm protein list has risen to 6871 proteins, and represents the most complete and reliable catalogue of human sperm proteins to date (Supplementary Table 1).

The in-depth analyses and data mining of the sperm molecular composition are contributing to increase our knowledge about sperm function. Previous enrichment analyses, focused on the identification of those biological pathways likely to be the most active in human sperm, had revealed metabolism (including protein and RNA metabolism), membrane trafficking, apoptosis, cell cycle, hemostasis, and meiosis as the most significant pathways in the whole male gamete (Amaral *et al.*, 2014a). However, there is growing evidence showing the role of the spermatozoon in embryogenesis and transmission of epigenetic marks important for early and future events of the offspring (Hammoud *et al.*, 2009; Brykczynska *et al.*, 2010; Vavouri and Lehner, 2011; Castillo *et al.*, 2014a, 2015; Wei *et al.*, 2014b; Fullston *et al.*, 2015), which are functions that are not revealed when enrichment analyses are performed using the whole set of sperm proteins (Castillo *et al.*, 2014b). Therefore, in the present review

we have focused our study of the sperm proteome on specific GO terms, rather than on the basis of enrichment analyses. Interestingly, a remarkable number of sperm proteins functionally related to different stages of fertilization, early embryogenesis, and gene expression modulation have been predicted (Supplementary Tables 3 and 4). This analysis provides valuable information that contributes to enhance the current understanding of the role of the male gamete in the generation of the new individual.

*Sperm proteins and fertilization*

Fertilization is a complex multistep process that relies in tight regulated events taking place in both gametes, the sperm cell and the oocyte. Of relevance, our analysis of the sperm proteome based on the use of public GO annotations revealed the presence of 103 sperm proteins that might be functionally related to different steps of the fertilization process, such as sperm capacitation, sperm acrosome reaction, sperm penetration, and sperm-egg fusion (Figure 2; Supplementary Table 3) (Brucker and Lipford, 1995; Visconti *et al.*, 1998; Bedford, 2004; De Jonge, 2005; Krawetz, 2005; Sutovsky and Manandhar, 2006; Visconti, 2009; Amaral *et al.*, 2014a; Dacheux and Dacheux, 2014; Cuasnicú *et al.*, 2016; Sullivan and Mieusset, 2016; Torabi *et al.*, 2017).

In particular, this subset of sperm proteins includes components with well-known roles in the acquisition of sperm motility and the hyperactivation occurring during sperm capacitation, such as the A-kinase anchor proteins 3 and 4 (AKAP3 and AKAP4), and the sperm specific cation channels CATSPERs (Ficarro *et al.*, 2003; Singh and Rajender, 2015; Williams *et al.*, 2015) (Supplementary Table 3). Also, the

male gamete contains proteins known to be indispensable for fertilization once it reaches the oocyte, such as the Izumo sperm-egg fusion protein 1 (IZUMO1; Figure 2), which directs sperm-egg fusion by binding its complementary oocyte protein Juno (Inoue *et al.*, 2005; Sutovsky, 2009; Bianchi *et al.*, 2014), and the sperm protein PLCzeta (PLCZ1), which induces the $Ca^{2+}$ oscillations in the oocyte that are required for egg activation, the formation of the maternal pronucleus, and the initiation of embryogenesis (Yoon and Fissore, 2007) (Figure 2).

Interestingly, published data from targeted gene deletions in mice support the essential role that have been predicted for some of the human sperm proteins associated to fertilization-related GO terms in this study. For instance, it has been shown that the acrosomal sperm protein Proprotein convertase subtilisin/kexin type 4 (PCSK4) null mice produce mutant sperm lacking the ability to proteolytically process the acrosin binding protein (ACRBP), impairing thus the capability of the sperm to bind to the zona pellucida, and to fertilize eggs (Gyamera-Acheampong *et al.*, 2006; Tardif *et al.*, 2012). Similarly, deficient mice of sperm acrosomal equatorial segment proteins, such as the equatorin (EQTN) and the sperm equatorial segment protein 1 (SPESP1; Figure 2), also produce sperm with reduced fertilizing capacity (Toshimori *et al.*, 1998; Wolkowicz *et al.*, 2003, 2008; Fujihara *et al.*, 2010; Hao *et al.*, 2014). In fact, the lack of SPESP1 also affects the correct amount and localization of several other sperm proteins involved in gamete fusion, such as the A Disintegrin And Metalloproteinase proteins (ADAMs), the EQTN, and the IZUMO1 (Cuasnicú *et al.*, 2016). Also remarkable is the identification of the sperm protein cysteine-rich secretory protein 1 (CRISP1), which is an epididymal androgen-regulated glycoprotein imported to sperm during sperm maturation that is involved in sperm motility, sperm penetration, and sperm-egg fusion (Cameo and

Blaquier, 1976; Krätzschmar *et al.*, 1996; Maldera *et al.*, 2014; Ernesto *et al.*, 2015) (Figure 2). Of relevance, CRISP1 interacts with egg complementary sites, such as the glycoprotein ZP3 and, in fact, CRISP1-deficient sperm are not able to penetrate into the oocyte (Da Ros *et al.*, 2015).

The presence of this body of experimental evidence showing the role of this subset of sperm proteins in fertilization is remarkable since it is validating the results from the GO annotations-based data analysis designed and underwent in the present review. Therefore, it supports the application of this strategy in the exploration of other potential functions of the sperm cell.

*Sperm proteins and the pre-implantational embryo development*

The analysis of the human sperm proteome profile has also allowed us to identify and highlight a subset of 93 proteins that may be functionally related to the formation of the zygote and following stages of embryo development prior implantation, which in humans takes place approximately at day 7 of development (Niakan *et al.*, 2012) (Supplementary Table 3). Early embryogenesis is characterized by dramatic changes in chromatin organization, including the replacement of paternal protamines by maternal histones (Oliva and Dixon, 1991; Wright, 1999; Oliva, 2006; Inoue *et al.*, 2011; Kong *et al.*, 2018), and the extensive epigenetic reprogramming of maternal and paternal genomes, which returns the zygote to a genetic state able to generate any cell type in the body (Reik and Walter, 2001; Reik *et al.*, 2001). This is exemplified by the post-fertilization DNA demethylation wave observed first in paternal and subsequently in maternal genome, except for both parental methylation imprinting

marks (Mayer *et al.*, 2000; Wei *et al.*, 2014a) (Figure 2). However, it is not until the third day of development, within the 4- and 8-cell stages (Figure 2), when the main embryonic genome activation (EGA) takes place (Niakan *et al.*, 2012), although a minor wave of transcription has been observed earlier, at the 2-cell stage (Vassena *et al.*, 2011) (Figure 2). Interestingly, the rapidly demethylation occurred in the paternal genome could enable this first minor wave of transcription (Santos and Dean, 2004). Afterwards, the embryo begins to synthesize proteins on its own moving towards the stages of morula and blastocyst, prior its subsequently implantation (Figure 2).

It is important to note that, due to the fact that the sperm cell is the result of complex series of cellular and molecular modifications, it becomes a challenge to distinguish proteins that are players in the process of early embryo development, from those that, in contrast, simply represent spermatogenesis leftovers with no relevant function in further stages. Although further research and novel approaches are demanded to decipher the functionality of the sperm proteins in humans, the studies already performed with animal models are contributing to shed light into this specific question. For that reason, we have added to our analysis the valuable phenotypic data contained in the MGI database (see Methods) as a tool to predict the potential function of these proteins in the human early embryogenesis (Figure 2; Supplementary Table 3). We found that the human sperm cell contains 59 different proteins whose depletion in mice through knock-out studies resulted in very marked impairments at different stages of the pre-implantational embryo development (Figure 2; Supplementary Table 3). Specifically, the male gamete delivers to the zygote 11 proteins related to embryo lethality during the first divisions from zygote to 8-cell stage, such as the transmembrane glycoprotein desmocollin 3 (DSC3) (Figure 2). Interestingly, the crucial

impact of DSC3 on the correct development of the embryo might be conducted by its function on the regulation of cell adhesion, which is necessary for the formation of blastomeres, occurring around day 1 of development in humans (Den *et al.*, 2006; Garrod and Chidgey, 2008) (Figure 2). Therefore, this developmental function could be attributed to the sperm-derived DSC3, since it may take place before the activation of the embryonic genome (around day 2 of development in humans; Figure 2). Similarly, the human spermatozoon also contains 29 proteins whose deficiencies induce impairments at the morula stage, and 19 proteins at the blastocyst formation (Figure 2; Supplementary Table 3). This is the case of the proteins lactosylceramide 1,3-N-acetyl-beta-D-glucosaminyltransferase (B3GNT5) and choline-phosphate cytidylyltransferase A (PCYT1A), which are both related to the formation of lipid membranes. Specifically, B3GNT5 catalyzes the biosynthesis of the lactoseries of glycosphingolipids (Henion *et al.*, 2001), and its depletion results in embryo lethality at morula stage due to alterations in cell adhesion and signaling processes (Biellmann *et al.*, 2008). PCYT1A, in contrast, is involved in the initiation of the synthesis of phosphatidylcholine, the most abundant phospholipid in mammalian cellular membranes (Vance and Vance, 2004). Remarkably, PCYT1A-null mice embryos were found to fail in the formation of blastocyst capable to achieve implantation (Wang *et al.*, 2005). This information might be extrapolated to humans, although it is important to take into account that the transcription of embryonic genes is already active at morula and subsequent stages. Therefore, further experimental evidences are necessary to confirm that these proteins and the rest of sperm proteins belonging to this group are also functional during human pre-implantational embryogenesis, as well as its potential specific paternal origin.

*Sperm proteins and the hypothesis of the epigenetic inheritance of acquired traits*

Previous studies on the impact of parental health on the offspring have been mainly focused on the female partner, and especially during the pregnancy period. This is due to the fact that several poor health offspring outcomes have been demonstrated from harmful maternal exposures or life style, either prior or during pregnancy (reviewed in Brion *et al.*, 2008; Feng *et al.*, 2014). However, recent findings have shown how the paternal life-history experiences have a greater influence on the future health of the offspring than previously thought. Some remarkable evidences supporting this hypothesis have been found in animal models, such as the detection of metabolism alterations on rodent offspring caused by the paternal diet (including caloric restrictions, low protein diets, or high fat diets) (Anderson *et al.*, 2006; Carone *et al.*, 2010; Ng *et al.*, 2010; de Castro Barbosa *et al.*, 2016). Other examples are the observation of a decreased fear response and the appearance of depressive symptoms in the offspring from traumatized male mice (Dietz *et al.*, 2011; Rodgers *et al.*, 2013; Gapp *et al.*, 2014), as well as the presence of cognitive impairments, and of higher levels of anxiety and aggressiveness in the offspring from males with postnatal exposures to nicotine and heroin, respectively (Farah Naquiah *et al.*, 2016; Renaud and Fountain, 2016). Interestingly, in some cases those harmful phenotypes could be found not only in the offspring (F1), which is due to intergenerational inheritance, but also in further generations, what is known as transgenerational inheritance.

Similar observational data as the indicated above for animal models has been generated in humans, also suggesting the existence of intergenerational and

transgenerational effects as a consequence of paternal life-history experiences (reviewed in Pembrey *et al.*, 2014). This is exemplified by the observed correlation between paternal overfeeding during mid-childhood (from 9 to 11 years old) and the lower longevity of sons and grandsons, due to an increased risk to suffer diabetes (Bygren *et al.*, 2001; Kaati *et al.*, 2002). Also, it has been observed an association between early paternal smoking habits and a higher BMI in the male progeny at 9 years old (Pembrey *et al.*, 2006).

Currently, several groups are working to elucidate the molecular mechanisms that could explain the above evidences, supporting the hypothesis of the transmission of some acquired traits through the paternal line (Sharma and Rando, 2017). Potential candidates have been proposed to provide the basis for this epigenetic inheritance, such as the sperm DNA methylation (de Castro Barbosa *et al.*, 2016; Donkin *et al.*, 2016; Soubry *et al.*, 2016), sperm histone marks including both the PTMs and the nucleohistone-nucleoprotamine pattern (Zeybel *et al.*, 2012; Vassoler *et al.*, 2013; Castillo *et al.*, 2015; Siklenka *et al.*, 2015), and the set of sperm small non-coding RNAs (sncRNAs) (Rassoulzadegan *et al.*, 2006; Grandjean *et al.*, 2015; Chen *et al.*, 2016a; Sharma *et al.*, 2016). Interestingly, all those epigenetic marks are maintained in the mature spermatozoa and, thus, they are able to escape from the epigenetic reprogramming occurring in the zygote (Borgel *et al.*, 2010; Chen *et al.*, 2016b), and could act as scaffold for the establishment of new epigenetic signatures in the embryo genome (Brunner *et al.*, 2014; Castillo *et al.*, 2014a, 2014b, 2015). In any case, those sperm epigenetic marks might be probably regulating gene expression at either transcriptional or post-transcriptional levels in the early embryo. For instance, it has been demonstrated that the translational RNAs (tRNAs) fragments altered in founder

males under a low protein diet are able to modulate the abundance of a subset of genes known to be expressed in pre-implantational embryos (Sharma *et al.*, 2016). Therefore, the altered phenotype observed in the offspring might be caused by an altered gene expression during early embryo development induced by changes in the sperm epigenetic profile.

However, in addition to the sperm DNA containing epigenetic marks, and the complex population of sperm sncRNAs (Krawetz *et al.*, 2011; Jodar *et al.*, 2013; Pantano *et al.*, 2015), the sperm cell also provides the oocyte with a large number of zinc finger- and bromodomain-containing proteins, transcription factors, histone modifiers, and other DNA- and RNA-related proteins that might also be critical for the regulation of gene expression in the early embryo, either at transcriptional or post-transcriptional levels (Castillo *et al.*, 2014a, 2014b, 2015). In fact, the gene-specific GO annotations-based analysis undertaken in this review has revealed a total of 560 sperm proteins with known roles in the regulation of gene expression in other cells or tissues (Table 2 and Supplementary Table 4). This subset of sperm proteins includes 1) transcription factors and transcription factor-related proteins (381 proteins), 2) chromatin modifiers able to modulate the DNA methylation pattern (25 proteins), 3) chromatin modifiers that might regulate histone PTMs (118 proteins), and 4) proteins that participate in the regulation of the transcription, processing, and function of non-coding RNAs (36 proteins) (Table 2; Supplementary Table 4). However, not all those proteins must have these potential roles in the early embryo and, also, some of them could simply represent spermatogenic leftovers. For instance, 69 of those 560 sperm proteins have a known role during spermatogenesis or sperm maturation, according to either their association to a sperm-related GO annotation or the observation of infertility

in their respective null-mice (Supplementary Table 4). This is the case of the sperm protein Probable ATP-dependent RNA helicase DDX4 (DDX4), whose corresponding null mice are infertile due to arrest of male meiosis (Kuramochi-Miyagawa *et al.*, 2010). In contrast, we found a total of 28 proteins, out of the 560 sperm proteins classified in this review as potentially involved in the regulation of gene expression during early embryogenesis. In fact, this subset of sperm proteins either showed associations to GO terms related to early embryogenesis or their targeted deletion in mice resulted in embryonic lethality before implantation (Supplementary Table 4). For instance, no 8-cell embryos were observed after target depletion of the nuclear autoantigenic sperm protein (NASP), which has been suggested to participate in both nucleosome remodeling and maintenance of the methylation pattern during the embryo pre-implantational epigenetic reprogramming (Mohan *et al.*, 2011).

To the best of our knowledge there is a lack of studies in the human sperm proteome contributing to explain the mechanisms implicated in the epigenetic inheritance of some acquired traits, although, as presented above, some sperm proteins might modify gene expression at the transcriptional and post-transcriptional levels in the early embryo (Figure 3). However, the proteomic data profiles from obese patients may be used to assess this hypothesis, since several authors suggested that obese males might transmit altered metabolic phenotypes to the offspring (Bygren *et al.*, 2001; Ng *et al.*, 2010; Rando, 2012). Of relevance, 3 sperm proteins related to the regulation of gene expression, according to the analysis undertaken in this review (Supplementary Table 4), were found deregulated in a comparative proteomic analysis of spermatozoa from obese and lean human males: the Eukaryotic Translation Elongation Factor 1 Alpha 1 (EEF1A), the CCR4-NOT Transcription Complex Subunit 1 (CNOT1) and the

Elongator Acetyltransferase Complex Subunit 3 (ELP3; Supplementary Table 4) (Liu *et al.*, 2015). Interestingly, CNOT1 is a scaffolding component of the mRNA deadenylase complex CCR4-NOT, which participates in processes such as mRNA degradation, miRNA-mediated repression, translational repression, and general regulation of transcription. Moreover, CCR4-NOT complex represses the expression of early trophectoderm transcription factors in embryonic stem cells, suggestive of its critical role in the maintenance of the totipotency of blastomeres during early embryogenesis (Zheng *et al.*, 2012). Also, ELP3 has been described as one of the responsible factors for paternal DNA demethylation upon oocyte fertilization (Okada *et al.*, 2010), among other functions such as the modification of tRNAs (Svejstrup, 2007). It is interesting to note that the modifications of tRNAs fragments seem to be also crucial for the transmission of altered metabolic phenotypes due to paternal high fat diet (Chen *et al.*, 2016b). The deregulation of the proteins CNOT1 and ELP3 in sperm proteome from obese males and the ability of those proteins to regulate epigenetic marks suggest them as potential candidates to transmit obesity-related paternal environment information to the new individual. Altogether, our findings reinforce the hypothesis of a potential role of some sperm proteins in the intergenerational epigenetic inheritance of acquired traits.

**Integrated analysis of the pre-implantational proteomes: deciphering the potential parental origin of the early embryo proteins**

The integrative analysis of the human sperm, oocyte and early embryo proteomes has been used as a complementary approach to determine which human sperm proteins may be involved in early embryo development. This may be inferred by the detection of blastocyst proteins having an unequivocal paternal origin. However,

two technical issues should be taken into account in this analysis. The first one is the nearly complete description of the sperm proteome as compared to the limited covered oocyte and embryo proteomes. As observed in Figure 3A, whereas 6871 proteins have been described in the sperm proteome, only 1376 and 1300 were indentified in the oocyte and blastocyst proteomes, respectively. Those differences might be probably due to the relative scarcity of available biological material from human oocytes and early embryos, which becomes technically insufficient for the identification of less abundant proteins (Jensen *et al.*, 2013; Kaihola *et al.*, 2016; Virant-Klun *et al.*, 2016). In contrast, the spermatozoa show two main advantages: 1) it can be easily purified in large quantities (Martinez-Heredia *et al.*, 2006; Oliva *et al.*, 2008, 2009; Codina *et al.*, 2015), and 2) it is a well compartmentalized cell that allows the proteomic assessment of different subfractions, such as the head, the chromatin, the tail, and the membranes (de Mateo *et al.*, 2011; Amaral *et al.*, 2013; Baker *et al.*, 2013; Castillo *et al.*, 2014a), allowing the identification of less abundant proteins.

The second issue is that, to the best of our knowledge, only the human embryo proteome on the blastocyst stage (5-6 days) is available in the literature (Jensen *et al.*, 2013; Kaihola *et al.*, 2016). As previously mentioned, the activation of the embryo genome occurs mainly between the 4-cell and 8-cell stage (Figure 2), implying that blastocyst proteins could already have an embryonic origin. However, several proteins have broad halve-lives ranging from 30 minutes to 8 days (Schwanhausser *et al.*, 2011), which suggests that some sperm and oocyte proteins may be still present and functional in the blastocyst stage. Additionally, it is also important to take into account that some proteins detected in the blastocyst could derive from the translation of the paternal and maternal mRNAs provided to the zygote, using the maternal translational machinery

(Swann *et al.*, 2012). In order to mitigate all these issues, we have also integrated into our proteome analyses the individual RNA-seq data from human sperm (Johnson *et al.*, 2015), oocyte, zygote, and embryos at 2-cell, 4-cell, 8-cell, morula and blastocyst stages (Dang *et al.*, 2016), as a complementary tool to identify potential paternally-derived embryo proteins. Also, in order to minimize the possibility of false positive identifications, we excluded from the present analysis the proteins that have also been identified in the proteomes of additional reproductive cells and fluids highly related to the oocyte and early embryo, such as the follicular fluid, cumulus cells and endometrium, since they may serve as a complementary protein source to the embryo proteome (See Methods and Supplementary Table 1).

Following these strict criteria, we managed to unequivocally classify 173 from the 1300 blastocyst proteins, according to their embryonic (18 proteins), maternal (47 proteins), and paternal (108 proteins) origin (Supplementary Table 5). As observed in Figure 3B, the transcriptional pattern of some blastocyst proteins classified as Group 1 (18 proteins) clearly reflects an embryonic origin by *de novo* transcription after the activation of the embryo genome (EGA). Specifically, those blastocyst proteins have not been detected in any of the cellular and fluid protein profiles of sperm, oocyte, follicular fluid, cumulus cells and mid-secretory endometrium (Supplementary Tables 1 and 5), and their corresponding RNAs were absent in the sperm, oocyte, and the stages of embryogenesis previous to the EGA (Supplementary Table 5). In contrast, we identified 155 blastocyst proteins with a potential maternal (47 proteins) or paternal (108 proteins) origin, either as proteins already present in the gametes and maintained intact until the blastocyst stage, or as the result of the potential translation of sperm and oocyte RNAs during the first stages of embryogenesis by the maternal translational

machinery (Supplementary Table 5). From the subset of blastocyst proteins predicted as maternally derived, 5 of them were exclusively detected in oocyte at protein level, while their corresponding RNAs were found absent in sperm, oocyte and all the embryo stages assessed. Therefore, the stability of those 5 embryo proteins during first steps of embryogenesis suggests their slow molecular turnover (Supplementary Table 5). In contrast, the remaining 42 maternal-derived blastocyst proteins identified in this review, follow a specific transcriptional pattern classified as Group 2 in Figure 3B. In particular, the RNAs encoding this subset of blastocyst proteins are found with high levels in the oocyte, and progressively drop during the different early embryogenesis stages, without any increase upon the EGA. This, together with the fact that those proteins and their corresponding RNAs are absent in the sperm, provides evidence for the maternal origin of these 42 embryo proteins (Figure 3B).

Unexpectedly, 108 blastocyst proteins were predicted to have a paternal origin, by either 1) their exclusive presence in the sperm protein profile combined with the absence of their corresponding transcripts in sperm, oocyte and all the different early embryogenesis stages (82 proteins), or 2) the solely detection of the corresponding RNA in sperm combined with their absence at protein level in both gametes or reproductive-associated cells or fluids (26 proteins; Group 3, Figure 3B). Interestingly, this group of paternal-derived blastocyst proteins were found enriched in the GO annotation "regulation of small GTPase mediated signal transduction" (Bonferroni corrected p-value<0.001). Dynamic analyses of gene expression during mouse pre-implantational development have revealed the activation of the genes involved in small GTPases-mediated signal transduction between the 8-cell and the morula stages (Hamatani *et al.*, 2004). Therefore, our results suggest that the male gamete might provide the pre-

implantational embryo with proteins that allow ensuring the appropriate gene expression pattern of GTPases involved in signal transduction processes. In addition, some of the blastocyst proteins inferred as paternal-derived were found related to already known sperm components required for pre-implantational embryogenesis, such as the centrosomal protein 135 (CEP135), which acts as a scaffolding protein during early centriole biogenesis (Ohta *et al.*, 2002). The roles of other potentially paternal-derived proteins are less well known in mammalian early embryogenesis. It is the case of the Rho GTPase Activating Protein 21 (ARHGAP21), whose homologue protein in *C. elegans* is crucial for the establishment of the radial polarity in the early embryo, process known in mammals as compactation (Anderson *et al.*, 2008) (Nance, 2014). Furthermore, some blastocyst proteins potentially derived from the sperm seem to be related to the regulation of epigenetic marks, such as the (A-T) mutated (ATM) kinase, which is involved in DNA damage response by phosphorylating numerous substrates including histones (Yamamoto *et al.*, 2012), and the DGCR8 Microprocessor Complex Subunit (DGCR8), which is crucial for primicroRNA processing to mature miRNA (Wang *et al.*, 2007).

This integrative analysis of "omic" data from reproductive cells and fluids provides the first step for the identification of embryo proteins with a potential paternal origin. The future validation of the impact of these 108 sperm proteins to the embryo, as well as the assessment of their crucial roles in pre-implantational embryogenesis and the epigenetic inheritance may shed new light into the real contribution of the father to the generation of a healthy offspring.

**The putative extra-testicular origin of a subset of human sperm proteins and the soma to embryo transmission hypothesis**

All human sperm proteomic studies conducted to date have been performed in purified ejaculated spermatozoa, and thus, in cells that have been in contact with the seminal fluid. The seminal fluid constitutes at least the 90% of the ejaculated volume and is composed by a mixture of secretions from the accessory sex glands (epididymis, prostate, and seminal vesicles), which are rich in lipids, glycans, oligosaccharides, inorganic ions, immune components, DNA, RNAs, miRNAs, proteins, and peptides, either free or encapsulated in extracellular vesicles (Saez *et al.*, 2003; Jones *et al.*, 2010; Ronquist *et al.*, 2011; Aalberts *et al.*, 2014; Drabovich *et al.*, 2014; Vojtech *et al.*, 2014; Chiasserini *et al.*, 2015; Jodar *et al.*, 2016, 2017).

During the past years, the existence of a communication between the sperm cells and the seminal fluid has been proposed, most probably through these extracellular vesicles (Sullivan and Saez, 2013; Jodar *et al.*, 2017). In fact, growing evidence is recently contributing to this hypothesis, such as the enrichment of RNAs from the seminal fluid extracellular vesicles in the peripheral membrane of mouse spermatozoa (Johnson *et al.*, 2015). On the one hand, from a functional point of view, the seminal fluid seems to be not only a medium to carry spermatozoa through male and female tracts, but also a source of nutrition and components that modulate sperm function, motility and fertilizing capacity (Saez *et al.*, 2003; De Jonge, 2005; Aalberts *et al.*, 2014). On the other hand, from the proteomic perspective, the high impact of the seminal fluid on the protein composition of the male gamete has also been shown. Specifically, both ejaculated sperm and seminal fluid were found to contain a remarkable number of common proteins in their respective proteome profiles (Jodar *et*

*al.*, 2017). Interestingly, a previous integrative analysis of proteins and transcripts present in human sperm, extracellular vesicles, and testes provided a list of sperm proteins with a potential extra-testicular origin (Jodar *et al.*, 2016).

In this review, we have improved the prediction of the putative tissue origin of the human sperm proteins by integrating the RNA and protein data available in the Human Protein Atlas database. This analysis was based in the premise that the process of translation in the mature ejaculated sperm is blocked and, therefore, those sperm proteins not expressed in any stage of the seminiferous tubules, neither at RNA nor at protein level, may be acquired from extra-testicular tissues. By doing this, we were able to identify with high confidence 165 different sperm proteins potentially provided by the fluids or extracellular vesicles from epididymis, prostate or seminal vesicles (Supplementary Table 6).

From a general perspective, the enrichment analysis focused on GO annotations related to biological processes showed that those potential accessory sex glands-derived sperm proteins were mainly involved on immune response, cell junction organization, response to stimulus, extracellular matrix disassembly, gene expression, and keratinization (Bonferroni corrected p-value < 0.05; Table 3). Interestingly, 7 of these potential extra-testicular sperm proteins have been related to the processes of fertilization (4 proteins) and pre-implantational embryo development (3 proteins; Supplementary Table 6). For instance, the beta-defensin 126 (DEFB126) is an epididymal protein known to be recruited to the sperm surface during the pass through the distal corpus and proximal cauda parts of the epididymis (Perry *et al.*, 1999) (Figure 4). Of note, a DEFB126 role in the efficient protection of the sperm from the female immune response has been reported (Yudin *et al.*, 2005), which is in agreement with the GO enrichment analysis underwent in this review with the accessory sex glands-derived

sperm proteins (Table 3). Additional epididymal proteins recruited by sperm are the epididymal sperm-binding protein 1 (ELSPBP1), which is involved in the correct acquisition of sperm motility (Parte *et al.*, 2012); the binder of sperm protein homolog 1 (BSPH1), which has a role in capacitation (Plante et al., 2014); and the cysteine-rich secretory protein 1 (CRISP1), which has been already highlighted because of its crucial role in sperm penetration and sperm-egg fusion (see section 2.1; Figure 4; Supplementary Table 6). Also remarkable is the fact that sperm seem to carry prostatic proteins that might be relevant for the success of ART. This is the case of the prostatic acid phosphatase (ACPP) which is found deregulated in idiopathic infertile couples that did not achieve pregnancy after artificial insemination (Xu *et al.*, 2012), suggesting that ACPP is crucial to reach and fertilize the oocyte. Similarly, a total of 9 sperm proteins functionally associated to the modulation of gene expression are suggested to be originated in the accessory sex glands (Figure 4; Supplementary Table 6), which is in agreement with recent evidence showing that sperm gain specific non-coding RNAs involved in the epigenetic inheritance of the paternal low protein diet phenotype through epididymosomes (Sharma *et al.*, 2016). In fact, the importation of sperm proteins and RNAs from accessory sex glands able to regulate gene expression in the embryo, once its genome is activated, might be a good strategy to provide environmental epigenetic information without the need to overcome the hemato-testicular barrier.

Unexpectedly, the results of our integrative analysis also suggest that not only the tissues from accessory sex glands are contributing to the sperm protein content. In fact, we found a subset of 286 sperm proteins with no RNA and protein detection neither in the testis nor in the accessory sex glands (Supplementary Table 7). Therefore, it can be suggested that those proteins could derive from the peripheral tissues outside the male reproductive tract, and be acquired by the sperm after testicular maturation.

These findings go in the same lane as the results obtained by Cossetti and colleagues, who showed the presence of a specific RNA from human melanoma in the sperm from mice that were previously subcutaneously inoculated with human melanoma cells (Cossetti *et al.*, 2014). Interestingly, the specific human transcript measured in Cossetti's study was also detected in the extracellular fraction of blood, indicating a potential active transport from mice dermis to the sperm cells through the extracellular vesicles pathway (Cossetti *et al.*, 2014). Our analysis revealed that some sperm proteins are specifically expressed at the protein level in a high variety of tissues, such as brain, lung, pancreas, and bone marrow and immune system, among others, but absent in testes and accessory sex glands at both transcript and protein level (Figure 4). From the functional perspective, proteins potentially derived from tissues other than testis and accessory sex glands were found mainly involved in keratinization, cell adhesion, immune response, programmed cell death, extracellular matrix organization, regulation of gene expression, and anion transport, among others (Bonferroni corrected p-value <0.05; Table 3).

Some of the proteins identified here as potentially originated in extra-testicular tissues were specifically found to be potentially related to processes of fertilization (2 proteins), early embryo development (2 proteins) and modulation of gene expression (12 proteins) (Supplementary Table 7). Interestingly, a remarkable number of sperm proteins probably originated by tissues from the immune system seem to be able to modulate gene expression by the regulation of the transcription and the histone modification pattern (Figure 4). It is the case of the protein-arginine deiminase type-4 (PADI4) which is involved in the modification of histone 1 (H1) by the conversion of arginine residues to the non-coded aminoacid citrulline. This modification induces H1 disassembly from the DNA, leading to a global chromatin decondensation. Although

this process was described in neutrophils during the innate immune response to infection (Neeli *et al.*, 2008), it has been recently shown a role in the mice early embryo, in order to undergo reprogramming, and to promote pluripontency and stem cell maintenance (Christophorou *et al.*, 2014).

Additional experiments are now required in order to validate the potential origin of these groups of sperm proteins that seem to be incorporated into the sperm cell during post-testicular processes. Although the success rate of pregnancies obtained after testicular sperm extraction (TESE) combined with intracytoplasmic sperm injection (ICSI) suggest that extra-testicular sperm proteins might not be crucial for early embryo development, the functionality of these extra-testicular proteins in the fertilization processes that are overcome by ICSI, and as potential modulators of the offspring's phenotype should be elucidated. This will contribute to unravel the mechanisms of the transmission of information within organs and generations, such as the transmission of parental olfactory experiences to the offspring (Dias and Ressler, 2014), which could be explained by the travel of epigenetic information from the peripheral tissues outside male reproductive tract to the sperm.

**Discussion and future directions**

The characterization of the sperm cell through MS-proteomic strategies combined with the integrative analysis of the data is providing a powerful approach to decipher the molecular aspects of sperm function (Amaral *et al.*, 2014a; Carrell *et al.*, 2016). In fact, the analysis reported in the current review, has increased our knowledge into the potential contribution of the father to the generation of healthy offspring. This has been possible thanks to the current availability of sperm protein catalogues,

allowing us to generate a nearly complete protein profile of the human male gamete. In addition, since the functionality of a cell is also influenced by its surroundings, impeding the cellular proteomic data to be analyzed in isolation, we have also proceeded to its exhaustive examination in combination with additional "omic" data and evidence from related tissues, cells and fluids. Due to the limitations associated to the research in human samples, data from studies underwent in animal models have been also included as a tool to predict the potential functions of the human sperm proteins.

In terms of the sperm function, the role of the male gamete during fertilization has been widely studied. In fact, our GO annotations-based analysis of the sperm proteome was in agreement with the many sperm proteins already known and proposed by others to be crucial for the different processes occurring through the sperm journey from the testis to the oocyte (Figure 2) (Toshimori *et al.*, 1998; Ficarro *et al.*, 2003; Inoue *et al.*, 2005; Fujihara *et al.*, 2010; Maldera *et al.*, 2014; Singh and Rajender, 2015; Cuasnicú *et al.*, 2016). However, less information is available so far about the involvement of the sperm proteins in the correct initiation and early progression of embryogenesis. The analysis shown in the present review has revealed remarkable groups of proteins that might be key players in post-fertilization processes, including not only those taking place during first stages of the pre-implantational embryo development, but also those able to modulate gene expression once the embryo is activated (Supplementary Tables 3 and 4). This raises the possibility that a subset of the sperm proteome could be a potential new player contributing to explain the hypothesis of the epigenetic inheritance of some acquired traits. Reinforcing this idea, we have also observed that some blastocyst proteins revealed as required for correct early embryogenesis are most probably paternal-derived (Supplementary Table 5).

Altogether, these findings show that the sperm proteome is not only limited to the proteins necessary for oocyte fertilization and spermatogenic leftovers, but also to a cargo of additional groups of proteins that are delivered to the embryo and may have critical immediate and future impacts.

Interestingly, our analyses also add further support to the idea that the protein composition of the mature spermatozoon is not concluded at testicular level, but it is only completed through the potential molecular communication between the sperm cell and the environment (Figure 4; Supplementary Table 6 and 7), as it has been already proposed (Sullivan and Saez, 2013; Jodar *et al.*, 2017). In fact, it can be hypothesized that, since sperm are not able to produce proteins *de novo* due to the blockage of its transcriptional and translational machineries, the protein importation from secretions of accessory sex glands and other peripheral tissues could be an efficient strategy to maintain the sperm proteomic profile in optimal conditions for its function at fertilization and potential future events in early embryogenesis. Moreover, these imported proteins during sperm maturation might provide environmental epigenetic information without the need to overcome the hemato-testicular barrier.

Overall the current data suggest that a subset of sperm proteins is crucial for the correct fertilization and beyond. These results should stimulate further experimental studies aimed to elucidate the roles of groups of specific sperm proteins in the processes of fertilization, embryo development, and epigenetic inheritance of acquired traits, both in normal and in altered conditions.

## Author's roles

JC and MJ: participation in study design, execution, analysis, manuscript drafting, critical discussion. RO: participation in study design, execution, manuscript drafting, critical discussion.


## Acknowledgements

We greatly thank Dr Alexandra Amaral for the critical revision of the manuscript, Tosca van Gelderen for the language revision, and Ada Soler-Ventura, Alberto de la Iglesia, and Ferran Barrachina for much advice and fruitful critical discussions.

## Funding

This work was supported by grants from the Spanish Ministry of Economy and Competitiveness (Ministerio de Economia y Competitividad; Fondos FEDER "Una manera de hacer Europa", PI13/00699, and PI16/00346), from Fundación Salud 2000 (SERONO 13-015), and from EUGIN-UB (EU-REP 2014) to RO. M.J is supported by a Postdoctoral Fellowship from the Government of Catalonia (Generalitat de Catalunya, pla estratègic de recerca i innovació en salut, PERIS 2016-2020, SLT002/16/00337). JC is supported by the Sara Borrell Postdoctoral Fellowship from the Spanish Ministry of Economy and Competitiveness (Ministerio de Economia y Competitividad, Acción Estratégica en Salud, CD17/00109).

## References


Aalberts M, Stout TAE, Stoorvogel W. Prostasomes: Extracellular vesicles from the prostate. *Reproduction* 2014:**147**;R1-14.

Amaral A, Castillo J, Estanyol JM, Ballescà JL, Ramalho-Santos J, Oliva R. Human Sperm Tail Proteome Suggests New Endogenous Metabolic Pathways. *Mol Cell Proteomics* 2013:**12**;330–342.

Amaral A, Castillo J, Ramalho-Santos J, Oliva R. The combined human sperm proteome: Cellular pathways and implications for basic and clinical science. *Hum Reprod Update* 2014a:**20**;40–62.

Amaral A, Paiva C, Attardo Parrinello C, Estanyol JM, Ballescà JL, Ramalho-Santos J, Oliva R. Identification of Proteins Involved in Human Sperm Motility Using High-Throughput Differential Proteomics. *J Proteome Res* 2014b:**13**;5670–5684.

Ambekar AS, Kelkar DS, Pinto SM, Sharma R, Hinduja I, Zaveri K, Pandey A, Prasad TSK, Gowda H, Mukherjee S. Proteomics of follicular fluid from women with polycystic ovary syndrome suggests molecular defects in follicular development. *J Clin Endocrinol Metab* 2015:**100**;744–753.

Ambekar AS, Nirujogi RS, Srikanth SM, Chavan S, Kelkar DS, Hinduja I, Zaveri K, Prasad TSK, Harsha HC, Pandey A, *et al.* Proteomic analysis of human follicular fluid: a new perspective towards understanding folliculogenesis. *J Proteomics* 2013:**87**;68–77.

Anderson DC, Gill JS, Cinalli RM, Nance J. Polarization of the C. elegans embryo by RhoGAP-mediated exclusion of PAR-6 from cell contacts. *Science* 2008:**320**;1771–1774.

Anderson LM, Riffle L, Wilson R, Travlos GS, Lubomirski MS, Alvord WG. Preconceptional fasting of fathers alters serum glucose in offspring of mice. *Nutrition* 2006:**22**;327–331.

Angelucci S, Ciavardelli D, Di Giuseppe F, Eleuterio E, Sulpizio M, Tiboni GM, Giampietro F, Palumbo P, Di Ilio C. Proteome analysis of human follicular fluid. *Biochim Biophys Acta - Proteins Proteomics* 2006:**1764**;1775–1785.

Arpanahi A, Brinkworth M, Iles D, Krawetz SA, Paradowska A, Platts AE, Saida M, Steger K, Tedder P, Miller D. Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. *Genome Res* 2009:**19**;1338–1349.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al.* Gene ontology: Tool for the unification of biology. *Nat Genet* 2000:**25**;25–29.

Azpiazu R, Amaral A, Castillo J, Estanyol JM, Guimerà M, Ballescà JL, Balasch J,

Oliva R. High-throughput sperm differential proteomics suggests that epigenetic alterations contribute to failed assisted reproduction. *Hum Reprod* 2014:**29**;1225–1237.

Baccetti B, Afzelius BA. The biology of the sperm cell. *Monogr Dev Biol* 1976;1–254.

Baker MA. Proteomics of post-translational modifications of mammalian spermatozoa. *Cell Tissue Res* 2016:**363**;279–287.

Baker MA, Naumovski N, Hetherington L, Weinberg A, Velkov T, Aitken RJ. Head and flagella subcompartmental proteomic analysis of human spermatozoa. *Proteomics* 2013:**13**;61–74.

Baker MA, Nixon B, Naumovski N, Aitken RJ. Proteomic insights into the maturation and capacitation of mammalian spermatozoa. *Syst Biol Reprod Med* 2012:**58**;211–217.

Baker MA, Reeves G, Hetherington L, Müller J, Baur I, Aitken RJ. Identification of gene products present in Triton X-100 soluble and insoluble fractions of human spermatozoa lysates using LC-MS/MS analysis. *Proteomics - Clin Appl* 2007:**1**;524–532.

Bedford JM. Enigmas of mammalian gamete form and function. *Biol Rev Camb Philos Soc* 2004:**79**;429–460.

Bianchi E, Doe B, Goulding D, Wright GJ. Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature* 2014:**508**;483–487.

Bianchi L, Gagliardi A, Campanella G, Landi C, Capaldo A, Carleo A, Armini A, De Leo V, Piomboni P, Focarelli R, *et al.* A methodological and functional proteomic approach of human follicular fluid en route for oocyte quality evaluation. *J Proteomics* 2013:**90**;61–76.

Biellmann F, Hülsmeier AJ, Zhou D, Cinelli P, Hennet T. The Lc3-synthase gene B3gnt5 is essential to pre-implantation development of the murine embryo. *BMC Dev Biol* 2008:**8**;109.

Borgel J, Guibert S, Li Y, Chiba H, Schübeler D, Sasaki H, Forné T, Weber M. Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* 2010:**42**;1093–1100.

Braga DPAF, Setti AS, Lo Turco EG, Cordeiro FB, Cabral EC, Cortezzi SS, Ono E, Figueira RCS, Eberlin MN, Borges E. Protein expression in human cumulus cells as an indicator of blastocyst formation and pregnancy success. *J Assist Reprod Genet* 2016:**33**;1571–1583.

Brion MJA, Leary SD, Lawlor DA, Smith GD, Ness AR. Modifiable maternal exposures and offspring blood pressure: A review of epidemiological studies of maternal age, diet, and smoking. *Pediatr Res* 2008:**63**;593–598.

Brucker C, Lipford GB. The human sperm acrosome reaction: physiology and regulatory mechanisms. An update. *Hum Reprod Update* 1995:**1**;51–62.

Brunner AM, Nanni P, Mansuy IM. Epigenetic marking of sperm by post-translational modification of histones and protamines. *Epigenetics and Chromatin* 2014:**7**;2.

Brykczynska U, Hisano M, Erkek S, Ramos L, Oakeley EJ, Roloff TC, Beisel C, Schubeler D, Stadler MB, Peters AH. Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat Struct Mol Biol* 2010:**17**;679–687.

Bukovsky A, Caudle MR, Svetlikova M, Wimalasena J, Ayala ME, Dominguez R. Oogenesis in Adult Mammals, Including Humans: A Review. *Endocrine* 2005:**26**;301–316.

Bygren LO, Kaati G, Edvinsson S. Longevity determined by paternal ancestors' nutrition during their slow growth period. *Acta Biotheor* 2001:**49**;53–59.

Cameo MS, Blaquier JA. Androgen-controlled specific proteins in rat epididymis. *J Endocrinol* 1976:**69**;47–55.

Carapito C, Duek P, Macron C, Seffals M, Rondel K, Delalande F, Lindskog C, Fréour T, Vandenbrouck Y, Lane L, *et al.* Validating Missing Proteins in Human Sperm Cells by Targeted Mass-Spectrometry- and Antibody-based Methods. *J Proteome Res* 2017:**16**;4340–4351.

Carone BR, Fauquier L, Habib N, Shea JM, Hart CE, Li R, Bock C, Li C, Gu H, Zamore PD, *et al.* Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* 2010:**143**;1084–1096.

Carrell DT. Epigenetics of the male gamete. *Fertil Steril* 2012:**97**;267–274.

Carrell DT, Aston KI, Oliva R, Emery BR, De Jonge CJ. The "omics" of human male infertility: integrating big data in a systems biology approach. *Cell Tissue Res* 2016:**363**;295–312.

Carrell DT, Emery BR, Hammoud S. The aetiology of sperm protamine abnormalities and their potential impact on the sperm epigenome. *Int J Androl* 2008:**31**;537–545.

Castillo J, Amaral A, Azpiazu R, Vavouri T, Estanyol JM, Ballesca JL, Oliva R. Genomic and proteomic dissection and characterization of the human sperm chromatin. *Mol Hum Reprod* 2014a:**20**;1041–1053.

Castillo J, Amaral A, Oliva R. Sperm nuclear proteome and its epigenetic potential. *Andrology* 2014b:**2**;326–338.

Castillo J, Estanyol J, Ballescà J, Oliva R. Human sperm chromatin epigenetic potential: genomics, proteomics, and male infertility. *Asian J Androl* 2015:**17**;601–609.

de Castro Barbosa T, Ingerslev LR, Alm PS, Versteyhe S, Massart J, Rasmussen M, Donkin I, Sjögren R, Mudry JM, Vetterli L, *et al.* High-fat diet reprograms the epigenome of rat spermatozoa and transgenerationally affects metabolism of the offspring. *Mol Metab* 2016:**5**;184–197.

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**;5382–5386.

Chang HM, Qiao J, Leung PCK. Oocyte-somatic cell interactions in the human ovary-novel role of bone morphogenetic proteins and growth differentiation factors. *Hum Reprod Update* 2016:**23**;1–18.

Chen JIC, Hannan NJ, Mak Y, Nicholls PK, Zhang J, Rainczuk A, Stanton PG, Robertson DM, Salamonsen LA, Stephens AN. Proteomic characterization of midproliferative and midsecretory human endometrium. *J Proteome Res* 2009:**8**;2032–2044.

Chen Q, Yan M, Cao Z, Li X, Zhang Y, Shi J, Feng GH, Peng H, Zhang X, Zhang Y, *et al.* Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* 2016a:**351**;397–400.

Chen Q, Yan W, Duan E. Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. *Nat Rev Genet* 2016b:**17**;733–743.

Chen Q, Zhang A, Yu F, Gao J, Liu Y, Yu C, Zhou H, Xu C. Label-Free Proteomics Uncovers Energy Metabolism and Focal Adhesion Regulations Responsive for Endometrium Receptivity. *J Proteome Res* 2015:**14**;1831–1842.

Chiasserini D, Mazzoni M, Bordi F, Sennato S, Susta F, Orvietani PL, Binaglia L, Palmerini CA. Identification and Partial Characterization of Two Populations of Prostasomes by a Combination of Dynamic Light Scattering and Proteomic Analysis. *J Membr Biol* 2015:**248**;991–1004.

Christophorou MA, Castelo-Branco G, Halley-Stott RP, Oliveira CS, Loos R, Radzisheuskaya A, Mowen KA, Bertone P, Silva JCR, Zernicka-Goetz M, *et al.* Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* 2014:**507**;104–108.

Codina M, Estanyol JM, Fidalgo MJ, Ballescà JL, Oliva R. Advances in sperm proteomics: best-practise methodology and clinical potential. *Expert Rev Proteomics* 2015:**12**;255–277.

Conti M, Franciosi F. Acquisition of oocyte competence to develop as an embryo: integrated nuclear and cytoplasmic events. *Hum Reprod Update* Advance Access published February 8, 2018, doi: 10.1093/humupd/dmx040.

Cossetti C, Lugini L, Astrologo L, Saggio I, Fais S, Spadafora C. Soma-to-germline transmission of RNA in mice xenografted with human tumour cells: Possible transport by exosomes. *PLoS One* 2014:**9**;e101629.

Coticchio G, Dal Canto M, Mignini Renzini M, Guglielmo MC, Brambillasca F, Turchi D, Novara PV, Fadini R. Oocyte maturation: gamete-somatic cells interactions, meiotic resumption, cytoskeletal dynamics and cytoplasmic reorganization. *Hum Reprod Update* 2015:**21**;427–454.

Cuasnicú PS, Da Ros VG, Weigel Muñoz M, Cohen DJ. Acrosome reaction as a preparation for gamete fusion. *Adv Anat Embryol Cell Biol* 2016:**220**;159–172.

Dacheux JL, Dacheux F. New insights into epididymal function in relation to sperm maturation. *Reproduction* 2014:**147**;R27-42.

Dahm R. Friedrich Miescher and the discovery of DNA. *Dev Biol* 2005:**278**;274–88.

Dang Y, Yan L, Hu B, Fan X, Ren Y, Li R, Lian Y, Yan J, Li Q, Zhang Y, *et al.* Tracing the expression of circular RNAs in human pre-implantation embryos. *Genome Biol* 2016:**17**;130.

Davies D V, Mann T. Functional development of accessory glands and spermatogenesis. *Nature* 1947:**160**;295.

Den Z, Cheng X, Merched-Sauvage M, Koch PJ. Desmocollin 3 is required for pre-implantation development of the mouse embryo. *J Cell Sci* 2006:**119**;482–489.

Dias BG, Ressler KJ. Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nat Neurosci* 2014:**17**;89–96.

Dietz DM, Laplant Q, Watts EL, Hodes GE, Russo SJ, Feng J, Oosting RS, Vialou V, Nestler EJ. Paternal transmission of stress-induced pathologies. *Biol Psychiatry* 2011:**70**;408–414.

Domínguez F, Garrido-Gómez T, López JA, Camafeita E, Quiñonero A, Pellicer A, Simón C. Proteomic analysis of the human receptive versus non-receptive endometrium using differential in-gel electrophoresis and MALDI-MS unveils stathmin 1 and annexin A2 as differentially regulated. *Hum Reprod* 2009:**24**;2607–17.

Donkin I, Versteyhe S, Ingerslev LR, Qian K, Mechta M, Nordkap L, Mortensen B, Appel EVR, Jørgensen N, Kristiansen VB, *et al.* Obesity and bariatric surgery drive epigenetic variation of spermatozoa in humans. *Cell Metab* 2016:**23**;369–378.

Drabovich AP, Saraon P, Jarvi K, Diamandis EP. Seminal plasma as a diagnostic fluid for male reproductive system disorders. *Nat Rev Urol* 2014:**11**;278–288.

Driancourt M a, Thuel B. Control of oocyte growth and maturation by follicular cells and molecules present in follicular fluid. A review. *Reprod Nutr Dev* 1998:**38**;345–362.

Eppig JJ. Oocyte-somatic cell interactions during oocyte growth and maturation in the mammal. *Dev Biol (N Y 1985)* 1985:**1**;313–47.

Erkek S, Hisano M, Liang CY, Gill M, Murr R, Dieker J, Schübeler D, Vlag J Van Der, Stadler MB, Peters AHFM. Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat Struct Mol Biol* 2013:**20**;868–875.

Ernesto JI, Weigel Muñoz M, Battistone MA, Vasen G, Martínez-López P, Orta G, Figueiras-Fierro D, De la Vega-Beltran JL, Moreno IA, Guidobaldi HA, *et al.* CRISP1 as a novel CatSper regulator that modulates sperm motility and orientation during fertilization. *J Cell Biol* 2015:**210**;1213–1224.

Estes SJ, Ye B, Qiu W, Cramer D, Hornstein MD, Missmer SA. A proteomic analysis

of IVF follicular fluid in women ≤32 years old. *Fertil Steril* 2009:**92**;1569–1578.

Farah Naquiah MZ, James RJ, Suratman S, Lee LS, Mohd Hafidz MI, Salleh MZ, Teh LK. Transgenerational effects of paternal heroin addiction on anxiety and aggression behavior in male offspring. *Behav Brain Funct* 2016:**12**;23.

Fawcett DW, Chemes HE. Changes in distribution of nuclear pores during differentiation of the male germ cells. *Tissue Cell* 1979:**11**;147–62.

Feng Y, Yu D, Yang L, Da M, Wang Z, Lin Y, Ni B, Wang S, Mo X. Maternal lifestyle factors in pregnancy and congenital heart defects in offspring: review of the current evidence. *Ital J Pediatr* 2014:**40**;85.

Ficarro S, Chertihin O, Westbrook VA, White F, Jayes F, Kalab P, Marto JA, Shabanowitz J, Herr JC, Hunt DF, *et al.* Phosphoproteome analysis of capacitated human sperm: Evidence of tyrosine phosphorylation of a kinase-anchoring protein 3 and valosin-containing protein/p97 during capacitation. *J Biol Chem* 2003:**278**;11579–11589.

Fowler PA, Tattum J, Bhattacharya S, Klonisch T, Hombach-Klonisch S, Gazvani R, Lea RG, Miller I, Simpson WG, Cash P. An investigation of the effects of endometriosis on the proteome of human eutopic endometrium: A heterogeneous tissue with a complex disease. *Proteomics* 2007:**7**;130–142.

Frapsauce C, Pionneau C, Bouley J, Delarouziere V, Berthaut I, Ravel C, Antoine J-M, Soubrier F, Mandelbaum J. Proteomic identification of target proteins in normal but nonfertilizing sperm. *Fertil Steril* 2014:**102**;372–80.

Frapsauce C, Pionneau C, Bouley J, de Larouzière V, Berthaut I, Ravel C, Antoine J-M, Soubrier F, Mandelbaum J. [Unexpected in vitro fertilization failure in patients with normal sperm: a proteomic analysis]. *Gynecol Obstet Fertil* 2009:**37**;796–802.

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**;1531–1536.

Fullston T, McPherson NO, Owens JA, Kang WX, Sandeman LY, Lane M. Paternal obesity induces metabolic and sperm disturbances in male offspring that are exacerbated by their exposure to an “obesogenic” diet. *Physiol Rep* 2015:**3**;e12336.

Gapp K, Jawaid A, Sarkies P, Bohacek J, Pelczar P, Prados J, Farinelli L, Miska E, Mansuy IM. Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nat Neurosci* 2014:**17**;667–669.

Garrido-Gomez T, Dominguez F, Lopez JA, Camafeita E, Quiñonero A, Martinez-Conejero JA, Pellicer A, Conesa A, Simón C. Modeling Human Endometrial Decidualization from the Interaction between Proteome and Secretome. *J Clin Endocrinol Metab* 2011:**96**;706–716.

Garrod D, Chidgey M. Desmosome structure, composition and function. *Biochim Biophys Acta - Biomembr* 2008:**1778**;572–587.

Gilchrist RB, Lane M, Thompson JG. Oocyte-secreted factors: regulators of cumulus

cell function and oocyte quality. *Hum Reprod Update* 2008:**14**;159–177.

Gilula NB, Epstein ML, Beers WH. Cell-to-cell communication and ovulation. A study of the cumulus-oocyte complex. *J Cell Biol* 1978:**78**;58–75.

Grandjean V, Fourré S, De Abreu DAF, Derieppe M-A, Remy J-J, Rassoulzadegan M. RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders. *Sci Rep* 2015:**5**;18193.

Green GR, Balhorn R, Poccia DL, Hecht NB. Synthesis and processing of mammalian protamines and transition proteins. *Mol Reprod Dev* 1994:**37**;255–263.

Gu B, Zhang J, Wu Y, Zhang X, Tan Z, Lin Y, Huang X, Chen L, Yao K, Zhang M. Proteomic analyses reveal common promiscuous patterns of cell surface proteins on human embryonic stem cells and sperms. *PLoS One* 2011:**6**;e19386.

Gyamera-Acheampong C, Tantibhedhyangkul J, Weerachatyanukul W, Tadros H, Xu H, Loo J-W van de, Pelletier R-M, Tanphaichitr N, Mbikay M. Sperm from Mice Genetically Deficient for the PCSK4 Proteinase Exhibit Accelerated Capacitation, Precocious Acrosome Reaction, Reduced Binding to Egg Zona Pellucida, and Impaired Fertilizing Ability1. *Biol Reprod* 2006:**74**;666–673.

Hamad A. Doctoral thesis: Molecular and physical interactions of human sperm with female tract secretions. 2017.

Hamad M, Shelko N, Montenarh M, Hammadeh ME. The impact of cigarette smoking on protamines 1 and 2 transcripts in human spermatozoa. *Hum Fertil* 2017;1–7.

Hamatani T, Carter MG, Sharov AA, Ko MS. Dynamics of global gene expression changes during mouse preimplantation development. *Dev Cell* 2004:**6**;117–131.

Hammoud SS, Nix DA, Hammoud AO, Gibson M, Cairns BR, Carrell DT. Genome-wide analysis identifies changes in histone retention and epigenetic modifications at developmental and imprinted gene loci in the sperm of infertile men. *Hum Reprod* 2011:**26**;2558–2569.

Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR. Distinctive chromatin in human sperm packages genes for embryo development. *Nature* 2009:**460**;473–478.

Hammoud SS, Purwar J, Pflueger C, Cairns BR, Carrell DT. Alterations in sperm DNA methylation patterns at imprinted loci in two classes of infertility. *Fertil Steril* 2010:**94**;1728–1733.

Hannan NJ, Stephens AN, Rainczuk A, Hincks C, Rombauts LJF, Salamonsen LA. 2D-DiGE Analysis of the Human Endometrial Secretome Reveals Differences between Receptive and Nonreceptive States in Fertile and Infertile Women. *J Proteome Res* 2010:**9**;6256–6264.

Hanrieder J, Nyakas A, Naessén T, Bergquist J. Proteomic Analysis of Human Follicular Fluid Using an Alternative Bottom-Up Approach. *J Proteome Res* 2008:**7**;443–449.

Hanrieder J, Zuberovic A, Bergquist J. Surface modified capillary electrophoresis combined with in solution isoelectric focusing and MALDI-TOF/TOF MS: A gel-free multidimensional electrophoresis approach for proteomic profiling-Exemplified on human follicular fluid. *J Chromatogr A* 2009:**1216**;3621–3628.

Hao J, Chen M, Ji S, Wang X, Wang Y, Huang X, Yang L, Wang Y, Cui X, Lv L, *et al.* Equatorin is not essential for acrosome biogenesis but is required for the acrosome reaction. *Biochem Biophys Res Commun* 2014:**444**;537–542.

Hashemitabar M, Bahmanzadeh M, Mostafaie A, Orazizadeh M, Farimani M, Nikbakht R. A proteomic analysis of human follicular fluid: Comparison between younger and older women with normal FSH levels. *Int J Mol Sci* 2014:**15**;17518–17540.

Henion TR, Zhou D, Wolfer DP, Jungalwala FB, Hennet T. Cloning of a Mouse β1,3 *N* -Acetylglucosaminyltransferase GlcNAc(β1,3)Gal(β1,4)Glc-ceramide Synthase Gene Encoding the Key Regulator of Lacto-series Glycolipid Biosynthesis. *J Biol Chem* 2001:**276**;30261–30269.

Hetherington L, Schneider EK, DeKretser D, Muller CH, Hondermarck H, Velkov T, Baker MA, Baker MA. Deficiency in Outer Dense Fiber 1 Is a Marker and Potential Driver of Idiopathic Male Infertility. *Mol Cell proteomics* 2016:**15**;3685–3693.

Holt WV, Fazeli A. Sperm selection in the female mammalian reproductive tract. Focus on the oviduct: Hypotheses, mechanisms, and new opportunities. *Theriogenology* 2016:**85**;105–112.

Inoue A, Ogushi S, Saitou M, Suzuki MG, Aoki F. Involvement of Mouse Nucleoplasmin 2 in the Decondensation of Sperm Chromatin after Fertilization. *Biol Reprod* 2011:**85**;70–77.

Inoue N, Ikawa M, Isotani A, Okabe M. The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature* 2005:**434**;234–238.

Intasqui P, Camargo M, Del Giudice PT, Spaine DM, Carvalho VM, Cardozo KHM, Cedenho AP, Bertolla RP. Unraveling the sperm proteome and post-genomic pathways associated with sperm nuclear DNA fragmentation. *J Assist Reprod Genet* 2013:**30**;1187–1202.

Jarkovska K, Martinkova J, Liskova L, Halada P, Moos J, Rezabek K, Gadher SJ, Kovarova H. Proteome mining of human follicular fluid reveals a crucial role of complement cascade and key biological pathways in women undergoing in vitro fertilization. *J Proteome Res* 2010:**9**;1289–301.

Jensen PL, Beck HC, Petersen J, Hreinsson J, Wånggren K, Laursen SB, Sørensen PD, Christensen ST, Andersen CY. Proteomic Analysis of Human Blastocoel Fluid and Blastocyst Cells. *Stem Cells Dev* 2013:**22**;1126–1135.

Jodar M, Selvaraju S, Sendler E, Diamond MP, Krawetz SA, Network for the RM. The presence, role and clinical use of spermatozoal RNAs. *Hum Reprod Update* 2013:**19**;604–624.

Jodar M, Sendler E, Krawetz SA. *The protein and transcript profiles of human semen*. 2016, pp.85–96.

Jodar M, Sendler E, Moskovtsev SI, Librach CL, Goodrich R, Swanson S, Hauser R, Diamond MP, Krawetz SA. Absence of sperm RNA elements correlates with idiopathic male infertility. *Sci Transl Med* 2015:**7**;295re6.

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.

Johnson GD, Mackie P, Jodar M, Moskovtsev S, Krawetz SA. Chromatin and extracellular vesicle associated sperm RNAs. *Nucleic Acids Res* 2015:**43**;6847–6859.

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

Jones JL, Saraswati S, Block AS, Lichti CF, Mahadevan M, Diekman AB. Galectin-3 is associated with prostasomes in human semen. *Glycoconj J* 2010:**27**;227–236.

De Jonge C. Biological basis for human capacitation. *Hum Reprod Update* 2005:**11**;205–214.

Jumeau F, Com E, Lane L, Duek P, Lagarrigue M, Lavigne R, Guillot L, Rondel K, Gateau A, Melaine N, *et al.* Human Spermatozoa as a Model for Detecting Missing Proteins in the Context of the Chromosome-Centric Human Proteome Project. *J Proteome Res* 2015:**14**;3606–3620.

Kaati G, Bygren L, Edvinsson S. Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. *Eur J Hum Genet* 2002:**10**;682–688.

Kaihola H, Yaldir FG, Hreinsson J, Hörnaeus K, Bergquist J, Olivier JDA, Åkerud H, Sundström-Poromaa I. Effects of Fluoxetine on Human Embryo Development. *Front Cell Neurosci* 2016:**10**;160.

Kim YH, Haidl G, Schaefer M, Egner U, Mandal A, Herr JC. Compartmentalization of a unique ADP/ATP carrier protein SFEC (Sperm Flagellar Energy Carrier, AAC4) with glycolytic enzymes in the fibrous sheath of the human sperm flagellar principal piece. *Dev Biol* 2007:**302**;463–476.

Kim YS, Kim MS, Lee SH, Choi BC, Lim JM, Cha KY, Baek KH. Proteomic analysis of recurrent spontaneous abortion: Identification of an inadequately expressed set of proteins in human follicular fluid. *Proteomics* 2006:**6**;3445–3454.

Kimmins S, Sassone-Corsi P. Chromatin remodelling and epigenetic features of germ cells. *Nature* 2005:**434**;583–589.

Kong Q, Banaszynski LA, Geng F, Zhang X, Zhang J, Zhang H, O'Neill CL, Yan P, Liu Z, Shido K, *et al.* Histone variant H3.3-mediated chromatin remodeling is essential for paternal genome activation in mouse preimplantation embryos. *J Biol Chem* 2018:**293**;3829–3838.

Krätzschmar J, Haendler B, Eberspaecher U, Roosterman D, Donner P, Schleuning WD. The human cysteine-rich secretory protein (CRISP) family. Primary structure and tissue distribution of CRISP-1, CRISP-2 and CRISP-3. *Eur J Biochem* 1996:**236**;827–36.

Krawetz SA. Paternal contribution: New insights and future challenges. *Nat Rev Genet* 2005:**6**;633–642.

Krawetz SA, Kruger A, Lalancette C, Tagett R, Anton E, Draghici S, Diamond MP. A survey of small RNAs in human sperm. *Hum Reprod* 2011:**26**;3401–3412.

de Kretser DM, Loveland KL, Meinhardt A, Simorangkir D, Wreford N. Spermatogenesis. *Hum Reprod* 1998:**13**;1–8.

Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Takamatsu K, Chuma S, Kojima-Kita K, Shiromoto Y, Asada N, Toyoda A, Fujiyama A, *et al.* MVH in piRNA processing and gene silencing of retrotransposons. *Genes Dev* 2010:**24**;887–892.

Kushnir MM, Naessén T, Wanggren K, Rockwood AL, Crockett DK, Bergquist J. Protein and steroid profiles in follicular fluid after ovarian hyperstimulation as potential biomarkers of IVF outcome. *J Proteome Res* 2012:**11**;5090–5100.

Kwon WS, Oh SA, Kim YJ, Rahman MS, Park YJ, Pang MG. Proteomic approaches for profiling negative fertility markers in inferior boar spermatozoa. *Sci Rep* 2015:**5**;13821.

Lee HC, Lee SW, Kyo WL, Lee SW, Cha KY, Kye HK, Lee S. Identification of new proteins in follicular fluid from mature human follicles by direct sample rehydration method of two-dimensional polyacrylamide gel electrophoresis. *J Korean Med Sci* 2005:**20**;456–460.

Lefièvre L, Chen Y, Conner SJ, Scott JL, Publicover SJ, Ford WCL, Barratt CLR. Human spermatozoa contain multiple targets for protein S-nitrosylation: An alternative mechanism of the modulation of sperm function by nitric oxide? *Proteomics* 2007:**7**;3066–3084.

Lewandowska AE, Macur K, Czaplewska P, Liss J, Łukaszuk K, Ołdziej S. Qualitative and Quantitative Analysis of Proteome and Peptidome of Human Follicular Fluid Using Multiple Samples from Single Donor with LC-MS and SWATH Methodology. *J Proteome Res* 2017:**16**;3053–3067.

Li LW, Fan LQ, Zhu WB, Nie HC, Sun BL, Luo KL, Liao TT, Tang L, Lu GX. Establishment of a high-resolution 2-D reference map of human spermatozoal proteins from 12 fertile sperm-bank donors. *Asian J Androl* 2007:**9**;321–329.

Lim H-J, Seok AE, Han J, Lee J, Lee S, Kang H-G, Cha BH, Yang Y. N-glycoproteomic analysis of human follicular fluid during natural and stimulated cycles in patients undergoing in vitro fertilization. *Clin Exp Reprod Med* 2017:**44**;63–72.

Liu A-X, Zhu Y-M, Luo Q, Wu Y-T, Gao H-J, Zhu X-M, Xu C-M, Huang H-F. Specific peptide patterns of follicular fluids at different growth stages analyzed by

matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Biochim Biophys Acta - Gen Subj* 2007:**1770**;29–38.

Liu Y, Guo Y, Song N, Fan Y, Li K, Teng X, Guo Q, Ding Z. Proteomic pattern changes associated with obesity-induced asthenozoospermia. *Andrology* 2015:**3**;247–259.

Machtinger R, Laurent LC, Baccarelli AA. Extracellular vesicles: Roles in gamete maturation, fertilization and embryo implantation. *Hum Reprod Update* 2016:**22**;182–193.

Maldera JA, Weigel Munoz M, Chirinos M, Busso D, GE Raffo F, Battistone MA, Blaquier JA, Larrea F, Cuasnicu PS. Human fertilization: epididymal hCRISP1 mediates sperm-zona pellucida binding through its interaction with ZP3. *Mol Hum Reprod* 2014:**20**;341–349.

Martin-DeLeon PA. Epididymosomes: Transfer of Fertility-modulating Proteins to the Sperm Surface. *Asian J Androl* 2015:**17**;720–725.

Martinez-Heredia J, Estanyol JM, Ballesca JL, Oliva R, Martínez-Heredia J, Estanyol JM, Ballescà JL, Oliva R. Proteomic identification of human sperm proteins. *Proteomics* 2006:**6**;4356–4369.

Martínez-Heredia J, de Mateo S, Vidal-Taboada JM, Ballescà JL, Oliva R. Identification of proteomic differences in asthenozoospermic sperm samples. *Hum Reprod* 2008:**23**;783–791.

Maruyama R, Singson A. Taking care of Dad's DNA. *Genome Biol* 2006:**7**;244.

de Mateo S, Castillo J, Estanyol JM, Ballescà JL, Oliva R. Proteomic characterization of the human sperm nucleus. *Proteomics* 2011:**11**;2714–2726.

de Mateo S, Martinez-Heredia J, Estanyol JM, Dominguez-Fandos D, Vidal-Taboada JM, Ballesca JL, Oliva R, Martínez-Heredia J, Estanyol JM, Domíguez-Fandos D, *et al.* Marked correlations in protein expression identified by proteomic analysis of human spermatozoa. *Proteomics* 2007:**7**;4264–4277.

Mattson BA, Albertini DF. Oogenesis: Chromatin and microtubule dynamics during meiotic prophase. *Mol Reprod Dev* 1990:**25**;374–383.

Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. Demethylation of the zygotic paternal genome. *Nature* 2000:**403**;501–502.

McDonough CE, Whittington E, Pitnick S, Dorus S. Proteomics of reproductive systems: Towards a molecular understanding of postmating, prezygotic reproductive barriers. *J Proteomics* 2016:**135**;26–37.

McReynolds S, Dzieciatkowska M, McCallie BR, Mitchell SD, Stevens J, Hansen K, Schoolcraft WB, Katz-Jaffe MG. Impact of maternal aging on the molecular signature of human cumulus cells. *Fertil Steril* 2012:**98**;1574–1580.e5.

Mezquita C. Chromatin composition, structure and function in spermatogenesis. *Revis Biol Celular* 1985:**5**;1–124.

Miescher F. Das Protamin, eine neue organische Base aus den Samenfäden des Rheinlachses. *Berichte der Dtsch Chem Gesellschaft* 1874:**7**;376–379.

Mohan KN, Ding F, Chaillet JR. Distinct Roles of DMAP1 in Mouse Development. *Mol Cell Biol* 2011:**31**;1861–1869.

Naaby-Hansen S, Diekman A, Shetty J, Flickinger CJ, Westbrook A, Herr JC. Identification of calcium-binding proteins associated with the human sperm plasma membrane. *Reprod Biol Endocrinol* 2010:**8**;6.

Naaby-Hansen S, Herr JC. Heat shock proteins on the human sperm surface. *J Reprod Immunol* 2010:**84**;32–40.

Nance J. Getting to know your neighbor: Cell polarization in early embryos. *J Cell Biol* 2014:**206**;823–832.

Neeli I, Khan SN, Radic M. Histone Deimination As a Response to Inflammatory Stimuli in Neutrophils. *J Immunol* 2008:**180**;1895–1902.

Ng S-F, Lin RCY, Laybutt DR, Barres R, Owens J a, Morris MJ. Chronic high-fat diet in fathers programs β-cell dysfunction in female rat offspring. *Nature* 2010:**467**;963–966.

Niakan KK, Han J, Pedersen RA, Simon C, Pera RAR. Human pre-implantation embryo development. *Development* 2012:**139**;829–841.

Nixon B, Mitchell LA, Anderson AL, Mclaughlin EA, O'bryan MK, Aitken RJ. Proteomic and functional analysis of human sperm detergent resistant membranes. *J Cell Physiol* 2011:**226**;2651–2665.

Ntostis P, Carter D, Iles D, Huntriss J, Tzetis M, Miller D. Potential sperm contributions to the murine zygote predicted by in silico analysis. *Reproduction* 2017:**154**;777–788.

Oh JW, Kim SK, Cho K-C, Kim M-S, Suh CS, Lee JR, Kim KP. Proteomic analysis of human follicular fluid in poor ovarian responders during in vitro fertilization. *Proteomics* 2017:**17**;1600333.

Ohta T, Essner R, Ryu JH, Palazzo RE, Uetake Y, Kuriyama R. Characterization of Cep135, a novel coiled-coil centrosomal protein involved in microtubule organization in mammalian cells. *J Cell Biol* 2002:**156**;87–99.

Okada Y, Yamagata K, Hong K, Wakayama T, Zhang Y. A role for the elongator complex in zygotic paternal genome demethylation. *Nature* 2010:**463**;554–558.

Oliva R. Protamines and male infertility. *Hum Reprod Update* 2006:**12**;417–435.

Oliva R, Dixon GH. Vertebrate protamine genes and the histone-to-protamine replacement reaction. *Prog Nucleic Acid Res Mol Biol* 1991:**40**;25–94.

Oliva R, Martínez-Heredia J, Estanyol JM, Martinez-Heredia J, Estanyol JM. Proteomics in the study of the sperm cell composition, differentiation and function. *Syst Biol Reprod Med* 2008:**54**;23–36.

Oliva R, de Mateo S, Estanyol JM. Sperm cell proteomics. *Proteomics* 2009:**9**;1004–1017.

Paasch U, Heidenreich F, Pursche T, Kuhlisch E, Kettner K, Grunewald S, Kratzsch J, Dittmar G, Glander H-J, Hoflack B, *et al.* Identification of Increased Amounts of Eppin Protein Complex Components in Sperm Cells of Diabetic and Obese Individuals by Difference Gel Electrophoresis. *Mol Cell Proteomics* 2011:**10**;M110.007187.

Pantano L, Jodar M, Bak M, Ballescà JL, Tommerup N, Oliva R, Vavouri T. The small RNA content of human sperm reveals pseudogene-derived piRNAs complementary to protein-coding genes. *RNA* 2015:**21**;1085–1095.

Parmar T, Gadkar-Sable S, Savardekar L, Katkam R, Dharma S, Meherji P, Puri CP, Sachdeva G. Protein profiling of human endometrial tissues in the midsecretory and proliferative phases of the menstrual cycle. *Fertil Steril* 2009:**92**;1091–1103.

Parte PP, Rao P, Redij S, Lobo V, D'Souza SJ, Gajbhiye R, Kulkarni V. Sperm phosphoproteome profiling by ultra performance liquid chromatography followed by data independent analysis (LC-MS(E)) reveals altered proteomic signatures in asthenozoospermia. *J Proteomics* 2012:**75**;5861–71.

Paule SG, Airey LM, Li Y, Stephens AN, Nie G. Proteomic Approach Identifies Alterations in Cytoskeletal Remodelling Proteins during Decidualization of Human Endometrial Stromal Cells. *J Proteome Res* 2010:**9**;5739–5747.

Pembrey M, Saffery R, Bygren LO. Human transgenerational responses to early-life experience: potential impact on development, health and biomedical research. *J Med Genet* 2014:**51**;563–572.

Pembrey ME, Bygren LO, Kaati G, Edvinsson S, Northstone K, Sjöström M, Golding J, Alspac T, Team S. Sex-specific, male-line transgenerational responses in humans. *Eur J Hum Genet* 2006:**14**;159–166.

Perry AC, Jones R, Moisyadi S, Coadwell J, Hall L. The novel epididymal secretory protein ESP13.2 in Macaca fascicularis. *Biol Reprod* 1999:**61**;965–72.

Pixton KL, Deeks ED, Flesch FM, Moseley FLC, Björndahl L, Ashton PR, Barratt CLR, Brewis IA. Sperm proteome mapping of a patient who experiencd failed fertilization at IVF reveals altered expression of at least 20 proteins compared with fertile donors: Case report. *Hum Reprod* 2004:**19**;1438–1447.

Poccia D. Remodeling of nucleoproteins during gametogenesis, fertilization, and early development. *Int Rev Cytol* 1986:**105**;1–65.

Rando OJ. Daddy issues: Paternal effects on phenotype. *Cell* 2012:**151**;702–708.

Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I, Cuzin F. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* 2006:**441**;469–474.

Redgrove KA, Anderson AL, Dun MD, McLaughlin EA, O'Bryan MK, Aitken RJ, Nixon B. Involvement of multimeric protein complexes in mediating the

capacitation-dependent binding of human spermatozoa to homologous zonae pellucidae. *Dev Biol* 2011:**356**;460–474.

Regiani T, Cordeiro FB, Da Costa LDVT, Salgueiro J, Cardozo K, Carvalho VM, Perkel KJ, Zylbersztejn DS, Cedenho AP, Lo Turco EG. Follicular fluid alterations in endometriosis: Label-free proteomics by MSEas a functional tool for endometriosis. *Syst Biol Reprod Med* 2015:**61**;263–276.

Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. *Science* 2001:**293**;1089–1093.

Reik W, Walter J. Genomic imprinting: Parental influence on the genome. *Nat Rev Genet* 2001:**2**;21–32.

Renaud SM, Fountain SB. Transgenerational effects of adolescent nicotine exposure in rats: Evidence for cognitive deficits in adult female offspring. *Neurotoxicol Teratol* 2016:**56**;47–54.

Richani D, Gilchrist RB. The epidermal growth factor network: role in oocyte growth, maturation and developmental competence. *Hum Reprod Update* 2018:**24**;1–14.

Rodgers AB, Morgan CP, Bronson SL, Revello S, Bale TL. Paternal stress exposure alters sperm microRNA content and reprograms offspring HPA stress axis regulation. *J Neurosci* 2013:**33**;9003–9012.

Ronquist GKG, Larsson A, Ronquist GKG, Isaksson A, Hreinsson J, Carlsson L, Stavreus-Evers A. Prostasomal DNA characterization and transfer into human sperm. *Mol Reprod Dev* 2011:**78**;467–476.

Da Ros V, Muñoz M, Battistone M, Brukman N, Carvajal G, Cohen D, Cuasnicú P, Curci L, Da Ros V, Gómez Elías M, *et al.* From the Epididymis to the Egg: participation of CRISP Proteins in Mammalian Fertilization. *Asian J Androl* 2015:**17**;711–715.

Saez F, Frenette G, Sullivan R. Epididymosomes and Prostasomes: Their Roles in Posttesticular Maturation of the Sperm Cells. *J Androl* 2003:**24**;149–154.

Saez F, Sullivan R. Prostasomes, post-testicular sperm maturation and fertility. *Front Biosci (Landmark Ed* 2016:**21**;1464–73.

Salustri A, Fulop C, Hascall VC, Camaioni A. Oocyte-Granulosa Cell Interactions. In: *The Ovary: Second Edition*. 2003, Elsevier Inc., pp.131–143.

Santos F, Dean W. Epigenetic reprogramming during early development in mammals. *Reproduction* 2004:**127**;643–651.

Saraswat M, Joenväärä S, Jain T, Tomar AK, Sinha A, Singh S, Yadav S, Renkonen R. Human Spermatozoa Quantitative Proteomic Signature Classifies Normo- and Asthenozoospermia. *Mol Cell Proteomics* 2017:**16**;57–72.

Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. Global quantification of mammalian gene expression control. *Nature* 2011:**473**;337–342.

Scotchie JG, Fritz MA, Young SL, Mocanu M, Lessey BA. Proteomic Analysis of the Luteal Endometrial Secretome. *Reprod Sci* 2009:**16**;883–893.

Secciani F, Bianchi L, Ermini L, Cianti R, Armini A, La Sala GB, Focarelli R, Bini L, Rosati F. Protein profile of capacitated versus ejaculated human sperm. *J Proteome Res* 2009:**8**;3377–3389.

Severino V, Malorni L, Cicatiello AE, D'Esposito V, Longobardi S, Colacurci N, Miraglia N, Sannolo N, Farina A, Chambery A. An integrated approach based on multiplexed protein array and iTRAQ labeling for in-depth identification of pathways associated to IVF outcome. Uversky VN (ed). *PLoS One* 2013:**8**;e77303.

Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, Belleannee C, Kucukural A, Serra RW, Sun F, *et al.* Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* 2016:**351**;391–396.

Sharma U, Rando OJ. Metabolic Inputs into the Epigenome. *Cell Metab* 2017:**25**;544–558.

Siklenka K, Erkek S, Godmann M, Lambrot R, McGraw S, Lafleur C, Cohen T, Xia J, Suderman M, Hallett M, *et al.* Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science* 2015:**350**;aab2006.

Singh AP, Rajender S. CatSper channel, sperm function and male fertility. *Reprod Biomed Online* 2015:**30**;28–38.

Siva AB, Kameshwari DB, Singh V, Pavani K, Sundaram CS, Rangaraj N, Deenadayal M, Shivaji S. Proteomics-based study on asthenozoospermia: Differential expression of proteasome alpha complex. *Mol Hum Reprod* 2010:**16**;452–462.

Skerget S, Rosenow MA, Petritis K, Karr TL. Sperm Proteome Maturation in the Mouse Epididymis. Baltz JM (ed). *PLoS One* 2015:**10**;e0140650.

Skerget S, Rosenow M, Polpitiya A, Petritis K, Dorus S, Karr TL. The Rhesus Macaque (Macaca mulatta) Sperm Proteome. *Mol Cell Proteomics* 2013:**12**;3052–3067.

Soubry A, Guo L, Huang Z, Hoyo C, Romanus S, Price T, Murphy SK. Obesity-related DNA methylation at imprinted genes in human sperm: Results from the TIEGER study. *Clin Epigenetics* 2016:**8**;51.

Stephens AN, Hannan NJ, Rainczuk A, Meehan KL, Chen J, Nicholls PK, Rombauts LJF, Stanton PG, Robertson DM, Salamonsen LA. Post-Translational Modifications and Protein-Specific Isoforms in Endometriosis Revealed by 2D DIGE. *J Proteome Res* 2010:**9**;2438–2449.

Sullivan R, Mieusset R. The human epididymis: Its function in sperm maturation. *Hum Reprod Update* 2016:**22**;574–587.

Sullivan R, Saez F. Epididymosomes, prostasomes, and liposomes: Their roles in mammalian male reproductive physiology. *Reproduction* 2013:**146**;R21-35.

Sullivan R, Saez F, Girouard J, Frenette G. Role of exosomes in sperm maturation during the transit along the male reproductive tract. *Blood Cells, Mol Dis*

2005:**35**;1–10.

Sun G, Jiang M, Zhou T, Guo Y, Cui Y, Guo X, Sha J. Insights into the lysine acetylproteome of human sperm. *J Proteomics* 2014:**109**;199–211.

Sutovsky P. Sperm-egg adhesion and fusion in mammals. *Expert Rev Mol Med* 2009:**11**;e11.

Sutovsky P, Manandhar G. Mammalian spermatogenesis and sperm structure: Anatomical and compartmental analysis. In: De Jonge CJ, Barratt CL (eds) *The Sperm Cell: Production, Maturation, Fertilization, Regeneration*. 2006, Cambridge University Press, New York, pp.1–30.

Sutton WS. The chromosomes in heredity. *Biol Bull* 1903:**4**;231–251.

Svejstrup JQ. Elongator complex: how many roles does it play? *Curr Opin Cell Biol* 2007:**19**;331–336.

Swann K, Windsor S, Campbell K, Elgmati K, Nomikos M, Zernicka-Goetz M, Amso N, Lai FA, Thomas A, Graham C. Phospholipase C-ζ-induced Ca 2+ oscillations cause coincident cytoplasmic movements in human oocytes that failed to fertilize after intracytoplasmic sperm injection. *Fertil Steril* 2012:**97**;742–747.

Swegen A, Grupen CG, Gibb Z, Baker MA, de Ruijter-Villani M, Smith ND, Stout TAE, Aitken RJ. From Peptide Masses to Pregnancy Maintenance: A Comprehensive Proteomic Analysis of The Early Equine Embryo Secretome, Blastocoel Fluid, and Capsule. *Proteomics* 2017:**17**;1600433.

Tardif S, Guyonnet B, Cormier N, Cornwall GA. Alteration in the processing of the ACRBP/sp32 protein and sperm head/acrosome malformations in proprotein convertase 4 (PCSK4) null mice. *Mol Hum Reprod* 2012:**18**;298–307.

Thacker S, Yadav SP, Sharma RK, Kashou A, Willard B, Zhang D, Agarwal A. Evaluation of sperm proteins in infertile men: A proteomic approach. *Fertil Steril* 2011:**95**;2745–2748.

The Gene Ontology Consortium. Expansion of the gene ontology knowledgebase and resources: The gene ontology consortium. *Nucleic Acids Res* 2017:**45**;D331–D338.

Torabi F, Bogle OA, Estanyol JM, Oliva R, Miller D. Zona pellucida-binding protein 2 (ZPBP2) and several proteins containing BX7B motifs in human sperm may have hyaluronic acid binding or recognition properties. *Mol Hum Reprod* 2017:**23**;803–816.

Toshimori K, Saxena DK, Tanii I, Yoshinaga K. An MN9 antigenic molecule, equatorin, is required for successful sperm-oocyte fusion in mice. *Biol Reprod* 1998:**59**;22–29.

Lo Turco EG, Souza GHMF, Garcia JS, Ferreira CR, Eberlin MN, Bertolla RP. Effect of endometriosis on the protein expression pattern of follicular fluid from patients submitted to controlled ovarian hyperstimulation for in vitro fertilization. *Hum Reprod* 2010:**25**;1755–1766.

Lo Turco EGG, Cordeiro FB, Lopes PH de C, Gozzo FCFC, Pilau EJ, Soler TB, da Silva BF, Del Giudice PT, Bertolla RP, Fraietta R, *et al.* Proteomic analysis of follicular fluid from women with and without endometriosis: New therapeutic targets and biomarkers. *Mol Reprod Dev* 2013:**80**;441–450.

Twigt J, Steegers-Theunissen RP, Bezstarosti K, Demmers JAA. Proteomic analysis of the microenvironment of developing oocytes. *Proteomics* 2012:**12**;1463–71.

Uhlén M, Björling E, Agaton C, Szigyarto CA-K, Amini B, Andersen E, Andersson A-C, Angelidou P, Asplund A, Asplund C, *et al.* A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics. *Mol Cell Proteomics* 2005:**4**;1920–1932.

Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, *et al.* Tissue-based map of the human proteome. *Science* 2015:**347**;1260419–1260419.

Vance JE, Vance DE. Phospholipid biosynthesis in mammalian cells. *Biochem Cell Biol* 2004:**82**;113–128.

Vandenbrouck Y, Lane L, Carapito C, Duek P, Rondel K, Bruley C, Macron C, Gonzalez de Peredo A, Couté Y, Chaoui K, *et al.* Looking for Missing Proteins in the Proteome of Human Spermatozoa: An Update. *J Proteome Res* 2016:**15**;3998–4019.

Vassena R, Boue S, Gonzalez-Roca E, Aran B, Auer H, Veiga A, Belmonte JCI. Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development* 2011:**138**;3699–3709.

Vassoler FM, White SL, Schmidt HD, Sadri-Vakili G, Christopher Pierce R. Epigenetic inheritance of a cocaine-resistance phenotype. *Nat Neurosci* 2013:**16**;42–47.

Vavouri T, Lehner B. Chromatin organization in sperm may be the major functional consequence of base composition variation in the human genome. *PLoS Genet* 2011:**7**;e1002036.

Vigodner M, Shrivastava V, Gutstein LE, Schneider J, Nieves E, Goldstein M, Feliciano M, Callaway M. Localization and identification of sumoylated proteins in human sperm: excessive sumoylation is a marker of defective spermatozoa. *Hum Reprod* 2013:**28**;210–223.

Virant-Klun I, Leicht S, Hughes C, Krijgsveld J. Identification of Maturation-Specific Proteins by Single-Cell Proteomics of Human Oocytes. *Mol Cell Proteomics* 2016:**15**;2616–2627.

Visconti PE. Understanding the molecular basis of sperm capacitation through kinase design. *Proc Natl Acad Sci* 2009:**106**;667–668.

Visconti PE, Galantino-Homer H, Moore GD, Bailey JL, Ning X, Fornes M, Kopf GS. The molecular basis of sperm capacitation. *J Androl* 1998:**19**;242–248.

Vojtech L, Woo S, Hughes S, Levy C, Ballweber L, Sauteraud RP, Strobl J, Westerberg K, Gottardo R, Tewari M, *et al.* Exosomes in human semen carry a distinctive

repertoire of small non-coding RNAs with potential regulatory functions. *Nucleic Acids Res* 2014:**42**;7290–7304.

Wang G, Guo Y, Zhou T, Shi X, Yu J, Yang Y, Wu Y, Wang J, Liu M, Chen X, *et al.* In-depth proteomic analysis of the human sperm reveals complex protein compositions. *J Proteomics* 2013:**79**;114–122.

Wang J, Qi L, Huang S, Zhou T, Guo Y, Wang G, Guo X, Zhou Z, Sha J. Quantitative Phosphoproteomics Analysis Reveals a Key Role of Insulin Growth Factor 1 Receptor (IGF1R) Tyrosine Kinase in Human Sperm Capacitation. *Mol Cell Proteomics* 2015:**14**;1104–1112.

Wang L, Magdaleno S, Tabas I, Jackowski S. Early embryonic lethality in mice with targeted deletion of the CTP:phosphocholine cytidylyltransferase alpha gene (Pcyt1a). *Mol Cell Biol* 2005:**25**;3357–3363.

Wang Y, Medvid R, Melton C, Jaenisch R, Blelloch R. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet* 2007:**39**;380–385.

Wang Y, Wan J, Ling X, Liu M, Zhou T. The human sperm proteome 2.0: An integrated resource for studying sperm functions at the level of posttranslational modification. *Proteomics* 2016:**16**;2597–2601.

Wasbrough ER, Dorus S, Hester S, Howard-Murkin J, Lilley K, Wilkin E, Polpitiya A, Petritis K, Karr TL. The Drosophila melanogaster sperm proteome-II (DmSP-II). *J Proteomics* 2010:**73**;2171–2185.

Wei Y, Su J, Liu H, Lv J, Wang F, Yan H, Wen Y, Liu H, Wu Q, Zhang Y. MetaImprint: an information repository of mammalian imprinted genes. *Development* 2014a:**141**;2516–2523.

Wei Y, Yang C-R, Wei Y-P, Zhao Z-A, Hou Y, Schatten H, Sun Q-Y. Paternally induced transgenerational inheritance of susceptibility to diabetes in mammals. *Proc Natl Acad Sci* 2014b:**111**;1873–1878.

Williams HL, Mansell S, Alasmari W, Brown SG, Wilson SM, Sutton KA, Miller MR, Lishko P V., Barratt CLR, Publicover SJ, *et al.* Specific loss of CatSper function is sufficient to compromise fertilizing capacity of human spermatozoa. *Hum Reprod* 2015:**30**;2737–2746.

Wolkowicz MJ, Digilio L, Klotz K, Shetty J, Flickinger CJ, Herr JC. Equatorial Segment Protein (ESP) Is a Human Alloantigen Involved in Sperm-Egg Binding and Fusion. *J Androl* 2008:**29**;272–282.

Wolkowicz MJ, Shetty J, Westbrook A, Klotz K, Jayes F, Mandal A, Flickinger CJ, Herr JC. Equatorial Segment Protein Defines a Discrete Acrosomal Subcompartment Persisting Throughout Acrosomal Biogenesis1. *Biol Reprod* 2003:**69**;735–745.

Wright SJ. Sperm Nuclear Activation during Fertilization. *Curr Top Dev Biol* 1999:**46**;133–178.

Wu Y-T, Wu Y, Zhang J-Y, Hou N-N, Liu A-X, Pan J-X, Lu J-Y, Sheng J-Z, Huang H-F. Preliminary proteomic analysis on the alterations in follicular fluid proteins from women undergoing natural cycles or controlled ovarian hyperstimulation. *J Assist Reprod Genet* 2015:**32**;417–427.

Xu W, Hu H, Wang Z, Chen X, Yang F, Zhu Z, Fang P, Dai J, Wang L, Shi H, *et al.* Proteomic characteristics of spermatozoa in normozoospermic patients with infertility. *J Proteomics* 2012:**75**;5426–5436.

Yamamoto K, Wang Y, Jiang W, Liu X, Dubois RL, Lin C-S, Ludwig T, Bakkenist CJ, Zha S. Kinase-dead ATM protein causes genomic instability and early embryonic lethality in mice. *J Cell Biol* 2012:**198**;305–313.

Yoon S-Y, Fissore RA. Release of phospholipase C  and [Ca2+]i oscillation-inducing activity during mammalian fertilization. *Reproduction* 2007:**134**;695–704.

Yu H, Diao H, Wang C, Lin Y, Yu F, Lu H, Xu W, Li Z, Shi H, Zhao S, *et al.* Acetylproteomic Analysis Reveals Functional Implications of Lysine Acetylation in Human Spermatozoa (sperm). *Mol Cell Proteomics* 2015:**14**;1009–1023.

Yudin AI, Generao SE, Tollner TL, Treece CA, Overstreet JW, Cherr GN. Beta-Defensin 126 on the Cell Surface Protects Sperm from Immunorecognition and Binding of Anti-Sperm Antibodies1. *Biol Reprod* 2005:**73**;1243–1252.

Zamah AM, Hassis ME, Albertolle ME, Williams KE. Proteomic analysis of human follicular fluid from fertile women. *Clin Proteomics* 2015:**12**;5.

Zeybel M, Hardy T, Wong YK, Mathers JC, Fox CR, Gackowska A, Oakley F, Burt AD, Wilson CL, Anstee QM, *et al.* Multigenerational epigenetic adaptation of the hepatic wound-healing response. *Nat Med* 2012:**18**;1592–1592.

Zhang H, Niu Y, Feng J, Guo H, Ye X, Cui H. Use of proteomic analysis of endometriosis to identify different protein expression in patients with endometriosis versus normal controls. *Fertil Steril* 2006:**86**;274–282.

Zhao C, Huo R, Wang F-Q, Lin M, Zhou Z-M, Sha J-H. Identification of several proteins involved in regulation of sperm motility by proteomic analysis. *Fertil Steril* 2007:**87**;436–438.

Zheng X, Dumitru R, Lackford BL, Freudenberg JM, Singh AP, Archer TK, Jothi R, Hu G. Cnot1, Cnot2, and Cnot3 maintain mouse and human ESC identity and inhibit extraembryonic differentiation. *Stem Cells* 2012:**30**;910–922.

Zhou T, Wang G, Chen M, Zhang M, Guo Y, Yu C, Zhou Z, Si W, Sha J, Guo X. Comparative analysis of macaque and human sperm proteomes: Insights into sperm competition. *Proteomics* 2015:**15**;1564–1573.

Zhu Y, Wu Y, Jin K, Lu H, Liu F, Guo Y, Yan F, Shi W, Liu Y, Cao X, *et al.* Differential proteomic profiling in human spermatozoa that did or did not result in pregnancy via IVF and AID. *Proteomics - Clin Appl* 2013:**7**;850–858.

Zuccotti M, Merico V, Cecconi S, Redi CA, Garagna S. What does it take to make a developmentally competent mammalian egg? *Hum Reprod Update* 2011:**17**;525–

540.

**Figure legends**

**Figure 1. Processes of spermatogenesis, oogenesis, and fertilization.** The process of spermatogenesis takes place in the testis and is initiated with the replication of spermatogonia (upper left). Spermatogonia differentiate into spermatocytes which undergo genetic recombination at meiosis producing haploid round spermatids. Round spermatids experience marked chromatin and morphological changes, including a extreme DNA condensation, the formation of the flagella and the acrosome, the elimination of most of the cytoplasm, and the adoption of a hydrodynamic shape, differentiating first into elongating spermatid, and finally into the testicular spermatozoon (left). Testicular spermatozoa are released to the epididymis, where they initiate a maturation process that includes the acquisition of motility, and other cellular functions and potentials. During the process of ejaculation, the mature sperm cell enters into close contact with the seminal fluid, which is constituted by accessory sex glands fluids and extracellular vesicles (left). The process of oogenesis takes place in the ovary (upper right). Oogonia replicate and differentiate into primary oocytes, which are present at birth, and remain arrested at profase I of meiosis. The developing follicle grows and progresses in order to give rise to the vesicular follicle, which contains the secondary oocyte, arrested at metaphase II of meiosis until ovulation (right). The entry of the sperm cell into the oocyte, through the penetration of the zona pellucida and oocyte plasma membrane, triggers the completion of oocyte meiosis (bottom).

**Figure 2. Sperm proteins functionally involved in the processes of fertilization and pre-implantational embryo development.** Examples of the subset of 103 sperm proteins involved in the process of fertilization, as revealed by the functional analyses based on gene-specific Gene Ontology annotations (left). The proteins indicated here are functionally related to the correct acrosome reaction, binding of sperm to zona pellucida, penetration of zona pellucida, fusion of sperm to egg plasma membrane, and egg activation (left). Upon fertilization, the pre-implantational embryo development starts with the fusion of male and female pronuclei in the zygote, which develops during approximately 6 days in order to reach the blastocyst stage (upper right; ICM: inner cell mass). In humans, the embryo genome activation (EGA) takes place between 4-cell and 8-cell embryo stages, although minor gene activation occurs before, between 2-cell and 4-cell embryo stages (middle right). In addition, paternal genome is demethylated within the first 24 h of embryo development, much faster than the maternal genome (middle right). The analysis of the phenotype data from gene-null mice enclosed in the Mouse Genome Informatics database revealed the potential critical role of 59 sperm proteins in different processes occurring during the first divisions of the embryo (up to 8-cell stage; 11 proteins), the formation of the morula (29 proteins), and the development of the blastocyst (19 proteins), which are indicated in boxes (bottom right).

**Figure 3. Integrative analysis of proteomic and transcriptomic profiles from gametes and blastocyst. A.** Venn diagram of the compiled protein profiles of human sperm, oocyte, and blastocyst. The venn diagram analysis shows 5362, 223 and 400 proteins exclusively detected in sperm, oocyte, and blastocyst, respectively. A total of

444 proteins are indentified in all three proteomes, 659 are shared between sperm and oocyte, 406 between blastocyst and sperm, and 50 between blastocyst and oocyte. **B.** Assessment of the potential origin of blastocyst proteins according to the corresponding transcriptional pattern observed in human gametes and all the cellular stages of the pre-implantational embryo development. The RNA abundance corresponding to blastocyst proteins has allowed the classification in three subgroups of a subset of them, according to their potential embryonic, maternal, or paternal origin: 1) Group 1: blastocyst proteins classified as embryonically derived (n=18), since their corresponding RNA was absent in both gametes, and the levels only increase after embryonic genome activation (EGA). The RNA levels of a component of this group, the claudin-6 (CLDN6), is shown in green. 2) Group 2: blastocyst proteins classified as maternally derived (n=42). Their corresponding transcripts levels were found with low levels in sperm and high levels in the oocyte. Additionally, the levels of those transcripts decrease during early stages of embryogenesis and they do not increase after EGA. The RNA levels of a component of this group, the protein FAM149A (FAM149A), is shown in yellow. 3) Group 3: blastocyst proteins classified as paternally derived (n=26). Their corresponding RNA was found with low levels in oocyte but high abundance in sperm. Moreover, the levels of those transcripts are low during early stages of embryogenesis and they do not increase after EGA. The RNA levels of a component of this group, the HMG box transcription factor BBX (BBX), is shown in blue.

**Figure 4: Human sperm proteins with potential extra-testicular origin.** Examples of the subset of sperm proteins potentially derived from extra-testicular tissues (165 proteins from accessory sex glands and 286 from other extra-testicular

tissues; see Supplementary Table 7), according to RNA-seq and protein data enclosed in the Human Protein Atlas database (right). Sperm proteins are classified based on their involvement in processes related to fertilization and pre-implantational embryo development (blue squares), and epigenetic regulation (grey squares), according to the Gene Ontology annotations data enclosed in the Gene Ontology Consortium database.

**Supplementary Tables legends**

**Supplementary Table 1. Compiled proteomes for human sperm, oocyte, embryo, cumulus cells, follicular fluid and mid-secretory endometrium.** Proteins are identified with the protein accession code from the UniProtKB/Swiss-Prot database and the gene identification from the Gene Cards database. A total of 6871, 1376, 1300, 179, 1394 and 2082 proteins were compiled for human sperm, oocyte, blastocyst, cumulus cells, follicular fluid, and mid-secretory endometrium, respectively.

**Supplementary Table 2. Gene Ontology terms included in the analyses of the sperm proteome.** Gene Ontology terms released from Gene Ontology Consortium database (Release date 2016-12-07).

**Supplementary Table 3. Human sperm proteins with roles in fertilization and embryo development.** Predicted functional roles for sperm proteins according to Gene Ontology annotations (Gene Ontology Consortion database), and the phenotype data of gene-null mice enclosed in the Mouse Genomic Informatics (MGI) database. A total of 103 proteins were associated to fertilization-related GO terms. A total of 93 proteins are functionally related to embryo development, 59 of which were found in the MGI database.

**Supplementary Table 4. Human sperm proteins potentially involved in the epigenetic inheritance of acquired traits.** A total of 560 sperm proteins are able to modulate gene expression. Those proteins are classified in 4 main groups according to the functional roles predicted by Gene Ontology annotations (Gene Ontology Consortium database): 1) Gene expression regulators including transcript factors and gene expression-related proteins (n=381), 2) DNA methylation modulators (n=25; 9 of them also could modulate histone posttranslational modifications (PTMs)), 3) Histone PTMs modulators (n=118), and 4) Regulators of sncRNAs biogenesis (n=36; 1 of them also could modulate histone PTMs). At least 69 of those 560 proteins have a known role in spermatogenesis whereas 28 have been related to early embryogenesis.

**Supplementary Table 5. Potential origin of blastocyst proteins.** The origin of 173 out of the 1300 proteins detected in the blastocyst was predicted according to their identification within the proteomes of human sperm, oocyte and blastocyst, combined to the published RNA levels detected in each gamete and stages of the pre-implantational embryo development. A total of 108 blastocyst proteins were classified as paternally derived due to 1) their exclusively presence in sperm proteome and the absence of the corresponding RNA in all the RNAseq datasets assessed (n=82), or 2) their absence in oocyte or embryo related cells and fluids proteomes, and their corresponding RNAs detected solely in sperm RNA-seq dataset (n=26). Similarly, 47 blastocyst proteins were classified as maternally derived due to 1) their exclusively presence in oocyte proteome and the absence of the corresponding RNA in all the RNAseq datasets assessed (n=5), or 2) their absence in sperm or embryo related cells and fluids proteomes, and their corresponding RNAs detected at high levels in the oocyte, which progressively drop during the different early embryogenesis stages, without any increase upon the

activation of the embryo transcription. Moreover, those RNAs were not detected in sperm (n=42). Finally, 18 blastocyst proteins were classified as embryonically derived due to their absence in any of the proteomes and RNA-seq dataset assessed, except for those RNA-seq data corresponding to early embryo stages in which embryo transcription has been activated.

**Supplementary Table 6. Subset of sperm proteins with putative origin in accessory sex glands (n=165).** The functional association of the sperm proteins to fertilization, embryo development, and epigenetics was predicted according to gene-specific GO annotations (Gene Ontology Consortium). The expression of the protein in tissues from accessory sex glands was predicted according to the RNA and protein levels enclosed in the Human Protein Atlas database.

**Supplementary Table 7. Subset of sperm proteins with putative origin in tissues outside the male reproductive tract (n=286).** The functional association of the sperm proteins to fertilization, embryo development, and epigenetics was predicted according to gene-specific GO annotations (Gene Ontology Consortium). The expression of the protein in tissues outside the male reproductive tract was predicted according to the protein levels enclosed in the Human Protein Atlas database.

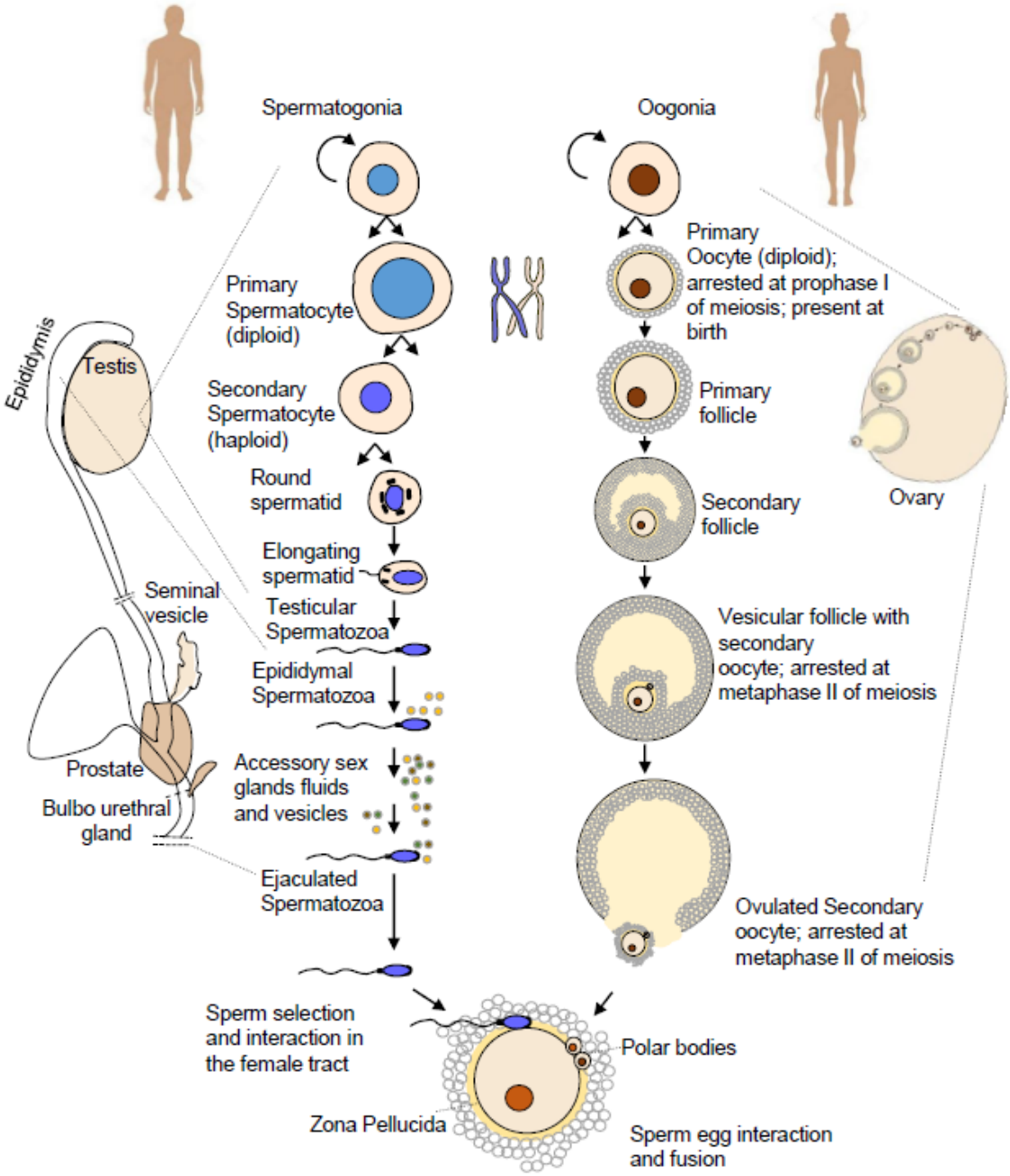

Spermatogonia
Oogonia
Primary Spermatocyte (diploid)
Primary Oocyte (diploid); arrested at prophase I of meiosis; present at birth
Epididymis
Testis
Secondary Spermatocyte (haploid)
Primary follicle
Round spermatid
Secondary follicle
Ovary
Elongating spermatid
Seminal vesicle
Testicular Spermatozoa
Vesicular follicle with secondary oocyte; arrested at metaphase II of meiosis
Epididymal Spermatozoa
Prostate
Accessory sex glands fluids and vesicles
Bulbo urethral gland
Ejaculated Spermatozoa
Ovulated Secondary oocyte; arrested at metaphase II of meiosis
Sperm selection and interaction in the female tract
Polar bodies
Zona Pellucida
Sperm egg interaction and fusion

**Figure 1**

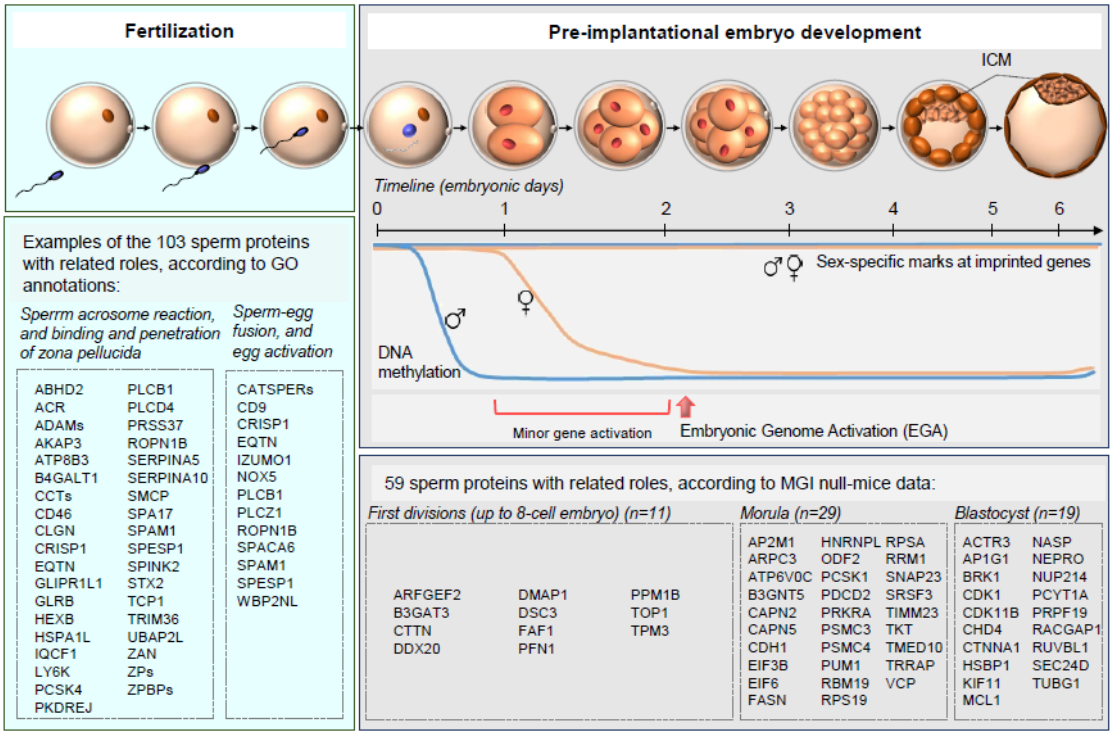
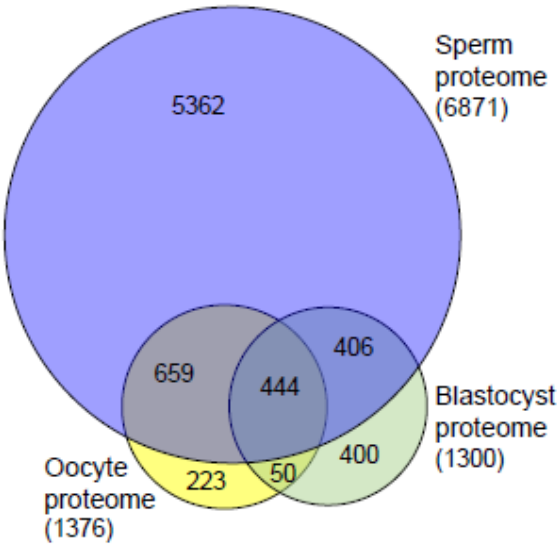

Fertilization
Pre-implantational embryo development
ICM
Timeline (embryonic days)
0
1
2
3
4
5
6
Sex-specific marks at imprinted genes
DNA methylation
Minor gene activation
Embryonic Genome Activation (EGA)
Examples of the 103 sperm proteins with related roles, according to GO annotations:
Sperrm acrosome reaction, and binding and penetration of zona pellucida
ABHD2
ACR
ADAMs
AKAP3
ATP8B3
B4GALT1
CCTs
CD46
CLGN
CRISP1
EQTN
GLIPR1L1
GLRB
HEXB
HSPA1L
IQCF1
LY6K
PCSK4
PKDREJ
PLCB1
PLCD4
PRSS37
ROPN1B
SERPINA5
SERPINA10
SMCP
SPA17
SPAM1
SPESP1
SPINK2
STX2
TCP1
TRIM36
UBAP2L
ZAN
ZPs
ZPBPs
Sperm-egg fusion, and egg activation
CATSPERs
CD9
CRISP1
EQTN
IZUMO1
NOX5
PLCB1
PLCZ1
ROPN1B
SPACA6
SPAM1
SPESP1
WBP2NL
59 sperm proteins with related roles, according to MGI null-mice data:
First divisions (up to 8-cell embryo) (n=11)
ARFGEF2
B3GAT3
CTTN
DDX20
DMAP1
DSC3
FAF1
PFN1
PPM1B
TOP1
TPM3
Morula (n=29)
AP2M1
ARPC3
ATP6V0C
B3GNT5
CAPN2
CAPN5
CDH1
EIF3B
EIF6
FASN
HNRNPL
ODF2
PCSK1
PDCD2
PRKRA
PSMC3
PSMC4
PUM1
RBM19
RPS19
RPSA
RRM1
SNAP23
SRSF3
TIMM23
TKT
TMED10
TRRAP
VCP
Blastocyst (n=19)
ACTR3
AP1G1
BRK1
CDK1
CDK11B
CHD4
CTNNA1
HSBP1
KIF11
MCL1
NASP
NEPRO
NUP214
PCYT1A
PRPF19
RACGAP1
RUVBL1
SEC24D
TUBG1

**Figure 2**

**A** Venn diagram of the proteomes from gametes and blastocyst proteins




**B** RNA abundance corresponding to blastocyst proteins

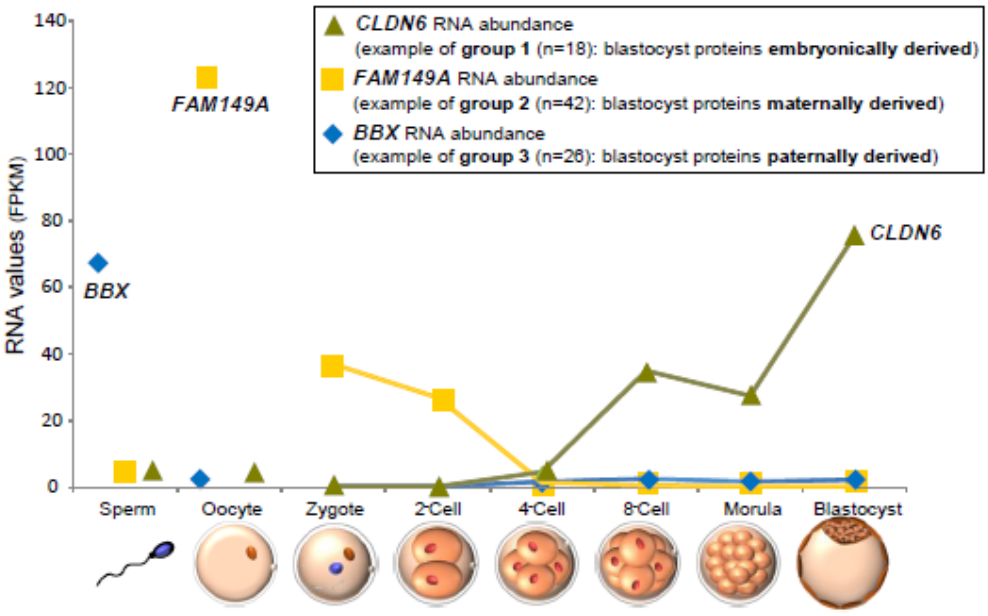



**Figure 3**

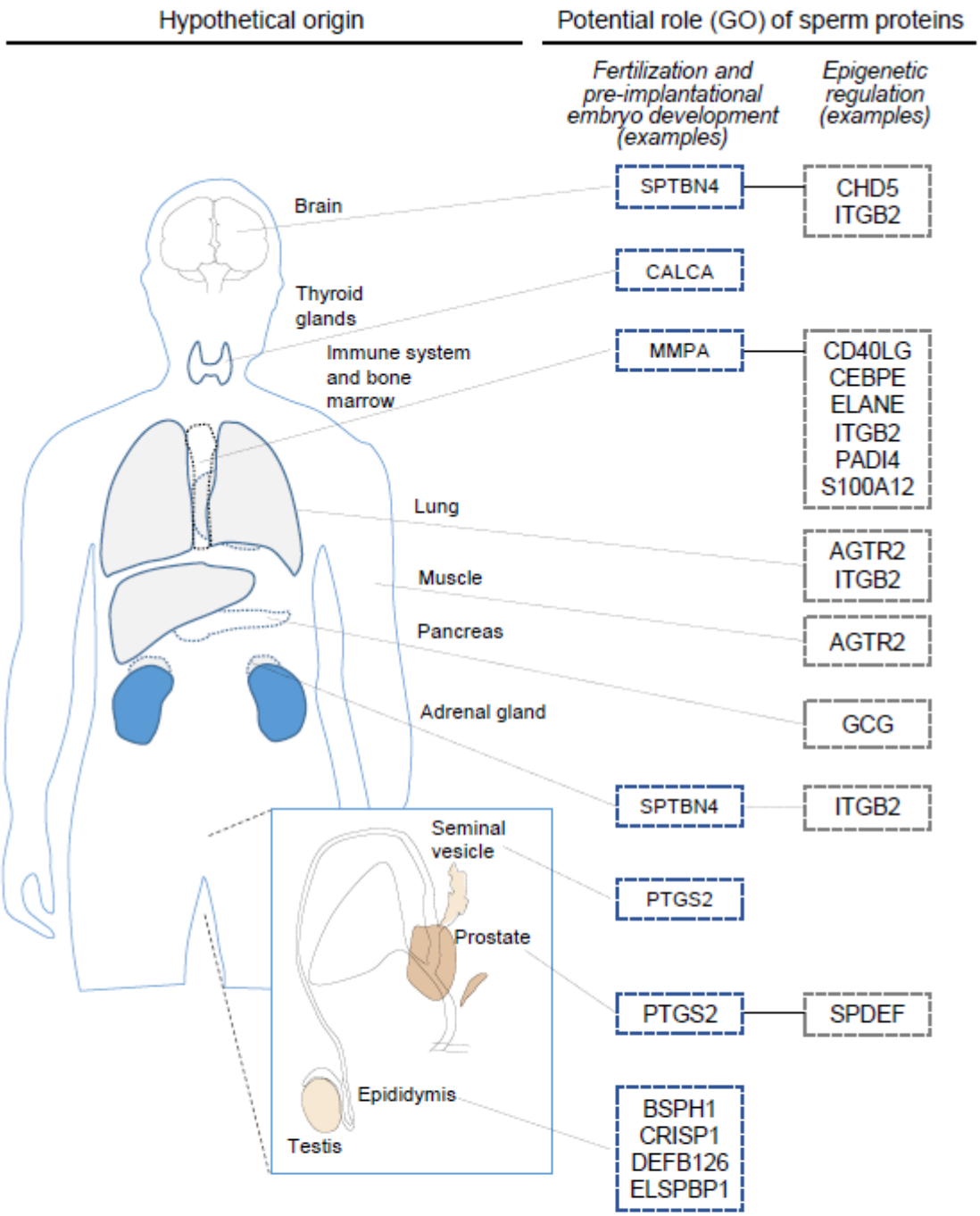

Hypothetical origin
Potential role (GO) of sperm proteins
Fertilization and pre-implantational embryo development (examples)
Epigenetic regulation (examples)
Brain
Thyroid glands
Immune system and bone marrow
Lung
Muscle
Pancreas
Adrenal gland
Seminal vesicle
Prostate
Epididymis
Testis
SPTBN4
CHD5
ITGB2
CALCA
MMPA
CD40LG
CEBPE
ELANE
ITGB2
PADI4
S100A12
AGTR2
ITGB2
AGTR2
GCG
SPTBN4
ITGB2
PTGS2
PTGS2
SPDEF
BSPH1
CRISP1
DEFB126
ELSPBP1

**Figure 4**