High-throughput sperm differential proteomics suggests that epigenetic alterations contribute to failed assisted reproduction

Rubén Azpiazu1,2,†, Alexandra Amaral1,2,3,†, Judit Castillo1,2, Josep Maria Estanyol4, Marta Guimerà5, Josep Lluís Ballesca5, Juan Balasch5, and Rafael Oliva1,2,*

1Human Genetics Research Group, IDIBAPS, Faculty of Medicine, University of Barcelona, Casanova 143, Barcelona 08036, Spain 2Biochemistry and Molecular Genetics Service, Hospital Clinic, Villarroel 170, Barcelona 08036, Spain 3Biology of Reproduction and Stem Cell Group, Centre for Neuroscience and Cell Biology, University of Coimbra, Largo Marqués de Pombal, Coimbra 3004-517, Portugal 4Proteomics Unit, Scientific Technical Services, University of Barcelona, Casanova 143, Barcelona 08036, Spain 5Clinic Institute of Gynaecology, Obstetrics and Neonatology, Hospital Clinic, Villarroel 170, Barcelona 08036, Spain

*Correspondence address. E-mail: roliva@ub.edu

Submitted on November 7, 2013; resubmitted on February 6, 2014; accepted on March 14, 2014

STUDY QUESTION: Are there quantitative alterations in the proteome of normozoospermic sperm samples that are able to complete IVF but whose female partner does not achieve pregnancy?

SUMMARY ANSWER: Normozoospermic sperm samples with different IVF outcomes (pregnancy versus no pregnancy) differed in the levels of at least 66 proteins.

WHAT IS KNOWN ALREADY: The analysis of the proteome of sperm samples with distinct fertilization capacity using low-throughput proteomic techniques resulted in the detection of a few differential proteins. Current high-throughput mass spectrometry approaches allow the identification and quantification of a substantially higher number of proteins.

STUDY DESIGN, SIZE, DURATION: This was a case–control study including 31 men with normozoospermic sperm and their partners who underwent IVF with successful fertilization recruited between 2007 and 2008.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Normozoospermic sperm samples from 15 men whose female partners did not achieve pregnancy after IVF (no pregnancy) and 16 men from couples that did achieve pregnancy after IVF (pregnancy) were included in this study. To perform the differential proteomic experiments, 10 no pregnancy samples and 10 pregnancy samples were separately pooled and subsequently used for tandem mass tags (TMT) protein labelling, sodium dodecyl sulphate–polyacrylamide gel electrophoresis, liquid chromatography tandem mass spectrometry (LC-MS/MS) identification and peak intensity relative protein quantification. Bioinformatic analyses were performed using UniProt Knowledgebase, DAVID and Reactome. Individual samples (n = 5 no pregnancy samples; n = 6 pregnancy samples) and aliquots from the above TMT pools were used for western blotting.

MAIN RESULTS AND THE ROLE OF CHANCE: By using TMT labelling and LC-MS/MS, we have detected 31 proteins present at lower abundance (ratio no pregnancy/pregnancy < 0.67) and 35 at higher abundance (ratio no pregnancy/pregnancy > 1.5) in the no pregnancy group. Bioinformatic analyses showed that the proteins with differing abundance are involved in chromatin assembly and lipoprotein metabolism (P values < 0.05). In addition, the differential abundance of one of the proteins (SRSF protein kinase 1) was further validated by western blotting using independent samples (P value < 0.01).

LIMITATIONS, REASONS FOR CAUTION: For individual samples the amount of recovered sperm not used for IVF was low and in most of the cases insufficient for MS analysis, therefore pools of samples had to be used to this end.

WIDER IMPLICATIONS OF THE FINDINGS: Alterations in the proteins involved in chromatin assembly and metabolism may result in
epigenetic errors during spermatogenesis, leading to inaccurate sperm epigenetic signatures, which could ultimately prevent embryonic development. These sperm proteins may thus possibly have clinical relevance.

**STUDY FUNDING/COMPETING INTEREST(S):** This work was supported by the Spanish Ministry of Economy and Competitiveness (Ministerio de Economía y Competitividad; FEDER BFU 2009–07118 and PI13/00699) and Fundación Salud 2000 SERONO 13-015. There are no competing interests to declare.

**Key words:** human sperm / differential proteomics / isotopic tandem mass tags / epigenetics / assisted reproduction

**Introduction**

The routine assessment of semen quality is currently based on sperm count, motility and morphology or, occasionally, on the study of sperm DNA damage (Domínguez-Fandos et al., 2007; Castillo et al., 2011; Simon et al., 2011). However, in most cases these phenotypic assessments do not provide a clue for the aetiology of male infertility. In addition, the current treatments using assisted reproduction techniques (ART), such as IVF or ICSI, are successful in only ~32% of the cases (de Mateo et al., 2009; Ferraretti et al., 2013). This situation also occurs in couples without a perceptible female factor and whose male partner is apparently normal (normozoospermic individuals). In fact, a currently unsolved issue in andrology is why some normozoospermic samples result in pregnancy after ART, while others do not. One of the reasons for the poor ART outcomes may be the existence of sperm alterations at the molecular level, which are not detected using the current routine clinic laboratory procedures (Oliva et al., 2009; Oliva and de Mateo, 2011; Oliva and Ballesca, 2012).

Changes in specific sperm protein levels are possibly involved in these potential molecular alterations and indeed several studies, using a variety of protein identification strategies, have tried to clarify this. Two kinds of studies are found in the literature: those targeted to the quantification of one or a few proteins using specific antibodies, enzymatic activity assays or direct detection (de Yebra et al., 1993; de Yebra and Oliva, 1993; Bench et al., 1998; de Yebra et al., 1998; Mengual et al., 2003; de Mateo et al., 2009; 2011b,c) and those using wider proteomic approaches relying on mass spectrometry (MS) to concurrently analyse as many sperm proteins as possible (de Mateo et al., 2007; Oliva et al., 2007, 2008; Botta et al., 2009; Liao et al., 2009; Secciani et al., 2009; Oliva and Castillo, 2011; Paasch et al., 2011; Thacker et al., 2011; Oliva, 2012). Differential proteomic studies are often carried out to compare sperm samples with defined aetiologies, such as asthenozoospermic versus normozoospermic (Zhao et al., 2007; Martínez-Heredia et al., 2008; Chan et al., 2009; Siva et al., 2010; Parte et al., 2012). On the other hand, proteomic studies aimed to investigate the correlation between sperm protein levels and ART results are scarce. The pioneer study of this kind used two-dimensional gel electrophoresis (2DE) to compare the proteome of sperm from one normozoospermic patient involved in failed IVF to that of three fertile donors, and resulted in the detection of 4 differential proteins (Pixton et al., 2004). The use of a more sensitive proteomic approach (2D fluorescence difference gel electrophoresis) to compare the proteomic mapping of sperm from 3 infertile patients (with no fertilization after IVF) to that of 3 fertile donors resulted in the identification of 12 differentially expressed proteins (Frapsauce et al., 2009). More recently, a 2DE comparison between 10 normozoospermic sperm samples associated with failed fertilization after artificial insemination and 10 sperm samples from healthy semen donors detected 24 differentially abundant proteins (Xu et al., 2012).

Although the outcomes obtained in the aforementioned studies are relevant, it is well established that proteomic approaches based on 2DE allow the identification of a minority of the sperm proteome when compared with the detection power of current high-throughput approaches based on liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS; Baker et al., 2007, 2013; de Mateo et al., 2011a, 2013; Amaral et al., 2013, 2014; Wang et al., 2013). An elegant strategy to perform relative protein quantification by LC-MS/MS consists of the use of isobaric tandem mass tags (TMTs), which are reactive towards lysine residues and free amino termini of peptides (Thompson et al., 2003; Dayon and Sanchez, 2012). Despite the efficiency of this approach, that has been applied to study proteome alterations in distinct somatic cells (Dayon et al., 2010; Sinclair and Timms, 2011; Raso et al., 2012; Paulo et al., 2013; just to give a few examples), the use of the TMT technology in reproductive biology context is scarce (Zhu et al., 2013).

The aim of this study was to conduct a quantitative comparison of the proteome of sperm samples from normozoospermic individuals with different ART outcomes (i.e. pregnancy group versus no pregnancy group; but for which a similar number of morphologically normal embryos were obtained) using TMT and LC-MS/MS. We reasoned that this approach would result in the identification of sperm proteomic differences that may affect embryonic development and thus have clinical significance.

**Materials and Methods**

**Chemicals**

All reagents were supplied from Sigma-Aldrich (St. Louis, MO, USA) unless stated otherwise.

**Ethical approval**

Patients undergoing IVF treatment were recruited from the Assisted Reproduction Unit of the Clinic Hospital of Barcelona, Spain. All patients signed informed consent forms and all human material was used in accordance with the appropriate ethical and Internal Review Board guidelines.

**Biological material selection**

A total of 31 couples with similar clinical male and female characteristics but different IVF outcomes were included in this study. The pregnancy group (n = 16) corresponds to couples who achieved pregnancy after IVF and the no pregnancy group (n = 15) to those who did not achieve pregnancy after IVF. All female partners were affected by tubal pathology, while having normal oocyte production and gestational capacity (FSH, LH and estradiol levels were normal). Routine seminal analysis was performed according to the
World Health Organization Guidelines (WHO, 2010) and all sperm samples used were normozoospermic. ICSI was not attempted in any of the couples because a substantial number of embryos were obtained using IVF and good-quality embryos were transferred in all the cases. Noteworthy, apart from the contrasting ART outcomes, no statistically significant differences were detected in any of the clinical variables in the two groups of samples (Table I, Supplementary data, Table S1).

Sperm preparation

Fresh semen samples were obtained by masturbation after 3–5 days of sexual abstinence. Samples were prepared on a density gradient using Pure-sperm™ (Nidacon, Mölndal, Sweden), according to the manufacturer’s protocol. The recovered Pure-sperm fraction not used for IVF was washed in Ham’s F-10 nutrient mixture (Life Technologies™, Paisley, UK) and stored at −20 °C in Ham’s F-10 supplemented with 7% (v/v) glycerol. Only samples containing no contaminating round cells were used.

Quantitative proteomics

The general strategy used is depicted in Figure 1. Two pools of sperm samples (pregnancy group and no pregnancy group) containing 150 million sperm each (15 million from 10 samples for each group) were used. The TMT duplex™ Isotopic Label Reagent Set (Thermo Scientific, Rockford, IL, USA) was used for labelling. Essentially, the gel slices were digested with 150 ng trypsin (Promega, Madison, WI, USA) at 37 °C overnight, according to the manufacturer’s recommendations. Tryptic peptides were separated by means of in-gel digestion, peptides were separated by one-dimensional liquid chromatography and analysed by MS/MS.

Protein solubilization

Proteins were solubilized in Lyss Buffer (7 M urea, 2 M thiourea, 1% (w/v) CHAPS, 1% (w/v) N-octyl-D-glucopyranoside, 18 mM dithiothreitol (DTT) and 2.4 mM phenylmethylsulphonyl fluoride (PMSF) for 1 h at room temperature, with constant shaking, as described previously (Martinez-Heredia et al., 2006, de Mateo et al., 2011a; Amaral et al., 2013).

TMT labelling

Proteins (100 μg) from the two pooled samples (pregnancy group and no pregnancy group) were precipitated with 80% (v/v) cold acetone and separately labelled according to the manufacturer’s recommendations. Briefly, proteins were incubated in 45 mM triethyl ammonium bicarbonate with 0.1% SDS, reduced in 9.5 mM tris(2-carboxyethyl)phosphine (TCEP) for 1 h at 55 °C and alkylated with 17 mM iodoacetamine for 30 min, at room temperature, in the dark. TMT label reagents (0.8 mg each) were equilibrated to room temperature, dissolved in 24 μl anhydrous dimethyl sulphoxide (DMSO) and added to the reduced and alkylated proteins (the pregnancy group was labelled with TMT-127 and the no pregnancy group with TMT-126). After incubation for 1 h at room temperature, reactions were quenched by incubating for 15 min in 0.28% hydroxylamine. Labelled samples were then combined in equal amounts.

SDS-PAGE

Acetone precipitated proteins were suspended in Laemmi buffer (60 mM Tris HCl, pH 6.8, 2.2% (w/v) SDS, 5% (v/v) glycerol, 0.1 M DTT), and incubated for 10 min at 90 °C. After cooling, a total of 150 μg of labelled proteins were separated by SDS-PAGE (12% acrylamide gel with a 3.9% acrylamide stacking gel). The gel was fixed with 50% methanol 10% acetic acid and stained with EZBlue™ Gel Staining Reagent for 5 min (until a protein smear was just visible).

LC-MS/MS

The entire gel lane was carefully cut into small pieces (~1 mm intervals) that were then processed for MS analysis as described before (Amaral et al., 2013). Essentially, the gel slices were digested with 150 ng trypsin (Promega, Madison, WI, USA) at 37 °C overnight, according to the manufacturer’s recommendations. Tryptic peptides were separated by means of nano liquid chromatography using a Proxeon EASY-nLC (Thermo Fisher Scientific, Waltham, MA, USA) with a flow rate of 300 nl/min, an EASY C18 trap column (5 μm, 120 A˚, 75 mm inner diameter × 2 cm length), and an EASY C18 analytical column (3 μm, 120 A˚, 75 mm inner diameter × 10 cm length). The following linear gradient, using Solvent B (97% acetonitrile, 0.1% formic acid) and Solvent A (3% acetonitrile, 0.1% formic acid), was employed: 5–35% buffer B (150 min); 35–100% buffer B (5 min). MS/MS analysis was performed using an LTQ Orbitrap Velos (Thermo Fisher Scientific) with a nanoelectrospray ion source with precursor ion selection in the

<table>
<thead>
<tr>
<th>Table I</th>
<th>Groups of samples used in the differential proteomic comparison (n = 10 ± 10).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pregnancy group</td>
</tr>
<tr>
<td>Male partner</td>
<td>Age</td>
</tr>
<tr>
<td></td>
<td>Semen volume (ml)</td>
</tr>
<tr>
<td></td>
<td>Sperm concentration (million sperm/ml semen)</td>
</tr>
<tr>
<td></td>
<td>Sperm motility (% progressively motile sperm)</td>
</tr>
<tr>
<td>Female partner</td>
<td>Age</td>
</tr>
<tr>
<td></td>
<td>Body mass index (kg/m²)</td>
</tr>
<tr>
<td></td>
<td>Basal FSH levels (U/l)</td>
</tr>
<tr>
<td></td>
<td>Basal LH levels (U/l)</td>
</tr>
<tr>
<td></td>
<td>Basal E2 levels (U/l)</td>
</tr>
<tr>
<td></td>
<td>Number of oocytes collected</td>
</tr>
<tr>
<td>Embryos</td>
<td>Number of embryos obtained</td>
</tr>
</tbody>
</table>
Protein identification and quantification

Data were processed using Proteome Discoverer 1.4 (Thermo Fisher Scientific). For database searching, processed data were submitted to the in-house Homo sapiens UniProtKB/Swiss-Prot database (release 2012-04; 88 323 protein entries) using SEQUEST version 28.0 (Thermo Fisher Scientific). The following search parameters were used: two maximum miss cleavage for trypsin; TMT-labelled lysine and methionine oxidation as dynamic modifications; cysteine carbamidomethylation as fixed modification; 25 ppm precursor mass tolerance; 20 mmu fragment mass tolerance and 20 ppm TMR integration tolerance. Criteria used to accept identification included a minimum of two peptides (and at least one unique peptide) matched per protein, with a false discovery rate of 0.01.

Quantitative analysis was performed using Proteome Discoverer software. The reporter ion intensities were corrected according to the isotopic purities provided by the manufacturer. The relative protein level ratios between the two groups were determined from the TMT fragment peak intensities and were calculated using the median intensities of reporter ions for each protein. The ratios were log2 transformed to reduce variance heterogeneity (Boehm et al., 2007; Gan et al., 2007) and further normalized using the value of the median ratio of all the proteins. A 95% confidence interval was calculated (median ± 1.96 SD) and all median ratios not contained in this interval were considered up- or down-regulated (Jung et al., 2009; Karp et al., 2010; Podwojski et al., 2012). Importantly, all the proteins detected as differential using this method had a ratio > 1.5 (more abundant proteins) or < below 0.67 (less abundant proteins). Additionally, in order to validate the results obtained, the MS spectra of all the differential proteins detected were manually checked.

Differential proteome analysis

Proteins identified in different amounts in the two groups were classified according to subcellular localization and biological function(s) using the information available at the UniProt Knowledgebase (UniProtKB/Swiss-Prot) website (http://www.uniprot.org). The list of differential proteins was also analysed using the bioinformatics tool DAVID v6.7 (Database for Annotation, Visualisation and Integrated Discovery; http://david.abcc.ncifcrf.gov/) in order to identify overrepresented Gene Ontology (GO) Biological Process terms (Huang et al., 2009a,b). The significance of gene-enrichment analysis was expressed by an EASE Score, corresponding to a modified Fisher exact P value (P values < 0.05 were considered significant). Furthermore, the Reactome database (http://www.reactome.org/) was used to perform overrepresentation analysis, with the aim of recognizing those biological pathways likely to be affected in sperm samples from couples that did not achieve pregnancy. The significance of the association between the protein list and a certain pathway was expressed by an EASE Score, corresponding to a modified Fisher exact test that the association between the proteins in the dataset and the pathway is explained by chance alone (P values < 0.001 were considered significant).

Western blotting

To further confirm the proposal that there may be altered levels of proteins involved in sperm chromatin assembly with a potential epigenetic role in the male gamete and validate the MS results, we have used western blotting to compare the expression of SRSF protein kinase 1 (SRPK1; detected at higher levels in the no pregnancy group), not only in the pools of samples used for MS, but also in additional individual samples. SRPK1 expression was also analysed using 11 samples (5 and 6 from no pregnancy and pregnancy groups, respectively), using rabbit anti-human SRPK1 polyclonal antibody (0.5 µg/ml) and 35 µg total protein/sample, essentially as described before (Amaral et al., 2013). For loading controls, a mouse anti-human alphatubulin monoclonal antibody (clone DM1A purified mouse immunoglobulin;
The comparison of the proteomes of the no pregnancy (case) and pregnancy (control) groups of sperm samples using TMT-126 and TMT-127 peak intensities (Fig. 2) resulted in the detection of 66 differentially expressed proteins (ratio no pregnancy group/pregnancy group < 0.67 or > 1.5; Tables II and III). Of the 66 differential proteins detected, 31 were found at lower levels in the no pregnancy group (ratio < 0.67; Table II), whereas 35 were found at higher levels (ratio > 1.5; Table III). Noteworthy, the groups of higher or lower abundant proteins seem to correspond to different categories. For example, a first analysis highlights that the sub-list of higher abundant proteins includes six core histones (H2A type I-A, H2A type I-C, H2A type 2-C, H2B type 1A, H3 and H4), while none was found in the sub-list of lower abundant proteins. On the other hand, the lower abundant proteins includes two with possible roles in early development (left-right determination factor 1 and growth/differentiation factor 15) and three proteins related to lipid metabolism (lipase member I, lipoprotein lipase and apolipoprotein).

**Determination of protamine 1 (P1)/protamine 2 (P2) ratio**

In order to monitor the protamination status on the two groups of sperm samples (and thus to assess overall sperm chromatin packaging and protection), the P1/P2 ratio was also determined. Towards this goal, the insoluble fractions of eight samples (four from each group) were processed as described previously (Castillo et al., 2011). Basically, extracted protamines, along with external standards of known protamine concentrations, were run on acid-urea polyacrylamide gels (acid-urea PAGE). Band intensities were quantified using Quantity One software (Bio-Rad Laboratornes, Hercules, CA, USA) and used to determine the amounts of P1 and P2 in each of the samples. Quantification calculations were based on a regression curve constructed with the protamine standards.

**Statistical analyses**

Statistical analyses were performed using IBM SPSS statistics software (version 21.0; SPSS Corp., Chicago, IL, USA). One-sample Kolmogorov–Smirnov test was used to check for normal distribution and the independent sample t-test was used to compare: (i) the male and female parameters between the two groups of couples included in the study; (ii) SRPK1 expression between the two groups of sperm samples; (iii) P1/P2 ratio between the two groups of sperm samples. A P value of < 0.05 was considered significant.

**Results**

**Normozoospermic sperm proteome**

LC-MS/MS of the 2 groups of sperm samples (comprising 10 different samples each, loaded onto a single gel lane) resulted in the identification of 1717 proteins (Supplementary data is available at the Peptide Atlas protein repository; Dataset Identifier: PASS00459; http://www.peptideatlas.org/PASS/PASS00459). A GO Biological Process term analysis of all the proteins identified is shown in the Supplementary data, Table SII. Of all the proteins identified, 137 proteins turn out to be novel proteins (i.e. proteins not reported in any of the human sperm proteomics studies performed so far; Supplementary data, Fig. S1). These include proteins from different sperm subcellular parts, such as the nucleus (e.g. histone H2B type F-M; transcription factor E3; small ubiquitin-related modifier 2) or the mitochondrion (e.g. NADH dehydrogenase (ubiquinone) 1 alpha subcomplex subunit 3; mitochondrial import inner membrane translocase subunit Tim 17-B; propionyl-CoA carboxylase alpha chain), as well as secreted proteins, that may participate in sperm–oocyte interactions (e.g. alpha-1B-glycoprotein; alpha-2-HS-glycoprotein; beta-2-glycoprotein 1).

**Sperm samples associated with failed/ successful ART outcomes differ in the relative abundance of several proteins**

The comparison of the proteomes of the no pregnancy (case) and pregnancy (control) groups of sperm samples using TMT-126 and TMT-127 peak intensities (Fig. 2) resulted in the detection of 66 differentially expressed proteins (ratio no pregnancy group/pregnancy group < 0.67 or > 1.5; Tables II and III). Of the 66 differential proteins detected, 31 were found at lower levels in the no pregnancy group (ratio < 0.67; Table II), whereas 35 were found at higher levels (ratio > 1.5; Table III). Noteworthy, the groups of higher or lower abundant proteins seem to correspond to different categories. For example, a first analysis highlights that the sub-list of higher abundant proteins includes six core histones (H2A type I-A, H2A type I-C, H2A type 2-C, H2B type 1A, H3 and H4), while none was found in the sub-list of lower abundant proteins. On the other hand, the lower abundant proteins includes two with possible roles in early development (left-right determination factor 1 and growth/differentiation factor 15) and three proteins related to lipid metabolism (lipase member I, lipoprotein lipase and apolipoprotein).

**The differential proteins belong to particular GO categories and are involved in specific cellular pathways**

The categorization of the differential proteins according to the information available at the Uniprot Knowledgebase resulted in very interesting outcomes (Fig. 3). Concerning cellular localization (Fig. 3A), it is interesting to appreciate that the sub-list of lower abundant proteins comprises various membrane/secreted proteins (12 out of 31 proteins; 39%). In contrast, the majority of proteins in the sub-list of higher abundant proteins are mitochondrial (23%), membranar/secreted (17%) or nuclear (16%). Similarly, the sets of higher abundant and lower abundant proteins are associated with different cellular functions (Fig. 3B). Worth mentioning, all differential proteins involved in nucleosome assembly were detected at higher levels in the no pregnancy group, whereas all cellular trafficking proteins were less abundant in this group (Fig. 3B). However, this first analysis was only indicative of a trend and had no statistical significance associated.

We then questioned whether the differentially detected proteins were involved in explicit biological processes and, in such a case, if the sets of up- and down-regulated proteins were distinct. To answer these questions, we have performed GO terms enrichment analysis using DAVID. The outcomes obtained confirmed that the sets of proteins detected in lower and higher levels in the no pregnancy group are involved in distinctive biological processes (Table IV). To this extent, sperm from the no pregnancy group have higher levels of proteins involved in DNA packaging, nucleosome organization and chromatin/nucleosome assembly and lower levels of proteins with roles in intracellular transport and sexual reproduction (P values < 0.05). Additionally, biological processes related to apoptosis were also detected as enriched in the no pregnancy group.

With the aim of determining if higher or lower abundant proteins could affect specific cellular pathways and to elucidate how would the differential proteins contribute to failed ART outcomes in no pregnancy samples, Reactome pathways analysis was also carried out (Table V). Confirming our initial observations, analysis of the group of proteins present at lower abundance using Reactome resulted in the detection of lipoprotein metabolism (and specifically high-density lipoprotein (HDL)-mediated lipid transport) as a significant pathway (P values < 0.0001). On the other hand, and corroborating the outcomes obtained with DAVID, Reactome recognized meiosis and chromosome maintenance as pathways in which the higher abundant proteins might be involved (P values < 0.001).
The protamine 1 (P1)/protamine 2 (P2) ratio is not significantly different in the two groups of samples. The results indicate that in the group of normozoospermic samples associated with failed ART outcomes the P1/P2 ratio is $1.71 \pm 0.37$ (mean $\pm$ SD) and in the pregnancy group the P1/P2 ratio is $1.57 \pm 0.28$ (mean $\pm$ SD). No statistically significant differences were detected between the two groups ($P = 0.573$).

The levels of proteins with a potential sperm epigenetic function may be altered in normozoospermic samples associated with failed ART outcomes. The outcomes of the three analyses performed (manual protein functional categorization according to the information available at the Uniprot Knowledgebase; DAVID GO Biological Process terms enrichment and Reactome pathways analysis) are all in agreement with the idea that normozoospermic samples associated with failed ART outcomes may have altered levels of proteins involved in sperm chromatin assembly (which may have an epigenetic role in the male gamete).

The results obtained confirmed that the level of SRPK1 was higher in the pooled no pregnancy samples compared with the pooled pregnancy samples (Fig. 4A) and this was further corroborated by the analysis of additional independent sperm samples (Fig. 4B). Indeed, there was a statistically significant difference in the relative amount of SRPK1 in the two sets of samples ($P$ value $< 0.01$).

**Discussion**

Unravelling the putative causes of the increasing number of clinical cases of idiopathic male infertility is certainly one of the biggest challenges in modern andrology. Although reaching such a goal may be very difficult due to the multitude of possible reasons and the fact that infertility is,
by definition, a couple’s (rather than an individual’s) problem, modern omics technologies are expected to contribute to the elucidation of the molecular mechanisms that are at the origin of this disorder. In this study we have used TMT protein labelling and MS to identify potential quantitative proteomic alterations in otherwise normal sperm samples, which may have contributed to embryonic development failure after

Table II  Lower abundant proteins in normozoospermic samples that did not achieve pregnancy (n = 31).

<table>
<thead>
<tr>
<th>UniprotKB/Swiss-Prot accession number</th>
<th>Protein name</th>
<th>Ratio: no pregnancy/pregnancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>B0QY90</td>
<td>Eukaryotic translation initiation factor 3 subunit l</td>
<td>0.25</td>
</tr>
<tr>
<td>P31947</td>
<td>Isoform 2 of I4-3-3 protein sigma</td>
<td>0.38</td>
</tr>
<tr>
<td>Q02383</td>
<td>Semenogelin-2</td>
<td>0.41</td>
</tr>
<tr>
<td>B4DI70</td>
<td>cDNA FLJ35309, highly similar to Galectin-3-binding protein</td>
<td>0.43</td>
</tr>
<tr>
<td>Q6XZB0</td>
<td>Lipase member l</td>
<td>0.43</td>
</tr>
<tr>
<td>P04279-2</td>
<td>Isoform 2 of Semenogelin-1</td>
<td>0.44</td>
</tr>
<tr>
<td>P02768</td>
<td>Serum albumin</td>
<td>0.44</td>
</tr>
<tr>
<td>B3KRV7</td>
<td>Lipoprotein lipase</td>
<td>0.45</td>
</tr>
<tr>
<td>P54866-2</td>
<td>Isoform short of Delta-1-pyrroline-5-carboxylate synthase</td>
<td>0.48</td>
</tr>
<tr>
<td>Q99988</td>
<td>Growth/differentiation factor 15</td>
<td>0.49</td>
</tr>
<tr>
<td>Q6BCY4-2</td>
<td>Isoform 2 of NADH-cytochrome b5 reductase 2</td>
<td>0.49</td>
</tr>
<tr>
<td>O43592</td>
<td>Exportin-T</td>
<td>0.49</td>
</tr>
<tr>
<td>O75610</td>
<td>Left-right determination factor l</td>
<td>0.50</td>
</tr>
<tr>
<td>P0C7M6</td>
<td>IQ domain-containing protein F3</td>
<td>0.50</td>
</tr>
<tr>
<td>B7Z1J9</td>
<td>Parkinson disease 7 domain-containing protein 1</td>
<td>0.51</td>
</tr>
<tr>
<td>P05109</td>
<td>Protein S100-A8</td>
<td>0.51</td>
</tr>
<tr>
<td>Q8N427</td>
<td>Thioredoxin domain-containing protein 3</td>
<td>0.52</td>
</tr>
<tr>
<td>Q96RL7-4</td>
<td>Isoform 4 of vacuolar protein sorting-associated protein 13A</td>
<td>0.52</td>
</tr>
<tr>
<td>Q5H943</td>
<td>Kita-kyushu lung cancer antigen l</td>
<td>0.53</td>
</tr>
<tr>
<td>P02649</td>
<td>Apolipoprotein E</td>
<td>0.54</td>
</tr>
<tr>
<td>F3J237</td>
<td>Serpin B6</td>
<td>0.54</td>
</tr>
<tr>
<td>F5GX33</td>
<td>Calcium-binding mitochondrial carrier protein Aralar2</td>
<td>0.55</td>
</tr>
<tr>
<td>Q92599-3</td>
<td>Isoform 3 of Septin-8</td>
<td>0.56</td>
</tr>
<tr>
<td>P0CG04</td>
<td>Ig lambda-1 chain C regions</td>
<td>0.56</td>
</tr>
<tr>
<td>Q95336</td>
<td>6-phosphogluconolactonase</td>
<td>0.56</td>
</tr>
<tr>
<td>O14980</td>
<td>Exportin-1</td>
<td>0.56</td>
</tr>
<tr>
<td>B4DY72</td>
<td>Heat shock protein 105 kDa</td>
<td>0.57</td>
</tr>
<tr>
<td>P51148</td>
<td>Ras-related protein Rab-5C</td>
<td>0.57</td>
</tr>
<tr>
<td>Q15084</td>
<td>Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit A</td>
<td>0.57</td>
</tr>
<tr>
<td>Q96KX2</td>
<td>F-actin-capping protein subunit alpha-3</td>
<td>0.59</td>
</tr>
<tr>
<td>B4DPM9</td>
<td>Serine hydroxymethyltransferase</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Table III  Higher abundant proteins in normozoospermic samples that did not achieve pregnancy (n = 35).

<table>
<thead>
<tr>
<th>UniprotKB/Swiss-Prot accession number</th>
<th>Protein name</th>
<th>Ratio: no pregnancy/pregnancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q53HC9</td>
<td>Protein TSSC1</td>
<td>2.44</td>
</tr>
<tr>
<td>F8WV32</td>
<td>Lysozyme C</td>
<td>2.37</td>
</tr>
<tr>
<td>Q9H095-2</td>
<td>Isoform 2 of IQ domain-containing protein G</td>
<td>2.37</td>
</tr>
<tr>
<td>Q0P670</td>
<td>Uncharacterized protein C17orf74</td>
<td>2.29</td>
</tr>
<tr>
<td>Q8N5K1</td>
<td>CDGSH iron-sulphur domain-containing protein 2</td>
<td>2.28</td>
</tr>
<tr>
<td>P0S164-2</td>
<td>Isoform H14 of myeloperoxidase</td>
<td>2.04</td>
</tr>
<tr>
<td>P04083</td>
<td>Annexin A1</td>
<td>1.98</td>
</tr>
<tr>
<td>E9PIC0</td>
<td>Succinate dehydrogenase [ubiquione] cytochrome b small subunit, mitochondrial</td>
<td>1.94</td>
</tr>
<tr>
<td>P24752</td>
<td>Acetyl-CoA acetyltransferase, mitochondrial</td>
<td>1.94</td>
</tr>
<tr>
<td>Q8IZ16</td>
<td>Uncharacterized protein C7orf6l</td>
<td>1.93</td>
</tr>
<tr>
<td>Q5SRN2</td>
<td>Uncharacterized protein C6orf10</td>
<td>1.92</td>
</tr>
<tr>
<td>Q96A08</td>
<td>Histone H2B type 1-A</td>
<td>1.91</td>
</tr>
<tr>
<td>P62805</td>
<td>Histone H4</td>
<td>1.87</td>
</tr>
<tr>
<td>O43837-2</td>
<td>Isoform A of isocitrate dehydrogenase [NAD] subunit beta, mitochondrial</td>
<td>1.86</td>
</tr>
<tr>
<td>Q5TEC6</td>
<td>Histone cluster 2, H3, pseudogene 2</td>
<td>1.86</td>
</tr>
<tr>
<td>B7Z1I2</td>
<td>Aspartate aminotransferase</td>
<td>1.86</td>
</tr>
<tr>
<td>P10644</td>
<td>cAMP-dependent protein kinase type I-alpha regulatory subunit</td>
<td>1.80</td>
</tr>
<tr>
<td>P16152</td>
<td>Carbonyl reductase [NADPH] I</td>
<td>1.80</td>
</tr>
<tr>
<td>Q96SB4</td>
<td>SRSF protein kinase 1</td>
<td>1.79</td>
</tr>
<tr>
<td>E7EPB3</td>
<td>60S ribosomal protein L14</td>
<td>1.79</td>
</tr>
<tr>
<td>Q5TGZ0</td>
<td>Mitochondrial inner membrane organizing system protein 1</td>
<td>1.77</td>
</tr>
<tr>
<td>P13646-3</td>
<td>Isoform 3 of Keratin, type I cytoskeletal 13</td>
<td>1.76</td>
</tr>
<tr>
<td>E5RJX2</td>
<td>40S ribosomal protein S20</td>
<td>1.74</td>
</tr>
<tr>
<td>Q96IX5</td>
<td>Up-regulated during skeletal muscle growth protein 5</td>
<td>1.73</td>
</tr>
<tr>
<td>Q16777</td>
<td>Histone H2A type 2-C</td>
<td>1.70</td>
</tr>
</tbody>
</table>

Continued
successful IVF. While other sperm (and oocyte) molecular alterations cannot be ruled out, our data add further support to the increasing body of evidence suggesting that a tight regulation of sperm protein levels might be needed for normal embryonic development. Very interestingly, several of the differential proteins detected using our approach point to a range of cellular alterations that would result in inaccurate sperm epigenetic signatures.

In fact, a number of the differential proteins detected are localized to the chromatin. Of note, 6 out of the 14 histones (43%) identified in this study (Table III) were detected at significantly higher levels in the chromatin. Of note, 6 out of the 14 histones (43%) identified in this study (Table III) were detected at significantly higher levels in the chromatin. Additionally, analysis of the sub-set of more abundant proteins indicates enrichment in nucleosome assembly process (DAVID and Reactome). The increased levels of histones in samples that did not result in pregnancy may indicate a deficient nucleohistone to nucleoprotamine transition during spermiogenesis, which could result in abnormal sperm imprinting and/or in reduced paternal chromatin protection (Govin et al., 2004; Torregrosa et al., 2006; de Mateo et al., 2011c; Oliva and Castillo, 2011; Oliva and Ballesca, 2012). However, the relative amounts of proteasomes (P1/P2 ratio) in the no pregnancy group seemed to be normal and comparable to the pregnancy group. Thus, and because it is well-established sperm DNA damage is associated with alterations in protamination (Aoki et al., 2005; de Mateo et al., 2007; Castillo et al., 2011), these outcomes support the idea that no substantial nuclear DNA damage is present in these samples. On the other hand, SRPK1 levels were higher in the no pregnancy group.

SRPK1 specifically phosphorylates serine residues within arginine-serine-rich substrate regions and shuttles between the nucleus and the cytoplasm, is highly expressed in the testis and, among other functions, may contribute to sperm chromatin condensation (for review see Giannakouros et al., 2011). Indeed, one of the testis-specific substrates of SRPK1 is P1 (Papoutsopoulou et al., 1999). Importantly, protamine phosphorylation might be critical for the correct binding of protamine to the sperm DNA (Oliva and Mezquita, 1986; Oliva and Dixon, 1991; Oliva, 2006; Jodar and Oliva, 2014). Therefore, a deregulation of protamine phosphorylation could result in an abnormal distribution of the genes in the histone and protamine sperm chromatin regions (believed to have an epigenetic role; Arpanahi et al., 2009; Hammoud et al., 2009, 2011; Jodar et al., 2011; Oliva and Ballesca, 2012; Paradowska et al., 2012; Castillo et al., 2013) and potentially interfere with embryo development, without affecting sperm fertilization ability.

Furthermore, any putative epigenetic modification occurring during spermatogenesis could result in metabolic adjustments and conversely, metabolic shifts may also induce chromatin alterations. Indeed, a growing body of data is showing that there is a crosstalk between a cell’s metabolome and its epigenome (for reviews see Gut and Verdin, 2013; Hino et al., 2013; Martinez-Pastor et al., 2013). For instance, the levels of HDL cholesterol in patients presenting with hypercholesterolemia were correlated with epigenome-wide changes (Guay et al., 2012a,b; Kim et al., 2012). To this extent, our data suggest that sperm samples that did not result in pregnancy after IVF have reduced levels of proteins with functions in lipid/lipoprotein metabolism: lipase member I (LIP), apolipoprotein E (APOE) and lipoprotein lipase (LPL). LPL is a member of the triglyceride lipase family known to be expressed in human testis (Nielsen et al., 2010) and LPI is a testis-specific protein expressed in the cytosol of testis cells that is involved in sperm chromatin condensation and spermatid nuclear migration (Oliva et al., 2005). Thus, these alterations may be indicative of a metabolic shift in the sperm.
phospholipase (Hiramatsu et al., 2003; Wen et al., 2003). APOE has a key role in the transport and catabolism of lipoprotein particles and interestingly, male heterozygous knockout mice for a related member of the apo-lipoprotein family (APOB) were infertile, although their sperm were apparently normal (a phenotype resembling our no pregnancy individuals; Huang et al., 1995). Moreover, a recent study suggests an association between APOA1 up-regulation and inhibition of a family of leucine-rich nuclear export signal, that is involved in nucleocytoplasmic transport during spermatogenesis (Whiley et al., 2012). In accordance, in which epimutations in the corresponding gene have been reported in seminiferous epithelium (Sun et al., 2009) and (iii) exportin 1, known to mediate the nuclear export of both RNAs and proteins bearing a leucine-rich nuclear export signal, that is involved in nucleocytoplasmic transport during spermatogenesis (Whiley et al., 2012). In accordance, the samples (the same used for LC-MS/MS). (B) Box plot corresponding to SRPK1 expression (normalized to alpha-tubulin) in independent samples (pregnancy group: n = 6; no pregnancy group: n = 5). SRPK1/alpha-tubulin was expressed in arbitrary units. Asterisk denotes a statistically significant difference between the two groups (P value < 0.01).

### Table IV GO Biological Process terms enrichment of proteins differentially expressed in the no pregnancy group.

<table>
<thead>
<tr>
<th>GO biological process term</th>
<th>GO identification</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower abundant proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulation of cellular component organization</td>
<td>GO: 0051128</td>
<td>1.9 × 10⁻²</td>
</tr>
<tr>
<td>Sexual reproduction</td>
<td>GO: 0019953</td>
<td>1.9 × 10⁻²</td>
</tr>
<tr>
<td>Cellular localization</td>
<td>GO: 0051641</td>
<td>2.7 × 10⁻²</td>
</tr>
<tr>
<td>Regulation of biological quality</td>
<td>GO: 0065008</td>
<td>3.3 × 10⁻²</td>
</tr>
<tr>
<td>Macromolecule localization</td>
<td>GO: 0033036</td>
<td>4.3 × 10⁻²</td>
</tr>
<tr>
<td>Regulation of cellular component size</td>
<td>GO: 0032535</td>
<td>4.6 × 10⁻²</td>
</tr>
<tr>
<td>Intracellular transport</td>
<td>GO: 0046907</td>
<td>4.9 × 10⁻²</td>
</tr>
<tr>
<td>Higher abundant proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nucleosome assembly</td>
<td>GO: 0006334</td>
<td>3.4 × 10⁻⁶</td>
</tr>
<tr>
<td>Chromatin assembly</td>
<td>GO: 0031497</td>
<td>3.9 × 10⁻⁶</td>
</tr>
<tr>
<td>Protein-DNA complex assembly</td>
<td>GO: 0065004</td>
<td>4.6 × 10⁻⁶</td>
</tr>
<tr>
<td>Nucleosome organization</td>
<td>GO: 0034728</td>
<td>5.0 × 10⁻⁶</td>
</tr>
<tr>
<td>DNA packaging</td>
<td>GO: 0006323</td>
<td>1.3 × 10⁻⁵</td>
</tr>
<tr>
<td>Macromolecular complex subunit</td>
<td>GO: 0043933</td>
<td>1.7 × 10⁻⁴</td>
</tr>
<tr>
<td>Macromolecular complex assembly</td>
<td>GO: 0065003</td>
<td>1.2 × 10⁻³</td>
</tr>
<tr>
<td>Chromosome organization</td>
<td>GO: 0051276</td>
<td>2.9 × 10⁻³</td>
</tr>
<tr>
<td>Cellular component assembly</td>
<td>GO: 0022607</td>
<td>4.2 × 10⁻³</td>
</tr>
<tr>
<td>Cellular component biogenesis</td>
<td>GO: 0044085</td>
<td>7.0 × 10⁻³</td>
</tr>
<tr>
<td>Negative regulation of apoptosis</td>
<td>GO: 0043066</td>
<td>9.6 × 10⁻³</td>
</tr>
<tr>
<td>Regulation of apoptosis</td>
<td>GO: 0042981</td>
<td>1.7 × 10⁻²</td>
</tr>
<tr>
<td>Organelle organization</td>
<td>GO: 0006996</td>
<td>2.2 × 10⁻¹</td>
</tr>
<tr>
<td>Cellular component organization</td>
<td>GO: 0016043</td>
<td>2.8 × 10⁻²</td>
</tr>
<tr>
<td>Oxidation reduction</td>
<td>GO: 0055114</td>
<td>4.5 × 10⁻²</td>
</tr>
</tbody>
</table>

### Table V Reactome pathway analysis results of the lower abundant and higher abundant proteins in the no pregnancy group as compared with the pregnancy group.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Pathway identifier</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower abundant proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipoprotein metabolism</td>
<td>REACT_6823</td>
<td>5.6 × 10⁻⁴</td>
</tr>
<tr>
<td>HDL-mediated lipid transport</td>
<td>REACT_13621</td>
<td>1.7 × 10⁻⁴</td>
</tr>
<tr>
<td>Higher abundant proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meiosis</td>
<td>REACT_111183</td>
<td>1.1 × 10⁻²</td>
</tr>
<tr>
<td>Meiotic recombination</td>
<td>REACT_27271</td>
<td>1.6 × 10⁻⁶</td>
</tr>
<tr>
<td>Meiotic synapsis</td>
<td>REACT_75792</td>
<td>2.7 × 10⁻⁶</td>
</tr>
<tr>
<td>Chromosome Maintenance</td>
<td>REACT_22172</td>
<td>2.9 × 10⁻⁵</td>
</tr>
<tr>
<td>Nucleosome assembly</td>
<td>REACT_22344</td>
<td>1.0 × 10⁻⁵</td>
</tr>
<tr>
<td>Telomere maintenance</td>
<td>REACT_7970</td>
<td>2.5 × 10⁻⁶</td>
</tr>
</tbody>
</table>

Figure 4 Western blot quantification of SRPK1 in sperm samples of infertile patients differing in assisted reproduction outcome. (A) Expression of SRPK1 in pregnancy (n = 10) and no pregnancy (n = 10) pooled samples (the same used for LC-MS/MS). (B) Box plot corresponding to SRPK1 expression (normalized to alpha-tubulin) in independent samples (pregnancy group: n = 6; no pregnancy group: n = 5). SRPK1/alpha-tubulin was expressed in arbitrary units. Asterisk denotes a statistically significant difference between the two groups (P value < 0.01).
both intracellular transport and sexual reproduction were GO biological process terms found to be enriched in the sub-set of less abundant proteins.

Ultimately, and although the role of paternal proteins in early embryonic development (with a few exceptions) is uncertain, it is tempting to speculate that the detected lower levels of certain sperm proteins, such as the left-right determination factor 1 (LEFTY1) and growth/differentiation factor 1 (GDF15) may prevent embryonic development. LEFTY1 is a member of the transforming growth factor (TGF) beta superfamily of proteins that may be involved in the left-right asymmetry determination of organ systems during development and seems to play important roles in decidualization and embryo implantation and to participate in stem cell pluripotency and cell fate decisions (Raya and Izpisúa Belmonte, 2006; Tabibzadeh and Hemmati-Brivanlou, 2006; Tabibzadeh, 2011). Our results also raise a potential involvement of this protein in sperm cell dysfunction. Similarly, GDF15, which also belongs to the TGF beta superfamily, seems to be able to promote cell proliferation and differentiation during mouse prostate development and has been described as an abundant cytokine in human seminal plasma (Böttner et al., 1999; Noorali et al., 2007; Soucek et al., 2010). Then again, several members of the TGF beta superfamily have been indicated as key mediators of testis development and spermatogenesis (Loveland et al., 2007; Fan et al., 2012).

To sum-up, sperm proteomic epigenetic alterations may be the origin of the stalling of embryo development after successful IVF in at least some idiopathic infertile couples. Our data suggest that proteomic alterations occurring during spermiogenesis may alter sperm epigenetic signatures, ultimately resulting in abnormal embryo development. These proteomic alterations seem to be subtle, not affecting major features of sperm quality, but putatively influencing sperm genetic chromatin distribution and imprinting. Ultimately, some of the differential proteins described in this study may be effective sperm biomarkers and the assessment of their expression in sperm (using, for example, a sperm-specific protein array) could ultimately be used to help anticipate ART results.

**Supplementary data**

Supplementary data are available at http://humrep.oxfordjournals.org/.

**Acknowledgements**

The proteomic analysis was carried out at the Technological and Scientific Centre at the University of Barcelona, a member of the ProteoRed network (http://www.proteored.org/). Maria José Fidalgo is thanked for many fruitful discussions.

**Authors’ roles**


**Funding**

This work was supported by grants to RO from the Spanish Ministry of Economy and Competitiveness (Ministerio de Economía y Competividad; FEDER BFU 2009–07118 and PI13/00699) and Fundación Salud 2000 SERONI 13–015. R.A. was supported by an IDIBAPS fellowship; A.A. was supported by the Portuguese National Science Foundation (Fundaçao para a Ciência e a Tecnologia; SFRH/BPD/63120/2009); J.C. was supported by a fellowship from the University of Barcelona.

**Conflict of interest**

The authors have no conflict of interest to declare.

**References**


Dayon L, Sanchez JC. Relative protein quantification by MS/MS using the Tandem Mass Tag Technology. Methods Mol Biol 2012;893:115–127 [Internet].


Raya A, Izpisúa Belmonte JC. Left-right asymmetry in the vertebrate embryo: from early information to higher-level integration. Nat Rev Genet 2006;7:283–293.


