Proteomics and the search for biomarkers of female reproductive diseases

Katie L Meehan, Adam Rainczuk, Lois A Salamonsen and Andrew N Stephens

Prince Henry's Institute of Medical Research, Level 4, PO Box 5152, Clayton, Victoria 3168, Australia

Correspondence should be addressed to K L Meehan; Email: katie.meehan@princehenrys.org

K L Meehan and A Rainczuk contributed equally to this work

Abstract

Over the past decade, high-throughput proteomics technologies have evolved considerably and have become increasingly more commonly applied to the investigation of female reproductive diseases. Proteomic approaches facilitate the identification of new disease biomarkers by comparing the abundance of hundreds or thousands of proteins simultaneously to find those specific to a particular clinical condition. Some of the best studied areas of female reproductive biology applying proteomics include gynaecological cancers, endometriosis and endometrial infertility. This review will discuss the progress that has been made in these areas and will highlight some of the emerging technologies that promise to contribute to better understanding of the female reproductive disease.

Introduction

Diseases of the female reproductive tract contribute significantly to the global burden of diseases and as such are an important target for the development of better diagnostic, prognostic and therapeutic strategies. The use of proteomics technologies, enabling the simultaneous comparison of hundreds or thousands of proteins to identify disease-specific biomarkers, offers an attractive approach to the identification and development of new clinically relevant tools. The increasing availability of proteomics technologies, development and application of better sample preparation techniques and recent advances in mass spectrometry (MS) instrumentation are now enabling identification and quantitation of lower abundance proteins involved in the disease. Such developments are likely to make a significant impact on reproductive biology research.

To date, the most common proteomics technologies applied to investigate the female reproductive diseases include ‘gel-based’ proteomics such as two-dimensional gel electrophoresis (2DE) or differential in-gel electrophoresis (DIGE). A technique known as surface-enhanced laser desorption/ionisation time-of-flight MS (SELDI TOF MS) has also been widely used. These have been applied to identify changes in protein abundance between normal and disease states, or to decipher the underlying molecular mechanisms of tumour chemoresistance. Sample types that have been commonly used include entire or laser capture microdissection (LCM) dissected tissues, minimally fractionated plasma/serum, uterine lavage, and both primary and immortalised cell systems. Other more advanced ‘gel-free’ proteomics technologies including multidimensional proteomics identification technology (Mudpit), quantitative MS-based proteomics and matrix-assisted laser desorption ionisation imaging (MALDI-IMS) have also been applied, although in a more limited context. Despite these ongoing efforts, research initiatives have yet to develop widely applicable clinical tools for the diagnosis of disease.

In this review, we provide an overview of proteomics technologies as applied to research targeting reproductive disease. We discuss the challenges of applying proteomics to the analysis of commonly used sample types in reproductive biology, and detail the progress made thus far in some of the best studied areas including gynaecological cancers, endometriosis (EOS) and endometrial infertility. In addition, several emerging technologies that promise to contribute to better understanding of the reproductive disease will be highlighted.

Proteomics technologies for biomarker discovery

Proteomic analyses permit the global comparison of proteins from almost any biological sample, enabling the identification of multiple proteins of interest within a single experiment. Differential protein expression between conditions (for example, healthy versus disease), relative protein abundance, the presence of post-translational modifications, sub-cellular localisation
and protein turnover, protein–protein interactions or other functional aspects may all be examined by applying proteomic tools within a well-defined hypothesis. As proteomics technologies continue to evolve and become more sophisticated, several key requirements for successful studies have become clear. Experimental reproducibility is the key to biomarker identification, and sample handling at all stages of the experiment is a crucial factor. Reproducible proteomics datasets can only be generated if samples are collected, processed and stored according to highly standardised and robust procedures. Furthermore, there is a significant need to characterise (by phenotyping, pathology testing or otherwise) the samples prior to analyses to ensure that representative sample types are being used. To this end, ongoing efforts such as the Human Plasma Proteome Project have sought to define specific parameters for the analysis of blood-based biomarkers (http://www.hupo.org/research/hppp/).

Studies conducted across multiple laboratories have defined a standard set of conditions for the collection, handling, storage and processing of plasma for proteomics studies, to overcome the inherent variability in serum introduced during clotting (for review see Rai & Vitzhum (2006)). Indeed, up to 40% of peptides present in serum are specifically found in serum, but not in plasma (Anderson 2010). Other recent studies have demonstrated that post-translational modifications – for example, the phosphorylation state of many proteins – can result from tissue injury at collection, changes in oxygenation or time before fixation or freezing (Espina et al. 2008, Zhou et al. 2009). The presence of highly abundant, disease-unrelated proteins in samples must also be accommodated. These abundant proteins tend to mask proteins of lower abundance – for example, cytokines and growth factors that play important biological roles in disease progression – and must be removed prior to analysis. Studies targeting low abundance proteins therefore require high quality, fractionated samples and a well-defined experimental question for success.

A number of the more common experimental methodologies and techniques applied in proteomics studies are outlined in Table 1. Overall, proteomic approaches may be broadly defined as either ‘gel-based’ or ‘gel-free’ (Fig. 1); furthermore, ‘label-based’ or ‘label-free’ approaches may be used in each case to quantitatively analyse the samples for changes in protein abundance.

Gel-based proteomics typically uses 2DE to separate proteins prior to analysis, generating unique proteomic patterns for each sample. Proteins may either be labelled prior to separation by 2DE (label-based) or be alternately stained following separation (label-free) to visualise protein profiles (Fig. 1). Pattern analysis is used to locate proteins of interest, following which the identity of the protein can be determined by MS. Gel-based approaches are limited by the size and hydrophobicity of the proteins being studied; however, they offer very high resolution and are typically easily accessible to almost any laboratory.

By contrast, gel-free proteomic approaches apply multiple chromatographic steps to separate peptides or proteins, followed by analysis and identification by MS (Fig. 1). Label-based proteomics, where proteins are either isotopically or chemically labelled prior to separation, is currently the most commonly applied technique for gel-free proteomics. Quantitation and comparison are facilitated by detection of the introduced label during MS analysis. However, label-free approaches are becoming increasingly more common. In this case, peptide separation is compared across chromatographic profiles, and quantitation is determined either by spectral counting or by peak intensity during MS. A discussion of MS technologies used in proteomics is beyond the scope of this article; these have been reviewed recently (Griffiths & Wang 2009).

Challenges of working with reproductive samples

Significant challenges arise when working with reproductive tissues, fluids and related biological materials. Biological fluids such as plasma or urine are easy and relatively non-invasive to collect, and are therefore favoured in clinical applications. However, these are often not ideal choices for discovery-based proteomics studies. The vast dynamic range (at least 6–12 orders of magnitude) and dominance of highly abundant plasma proteins make the analysis and identification of low molecular mass proteins from plasma difficult (Anderson 2010). In addition, if these samples are distant from disease lesions, disease-specific markers may only be present at very low abundance. To overcome these caveats, it has been proposed that the samples proximal to disease lesions be used (Hanash et al. 2008). For example, ascitic fluid that bathes ovarian tumours has been used in biomarker studies of ovarian cancer (Faca et al. 2008). Similarly, uterine and peritoneal fluids have more recently been investigated in several studies of EOS and infertility (Hannan et al. 2009a). Nevertheless, all of these fluids contain abundant blood proteins that must be removed prior to analysis (Hannan et al. 2009a).

Fractionation is therefore an essential component for any analysis of biological fluid. A number of strategies are applicable (Table 1); some of these include isoelectric focusing, chromatography, immunodepletion (Ametzazurra et al. 2009, Hannan et al. 2009a) or affinity depletion (e.g. glycoprotein enrichment via lectin chromatography (Imre et al. 2008, Li et al. 2008a, Abbott et al. 2009, Di Michele et al. 2009b).

As an alternative, many studies choose to analyse tissue samples for markers of disease. While relatively simple to collect (e.g. via biopsy or curettage), biomarkers found in tissue are not directly translatable for clinical use (unlike biological fluids). In addition, the proteomic analysis of tissue is dominated by the
Proteomics strategies applied to gynaecological disease

In the section that follows, we provide an overview of proteomics technologies as applied to three areas of intense interest in the field; gynaecological cancers, EOS and endometrial infertility. While biomarker discovery is the main theme of most studies discussed, prediction of response to therapies and prognostic outcomes are mentioned.

Epithelial ovarian cancer

Epithelial ovarian cancer (EOC) is currently the fifth leading cause of cancer-related deaths in women (Jemal et al. 2008), with around 21, 550 new diagnoses expected in the US in 2010 (http://www.cancer.gov/cancertopics/types/ovarian). Typically diagnosed at a late stage due to lack of symptoms or adequate screening strategies, patients with advanced disease have <30% chance of 5-year survival (Chan et al. 2006). Currently, CA 125 is the only clinically utilised marker for EOC diagnosis; however, this is neither a suitable marker for routine detection of early-stage disease nor for regular population screening. There is therefore a pressing need to identify a marker of early-stage disease, capable of diagnosing early-stage EOC and thus improving the 5-year survival.

Several research groups worldwide seek to address this issue; for an extensive review, see Cadron et al. (2009).

Limited gel-based studies have been applied to the discovery of ovarian cancer biomarkers. Bengtsson et al. (2007) used LCM and DIGE to analyse tissue samples from 64 patients with either benign or malignant disease. Over 200 changes in protein abundance between malignant and non-malignant tissues were identified; immunohistochemistry confirmed the cancer-specific changes for five of these proteins (Bengtsson et al. 2007). In another study, Zhao et al. (2008) applied DIGE to serum samples from women with or without ovarian cancer and identified 41 proteins with different expression profiles. A combination of transferrin, haptoglobin and CA 125 was found to potentially improve the sensitivity and specificity of diagnosis of ovarian epithelial cancers (Zhao et al. 2008). DIGE has also been applied in other contexts – for example, to study the mechanisms conferring chemoresistance in ovarian cancers (Cicchillitti et al. 2009, Di Michele et al. 2009a, Kim et al. 2009).

Gel-free quantitative proteomics technologies have been more extensively applied to ovarian cancer biomarker research to date. ‘Mudpit’ studies, involving the identification of large numbers of proteins by MS but without quantitative information, have been used to characterise immortalised and primary ovarian cells and their secretory products (Gunawardana et al. 2009, Kuk et al. 2009, Levanon et al. 2010). Of more interest are those quantitative studies attempting to identify disease markers. Stable isotope labelling of amino acids in cell culture (SILAC), whereby cells are grown in the presence of isotopically labelled amino acids and the abundance of labelled proteins are compared by MS, was applied by Faca et al. (2008) to analyse ovarian cancer cell lines and ovarian tumour cells enriched from ascites fluid. Over 6400 secretory and cell surface proteins were identified (Faca et al. 2008); subsequent bioinformatic analysis identified functional groups of proteins involved in a) accelerated detoxification of drug substrates, b) inhibition of apoptosis and associated cytoskeletal remodelling and c) modulation of basal metabolism.

In a more recent study, immunodepleted plasma from a mouse xenograft model was assessed for the presence of circulating human peptides (Pitteri et al. 2009). Isotopic labelling was used to quantitate the presence of human-derived peptides, which could only have been produced by the implanted ovarian tumour. From 106 proteins of interest identified, 58 were secreted or shed from tumour cells; five of these were confirmed by independent means at increased levels in an independent set of plasma samples (Pitteri et al. 2009). This is one of the most exciting studies in the field to date, and this demonstrates the power of quantitative proteomics for the identification of biomarkers of ovarian cancer.

Several studies have also applied SELDI TOF MS to compare proteomic profiles in plasma and urine from ovarian cancer patients (for a review, see Engwegen et al. (2006)). While originally intended as a biomarker discovery tool, SELDI TOF MS is now widely used in high-throughput screening strategies to evaluate biomarkers identified by other approaches (Wang et al. 2008, Rism et al. 2009). This is primarily due to the inability to directly identify biomarker candidates by...
Table 1 Commonly applied proteomics strategies for sample fractionation, analysis and validation to identify markers of reproductive diseases.

<table>
<thead>
<tr>
<th>Type</th>
<th>Function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fractionation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immunoaffinity depletion resins</td>
<td>Remove 6–20 highly abundant proteins from complex biological samples</td>
<td>Stempler et al. (2008)</td>
</tr>
<tr>
<td>Isoelectric focusing (IEF) and OFFGEL</td>
<td>Fractionates proteins in pH gradients according to their pI</td>
<td>Hubner et al. (2008) and Ermoult et al. (2010)</td>
</tr>
<tr>
<td>Affinity enrichment – glycoproteins or phosphoproteins</td>
<td>Affinity columns for enrichment, e.g. lectins (binds glycoproteins), titanium oxide (TiO) and immobilised metal affinity columns (IMAC) (bind phosphopeptides)</td>
<td>Drickamer &amp; Taylor (2002) and Zaia (2008)</td>
</tr>
<tr>
<td>Size exclusion/affinity nanoparticles (SEAN)</td>
<td>Selective binding and capture of positively charged, low-molecular weight proteins</td>
<td>Rainczuk et al. (2010)</td>
</tr>
<tr>
<td>Array technologies</td>
<td>Specific binding of protein subsets based on array surface affinity. Protein, antibody, and carbohydrate arrays are available. Rapid screening of protein subsets</td>
<td>Wingren &amp; Borrebaeck (2009) and Wingren et al. (2009)</td>
</tr>
<tr>
<td>Label-based comparative proteomics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Differential in-gel electrophoresis (DIGE)</td>
<td>Quantitation using fluorescent CyDyes following two-dimensional gel electrophoresis (IEF and SDS-PAGE). Up to three different samples (two samples and one reference sample) can be compared on a single gel</td>
<td>Unlu et al. (1997), Wilson et al. (2005), Miller et al. (2006) and Timms &amp; Cramer (2008)</td>
</tr>
<tr>
<td>Stable isotope labelling of amino acids in cell culture (SILAC)</td>
<td>Immortalised cells are cultured in the presence of stable amino acid isotopes, which are directly incorporated into proteins. Experiments quantitatively comparing two cell conditions (control versus treatment)</td>
<td>Ong et al. (2002)</td>
</tr>
<tr>
<td>Acid cleavable isotope-coded affinity tag (iCAT) labelling and MS</td>
<td>Isotopic labelling of cysteine-containing peptides using iCAT reagents. Quantitatively comparison compared using MS</td>
<td>Gygi et al. (1999), Moseley (2001), Meehan &amp; Sadar (2004), DeSouza et al. (2005a) and Pan et al. (2009)</td>
</tr>
<tr>
<td>Isotope-coded protein labelling (iCPL) and MS</td>
<td>Labels amino groups at the protein level using nicotinoyloxy succinimide (Nic-NHS). Relative quantification is achieved by comparison of the relative abundance of peak intensity of differentially labelled peptide pairs</td>
<td>Schmidt et al. (2005)</td>
</tr>
<tr>
<td>Isobaric tags for relative and absolute quantitation (iTRAQ) and MS</td>
<td>Labels N-termini and lysine side chains of peptides from up to eight different biological samples with eight specific isobaric isobaric-independent reagents that during MS, give rise to eight unique reporter ions that are subsequently used to quantify the different samples</td>
<td>Latterich et al. (2008)</td>
</tr>
<tr>
<td>Acrylamide labelling and MS</td>
<td>Labels 13C acrylamide of cysteine residues within proteins and peptides</td>
<td>Faca et al. (2006) and Griffiths &amp; Wang (2009)</td>
</tr>
<tr>
<td>Label-free proteomics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mudpit</td>
<td>Involves the combined use of two-dimensional chromatography (SCX and RP) on-line with a MS</td>
<td>Link et al. (1999)</td>
</tr>
<tr>
<td>Spectral counting by MS</td>
<td>Number of peptides typically correlates with protein abundance. Comparison between different samples is achieved by comparing the number of identified MS spectra from the same protein in each MS dataset</td>
<td>Liu et al. (2004), Zybaiov et al. (2005), Pan et al. (2008), Wang et al. (2009) and Zegels et al. (2009)</td>
</tr>
<tr>
<td>Validation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiplexed assay technologies</td>
<td>Validation of multiple (up to 100) analytes simultaneously by immunoassay</td>
<td>Fu et al. (2010)</td>
</tr>
<tr>
<td>Multiple reaction monitoring (MRM)</td>
<td>Allows for quantitation of known proteins within any given sample via measurement of fragment ions within a triple-quadrupole MS</td>
<td>Kitteringham et al. (2009) and Yocum &amp; Chinnaiyan (2009)</td>
</tr>
</tbody>
</table>
SELDI TOF MS; poor study design, variability introduced through sample handling and inconsistencies in data analysis have also made SELDI TOF MS a less popular approach for biomarker discovery (De Bock et al. 2010). Nevertheless, SELDI TOF MS remains an important proteomic tool. In one noteworthy study, SELDI TOF MS was applied to identify discriminatory patterns between malignant and benign pelvic masses in proteins present in urine. Twenty significantly different proteins/peaks were visualised, with the three most significant subsequently identified as fibrinogen-α fragment, collagen-α1 (III) fragment and fibrinogen-β NT fragment (Petri et al. 2009). Although insufficiently sensitive or specific enough for stand-alone diagnosis, the three peaks identified conferred enhanced predictive ability for ovarian cancer over CA 125 alone (Petri et al. 2009).

A key protein class attracting interest in ovarian cancer biomarker research are the glycoproteins. In addition to the large glycoprotein CA 125, numerous other glycoproteins have been proposed as potential markers for ovarian cancer including osteopontin, human epididymis protein 4 (HE4; recently approved for clinical use), CA 15-3, CA 19-9, inhibin and several members of the kallikrein family (kallikreins 5, 6, 8,10, 11 and 14; Giuntoli et al. 1998, Kim et al. 2002, Drapkin et al. 2005, Borgono et al. 2006, Luo et al. 2006, Shan et al. 2006, McIntosh et al. 2007, Robertson et al. 2007, Shihle et al. 2007). Using specific affinity enrichment for glycoproteins (Table 1), recent analyses of the serum glycoproteome to identify biomarkers of ovarian cancer have been undertaken (Abbott et al. 2007). A number of tumour-specific glycan changes have been identified to date that can distinguish ovarian cancer patients from controls (Li et al. 2008a, Abbott et al. 2009), or predict response or resistance to chemotherapy (Di Michele et al. 2009b). Other studies have identified several aberrantly glycosylated proteins with potential as biomarkers for ovarian cancer (Imre et al. 2008). While promising, further work is needed to translate these findings into useful clinical tests and to extend the number of glycosylated targets for analysis.

Despite these encouraging results, no biomarkers of ovarian cancer have yet been identified that can reliably diagnose early-stage EOC, nor discriminate between benign and malignant diseases with suitable accuracy for clinical use. Further research is necessary to identify an acceptable marker of ovarian cancer.

**Endometrial cancer**

While endometrial cancer (EC) is the most common gynaecological malignancy in the United States (Jemal et al. 2008), most ECs are diagnosed and treated at an early stage and have a good prognosis (Jemal et al. 2008, Fiorelli et al. 2008). The incidence and mortality rates for EC continue to rise (von Gruenigen et al. 2005, Jemal et al. 2008), believed to be due to a combination of factors including increasing rates of obesity, early menstruation, late menopause, tamoxifen use and high oestrogen levels. Notably, there has been no statistical improvement in mortality rates for the past 25 years (Blair & Casas 2009). Currently, no biomarker exists for the detection of early-stage EC; diagnosis therefore relies on the presence of common symptoms, which are not always evident in all patients (Blair & Casas 2009). Relapsed or late-stage ECs are associated with high levels of morbidity and mortality. Characterisation of sensitive and specific biomarkers would reduce the incidence and mortality from EC, improve risk assessment, facilitate screening and enable better choice of treatment for EC patients (Lacey & Chia 2009).

Using gel-based proteomics, Li et al. (2008b, 2008c) identified and validated both calcyphosine and cyclophilin A as potential markers of EC with likely involvement in carcinogenesis. Importantly, up-regulation of cyclophilin A conferred resistance to chemotherapeutic-induced apoptosis in cancer cells (Choi et al. 2007, Gonzalez-Santiago et al. 2007). Cyclophilin A inhibition may therefore be useful in improved management and prognosis for patients with ECs (Choi et al. 2007, Gonzalez-Santiago et al. 2007). Similarly, a more recent study using DIGE identified several candidate markers of myometrial infiltration, and proposed potential targets for therapies directed against the initial steps of endometrial metastasis (Monge et al. 2009a, 2009b). Up to 20% of patients diagnosed with ‘early stage’ EC relapse, despite the apparent absence of locally advanced or metastatic disease at the time of treatment (Engelsen et al. 2009). Improved methods for the
pre-surgical staging and prognostic evaluation of ECs would therefore be of considerable clinical benefit.

A number of gel-free studies have also identified potential early-stage biomarkers of EC (DeSouza et al. 2005a, 2007, 2008, Kozak et al. 2005, Casado-Vela et al. 2009). SELDI TOF MS studies comparing serum samples between EC and cancer-free patients have also identified potential candidate biomarkers including chaperonin 10 (DeSouza et al. 2005a, Guo et al. 2005), calgranulin A (Guo et al. 2005) and two forms of apolipoprotein-A1 (Takano et al. 2010). Further research is needed to elucidate the sequence of additional interesting peptides identified from SELDI TOF MS studies (Zhu et al. 2008), and to validate and determine the sensitivity and specificity of resulting biomarkers. Indeed, a recent independent validation study of previous proteomics work (Kozak et al. 2005) assessed the discriminatory ability of three candidate biomarkers using over 400 serum samples. Although strong predictive trends were observed, the authors concluded that further large population-based analyses were necessary to validate their data (Farias-Eisner et al. 2010).

Ongoing analysis, including validation of protein expression changes in large and independent sample sets, is required for all of these studies. The combination of biomarker data with known risk factors for EC (for example, HNPCC heterozygosity, obesity or high oestrogen levels) is likely to prove important for the development of better clinical tools for EC diagnosis and treatment.

**Cervical cancer**

Cervical cancer is the second most lethal cancer among women worldwide with 288 000 deaths annually (http://www.who.int/vaccine_research/diseases/hpv/en/). Cervical cancer incidence and mortality rates have decreased significantly since the introduction of the conventional Pap test (Papanicolaou 1942) and the more recent ThinPrep Pap test (Weintraub & Morabia 2000). Recently, the cervical cancer vaccine GARDASIL (effective against four types of human papillomavirus) was commercialised. While clearly of major benefit, GARDASIL does not prevent all types of cervical cancer; in addition, the sensitivity and specificity of Pap tests rely upon the ability to visually detect and accurately identify characteristic cellular changes (Broomall et al. 2010). The need for sensitive and specific screening regimes therefore remains (Monge et al. 2009a). The discovery of new biomarkers for cervical cancer offers one way to improve sensitivity, specificity and ultimately patient prognosis.

To date, there have been few reports applying proteomics to identify cervical cancer biomarkers. Three 2DE studies analysing vaginal carcinoma reported over 120 proteins between them as ‘different’ in squamous cervical cancer tissues (Hellman et al. 2004, Bae et al. 2005, Zhu et al. 2009); however, no biomarker panels emerged from these datasets. In another study, LCM was used to isolate high-grade, pre-cancerous cells from ThinPrep-prepared slides. Using an MS-based approach, over 1000 proteins were identified; 150 were
up-regulated and 70 were down-regulated in high-grade dysplastic cells compared with normal cervical epithelial cells (Gu et al. 2007). The ability of these candidates to differentiate between underlying high-grade and mild dysplasia, either on their own or in combination with Pap test, requires further validation and testing. Nevertheless, this is an exciting finding that promises to improve the efficacy of cervical cancer screening.

Proteomic analysis of plasma proteins by SELDI TOF MS has also been used in an attempt to differentiate between in situ and invasive cervical cancers (Lin et al. 2006). However, the identity of the differentiated proteins was not determined: these may be of value in the understanding of cervical cancer invasion and in the development of novel therapeutic intervention (Lin et al. 2006). This again highlights the need for proteomics studies to provide both identification and quantitation strategies to expedite translation of basic science into the clinic.

**Endometriosis**

EOS is a chronic disorder that occurs when endometrial fragments or cells form ectopic lesions outside the endometrium, classically on surfaces within the peritoneal cavity, where they remain hormonally responsive, stimulate angiogenesis and are highly invasive. EOS is generally accompanied by pelvic pain and dysmenorrhoea, and is significantly associated with infertility (http://www.nlm.nih.gov/medlineplus/endometriosis.html#cat5). Gene profiling and other studies suggest that progesterone resistance of the eutopic endometrium may be an underlying feature of EOS (Burney et al. 2007).

The gold standard for diagnosis of EOS is a surgical procedure that is typically performed as an outpatient surgical procedure. As a result, there is generally considerable delay in diagnosis – around 8 years from the onset of symptoms (Hadfield et al. 1996). There is therefore an urgent, unmet need for a minimally invasive diagnostic test for EOS.

To date, the two most commonly applied technologies in the investigations of EOS are SELDI TOF MS and 2DE. Using SELDI TOF MS to analyse proteins in serum, Seeber et al. (2009) identified 57 peaks that achieved the diagnostic targets of either >90% specificity with >20% sensitivity or vice versa, for the diagnosis of EOS in a test group. When combined with their best performing markers from a previous study, it was determined that 73% of all the subjects would have been diagnosed with 94% accuracy (Seeber et al. 2008). However, the test group included only subjects with stage III–IV EOS; it is unclear whether the test would also detect early-stage disease. A similar study (Liu et al. 2007) also found discriminatory patterns between EOS and control patients, but no protein identities were determined.

SELDI TOF MS has also been applied to compare eutopic endometrium from women without and with EOS, along with peritoneal lesions and normal peritoneal biopsies. These studies demonstrated that endometrial proteins of low abundance and molecular mass (2.8–12.3 kDa) were decreased in women with EOS (Kyama et al. 2006). The greatest differences were found between peritoneal lesions and normal peritoneum. In particular, the increased abundance of a cluster of proteins ~23 kDa, identified as transgelin, was observed in lesions (Kyama et al. 2006). Subsequent works demonstrated that this cluster is regulated by human chorionic gonadotrophin in normal eutopic endometrium, but not in women with EOS (Brosens et al. 2009).

In studies of EOS, 2DE has been extensively applied to both tissue biopsies and fluids. Proteins identified as aberrant in eutopic endometrium from women with and without EOS included secretory proteins (apolipoprotein A2), redox regulators, (peroxiredoxin 2), chaperonins (heat shock protein 90 and annexin A2), and proteins associated with DNA metabolism and catabolism (Fowler et al. 2007). More recently, the application of DIGE identified and validated vimentin, PRDX6 and RNase/angiogenin inhibitor 1 (RNH1) as decreased in eutopic endometrium of women with EOS (Stephens et al. 2010). An important finding of this study was the presence of multiple charge and size isoforms of PRDX6, including one isoform specific to the endometrium of women with EOS and undetectable in normal endometrium (Stephens et al. 2010). Such unique isoforms could enhance discrimination between the disease states, and thus prove excellent candidates for diagnostic purposes.

Sample types other than tissues and plasma are now being increasingly analysed for new markers of EOS. Several studies have investigated peritoneal fluid using 2DE (Kyama et al. 2006, Ferrero et al. 2007, 2008). Although cyclical changes were not detected, all of these studies found differences in protein abundance between women with or without EOS (Ferrero et al. 2009). Furthermore, several protein isoforms differed between women with stage I–II disease compared with stage III–IV disease (Ferrero et al. 2008). Most of the proteins identified were involved in the immune response, consistent with our understanding of the inflammatory nature of the disease and the increased peritoneal fluid content of leukocytes associated with EOS.

Endometrial fluid aspirate or endometrial lavage, containing secretions from eutopic endometrium, is increasingly believed to contain potential markers of EOS. However, >90% of proteins identified in lavage fluid are derived from serum, and they mask the less abundant proteins (Hannan et al. 2009b). Following immunodepletion of major serum proteins from aspirates of women with early or advanced EOS and those without disease, 31 discriminatory proteins were identified (Ametzazurra et al. 2009). One of these was glycodelin, a well-established product of the endometrial epithelium, with contraceptive and immunosuppressive
properties (Seppala et al. 1998). Consistent with this finding were the results of other studies demonstrating reduced glycodelin gene expression (Kao et al. 2003) and protein levels in tissues (Fowler et al. 2007). Very recently, 2DE of urine from women with or without EOS detected one highly up-regulated protein (five isoforms of cytokeratin 19) associated with the disease (Tokushige et al. 2010). While further investigation of this finding is needed, it would indeed be exciting should such an abundant urine marker prove to be useful for early detection of EOS and for prognostic evaluation following surgery. These differentially expressed proteins need to be fully validated to determine their cellular source, their specificity and sensitivity for detection of EOS, and their potential to differentiate between subsets of the disease in large cohorts of women.

**Endometrial receptivity and the diagnosis of infertility**

The human endometrium is receptive for implantation of a blastocyst for only 4–5 days of each menstrual cycle. Failure of implantation is a major reason for infertility in women, and the inability to achieve a state of receptivity is responsible for much of the failure of reproductive technologies (Salamonsen et al. 2009). In addition to the ‘normal’ disturbance of critical regulators in women presenting with infertility is the disruptive influence of hormonal treatment on women before embryo transfer in an IVF cycle. The current lack of a robust test for endometrial receptivity impairs clinicians’ ability to diagnose infertility of endometrial origin, confounding decision-making in the application of artificial reproductive technologies – for example, whether or not to transfer an embryo in any particular cycle. Gene array analyses have identified genes that are regulated across the menstrual cycle (Giudice 2004, Ponnampalam et al. 2006) and in women undergoing IVF (Horcajadas et al. 2007), but no clear genetic markers of the receptive phase have emerged. New proteomics technologies provide a clear opportunity for biomarker discovery aimed at improving patient outcomes from IVF.

Proteomic analysis of cycling human endometrium by DIGE identified 41 different gene products as differentially regulated between the mid-proliferative (non-receptive) and mid-secretory (receptive) phases of the menstrual cycle (Chen et al. 2009). Validation of three of these by immunohistochemistry defined their cellular source of origin as being epithelial (Rho-DGila and C1C1) or stromal (PGRMC1). A key finding of this study was that the protein abundance changes generally did not reflect published gene expression changes (Chen et al. 2009), a finding confirmed in other studies of endometrium (Stephens et al. 2010). In another study using DIGE to examine biopsy material from fertile women on day 2 (non-receptive) or day 7 (receptive) after the LH surge, only two proteins (Annexin A2 and Stathmin 1) were consistently different (Domínguez et al. 2009). Furthermore, in a functional model of refractoriness at LH+7 due to the presence of an intrauterine device (IUD), the proteins remained at levels observed in the LH+2 samples, but the receptive pattern was regained 3 months after IUD removal. PGRMC1 levels were also decreased in association with receptivity, as observed by Chen et al. (2009).

Interestingly, most of the differentially expressed proteins identified using gel-based methods were different from those identified in gel-free studies of endometrial tissues using isotope-coded affinity tags (ICAT) and online MS analysis of similar endometrial tissues (deSouza et al. 2005b). Indeed, in this study only two proteins showed unquestionable differential expression in the secretory phase: -glutamate NMCA receptor subunit z 1 precursor and FRAT1. While the focus of the studies was different, this demonstrates the need to use overlapping techniques afforded by gel-based and gel-free proteomic approaches for biomarker discovery.

The identification of only a very few proteins as differentially expressed in gel-based studies reflects the dominance of abundant structural and housekeeping proteins present in endometrial tissues. Pre-fractionation of such samples is therefore required to remove the most abundant proteins and reveal those of low abundance.

The endometrial secretome offers an alternate sample choice of far less complexity than the tissues, since it contains no structural proteins and can be easily pre-fractionated to remove abundant plasma proteins (Hannan et al. 2009b). Until recently, little was known of the composition of the endometrial secretome. Three recent studies (Casado-Vela et al. 2009, Hannan et al. 2009a, Scotchie et al. 2009) have begun to address this issue. Casado-Vela et al. (2009) applied three complementary proteomics strategies to analyse uterine aspirates from women in the secretory phase: HPLC tandem MS (MS/MS), denaturing one-dimensional gel electrophoresis followed by HPLC/MS/MS and 2DE followed by MALDI-TOF MS. In total, 803 different proteins were identified in aspirates; of these, several of the 22 most abundant were serum proteins. Among the known endometrial secreted proteins found in the aspirate were IL18, MMP9, mucins-1 and -16, vitamin D-binding protein and glycodelin (Casado-Vela et al. 2009). However, changes with the menstrual cycle phase or in infertile women were not examined.

As an alternate approach, lavage of the uterine cavity with saline has also been used as a sample type to identify secretory proteins in the uterine cavity. Given the abundant glycocalyx on the luminal endometrial surface, it might be expected that lavage could isolate proteins loosely bound to the highly glycosylated surface that would not be observed by simple aspirate. Scotchie et al. (2009) analysed lavage fluid following depletion of abundant serum proteins and compared
the secretome on days LH + 4 and LH + 9 (early and mid-secretory phases; Scotchie et al. 2009). Analysis by DIGE identified a total of 152 unique proteins (from a total of 468 proteins observed, including multiple fragments or derivatives of the same parent protein), including 64 proteins not being previously detected in the endometrium. Interestingly, a number of proteins thought to be critical to implantation and present in uterine fluid were not detected, suggesting very low relative abundance of these proteins. Some of the same proteins including α1-antitrypsin and transferrin were also found to be increased in endometrial tissues, uterine fluid and secretions from mid-secretory phase tissues in another study using 2DE (Parmar et al. 2009).

A more extensive study, applying DIGE to uterine lavage fluid following immunodepletion of abundant serum proteins, has identified differences between receptive and non-receptive states in fertile and infertile women (Hannan et al. 2009a). Seven proteins were significantly increased in the mid-secretary phase compared with the mid-proliferative phase: these included three isoforms of α2-macroglobulin (A2M), α1-antitrypsin, and activin receptor type-2B. When comparisons were made between mid-secretory phase fluid from fertile versus infertile women, 6 proteins were significantly reduced and 12 were increased in the infertile cohort. Interestingly, different isoforms of some proteins (α1-antichymotrypsin) both increased and decreased suggesting alternate processing or regulatory mechanisms. Immunohistochemistry on tissues from the same women demonstrated that antithrombin III (ANT3) was present in both epithelial and decidualised stromal cells, and showed abundance changes in agreement with proteomic findings. Likewise, A2M was elevated in epithelial cells in endometrium from infertile versus fertile women and in proliferative phase tissues.

These studies demonstrate the complexity of the human uterine secretome during the period in which endometrial receptivity is attained, and that the secretome is altered in women with unexplained infertility. Many of the regulatory proteins anticipated to be present in uterine fluid were not detected. However, when either lavage fluid (Hannan et al. 2009a) or aspirate (Boomsma et al. 2009) was subjected to multiplex analysis for a wide range of cytokines and chemokines, both classes of molecules were detected and varied according to cycle stage. While different analytes were examined in these two studies, the range of concentrations (from <1 to 1000 pg/ml) were similar, and both groups observed differences between cycle phases and between infertile and fertile women. The combination of such sensitive techniques for known molecules with the broader range of proteins identified using global proteomics technologies is likely to provide useful markers of infertility in the near future.

New and emerging technologies for the proteomic analysis of reproductive disease

A number of new proteomics technologies and strategies exist that have not yet been widely applied to reproductive disease research. A recurring problem for proteomics studies is large dynamic range of protein abundances observed in plasma and other biological fluids containing serum-derived proteins. Similarly, the analysis of tissue samples is also hampered by the presence of highly abundant structural and housekeeping proteins. New strategies for the enrichment of less abundant proteins are required in order to facilitate the efficient discovery of ‘low abundance’ biomarkers. Both existing and emerging technologies for the detection and quantitation of new biomarkers promise to further our understanding of reproductive biology, and are easily applicable to the reproductive disease research.

While traditional biochemical approaches continue to be the most widely applied strategies for sample fractionation (Table 1), new approaches are emerging that offer innovative solutions to the analysis of low abundance proteins from complex biological samples. One novel application involves the use of dual function core-shell hydrogel nanoparticles to capture and concentrate small, low abundance proteins from biological fluids (Luchini et al. 2008, Rainczuk et al. 2010). These nanoparticles perform both selective charge-based binding and size exclusion chromatography in a single batch step (Fig. 2); this not only reduces variation due to sample handling, but is also very efficient at capturing small, low abundance proteins. The combination of nanoparticles followed by gel-free proteomic approaches provides an excellent fractionation strategy for plasma and other complex reproductive samples. The approach is also rapid, cheap and easily accessible to any laboratory.

The analysis of differential protein expression in complex biological samples requires strategies for rapid, highly reproducible and accurate protein quantitation. SELDI TOF MS, so commonly used in studies of reproductive diseases, does not provide the protein identities necessary for subsequent validation studies. Accordingly, some of the most commonly applied approaches for quantitation and identification of biomarkers include the use of isotopic (ICAT, SILAC), fluorescent (DIGE) and other labelling techniques (Table 1). Somewhat surprisingly, few studies have used an approach termed isobaric tags for relative and absolute quantification (iTRAQ). In this system, isobaric mass tags are used to label N-termini and lysine side chains of peptides prior to fractionation; up to eight different samples may be combined prior to analysis. Both protein identities and quantitative information are obtained by MS in a single experiment, generating very large, information-rich datasets rapidly. A wider variety of
proteins (such as very hydrophobic/hydrophilic, small or large proteins) can be assessed using this strategy, compared to other approaches (for example, 2DE). The quantitative and comprehensive nature of the iTRAQ approach makes it a desirable high-throughput alternative, and we expect that it will become more widely applied by researchers in the field.

The follow-up of candidate biomarkers, identified by proteomics strategies, is a key requirement for translation of potential biomarkers into clinical diagnostics. Immunoassays, western blot, immunohistochemical staining and real-time PCR have been extensively applied in this context. New multiplex immunoassay technologies such as Luminex or AlphaScreen (reviewed in Vignali (2000) and Taouji et al. (2009)), permitting measurement of multiple analytes in a single ELISA format, offer an ideal complement to proteomics studies requiring validation of multiple candidates (Dunbar & Jacobson 2000). Commercially available kits allow the simultaneous analysis of up to 42 different cytokines in a single assay (Richens et al. 2010). Alternately, researchers may couple antibodies against candidates of interest to the beads used in each system.

Targeted MS-based assays are also becoming more widely used for biomarker validation. Multiple reaction monitoring (MRM) allows the detection and quantitation of specific peptides in a complex sample by measuring known fragment ions in a triple-quadrupole MS (Huttenhain et al. 2009). With the theoretical ability to multiplex large numbers of analytes in a single sample, MRM provides an alternative for validation where traditional ELISA-based assays are not possible – for example, to detect a specific modification on a peptide or protein. The methods and technologies involved in the application of MRM have been recently reviewed (Huttenhain et al. 2009, Schmidt et al. 2009).

An exciting new technology spanning both discovery and validation is the use of MALDI-IMS, which facilitates the direct analysis of protein distribution and abundance in thin tissue sections (Cornett et al. 2007). In IMS, proteins are directly ionised from the surface of a tissue sample; a characteristic histological pattern is derived for hundreds or thousands of individual proteins simultaneously, which may be further investigated and identified in the same tissue section. IMS has already been applied to generate in situ proteome profiles at implantation sites in mice (Burnum et al. 2008), and to analyse ovarian cancer tissue sections to identify proteins at tumour margins (Lemaire et al. 2007, Kang et al. 2010, Meriaux et al. 2010). In particular, Meriaux et al. (2010) recently used IMS to identify five distinct lipid regions associated with specific cancer morphology. Correlation with the known fatty acid compositions of ceramide and sphingomyelin isolated from ovarian tissues (Dyatlovitskaya et al. 1997, Rylova et al. 1998) led them to suggest the existence of a cytokine–lipid pathway linked to resistance to apoptotic stimuli (Meriaux et al. 2010). The results from this study provided novel mechanistic insights into ovarian cancer pathogenesis. Accordingly, IMS could easily be applied in other reproducative biomarker research, facilitating better classification of disease lesions and their surrounding margins, characterisation of underlying molecular pathogenesis or the discovery of new markers of disease.

**Conclusion**

Reproductive diseases in women represent a significant global health burden, and proteomic approaches offer one avenue to discover new molecules allowing more sensitive and specific detection or diagnostic strategies. For successful biomarker studies, the choice of sample is critical in determining downstream strategies. Existing studies strongly suggest that the choice of biological sample needs to be extended beyond plasma or tissues; alternatives such as uterine lavage or peritoneal fluids are increasingly being...
applied in the context of reproductive diseases. Highly standardised techniques for collection, processing and storage are also essential. The dynamic range of protein abundances inherent to biological systems also necessitates that suitable fractionation is undertaken, to decrease the complexity of samples prior to analysis. Improvements in sample pre-fractionation and preparation, combined well-designed study hypotheses and rigorous experimental procedures will continue to generate new potential markers of the reproductive diseases.

While existing studies have identified large numbers of potential biomarkers for a number of clinically important diseases of the female reproductive tract, there remains ample scope to apply both existing and emerging proteomics technologies in the context of reproductive disease research. New technologies are improving our ability to detect and quantitate changes at very low levels in biological samples, furthering insight into the unique molecular and physiological changes underlying disease development and progression. Nevertheless, robust validation of the observed changes is also required for the translation of potential biomarkers into clinically relevant, diagnostic and prognostic tools. Such studies require not only confirmation of changes in a limited sample size; large-scale cohort studies and appropriate statistical testing are essential. New approaches for validation, enabling higher throughput approaches than previously possible, will make this a more feasible strategy.

The application of proteomics technologies has not yet provided widely applicable clinical tools for the diagnosis of reproductive diseases. Nevertheless, good progress is being made, and rapidly emerging technologies for protein detection, coupled with pre-fractionation techniques and developments in bio-informatics, are now enabling deeper penetration of the proteome of tissues and biological fluids. The meticulous application of these new techniques to the most appropriate samples will ultimately provide the ‘holy grail’ of reliable, sensitive and specific markers of reproductive diseases.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the review reported.

Funding

Work in the Salamonsen laboratory is supported by the Ovarian Cancer Research Foundation.

www.reproduction-online.org

References


Blair AR & Casas CM 2009 Gynecologic cancers. Primary Care 36 115–130.


Proteomics and the search for biomarkers


Received 17 May 2010
First decision 23 June 2010
Accepted 13 July 2010


