Identification of genital tract markers in the human seminal plasma using an integrative genomics approach

A.D. Rolland¹, R. Lavigne¹,2, C. Dauly³, P. Calvel¹, C. Kervarrec¹, T. Freour⁴, B. Evrard¹, N. Rioux-Leclercq⁵,6, J. Auger⁷, and C. Pineau¹,2,*

¹Inserm, U1085, IRSET; University of Rennes I, Campus de Beaulieu, Rennes cedex F-35042, France ²Proteomics Core Facility Biogenouest, Campus de Beaulieu, Rennes cedex F-35042, France ³Thermo Fisher Scientific, 16 Avenue du Québec, Courtaboeuf cedex F-91963, France ⁴Médecine et Biologie du Développement et de la Reproduction, 38 Boulevard Jean Monnet, CHU Nantes, Nantes F-44000, France ⁵Department of Pathology, CHU Pontchaillou, Rennes, France ⁶CNRS/UMR6290; BioSit; Faculté de Médecine, University of Rennes I, Rennes, France ⁷Service d’Histologie-Embryologie, Biologie de la Reproduction/CECOS, Hôpital Cochin, Paris F-75014, France

*Correspondence address. Inserm U1085 – IRSET, Proteomics Core Facility Biogenouest, Campus de Beaulieu, 35042 Rennes cedex, France. Tel: +33 2 2323 5279; Fax: +33 2 2323 5282; E-mail: charles.pineau@rennes.inserm.fr

Submitted on June 27, 2012; resubmitted on August 22, 2012; accepted on September 3, 2012

STUDY QUESTION: Can protein biomarkers of the male genital tract be identified in human seminal plasma?

SUMMARY ANSWER: We identified potential biomarkers for each of the organs participating in the secretions of the human seminal plasma.

WHAT IS KNOWN ALREADY: The seminal plasma fulfills critical functions for fertility by providing spermatozoa with a protective milieu, promoting their final maturation and modulating the immune responsiveness of the female reproductive tract. It is also considered to be a promising source of biomarkers of male infertility and/or pathologies of the male genital tract.

STUDY DESIGN, SIZE, DURATION: This study combines proteomic analyses of normal seminal plasma together with transcriptomic gene expression profiling of human healthy tissues.

MATERIALS, SETTING, METHODS: Non-liquefied seminal plasma proteins from a healthy donor were prefractionated using two sequential Proteominer™ libraries. Eight subproteome fractions were collected, trypsin digested and subjected to three successive mass spectrometry analyses for peptide characterization. The list of identified proteins was compared with and merged with other available data sets of the human seminal plasma proteome. The expression of corresponding genes was then investigated using tissue transcriptome profiles to determine where, along the male reproductive tract, these proteins were produced. Finally, tissue specificity of a selected subset of biomarker candidates was validated on human tissues.

MAIN RESULTS AND THE ROLE OF CHANCE: We first performed a proteomic analysis of the human seminal plasma and identified 699 proteins. By comparing our protein list with other previous proteomic data sets, we found that 2545 unique proteins have been described so far in the human seminal plasma. We then profiled their expression at the gene level and identified 83 testis, 42 epididymis, 7 seminal vesicle and 17 prostate candidate protein markers. For a subset of testis-specific candidates, i.e. TKTL1, LDHC and PGK2, we further validated their germ cell expression and demonstrated that such markers could distinguish between semen from fertile and infertile men.

LIMITATIONS, REASONS FOR CAUTION: While some of the markers we identified are well-known tissue-specific products, further dedicated studies to validate the biomarker status of new candidates will be required. Additionally, whether or not the abundance of these proteins is indeed decreased in some specific pathological situations remains to be determined.

WIDER IMPLICATIONS OF THE FINDINGS: Using an integrative genomics approach, we identified biomarker candidates for each of the organs participating in the seminal plasma production. In this study, we essentially focused on germ cell markers and their potential application for the diagnosis of male infertility. Other types of markers also deserve a focused attention given their potential predictive value for various reproductive disorders, notably for prostate cancers.

STUDY FUNDING/COMPETING INTEREST(S): This work was supported by the Proteomics Core Facility at Biogenouest and was funded by Conseil Régional de Bretagne, IBISA and Agence de la Biomédecine grants. The authors declare that there exists a competing financial interest.

© The Author 2012. Published by Oxford University Press on behalf of the European Society of Human Reproduction and Embryology. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
Introduction

Infertility is a common problem throughout the world. Approximately 15% of all couples encounter fertility problems during their reproductive lifetime and in about half of the cases the cause is of male origin (Bruckert, 1991; Juul et al., 1999). Despite the possible deleterious role of lifestyle, environmental and/or genetic factors evoked on male reproduction, the etiologies of most infertility cases remain unknown. Routine semen analysis may indicate defective sperm production or diminished sperm motility as well as an abnormal sperm morphology resulting from a defective spermatogenesis. It may also reveal the absence of spermatozoa in semen, a situation termed azoospermia which concerns ~8% of men with fertility problems (Rowe et al., 2000). It is the consequence of either an obstructed or discontinuous male genital tract such as the bilateral agenesis of the vas deferens (obstructive azoospermia, OA) or a failure of the testis to initiate or maintain spermatogenesis due to endogenous or exogenous abnormalities (non-obstructive azoospermia, NOA).

The human seminal plasma is a complex biological fluid that originates from the testes (1–2%), the epididymides (2–4%), the seminal vesicles (65–75%), the prostate gland (25–30%) and the bulbourethral or Cowper’s glands (<1%). In addition to its obvious role in transporting male gametes, the seminal plasma provides a nutritive and protective milieu for spermatozoa. It is also actively involved in their final maturation as well as in modulating the immune responsiveness of the female reproductive tract (for reviews, see Robertson, 2005; Rodriguez-Martinez et al., 2011). To fulfill its different functions, the seminal plasma is composed of a complex range of inorganic and organic constituents. For instance, it contains zinc that stabilizes the sperm chromatin, basic amines (e.g. putrescine, spermine, spermidine and cadaverine) that counteract the acidic environment of the vaginal canal, prostaglandins that are involved in suppressing an immune response in the female against the foreign semen and fructose that serves as the main energy source of sperm cells (for reviews, see Fung et al., 2004; Pilch and Mann, 2006). Briefly, fresh ejaculates were recovered on ice and immediately centrifuged at 14 000 g for 5 min at 4 °C. The supernatants were supplemented with protease inhibitors (Complete™ Protease Inhibitor Cocktail; Roche Applied Science) and ultracentrifuged at 105 000 g for 60 min at 4 °C. The protein concentration was then determined in pooled ejaculates using the bicinchoninic acid (BCA) protein assay (Sigma-Aldrich). For western blot analyses, seminal plasmas from patients undergoing infertility check-up were collected following the current World Health Organization guidelines (Rowe et al., 2000). After liquefaction at room temperature for 30 min, seminal plasmas were centrifuged at 14 000g for 10 min at 4 °C, complemented with anti-proteases (Complete™ Protease Inhibitor Cocktail; Roche Applied Science) and ultracentrifuged at 105 000g at 4 °C for 60 min. The protein concentration was then determined using the BCA protein assay (Sigma-Aldrich) prior storage at −80 °C.

Methods

Materials

The solid-phase combinatorial peptide Library-1 (ProteoMiner™) and carboxylated version (Library-2) were both from Bio-Rad Laboratories (Hercules, CA, USA) as were materials for electrophoresis such as plates and reagents. Sequencing grade trypsin was from Promega (Madison, WI, USA). All other chemicals were from Sigma-Aldrich (Saint Quentin Fallavier, France).

Human seminal plasma collection

Proteomic analysis was conducted on pooled ejaculates from a single healthy donor, which were processed as described elsewhere (Pilch and Mann, 2006). Briefly, fresh ejaculates were recovered on ice and immediately centrifuged at 14 000 g for 5 min at 4 °C. The supernatants were supplemented with protease inhibitors (Complete™ Protease Inhibitor Cocktail; Roche Applied Science) prior to ultracentrifugation at 105 000 g for 60 min at 4 °C. For each ejaculate, the clear supernatant resulting from ultracentrifugation was stored at −80 °C until use. Protein concentration was finally determined in pooled ejaculates using the bicinchoninic acid (BCA) protein assay (Sigma-Aldrich).

For western blot analyses, seminal plasmas from patients undergoing infertility check-up were collected following the current World Health Organization guidelines (Rowe et al., 2000). After liquefaction at room temperature for 30 min, seminal plasmas were centrifuged at 14 000g for 10 min at 4 °C, complemented with anti-proteases (Complete™ Protease Inhibitor Cocktail; Roche Applied Science) and ultracentrifuged at 105 000g at 4 °C for 60 min. The protein concentration was then determined using the BCA protein assay (Sigma-Aldrich) prior storage at −80 °C.
Combinatorial peptide ligand library fractionation of seminal plasma proteins

The seminal plasma (472 mg proteins) was loaded onto a column containing 1 ml of NH₂-Library (Fractions E). The column effluent was continuously injected in a second column packed with COOH-Library (Fractions S). The columns connected in series were then washed with a phosphate-buffered saline until UV baseline of the effluent of the second column was reached. After the wash, each individual column was subjected to four distinct elutions (Fractions 1–4). The eight eluates (Fractions E1–E4 and S1–S4) were immediately neutralized, desalted by dialysis at 4°C against a 10-mM ammonium carbonate solution (cut-off of dialysis membrane was 1000 Da) and then lyophilized. For a more detailed protocol, see the Supplementary data.

Nano-LC-MS/MS analysis of human seminal plasma proteins

Eluates from the Proteominer™ fractionation were resuspended in ammonium bicarbonate to adjust the sample concentration to 1 mg/ml, then 5 μg of proteins of each fraction were reduced and alkylated. Proteins were digested using modified trypsin (Promega, Madison, WI, USA). Finally, the peptide mixtures were analyzed by nano-LC-MS/MS on a LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, USA). Finally, the peptide mixtures were analyzed by nano-LC-MS/MS on a LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). For a full description of the analytical parameters, see the Supplementary data.

Data processing and analysis

The Proteome Discoverer software (version 1.2; Thermo Fisher Scientific) with the SEQUEST® and Mascot search algorithms was used for protein identification. MS/MS spectra were searched against Uniprot human database (Release 2011_05; 20 239 sequences; 11 284 379 residues) for peptide characterization and protein identification. The Proteome Discoverer software was also used to generate lists of identified peptides in order to exclude them for subsequent LC-MS/MS analyses. A total of three acquisitions per sample were performed using this strategy. For the third analysis, results files from the first two runs were combined in the Proteome Discoverer software to generate a second exclusion list that contained peptides identified from the first and second acquisitions. For a full description of analytical parameters, see the Supplementary data.

Mapping of protein identifiers from other published data sets

Protein identifiers (IDs) from previous studies (GI numbers; Fung et al., 2004; Thimon et al., 2008; UniProt Entry names (Utleg et al., 2003; Drake et al., 2009); UniProt Accession numbers (Pilch and Mann, 2006; Polialkov et al., 2009); International Protein Index (IPI) numbers (Wang et al., 2009; Batruch et al., 2011) were retrieved and converted into current UniProt Accession numbers using the UniProt Batch retrieval and ID mapping modules (http://www.uniprot.org; release 2011_12). All UniProt accession numbers were further linked to EntrezGene IDs that were mapped to corresponding probe sets on HG-U133 Plus 2.0 GeneChip® arrays using the Affymetrix CSV file (http://www.affymetrix.com/index.affx; release 32, 9 June 2011). Combined data from the nine data sets are available in the Supplementary data, File S1.

Microarray tissue profiling experiments

Human seminal vesicles data set

Normal seminal vesicles were obtained at the Rennes University Hospital from patients who underwent radical prostatectomy. The protocol was approved by the local ethics committee of Rennes, and informed consent was obtained from the donors. Total RNAs were extracted using Trizol reagent (Invitrogen, Cergy Pontoise, France), purified with phenol–chloroform and cleaned-up using RNeasy Total RNA Isolation Kit (Qiagen, Courtaboeuf, France). Total RNAs (500 ng) were used as templates in an in vitro transcription reaction using the GeneChip® 3′ IVT Express Kit (Affymetrix, Santa Clara, CA, USA) according to the manufacturer’s instructions. Purified labeled cRNAs (20 μg) were fragmented and hybridized to HG-U133 Plus 2.0 GeneChip® arrays (Affymetrix). GeneChips were then washed using a GeneChip® Fluidics Station 450 (Affymetrix) and stained according to the manufacturer’s instructions. Corresponding CEL files are available on the NCBI Gene Expression Omnibus public repository (GSE17340).

Normalization and filtration procedures

CEL files corresponding to human prostate (GSE7307), epididymides (GSE7808; Thimon et al., 2007) and testicular samples (E-TABM-130; Chalmel et al., 2007) were downloaded from the NCBI GEO and from the EBI ArrayExpress public repositories. They were uploaded into the AMEN software (http://sourceforge.net/projects/amen/) together with our seminal vesicle data set and submitted to the RMA normalization procedure (Irizarry et al., 2003). Transcripts detected in a single tissue were selected as follows: signal intensities had to be above the median value (4.94531056811573) in that tissue and under this value in all other tissues. A fold change of ≥3 between the given sample and all other tissues was further required. Selected probesets were finally submitted to a statistical filtration (Llima statistical test, false discovery rate of <1%; Wettenhall and Smyth, 2004).

Gene ontology term analysis

UniProt accession numbers corresponding to all proteins identified within the seminal plasma were compared with the Human IPI proteome file to search for over- and under-represented functional annotation terms using the AMEN software. Similarly, genes identified as being expressed in a tissue specific way were compared with those of the entire microarray. For a full description of methodology, see the Supplementary data.

Validation experiment

Western blot analyses and/or immunohistochemistry experiments were performed as previously described (Com et al., 2006). The primary antibodies used for western blots were the mouse monoclonal anti-TKTL1 antibody (Ref WH000827771; Sigma-Aldrich) at a final dilution of 1:200, the rabbit monoclonal anti-LDH-C (Ref ab52747; Abcam, Paris, France) at a final dilution of 1:10 000, and the rabbit polyclonal anti-PGK2 antibody (Ref AV53820; Sigma-Aldrich). For immunohistochemistry, antibodies directed against TKTL1, LDH-C, PGK2 and S100A7 (Ref HPA006997; Sigma-Aldrich) were all used at a final dilution of 1:200.

Results

Fractionation and nano-LC-MS/MS analysis of the human seminal plasma

The large dynamic range of protein species present in biological fluids often prevents access to low-copy-number proteins in proteomic studies. To access such proteins, we used the Proteominer™ library technology to prefractionate seminal plasma prior to beginning proteome analysis. Non-liquefied seminal plasma was first loaded onto two different peptide libraries. The retained proteins were then subjected to four different elutions, which generated eight subproteome...
Identification of biological markers of the human male reproductive tract

The human seminal plasma is mainly composed of secretions from the seminal vesicles, the prostate, the epididymides and the testes. To decipher the respective contribution of each organ to the seminal plasma content, we performed a tissue-profiling analysis using publicly available gene expression data sets (testis, epididymis and prostate) and one data set that we generated for this study (seminal vesicle, GSE17340). The HG-U133 Plus 2.0 GeneChip® arrays notably contain 5445 probe sets that allow the investigation of gene expression for 2328 seminal plasma proteins (Fig. 1B, Supplementary data, File S1). We first identified genes that are specifically expressed (see the section Materials and methods for expression cut-offs) in each organ: the testis (1980 genes), the epididymis (389 genes), the seminal vesicle (37 genes) and the prostate (54 genes) (Supplementary data, Fig. S1A). The relevance of these genes was confirmed by highlighting enriched functions that are consistent with the biology of the respective organs, e.g. ‘spermatogenesis’ or ‘meiosis’ for the testis or ‘prostate gland development’ for the prostate (Supplementary data, Fig. S1B). We then focused on subsets of genes that were expressed in a tissue-specific manner and for which the corresponding gene product had also been identified in the human seminal plasma. We thus highlighted 83 testis, 42 epididymis, 7 seminal vesicle and 17 prostate proteins that represent potential markers in the human seminal plasma (Fig. 2, Supplementary data, File S2).

Validation of markers of male genital tract organs

Several of the proteins identified above are already known to be specific products of one of the four tissues investigated, e.g. SEMG1 and SEMG2 for the seminal vesicles, and FOLH1 and TGM4 for the prostate. In order to validate the biomarker status of some of our candidate proteins, we performed western blot analyses and/or immunohistochemistry using the few commercially available antibodies. TKT1L1, LDHC and PGK2 were consistently detected in testicular extracts and in normal seminal plasma only (Fig. 3A). Immunohistochemical studies further confirmed the testis-specific expression of TKT1L1, LDHC and PGK2. TKT1L1 was found expressed throughout germ cell development, with spermatogonia exhibiting the strongest labeling. LDHC was expressed in both meiotic spermatocytes and post-meiotic spermatids, whereas PGK2 was specifically observed in elongated spermatids (Fig. 3B). In the head segment of the epididymis, PGK2 and to a lesser extent LDHC immunoreactivities were present in epithelial cells, known to be involved in protein reabsorption (Fig. 3B). Of note is that while one could expect to observe labeled epididymal spermaatozoa (which are known to express PGK2 and LDHC), this was not true here because of extensive washes performed prior to tissue fixation and embedding. On the other hand, the expression of the epididymis-specific protein S100A7 was validated: whereas no staining was observed in the testis, the seminal vesicle, the prostate and the head of the epididymis, a signal was detected in the body, and reached highest levels in the epithelial cells of the tail (Fig. 3B). Finally, we assessed the presence of potential testicular germine markers in seminal plasma using sperm samples from fertile donors and from patients exhibiting various conditions (Fig. 3C). Whereas TKT1L1, LDHC and PGK2 were consistently...
Figure 1  In-depth analysis of the human seminal proteome. (A) Numbers of identified proteins in the human seminal plasma from multiple injections of each Proteominer™ fraction and dynamic exclusion of previously identified peptides. The numbers of unique proteins characterized in the raw human seminal plasma (Raw), in the final flow through (FT), and the eight Proteominer™ fractions (E1–E4 and S1–S4) are plotted. The numbers of new proteins identified in the second and third nanoLC-MS/MS analyses (injections 2 and 3, respectively) are also presented. (B) Comparison of the human seminal plasma proteomic studies. Protein IDs from other proteomic analyses of the human seminal plasma (GI numbers (Fung et al., 2004; Thimon et al., 2008); UniProt Entry names (Utleg et al., 2003; Drake et al., 2009); UniProt Accession numbers (Pilch and Mann, 2006; Poliakov et al., 2009) and IPI numbers (Wang et al., 2009; Batruch et al., 2011)) were retrieved and converted into UniProt accession numbers using the UniProt batch retrieval and ID mapping modules (http://www.uniprot.org; release 2011_12). UniProt accession numbers were finally linked to EntrezGene IDs and associated Probe sets from the HG-U133 Plus 2.0 GeneChip® arrays (Affymetrix) in order to perform the tissue-profiling analysis. (C) Functional mining of the human seminal plasma proteome. All proteins identified in the human seminal plasma proteome and their associated gene ontology (GO) terms were compared with those of the entire human proteome data set. ‘Molecular function’ and ‘cellular component’-enriched or -depleted GO terms were identified using the Fisher exact probability test. Only GO terms with a P-value of $\leq 10^{-5}$ and for which the number of proteins bearing the annotation term was $\geq 30$ were considered as statistically enriched. To avoid redundancy between closely related terms, we applied an ontology-specific information rate of $\geq 0.4$ (Chalmel and Primig, 2008). Rectangles indicate the observed (left) and expected (right) numbers of proteins bearing a specific GO term, whereas the number of proteins exhibiting this GO term in the entire human proteome is given next to the rectangle. Statistical enrichment (red) or depletion (blue) is indicated according to the scale bar.
detected in normal seminal plasmas, these proteins were either undetectable or barely detectable in seminal plasmas from patients with NOA, OA and vasectomized men (Fig. 3C).

**Discussion**

Deciphering the seminal plasma proteome is important for gaining new insights into the male gamete biology. Indeed, proteins identified thus far in the seminal plasma cover a large array of functions, ranging from energy production or immune response, to cell recognition (for reviews, see Robertson, 2005; Rodriguez-Martinez et al., 2011).

However, only few are unambiguously linked to (in)fertility. With the goal of constructing a comprehensive annotation of this biological fluid, we conducted an in-depth proteomic analysis of the human seminal plasma and identified a total of over 2500 gene products either directly or by inference from previously published works. We also designed an integrative genomics approach to coalesce the proteomic data together with the transcriptomic profiles of the organs involved in the production of this biofluid. Using this innovative method, we identified multiple potential organ marker proteins that could be used to create molecular signatures of normal seminal plasma and serve as diagnostic tool for male reproductive disorders.
In this study, we focused first on the proteomic analysis of the human seminal plasma. We combined sample pre-fractionation using peptide ligand library columns with subsequent shotgun nano-LC-MS/MS analysis of the tryptic peptides corresponding to each elution fraction and identified 669 proteins. Ligand libraries have been shown to be a key technology for capturing low-abundance proteins of the so-called ‘deep’ proteome (for reviews, see Righetti et al., 2006; Righetti and Boschetti, 2008; Boschetti and Righetti,

Figure 3  Validation of candidate markers of the urogenital tract. (A) Western blot analysis of TKTL1, LDHC and PGK2 in human tissues. Protein extracts (20 μg) from human testis (T), head, body and tail of the epididymis (EpH, EpB and EpT, respectively), seminal vesicle (SV), prostate (P) and normal seminal plasma (SP) were separated on 12% polyacrylamide gels, electrotransferred onto PVDF membranes and probed with specific antibodies. TKTL1, LDHC and PGK2 were detected only in the testicular protein extract and in normal seminal plasma. (B) Immunolocalization of TKTL1, LDHC, PGK2 and S100A7 along the male urogenital tract. Tissues were fixed by immersion in a formalin fixative and embedded in paraffin. Sections (5 μm) were dewaxed, rehydrated and incubated with antibodies against TKTL1, LDHC, PGK2 or S100A7. Complexes were revealed using a strepavidin–peroxidase amplification combination and sections were counterstained with Masson hematoxylin. TKTL1, LDHC and PGK2 immunoreactivities were detected only in the testis. TKTL1 immunostaining was present in all germ cells, with spermatogonia (arrowheads) exhibiting the strongest staining. LDHC was detected in meiotic spermatocytes (arrowheads) and post-meiotic spermatids (arrows), and PGK2 was detected only in post-meiotic spermatids (arrows). Some residual immunostaining for PGK2 and to a lesser extent for LDHC was observed in epithelial cells in the epididymis head sections (arrowheads). S100A7 was detected only in the body and the tail of the epididymis (arrowheads). Scale bars: 100 μm. (C) Screening for the testicular germline markers TKTL1, LDHC and PGK2 in the seminal plasma of normal and infertile donors. Seminal plasmas (50 μg protein), from fertile donors (lanes 1–5 and 10–13), from patients with NOA (lanes 7, 8, 14, 15 and 18) or OA (lanes 9 and 16) or from vasectomized men (lanes 6, 17), were separated on 12% polyacrylamide gels, electrotransferred onto PVDF membranes and probed with specific antibodies. TKTL1, LDHC and PGK2 proteins were all detected in seminal plasmas from fertile donors only.
identification of additional proteins in some individual fractions and in reinforcing the data sets obtained from published studies (Utleg et al., completed, the list of identified proteins was merged with the existing peptide characterization. To performing a thorough sampling of complex protein mixtures was re-injecting the sample for a second analysis by LC-MS/MS in which poorly represented species in each elution fraction. This involved re-injecting the sample for a second analysis by LC-MS/MS in which the previously identified peptides were excluded. Such an approach to performing a thorough sampling of complex protein mixtures was first described by Rudomin et al. (2009). In the present study, this procedure was performed twice for a total of three LC-MS/MS acquisitions per sample. This resulted in the identification of >50% additional proteins in some individual fractions and in reinforcing the identification of ~35% of proteins initially identified through a single peptide characterization.

Once our analysis of the human seminal plasma proteome was completed, the list of identified proteins was merged with the existing data sets obtained from published studies (Uteeg et al., 2003; Fung et al., 2004; Pilch and Mann, 2006; Thimon et al., 2008; Drake et al., 2009; Poliakov et al., 2009; Wang et al., 2009; Batruch et al., 2011). Some of these studies were published several years ago with often MS data not publicly available, which prevented us from performing protein identification against current databases and, most importantly, according to current guidelines. Instead, we had to convert and update all protein IDs provided in these studies into current Uniprot IDs prior to comparing them. We then found that >2500 non-redundant proteins were identified in the human seminal plasma so far. The different data sets only partially overlapped. A first explanation could be that differences in sample preparation or origin (i.e. liquefied versus non-liquefied plasma seminal, prostatic, epididymosomes) and in fractionation methods employed prior to analysis (i.e. 1D gel, 2D gel, ligand libraries) resulted in different protein enrichments. A second explanation would be that only very abundant seminal plasma proteins could be consistently identified across studies, while the identification of low-abundance ones would be more random. This latter hypothesis was confirmed when we used the number of characterized peptides as a reflection of protein abundance and found that proteins retrieved in multiple studies had been identified through a much higher number of peptides than those identified in only a single study (data not shown). While a small complement of low-abundance proteins certainly remains to be identified, the present data set is probably close to the complete seminal plasma proteome and constitutes a fundamental base to gain insights into this biofluid.

While the deciphering of the human seminal plasma proteome has greatly progressed over the past decade, still too little is known regarding the respective contribution of each organ to this biofluid. This lack of knowledge dramatically impairs our full understanding of how these proteins may interact together at the time of ejaculation and achieve their function(s) once mixed together. In order to predict the tissue origin for each of these proteins, we integrated proteomic data sets together with transcriptome gene expression data for each of the organs involved in the biofluid production. We aimed at demonstrating that it was possible with such an integrative approach to highlight proteins of interest without performing the so-called ‘differential studies’ generally employed to compare normal and pathological biofluids for biomarker identification projects. Through our analysis, we identified numerous proteins whose gene expression was restricted to the testes, epididymides, seminal vesicles and prostate. The monitoring of such proteins in the seminal plasma would potentially offer a non-invasive diagnosis tool for various reproductive disorders associated with each of these organs. As an example, the gene expression levels of at least three of the prostate markers identified in the present study (i.e. ALOX15B, DSC3 and TGM4) appear to be down-regulated in patients exhibiting benign or aggressive prostate tumors (Varambally et al., 2005). Thus, the expected decreased concentration of the corresponding proteins in the seminal plasma may possibly help early diagnosis of such pathologies.

In men with NOA, one or several invasive testicular biopsies are proposed to the patient for eventually recovering a few mature germ cells and entering an intracytoplasmic sperm injection program. These invasive biopsies may impair future spermatogenesis or induce testicular atrophy and only lead to sperm cell recovery in ~50% of cases (for review, see Ishikawa, 2012). Several indicators have been proposed to be potential indirect markers for the presence of spermatozoa in the testis. These include examination parameters (i.e. 2D:4D finger length ratio; Auger and Eustache, 2011) and proteins secreted by Sertoli cells such as inhibin A and B or AMH. However, none of these factors could be reliably used to predict successful biopsy outcomes (Anawalt et al., 1996; Illingworth et al., 1996; Anderson et al., 1998; Fenichel et al., 1999; von Eckardstein et al., 1999; Brugo-Olmedo et al., 2001; El Garem et al., 2002; Guthauser et al., 2002; Vernaeve et al., 2002; and for review, see Carpi et al., 2009). To date the only valuable method for answering whether or not spermatozoa can be found in the testes of NOA patients is actually the surgery with testis biopsy itself. This crucially calls for the discovery and validation of specific seminal plasma biomarkers that are absolutely required by clinicians and will help in predicting sperm-positive testicular biopsies. From our protein candidate list, we notably found 83 seminal plasma proteins for which the expression of the corresponding genes was only detected in the testis. Importantly, most of these genes were specifically expressed in the germ line, as evidenced by their high expression levels in isolated spermatocyte and spermatid samples. When investigated by immunohistochemistry, some of the corresponding proteins appeared to be expressed in the testis throughout germ cell maturation (i.e. TKTL1), whereas others were restricted to meiotic and/or post-meiotic germ cells (i.e. LDHC and PGK2). Interestingly, the presence/absence of a combination of such germ cell protein markers in the seminal plasma could be correlated with the status of spermatogenesis in NOA patients. Ideally, the seminal plasma of a Sertoli-Cell-Only syndrome patient would be negative for all three markers, patients with maturation arrests at either the spermatogonia or spermatocyte stages would be positive for TKTL1 or TKTL1 and LDHC, respectively, while patients with complete spermatogenesis would be positive for...
TKTL1, LDHC and PGK2. Additionally, since both PGK2 and LDHC are required for sperm mobility and male fertility (Odet et al., 2008; Danshina et al., 2010), the detection of these proteins in the human seminal plasma would guarantee not only the retrieval of testicular spermatozoa, but also their good quality. Using western blots, we clearly detected these three germ cell markers in normal seminal plasmas and were thus able to successfully distinguish those from seminal plasmas devoid of spermatozoa. Routinely, however, the measurement of germ cell markers in the seminal plasma may appear impractical. Indeed, in some pathological NOA situations only a few nests of residual spermatogenesis can be present in various areas of the testes. As a consequence the expected germ cell protein markers could be found at very low amounts in the seminal plasma. Therefore, the development of a more sensitive diagnostic antibody array (Fertichip™) for measuring several of those identified germ cell markers in normal and pathological seminal plasmas using specific monoclonal antibodies is ongoing in our laboratory.

Recently, two proteomic studies with a similar aim as ours were conducted by the same group (Batruch et al., 2011; Drabovich et al., 2011). In the first one, Batruch et al. (2011) directly compared the seminal plasma proteome of fertile donors with that of vasectomized men. The authors assumed that proteins specific to the seminal plasmas of fertile donors originate from the testis and/or the epididymis, which was partially confirmed by database searching and literature mining. According to our tissue profiling experiment, the genes corresponding to 15 of the 32 potential protein markers identified by Batruch et al. (2011) were indeed found to be expressed in only one or both organs (data not shown). Importantly, the remaining ones were also expressed in other tissues (e.g. the seminal vesicles and/or the prostate) illustrating that the lack of a protein being identified in a sample is not always a convincing indication that the protein is not expressed. In a more recent follow-up study, Drabovich et al. (2011) used their biomarker candidate list to develop a multiplex assay based on the selected reaction monitoring technology with the objective of discriminating OA from NOA patients. This study was rather convincing, as their approach appeared efficient and robust to distinguish normal seminal plasmas from that of NOA or vasectomized men. Even more promising were 11 markers that allowed the discrimination of NOA from vasectomized men and could potentially distinguish between different NOA patients (Drabovich et al., 2011). However, we could demonstrate from our transcriptomic data set that these 11 markers are expressed in the epididymis but not, or only at baseline levels, in the testis (data not shown). As a consequence, these 11 proteins can indeed distinguish NOA from vasectomized men, not because of the absence of spermatozoa in vasectomized samples, but because of the absence of epididymal secretions. Their use for predicting the status of spermatogenesis in NOA patients may thus be questionable.

In the present study, we used an integrative genomics approach involving in-depth proteomic characterisation and bioinformatic mining of the human seminal plasma together with transcriptomic analysis to predict tissue origin of the protein constituents of this biofluid. This concept study led to the identification of several biomarkers specific to each of the organs involved in the seminal plasma production, including specific proteins of post-meiotic germ cells whose presence in seminal plasmas might be monitored by clinicians to better diagnose infertile patients, in particular those with NOA.

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

Acknowledgements
We thank Drs F. Chalmel for advice concerning the use of the AMEN software, L. Guerrier for help with the Proteominer™ technology and B. Jégou for advice.

Authors’ roles
A.D.R. and C.P. were involved in study concept and design. A.D.R., R.L., C.D., P.C., C.K. and B.E. contributed to data acquisition. A.D.R., R.L., P.C. and C.P. were involved in analysis and interpretation of data. T.F., N.R.-L. and J.A. provided clinical samples and participated in critical discussion. A.D.R., R.L. and C.P. contributed to drafting the manuscript.

Funding
This work was supported by the Proteomics Core Facility at Biogone-ouest and was funded by Conseil Régional de Bretagne, IBSA and Agence de la Biomédecine grants. A.D.R. holds a fellowship from Inserm and Région Bretagne. P.C. holds a fellowship from Organon. We thank Dr Y. Vandenbrouck and the technical staff of the webportal ProteHome for hosting and making available our data.

Conflict of interest
The authors declare that there exists a competing interest in this work that is related to a patent application on the use of identified germ cell-specific proteins in an antibody-based assay (Fertichip™) to predict successful testicular biopsy outcomes in human NOA.

References


von Eckardstein S, Simoni M, Bergmann M, Weinbauer GF, Gassner P, Schepers AG, Nieschlag E. Serum inhibitin B in combination with serum FSH alone for impaired spermatogenesis in men, but
