Human Reproduction, Vol.27, No.7 pp. 2020–2029, 2012
Advanced Access publication on May 3, 2012 doi:10.1093/humrep/des127

human reproduction

ORIGINAL ARTICLE *Infertility*

Combined mRNA microarray and proteomic analysis of eutopic endometrium of women with and without endometriosis

A. Fassbender[1], N. Verbeeck[2,3,4], D. Börnigen[2,4], C.M. Kyama[1,5], A. Bokor[1,6], A. Vodolazkaia[1], K. Peeraer[1], C. Tomassetti[1], C. Meuleman[1], O. Gevaert[2,7], R. Van de Plas[8], F. Ojeda[2,3], B. De Moor[2,4], Y. Moreau[2,4], E. Waelkens[3,9], and T.M. D'Hooghe[1,10,*]

[1]Department of Obstetrics and Gynecology, Leuven University Fertility Center, UZ Gasthuisberg, 3000 Leuven, Belgium [2]Department of Electrical Engineering-ESAT, SCD-SISTA, KU Leuven, 3001 Leuven, Belgium [3]SYBIOMA, Facility for Systems Biology Based Mass Spectrometry, O&N2, KU Leuven, 3000 Leuven, Belgium [4]IBBT Future Health Department, Kasteelpark Arenberg 10, Box 2446, 3001 Leuven, Belgium [5]Department of Medical Lab Sciences, Institute of Tropical Medicine and Infectious Diseases, Jomo Kenyatta University of Agriculture & Technology, P.O. Box 62000-00200 City Square, Nairobi Kenya [6]Department of Obstetrics and Gynecology, Semmelweis University, Budapest, Hungary [7]Stanford University School of Medicine, Center for Cancer Systems Biology (CCSB) & Department of Radiology, Stanford, California [8]Vanderbilt University School of Medicine, Dept. of Biochemistry, Mass Spectrometry Research Center, Nashville, TN 37232, USA [9]Department of Cellular and Molecular Medicine, KU Leuven, O&N1, Herestraat 49 - bus 901, 3000 Leuven, Belgium [10]Division of Reproduction, Institute of Primate Research, PO BOX 24481, 00502 Karen, Nairobi, Kenya

*Correspondence address. Tel: +32-16-343624; Fax: +32-16-344368; E-mail: thomas.dhooghe@uz.kuleuven.ac.be

Submitted on November 17, 2011; resubmitted on March 7, 2012; accepted on March 15, 2012

BACKGROUND: An early semi-invasive diagnosis of endometriosis has the potential to allow early treatment and minimize disease progression but no such test is available at present. Our aim was to perform a combined mRNA microarray and proteomic analysis on the same eutopic endometrium sample obtained from patients with and without endometriosis.

METHODS: mRNA and protein fractions were extracted from 49 endometrial biopsies obtained from women with laparoscopically proven presence ($n = 31$) or absence ($n = 18$) of endometriosis during the early luteal ($n = 27$) or menstrual phase ($n = 22$) and analyzed using microarray and proteomic surface enhanced laser desorption ionization-time of flight mass spectrometry, respectively. Proteomic data were analyzed using a least squares-support vector machines (LS-SVM) model built on 70% (training set) and 30% of the samples (test set).

RESULTS: mRNA analysis of eutopic endometrium did not show any differentially expressed genes in women with endometriosis when compared with controls, regardless of endometriosis stage or cycle phase. mRNA was differentially expressed ($P < 0.05$) in women with (925 genes) and without endometriosis (1087 genes) during the menstrual phase when compared with the early luteal phase. Proteomic analysis based on five peptide peaks [2072 mass/charge (m/z); 2973 m/z; 3623 m/z; 3680 m/z and 21133 m/z] using an LS-SVM model applied on the luteal phase endometrium training set allowed the diagnosis of endometriosis (sensitivity, 91; 95% confidence interval (CI): 74–98; specificity, 80; 95% CI: 66–97 and positive predictive value, 87.9%; negative predictive value, 84.8%) in the test set.

CONCLUSION: mRNA expression of eutopic endometrium was comparable in women with and without endometriosis but different in menstrual endometrium when compared with luteal endometrium in women with endometriosis. Proteomic analysis of luteal phase endometrium allowed the diagnosis of endometriosis with high sensitivity and specificity in training and test sets. A potential limitation of our study is the fact that our control group included women with a normal pelvis as well as women with concurrent pelvic disease (e.g. fibroids, benign ovarian cysts, hydrosalpinges), which may have contributed to the comparable mRNA expression profile in the eutopic endometrium of women with endometriosis and controls.

Key words: eutopic endometrium / surface enhanced laser desorption ionization-time of flight mass spectrometry / microarray / endometriosis / biomarker

Introduction

Endometriosis is defined as the presence of endometrial gland and stroma tissue outside the uterine cavity. Endometriosis affects 6–10% of women of reproductive age in the general population; however, its prevalence is 35–50% in women with pain, infertility or both (Sensky and Liu, 1980; Houston, 1984; Cramer and Missmer, 2002; Giudice and Kao, 2004). At present, the only way to conclusively diagnose endometriosis is through laparoscopic inspection, preferably with histological confirmation (Kennedy *et al.*, 2005). This contributes to the diagnostic delay of endometriosis between the onset of symptoms and a diagnosis of 8–11 years (Hadfield *et al.*, 1996; Arruda *et al.*, 2003; Husby *et al.*, 2003). In a clinical practice dealing with women with subfertility with or without pain, a non-invasive test of endometriosis with high sensitivity would allow us to identify those women with endometriosis who could benefit from laparoscopic surgery reported to improve these symptoms, i.e. increase fertility and decrease pain if present preoperatively (Kennedy *et al.*, 2005; D'Hooghe *et al.*, 2006). As published evidence suggests that endometriosis can be progressive in 50% (D'Hooghe and Debrock, 2002), early non-invasive diagnosis of cases has the potential to allow early treatment and prevent progression. Ideally, normalization of the plasma levels of endometriosis biomarkers during/after treatment would also correlate with decreased pelvic pain and increased fertility.

Currently, such a test is not available for the diagnosis of endometriosis (Kennedy *et al.*, 2005; May *et al.*, 2010).

Several investigators have reported differential gene expression in the eutopic endometrium from women with and without endometriosis using mRNA microarray analysis (Kao *et al.*, 2003; Absenger *et al.*, 2004; Burney *et al.*, 2007; Sherwin *et al.*, 2008). The current study was performed to test the hypothesis that differential gene and protein expression in the menstrual and early luteal endometria between women with endometriosis and controls could be used to allow a better understanding of the pathogenesis of endometriosis and the development of a semi-invasive diagnostic test for endometriosis. We selected menstrual endometrial samples because (retrograde) menstruation is considered to be a key event in the pathogenesis of endometriosis and these samples have not been studied before using proteomics and mRNA microarrays in the context of endometriosis. We also selected endometrial samples obtained during the early luteal phase because our previous research indicated that analysis of the late luteal transcriptome of an endometrium sample is not likely to form the basis of a minimally invasive diagnostic test for endometriosis (Sherwin *et al.*, 2008).

Materials and Methods

This study (microarray and proteomic analyses) was not a blinded study.

Patients and samples

All endometrium samples used for this study ($n = 49$) had been collected previously (retrospective study) using a pipelle sampler (Pipelle Laboratories CCD, Paris, France) from women undergoing laparoscopies for infertility and/or pain and had been frozen in our biobank at −80°C (Table I). All patients had signed a written informed consent prior to recruitment and the study protocol was approved by the institutional ethical and review board of University Hospital Gasthuisberg, Leuven,

Table I Clinical characteristics of study population.

	All study population ($n = 49$)	Controls ($n = 18$)	Endometriosis ($n = 31$)	Stages I–II ($n = 16$)[a]	Stages III–IV ($n = 15$)[a]
Age (years)					
Mean (SD)	31.96 ± 4.54	33.83 ± 5.07	30.87 ± 3.89	30.63 ± 4.03	31.13 ± 3.52
Median (range)	32 (24–43)	33.5 (24–43)	30 (24–41)	31 (24–41)	31 (24–39)
n with both infertility and pain	27	9	18	6	12
n with subfertility only	19	8	11	8	3
n with pain only	3	1	2	2	
Type of pain					
Dysmenorrhoea (*n*)	27	9	18	6	12
Dyspareunia (*n*)	11	4	7	4	3
Chronic pelvic pain (*n*)	7	2	5	2	3
Dyschezia (*n*)	3	1			2
Cycle information					
Regular cycle (*n*)	35	12	23	10	13
Irregular cycle (*n*)	10	4	6	6	
Controls with a normal pelvis		9			
Controls with other pelvic pathology		9 (adhesions, PID, hydrosalpinx, ovarian cyst)[b]			

[a]Stages I–II refer to minimal–mild and Stages III–IV moderate–severe endometriosis.
[b]Hydrosalpinx (1x); pelvic inflammatory disease (PID) related adhesion (1x); adhesion not related to PID (2x); ovarian cyst (3x); parasalpingeale cysten (2x).

Belgium. Endometrial samples from patients using the oral contraceptive pill (combined or progesterone only), patients on chronic medication and patients operated within 6 months prior to the time of sample collection were excluded.

Initially, we aimed to study 10 endometrium samples in each subgroup, resulting in a total of 60 samples of women with and without endometriosis. However, owing to time constraints (Dr Fassbender A. Phd defense deadline) we were only able to use 49 endometrial samples (collected for our biobank between 2005 and 2009) from 31 women with endometriosis [menstrual ($n = 14$) and early luteal) ($n = 17$) phases] and from 18 controls without endometriosis [menstrual ($n = 8$) and early luteal ($n = 10$) phases].

All women with endometriosis ($n = 31$) were classified as either minimal–mild (Stage I–II, $n = 16$) or moderate–severe (Stage III–IV, $n = 15$) according to the most recent classification by the American Society of Reproductive Medicine (ASRM, 1997). The absence of endometriosis was documented by laparoscopy in the 18 control patients.

Extraction of mRNA and microarray preparation

Total RNA was isolated from endometrium tissue samples using TRIzol reagent (Invitrogen, Carlsblad, CA, USA) following the manufacturer's protocol (Fig. 1). The RNA preparations were then DNase treated and purified using the RNeasy Mini kit (Qiagen, Valencia, CA, USA). Samples were stored in RNase-free H_2O and the RNA concentration and purity were determined spectrophotometrically using Nanodrop ND-1000 (Nanodrop Technologies) and RNA integrity was assessed using a Bio-analyzer 2100 (Agilent, Santa Clara, CA, USA). Using the Ambion WT Expression Kit, per sample 100 ng of total RNA, spiked with bacterial poly-A RNA positive control (Affymetrix, Santa Clara, CA, USA), was converted to double-stranded cDNA in a reverse transcription reaction. Subsequently, the sample was converted and amplified to antisense cRNA in an *in vitro* transcription reaction and was converted to single-stranded sense cDNA. Finally, samples were fragmented and labeled with biotin in a terminal labeling reaction, according to the Affymetrix WT Terminal Labeling Kit. A mixture of purified and fragmented biotinylated cDNA and hybridization controls (Affymetrix) was hybridized on Affymetrix GeneChip Human Gene 1.0 ST Arrays followed by staining and washing in a GeneChip® fluidics station 450 (Affymetrix), according to the manufacturer's procedures. To assess the raw probe signal intensities, chips were scanned using a GeneChip® scanner 3000 (Affymetrix).

Preparation of endometrium samples for SELDI-TOF MS

As highly abundant proteins, such as hemoglobin, are expressed at high levels in endometrium (Dassen *et al.*, 2008), depletion of these proteins is needed first to allow detection of less abundant proteins more relevant for the discovery of new peptides (Fig. 1). In a pilot study we compared

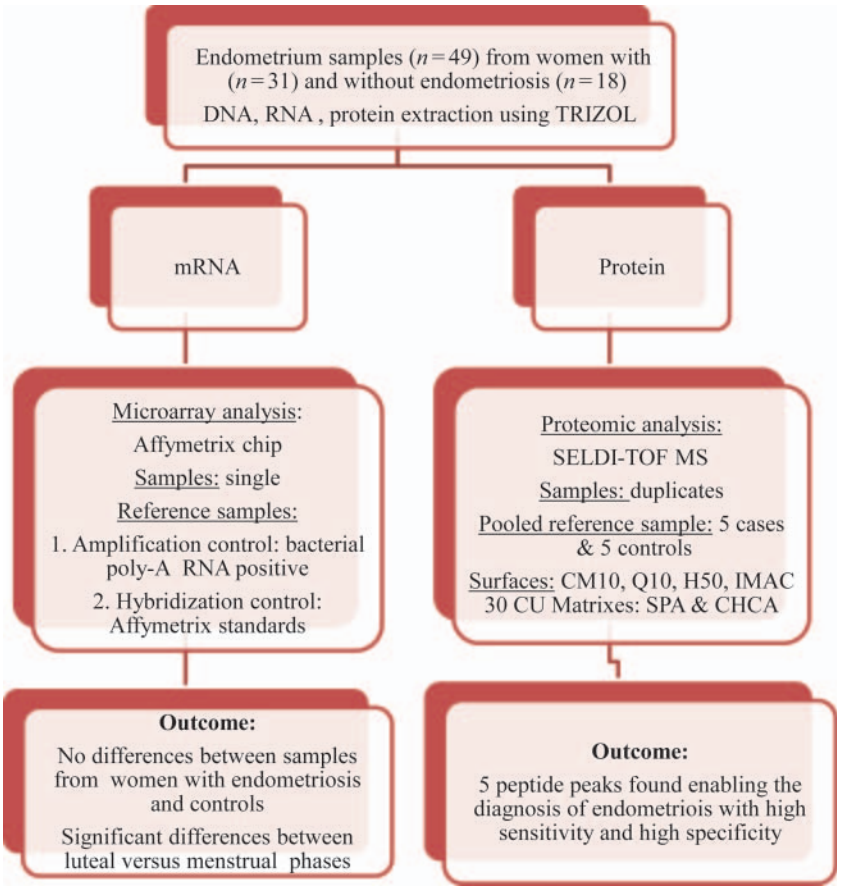

Figure 1 Workflow of the experiment. This workflow diagram represents the following: sample size, technique used and outcome of the studies. SELDI-TOF MS, surface enhanced laser desorption ionization-time of flight mass spectrometry.

two different depletion methods: Cibacron Blue method and Proteominer kit (Bio-Rad, Hercules, CA, USA) using endometrial samples. Proteominer kit (Bio-Rad) resulted in a more enriched surface enhanced laser desorption ionization-time of flight mass spectrometry (SELDI-TOF MS) spectra and in less sample loss. According to manufacturer's instructions, we added 500 μl of endometrium sample to the column. The elution contained the protein or peptide of interest and was stored at −80°C until further analysis. The eluted fraction was screened using SELDI-TOF MS on four different surface chemistries (CM10, Q10, IMAC 30 and H50).

Calculation of intra- and inter-individual coefficient of variation

In order to enhance reproducibility across the different chemical surfaces used, a reference sample was spotted in duplicate on each type of surface to calculate experimental intra- and inter-coefficient of variation (CV) and to optimize the array reading parameters (such as laser intensity, focus mass and mass range). The reference sample was taken from a pool (5 ml) of plasma samples (500 μl per patient) from five women with endometriosis and from five controls without endometriosis, selected at random.

Profiling of endometrium proteins on the SELDI-TOF MS

SELDI-TOF MS employs 8–16 spot chips and each spot contains a solid-phase chromatographic surface for binding proteins under particular binding conditions (Poon, 2007). There are several types of ProteinChip arrays with different chromatographic properties, including hydrophobic, hydrophilic, anion- and cation exchange and metal affinity. These properties enable them to capture different subsets of proteins according to their physicochemical properties (Poon, 2007).

To increase the number of detectable proteins, four different chip surfaces with distinct chromatographic properties and binding affinities were used. Briefly, ProteinChip array spots of IMAC 30 were first twice pre-activated with copper sulfate solution for 5 min on a shaker. The surface was washed three times with Milli-Q water, a neutralizing solution (Sodium acetate pH 4.0) was added and incubated for 5 min and the surface was rinsed with Milli-Q water twice. The surface was then equilibrated with 150 μl binding buffer (0.1 mol/l phosphate, 0.5 mol/l NaCl, pH 7.0) while shaking for 5 min at room temperature. H50 was first pre-activated by applying 5 μl of 50% acetonitrile and incubating in a humid chamber for 5 min, then equilibrated twice in 150 μl of binding buffer (10 acetonitrile, 0.1% trifluoroacetic acid (TFA)) while shaking for 5 min at room temperature. ProteinChip array spots of CM10 and Q10 were equilibrated with 150 μl binding buffer (CM10: 50 mM NaOAC, pH 4.0; Q10: 50 mM Tris HCl, pH8.0) while shaking for 5 min at room temperature to pre-activate binding surfaces. For all four surfaces (CM10, Q10, H50 and IMAC 30 Cu) the buffer was removed and 10 μl of each individual sample (15 μg per spot), diluted with surface-type dependent binding buffer were loaded onto each spot in duplicate and incubated for 30 min at room temperature while being shaken. The unbound proteins/peptides on the ProteinChip array surfaces were washed away with appropriate buffer (see above) three times for 5 min while being shaken, rinsed twice in 150 μl Milli-Q water and air dried. The cassette was then centrifuged (upside down lying on the Whatman paper) at 1000*g* for 2 min. Mass spectra of the retained proteins were obtained by ionizing the proteins using two types of energy-absorbing molecule: alpha-cyano-4-hydroxy cinnamic acid (CHCA) for the small molecules (<15 kDa), and sinapinic acid (SPA) for the larger molecules (both CHCA and SPA were obtained from Bio-Rad). CHCA (1 μl of 20% solution) was applied twice to the retained proteins on the spots. Fifty percent of SPA was applied in two consecutive steps in volumes of 1 μl.

Analyses of the retained peptides/proteins were performed with a ProteinChip System, Series 4000 SELDI-TOF-MS instrument (Bio-Rad). Mass accuracy was calibrated externally with the all-in-one peptide molecular mass standard (Bio-Rad) for the mass range of 1.6–20 kDa and with the all-in-one protein molecular mass standard (Bio-Rad, USA) for the mass range of 8–150 kDa.

Data analysis

mRNA microarray analysis

The mRNA microarray analysis was based on the robust multichip average (RMA) expression values which were obtained with the xps package 1.7.2 (Stratowa, 2003) of Bioconductor (http://www.bioconductor.org; Gentleman *et al*., 2004). We compared the RMA expression values of the different conditions via the moderated *t*-statistic with the limma package 3.2.1 of Bioconductor (Smyth, 2004, 2005). The moderated *t*-statistic applies an empirical Bayesian strategy to compute the gene-wise residual SDs and thereby increases the power of the test, especially beneficial for smaller data sets. The resulting *P*-values were corrected for multiple testing with Benjamini and Hochberg (1995) to control the false discovery rate. We selected probe sets with an absolute fold change >2 and a corrected *P*-value <0.05. Functional annotation, such as the Database for Annotation, Visualization and Integrated Discovery (DAVID), was used to analyze pathway processes (http://david.abcc.ncifcrf.gov). DAVID provides a comprehensive set of functional annotation tools for investigators to understand the biological meaning behind large lists of genes.

Proteomics analysis

The SELDI-TOF mass spectra were baseline corrected and normalized on the basis of total ion current using the ProteinChip data manager software 3.5 (Bio-Rad, Hercules, CA, USA) and smoothed using a least-squares polynomial filter in Matlab 7 (Natick, MA, USA).

Differentially expressed mass peaks with *P*-value >0.15 were removed from the analysis. Data were analyzed using custom scripts written in Matlab (Natick).

For each of the different conditions (namely chip type-matrix type-cycle phase), three different setups were examined: control versus minimal–mild, control versus moderate–severe and control versus minimal–severe. For each of the three different setups, a 'combined' spectrum was calculated combining relevant mass over charge (*m*/*z*) values from each of the two conditions being compared. Peak picking was then performed on this combined spectrum. The peaks were quantified using the peak height. Peaks <2 kDa were excluded from data analysis, as recommended by the manufacturing company (Bio-Rad) and previous investigators (Ding *et al*., 2010; Wang *et al*., 2010).

In order to obtain robust results, the analysis was performed using repeated random subsampling cross-validation (Bouckaert and Frank, 2004). The random subsampling method, or Monte Carlo cross-validation, is an alternative to *k*-fold cross-validation and can be found in many statistics books (e.g. Hastie and Friedman, 2009). It essentially makes what one would call in classical statistics 'combinations' of the samples to divide them into 'training' and 'test' sets. First, the samples were randomly split into two stratified parts, the training set (70% of the total samples) and the test set (30% of the total samples). Data obtained from the training set were used to identify a pattern that discriminates between the presence and absence of disease. Potential biomarkers selected from the training set were evaluated in the test set (30%) to determine sensitivity and specificity. For our repeated random subsampling, this random split

into training and test set was repeated 100 times, as this method has been reported to produce reproducible results (Bouckaert and Frank, 2004).

The presented algorithm is a feature selection algorithm, aiming at the selection of those features, i.e. mass spectrometry peaks, which best discern between endometriosis and control groups, because these peaks are the best for classification purposes. As the number of peaks that resulted from the peak picking was relatively high (on an average 130 peaks per condition), feeding these peaks directly into the feature selection algorithm would have two major disadvantages: very long calculation times owing to the high number of repeated subsamplings and model trainings, and increased risk of selecting a possibly good but suboptimal set of peaks. Therefore, we chose to eliminate those peaks with a *P*-value >0.15 that would not contribute to the selection of the optimal set of peaks. The *P*-value was determined using the Wilcoxon rank-sum test, testing whether peaks in the diseased samples were differentially expressed when compared with the control samples. Using this method, we were able to decrease the number of peaks fed to the algorithm by a factor of 7 to, on average, about 18 peaks per condition. The remaining peaks were then used to construct a least squares-support vector machines (LS-SVM) model, using leave one out cross-validation (LOOVC), which is a special case of *k*-fold cross-validation, where *k* is equal to the number of samples. LS-SVMs are well-known supervised machine learning algorithms, related to artificial neural networks, which show good classification properties (Suykens *et al.*, 2011). In our LS-SVM algorithm, each of the input peaks is ranked in terms of classification power, meaning that it will give the highest rank to the peak which can best distinguish the two groups.

The algorithm works in an iterative way. In short, it begins with the full set of peaks, reduces this to the best subset of 20 peaks and finally ends up with the best subset of 5 peaks for classification. In the first iteration, a model is constructed for each of the 100 splits, and for each of these 100 models the 5 highest ranked peaks are stored, resulting in 500 peptide peaks of interest. Out of these 500 peptide peaks, the 20 most frequently observed peaks are selected. In the second iteration, this subset of 20 peaks is used to construct a new LS-SVM model for each of the data obtained from the training set. Out of these 20 most frequently observed peaks, the 5 most frequently observed peaks were selected. In the final iteration this subset of five peaks is used to construct the final LS-SVM model in the data obtained from the training set. This model was applied to the data obtained from the test set (30% of samples) for validation in order to determine the sensitivity and specificity in each of the 100 splits. Finally, the average performance of the model was calculated over the 100 splits.

Results

Microarray results

When gene expression of endometrium was compared between endometriosis patients and controls, no significant differences were found, regardless of endometriosis stage or cycle phase (after multiple testing) as shown in Table II. In women with endometriosis, 471 and 454 genes were down-regulated and up-regulated, respectively ($P < 0.05$ after multiple testing), in the menstrual phase endometrium compared with the early luteal phase endometrium.

When comparing menstrual versus early luteal phase of control samples, 1087 genes were differentially expressed ($P < 0.05$; Table II). The up-regulated genes of menstrual phase endometrial samples compared with the luteal phase endometrial samples were shown to be involved in a pathway of wound healing, blood coagulation, hemostasis, chemotaxis and extracellular matrix as enriched categories. The down-regulated genes of menstrual phase endometrial samples compared with the luteal phase endometrial samples showed a pathway of carboxylic acid metabolic process, oxoacid metabolic process and cellular amino acid catabolic process as enriched categories.

Table II Number of significantly differentially expressed genes in the eutopic endometrium of women with endometriosis versus without endometriosis (control) after multiple testing.

Study population	Multiple testing, $P < 0.05$, log ratio < -1	Log ratio >1
Minimal–severe versus control	0	0
Minimal–mild versus control	0	0
Moderate–severe versus control	0	0
Moderate–severe versus minimal–mild	0	0
Menstrual minimal–severe versus control	0	0
Luteal minimal–severe versus control	0	0
Minimal–severe menstrual versus luteal	471	454
Control menstrual versus luteal	466	621
Menstrual minimal–mild versus control	0	0
Menstrual moderate–severe versus control	0	0
Luteal minimal–mild versus control	0	0
Luteal moderate–severe versus control	0	0

When menstrual phase samples were compared with early luteal phase samples, a total of 683 genes were differentially expressed in women with and without endometriosis (Supplementary data, Table SI), whereas 242 genes were differentially expressed only in women with endometriosis (Supplementary data, Table SII) and 404 genes were differentially expressed only in controls (Supplementary data, Table SIII). This differential gene expression was related to the following biological features: wound healing, blood coagulation, hemostasis, chemotaxis extracellular matrix, carboxylic acid metabolic process, oxoacid metabolic process and cellular amino acid catabolic process (both women with and without endometriosis, 683 genes); cellular fractions (membrane fraction, insoluble fraction and vesicular fraction) referring to disrupted cells (women with endometriosis only, 242 genes); and blood vessel morphogenesis, angiogenesis, blood vessel development and vasculature development (controls without endometriosis only, 404 genes). In women with endometriosis the 10 genes with the highest differential expression in the menstrual phase, when compared with the luteal phase, were associated with ion, cation and metal ion transport. In women without endometriosis,

the 10 genes with highest differential expression in the menstrual phase, when compared with the luteal phase, were associated with the regulation of cell proliferation, programmed cell death, apoptosis and organ development.

Proteomics results

Multivariate analysis was used to investigate our endometrium proteomics data. The range of differentially expressed peaks varied between 8 and 239 peaks, depending on chip type, matrix, stage of endometriosis and phase of the cycle. Using the IMAC CHCA data obtained from the training set, an LS-SVM model constructed using five peptide peaks in the luteal phase endometrium (2071 *m/z*; 2166 *m/z*; 2228 *m/z*; 3649 *m/z* and 40367 *m/z*) allowed the diagnosis of minimal–mild endometriosis [sensitivity, 94; 95% confidence interval (CI): 82–100; specificity, 100; 95% CI: 83–100 and positive predictive value (PPV), 100%; negative predictive value (NPV), 93.5%] in the data obtained from the test set. Using the CM10 SPA data obtained from the 'training set', an LS-SVM model constructed using five other peptide peaks in the luteal phase endometrium (3274 *m/z*; 7455 *m/z*; 13552 *m/z*; 39889 *m/z* and 42108 *m/z*) allowed the diagnosis of moderate–severe endometriosis (sensitivity, 92; 95% CI: 76–100; specificity, 84; 95% CI: 64–96 and PPV, 70.8%; NPV, 94.3%) in the data obtained from the test set. Using IMAC CHCA data obtained from the 'training set', an LS-SVM model constructed using five other peptide peaks in the luteal phase endometrium [2072 *m/z*; 2973 *m/z*; 3623 *m/z*; 3680 *m/z* and 21133 *m/z*] allowed the diagnosis of minimal–severe endometriosis (sensitivity, 91; 95% CI: 74–98; specificity, 80; 95% CI: 66–97 and PPV, 87.9%; NPV, 84.8%) in the data obtained from the test set.

We found an intra-assay CV of 9 and 10%, and inter-assay CV of 8 and 11% using the reference sample spotted on the CM10 SPA or IMAC CHCA SELDI surface, respectively.

Discussion

To the best of our knowledge, our study presents the largest mRNA microarray data analysis of menstrual endometrium of women with and without endometriosis and shows that mRNA expression is different, and more related to pathways of wound healing, blood coagulation, hemostasis, chemotaxis and extracellular matrix, in the menstrual endometrium than in the luteal endometrium in all study subjects, with or without endometriosis. The strength of our study is its design in accordance with the quality assessment of diagnostic accuracy studies guidelines (Whiting *et al.*, 2003; May *et al.*, 2010) with respect to control group, stages of endometriosis and cycle phase selection. In line with these guidelines, we selected our controls from women with symptoms consistent with endometriosis but without laparoscopic evidence of endometriosis (Whiting *et al.*, 2003; May *et al.*, 2010). A potential pitfall of our study design is the fact that our control group included women with a normal pelvis ($n = 8$) as well as women with concurrent non-endometriotic pelvic disease ($n = 10$, including fibroids, benign ovarian cysts, hydrosalpinges, etc.), which may have contributed to the lack of mRNA differences observed in the eutopic endometrium of women with endometriosis and controls, and on the sensitivity/specificity of the biomarkers identified in the proteomics section. In future research it would also be useful to include an asymptomatic control group of proven fertility (i.e. patients with tubal ligation) with or without endometriosis in order to test the hypothesis that mRNA/proteomics expression in the eutopic endometrium is affected by the absence or presence of pain and subfertility.

Our data are in line with earlier reports showing important differences in mRNA expression between the follicular and luteal phase eutopic endometrium, and between early, mid- and late luteal phases and across the entire menstrual cycle (Table III; Ponnampalam *et al.*, 2004; Punyadeera *et al.*, 2005; Talbi *et al.*, 2006; Burney *et al.*, 2007; Aghajanova and Giudice, 2010), to the extent that cycle phase had a more pronounced effect on mRNA expression of eutopic endometrium than the presence or absence of endometriosis (Aghajanova and Giudice, 2010). Interestingly, the expression of the fibrinogen beta chain gene in the eutopic endometrium was significantly increased in menstrual tissue when compared with the luteal endometrium in women without endometriosis but not in women with endometriosis. This observation is in line with data from our plasma proteomics study showing increased expression of fibrinogen beta chain in plasma from women without endometriosis when compared with those with endometriosis (Fassbender *et al.*, 2012). Decreased fibrinogen beta chain levels have also been observed in uterine flushings from baboons

Table III Overview of mRNA microarray studies of menstrual endometrium.

Reference	Sample number	Cycle phase	Microarray platform	Results
Punyadeera *et al.* (2005)	$n = 24$ (female volunteers) $n = 4$, pooled samples for microarray $n = 20$, validation with RT–PCR	Menstrual versus luteal phase	Affymetrix HU-133A chips	794 transcripts differentially expressed 1. Secretoglobin family 1D 2. MMP-1 ↑ Luteal phase
Ponnampalam *et al.* (2004)	$n = 43$, normal cycling women	Menstrual phase, early, mid-, late follicular and early, mid–late-luteal phases	Glass microarrray slides from Peter Mac Callum Cancer Institute	571 differentially expressed genes A strong relationship between the grouping of endometrial samples based on gene expression profiles and their histopathological cycle stages

with induced endometriosis when compared with controls (www.patentstorm.us/patents/7794958).

In our study we did not observe differential mRNA expression in women with and without endometriosis during the early luteal phase, as opposed to other investigators (Burney et al., 2007) who reported up- and down-regulation of many genes in the early luteal phase. Similar to our observation, no genes were differentially expressed in a consistent manner in the eutopic endometrium from patients with deep endometriosis compared with controls over the course of the menstrual cycle in both epithelial and stromal cells (Matsuzaki et al., 2005). In contrast, differential mRNA expression of the eutopic endometrium was observed in women with endometriosis when compared with controls during the mid-luteal (Kao et al., 2003) and late luteal phases (Sherwin et al., 2008) (Table IV). Microarray results reported in different studies can vary as a result of many factors, including patient selection, sample size and methodological details of array. In our current study, we corrected the data for multiple testing as in our previous study (Sherwin et al., 2008). However, this was not carried out in other studies (Kao et al., 2003; Matsuzaki et al., 2005) except for one study (Burney et al., 2007).

In the present study proteomic and mRNA microarray analysis of the same endometrial sample was performed after TRIzol treatment (Fassbender et al., 2010) but comparison between the two analyses was not possible owing to the lack of any differentially expressed genes between women with and without endometriosis after multiple testing.

Our proteomic endometrial analysis allowed the diagnosis of minimal–severe, minimal–mild and moderate–severe endometriosis with the use of five selected peaks in the range of (2–21 kDa), (2–40 kDa) and (3–42 kDa), respectively, with high sensitivity and specificity with a model based on the training set and validated in a test set.

The question might arise, why did we choose five peaks for diagnostic modelling? On the one hand, previous studies have shown that a panel of biomarkers can improve the sensitivity and specificity of a diagnostic test compared with the diagnostic performance of any single biomarker (Robin et al., 2009; May et al., 2010). On the other hand there is the risk of overfitting the created model to the data set when too many biomarkers are selected, especially when using a limited number of samples. Therefore, we decided to take into account multiple biomarkers, while keeping the maximum number of biomarkers capped at 5 to prevent overfitting. A list of five biomarkers limits the number of candidates for further investigation into the pathogenesis of endometriosis and can be used in a cost-effective way in a clinical application (e.g. enzyme-linked immunosorbent assay). When compared with previous research on endometriosis biomarkers using SELDI-TOF MS (Ding et al., 2010; Wang et al., 2010), our study has several strengths. First, our patient population was better characterized with respect to cycle phase. Previous investigators did not include any details of the cycle phase (Ding et al., 2010; Wang et al., 2010) and included controls without laparoscopic information (Wang et al., 2010; Table V). Second, we constructed a statistical LS-SVM model based on 70% of our samples divided into training set and test set by repeated random sub-sampling cross-validation data set, which strongly increases the chance of finding a biomarker that is valid in new data sets (Xua and Liang, 2001).

Table IV Overview of mRNA microarray analysis of the eutopic endometrium in women with endometriosis compared with controls.

Reference	Sample number	Cycle phase	Microarray platform	Results
Current study, Fassbender et al. (2011)	$N = 49$ Minimal–mild ($n = 16$) Moderate–severe ($n = 15$) Controls ($n = 18$)	Early luteal phase ($n = 27$) and menstrual phase ($n = 22$)	Affymetrix	No genes differentially expressed in women with endometriosis compared with controls
Sherwin et al. (2008)	$N = 16$ eutopic endometrium Minimal–mild ($n = 5$) Moderate–severe ($n = 5$) Controls ($n = 6$)	Late luteal phase	Agilent	8 genes up-regulated >1.75 fold ($P < 0.01$) and 1 gene down-regulated
Burney et al. (2007)	$N = 37$ Moderate–severe ($n = 21$) Controls ($n = 16$)	Follicular ($n = 6$) Early luteal ($n = 6$) Mid-luteal ($n = 9$)	Affymetrix	87 transcripts were altered more than 4-fold such as FOXO1A, MIG6, CYP26A1
Matsuzaki et al. (2005)	$N = 24$ Minimal–severe ($n = 12$) Controls ($n = 12$) Extracted epithelial and stromal cells from women with and without endometriosis using laser capture microdissection	Late follicular ($n = 6$) Early, mid-, late luteal ($n = 18$)	Clontech Atlashuman Array 1.2 cDNA expression array	No gene was differentially expressed in a consistent manner in the eutopic endometrium
Absenger et al. (2004)	Endometriosis ($n = 43$) Controls ($n = 48$)	Follicular or luteal phase	Affymetrix	95 >1.5-fold ↓64 ↑ 31 ↑Cyr61 in the luteal phase
Kao et al. (2003)	$N = 20$ Mild–moderate ($n = 8$) Controls ($n = 12$)	Mid-luteal phase ($n = 20$)	Affymetrix	91 increased and 115 decreased more than 2-fold

Table V Overview of proteomic SELDI-TOF MS analysis of eutopic endometrium in women with endometriosis compared with controls.

Reference	Sample number	Cycle phase	Surface	Results kDa	Sensitivity (%)	Specificity (%)	Validation in a blinded study population
Current study, Fassbender *et al.* (2011)	$N = 49$ Minimal–mild ($n = 16$) Moderate–severe ($n = 15$) Controls ($n = 18$)	Early luteal phase ($n = 27$) and menstrual phase ($n = 22$)	IMAC 30; CM10; Q10; H50	Minimal–severe versus controls: 2072 *m/z*; 2973 *m/z*; 3623 *m/z*; 3680 *m/z*; 21133 *m/z* Minimal–mild versus controls: 2071 *m/z*; 2166 *m/z*; 2228 *m/z*; 3649 *m/z*; 40367 *m/z* Moderate–severe versus controls: 3274 *m/z*; 7455 *m/z*; 13552 *m/z*; 39889 *m/z*; 42108	91 94 92	80 100 84	No
Kyama *et al.* (2011)	$N = 29$ Minimal–mild ($n = 9$) Moderate–severe ($n = 10$) Controls ($n = 10$)	Luteal phase ($n = 29$)	CM10; H50; IMAC 30; Q10	Minimal–severe versus controls: 8.650 *m/z*; 8.659 *m/z*; 13.910 *m/z*; 5.183 *m/z*; 1.949 *m/z* Minimal–mild versus control: 1.924 *m/z*; 2.504 *m/z*; 90.675 *m/z*, *T-Plastin*; 39.956 *m/z*, *Annexin 5* Moderate–severe versus controls: 10.110 *m/z*, 5.828 *m/z*; 12.172 *m/z*; 4.279 *m/z*	89.5 100 80	90 100 70	No
Ding *et al.* (2010) Mitochondrial protein expression	$N = 53$ Minimal–mild ($n = 19$) Moderate–severe ($n = 5$) Controls ($n = 29$)	Not mentioned	CM10	Minimal–severe versus controls: 15.334 *m/z*; 15.128 *m/z*; 16.069 *m/z* Minimal–mild; moderate–severe versus controls not mentioned	87.5	86.2	No
Wang *et al.* (2010)	$N = 26$ Minimal–mild ($n = 8$) Moderate–severe ($n = 5$) Controls ($n = 13$)	Not mentioned	H4	Minimal–severe versus controls: 6.898 *m/z*; 5.891 *m/z*; 5.385 *m/z*; 6.448 *m/z*; 5.425 *m/z* Minimal–mild; moderate–severe versus controls not mentioned	91.7	90	No
Fassbender *et al.* (2010)	$N = 16$ Minimal–mild ($n = 5$) Moderate–severe ($n = 5$) Controls ($n = 6$)	Luteal phase ($n = 16$)	CM10; IMAC 30	Minimal–severe versus controls: 32 peaks differentially expressed Minimal–mild versus controls: 23 peaks Moderate–severe versus controls: 25 peaks			No
Kyama *et al.* (2006)	$N = 9$ Eutopic EM mild ($n = 3$) Paired eutopic EM and peritoneum and peritoneal endometriotic lesion controls ($n = 3$)	Luteal phase ($n = 9$)	CM10; H50; IMAC 30; Q10	Mild versus control: 2.8–12.3 kDa was 3–24 times lower in the eutopic endometrium of women with endometriosis than controls			No

Our model now needs validation in another study population but it is superior to previous studies analyzing SELDI-TOF MS data using ProteinChip data manager software (Bio-Rad, Hercules, CA, USA) and support vector machines (SVM) without sample randomization and without division into a training and test set, and without a separate validation step (Wang *et al.*, 2010). In our study, we used LOOVC only for the analysis of data obtained from the training set in order to build our model, as this method of cross-validation has the tendency of overestimating the performance of the constructed model (Xua and Liang, 2001). The true model performance was validated in data obtained from the test set data set, in contrast with other investigators (Ding *et al.*, 2010) who used LOOCV on the entire data set, which could have led to an overestimation of the performance of the model.

Comparison of our current study with data from our three previously published endometrium SELDI-TOF MS papers (Kyama *et al.*, 2006; Fassbender *et al.*, 2010; Kyama *et al.*, 2011) revealed that the same 58 peaks (2–33 kDa) were differentially expressed in all four studies. The 3623 Da peak was selected to construct the diagnostic model for minimal to severe endometriosis, and was also differentially expressed in our previous SELDI-TOF MS study (Kyama *et al.*, 2006). The next step is to identify this 3623 kDa peak and other peaks selected in our model.

In conclusion, mRNA expression of eutopic endometrium was comparable in women with and without endometriosis but differed in the menstrual phase endometrium when compared with the luteal phase endometrium. Proteomic analysis of the luteal phase endometrium allowed the diagnosis of minimal–severe, minimal–mild or moderate–severe endometriosis with high sensitivity and specificity using a model based on the analysis of data from a training set which was then applied for validation to data obtained from a test set as part of preclinical studies. Such a test would be useful, especially in women with endometriosis which is not visible during gynaecological ultrasound examination. Indeed, women with ovarian endometriomas or deep rectovaginal nodules should be excluded from biomarker studies since these forms of the disease can be diagnosed by transvaginal sonography. However, a non-invasive diagnostic test should be used only for symptomatic women with endometriosis. We need to avoid its use as a broad screening tool because surgical treatment of asymptomatic women is questionable and may do more harm than good, as reported previously (Evers and Van Steirteghem, 2009).

Future studies are needed to confirm our preliminary data and to perform a comparison with the established biomarker, such as CA-125.

Supplementary data

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

Acknowledgements

We would like to thank the staff and nurses of the fertility centre and Petra Stevens for collecting samples. Microarray analyses were performed by the VIB MicroArray Facility (www.microarrays.be). We would also like to thank their team for excellent guidance throughout the experiment. We would further like to thank Katrien Drijkoningen, Katrien Luyten and Rieta Van Bree for assistance in the experiment.

Authors' roles

All authors declare to have participated in the study as mentioned below and who had seen and approved the final version. A.F., C.M.K., N.V., A.B., O.G., R.P., E.W. and T.M.H. were involved in study concept and design. A.F. were involved in acquisition of data. A.F., A.V., D.B., O.G., F.O., E.W. and T.M.H. contributed to analysis and interpretation of data. A.F., N.V., D.B., E.W. and T.M.H. contributed to drafting of the manuscript. A.F., A.V., D.B., C.M.K., A.V., A.B., O.G., R.P., F.O., E.W., C.T., K.P., C.M., Y.M., B.M. and T.M.H. contributed to critical revision of the manuscript for important intellectual content. T.M.H. had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Funding

This work was supported by grants from the Leuven University Council (Dienst Onderzoekscoordinatie, K.U.Leuven, Leuven, Belgium), the Flemish Fund for Scientific Research (FWO), Leuven – Belgium and K.U.Leuven Interfaculty Council for development Co-operation, Leuven, Belgium.

Research funded by a Ph.D. grant of the Agency for Innovation by Science and Technology (IWT)

Research supported by

- Research Council KUL: ProMeta, GOA MaNet, GOA 08/16, CoE EF/05/007 SymBioSys en KUL PFV/10/016 SymBioSys, START 1, several PhD/postdoc & fellow grants, IAP P6/28
- Flemish Government:
  - FWO: PhD/postdoc grants, projects, G.0553.06 (VitamineD), G.0302.07 (SVM/Kernel), G.0871.12N (Neural circuits) research communities (ICCoS, ANMMM, MLDM); G.0733.09 (3UTR); G.082409 (EGFR)
  - IWT: PhD Grants; SBO-MoKa, TBM-IOTA3
  - FOD:Cancer plans
  - IBBT
- Belgian Federal Science Policy Office: IUAP P6/25 (BioMaGNet, Bioinformatics and Modeling: from Genomes to Networks, 2007–2011) ;
- EU-RTD: ERNSI: European Research Network on System Identification; FP7-HEALTH CHeartED

The scientific responsibility is assumed by its authors.

Conflict of interest

None declared.

References

Absenger Y, Hess-Stumpp H, Kreft B, Kratzschmar J, Haendler B, Schutze N, Regidor PA, Winterhager E. Cyr61, a deregulated gene in endometriosis. *Mol Hum Reprod* 2004;**10**:399–407.

Aghajanova L, Giudice LC. Molecular evidence for differences in endometrium in severe versus mild endometriosis. *Reprod Sci* 2010; **18**:229–251.

Arruda MS, Petta CA, Abrao MS, Benetti-Pinto CL. Time elapsed from onset of symptoms to diagnosis of endometriosis in a cohort study of Brazilian women. *Hum Reprod* 2003;**18**:756–759.

ASRM. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. *Fertil Steril* 1997;**67**:817–821.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 1995; **57**:289–300.

Bouckaert RR, Frank E. Evaluating the replicability of significance tests for comparing learning algorithms. In: Dai H, Srikant R, Zhang C (eds). *Proceedings 8th Pacific-Asia Conference, PAKDD 2004, Sydney, Australia, May 26–28*, 2004, 3–12, Berlin: Springer.

Burney RO, Talbi S, Hamilton AE, Vo KC, Nyegaard M, Nezhat CR, Lessey BA, Giudice LC. Gene expression analysis of endometrium reveals progesterone resistance and candidate susceptibility genes in women with endometriosis. *Endocrinology* 2007;**148**:3814–3826.

Cramer DW, Missmer SA. The epidemiology of endometriosis. *Ann N Y Acad Sci* 2002;**955**:11–22.

Dassen H, Kamps R, Punyadeera C, Dijcks F, de Goeij A, Ederveen A, Dunselman G, Groothuis P. Haemoglobin expression in human endometrium. *Hum Reprod* 2008;**23**:635–641.

D'Hooghe TM, Debrock S. Endometriosis, retrograde menstruation and peritoneal inflammation in women and in baboons. *Hum Reprod Update* 2002;**8**:84–88.

D'Hooghe TM, Mihalyi AM, Simsa P, Kyama CK, Peeraer K, De Loecker P, Meeuwis L, Segal L, Meuleman C. Why we need a noninvasive diagnostic test for minimal to mild endometriosis with a high sensitivity. *Gynecol Obstet Invest* 2006;**62**:136–138.

Ding X, Wang L, Ren Y, Zheng W. Detection of mitochondrial biomarkers in eutopic endometria of endometriosis using surface-enhanced laser desorption/ionization time-of-flight mass spectrometry. *Fertil Steril* 2010;**94**:2528–2530.

Evers JL, Van Steirteghem AC. All that glistens is not gold. *Hum Reprod* 2009;**24**:2972–2973.

Fassbender A, Simsa P, Kyama CM, Waelkens E, Mihalyi A, Meuleman C, Gevaert O, Van de Plas R, de Moor B, D'Hooghe TM. TRIzol treatment of secretory phase endometrium allows combined proteomic and mRNA microarray analysis of the same sample in women with and without endometriosis. *Reprod Biol Endocrinol* 2010;**8**:123.

Fassbender A, Waelkens E, Verbeeck N, Kyama CM, Bokor A, Vodolazkaia A, Van de Plas R, Meuleman C, Peeraer K, Tomassetti C *et al.* Proteomics analysis of plasma for early diagnosis of endometriosis. *Obstet Gynecol* 2012;**119**:276–285.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;**5**:R80.

Giudice LC, Kao LC. Endometriosis. *Lancet* 2004;**364**:1789–1799.

Hadfield R, Mardon H, Barlow D, Kennedy S. Delay in the diagnosis of endometriosis: a survey of women from the USA and the UK. *Hum Reprod* 1996;**11**:878–880.

Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, USA: Springer, 2009.

Houston DE. Evidence for the risk of pelvic endometriosis by age, race and socioeconomic status. *Epidemiol Rev* 1984;**6**:167–191.

Husby GK, Haugen RS, Moen MH. Diagnostic delay in women with pain and endometriosis. *Acta Obstet Gynecol Scand* 2003; **82**:649–653.

Kao LC, Germeyer A, Tulac S, Lobo S, Yang JP, Taylor RN, Osteen K, Lessey BA, Giudice LC. Expression profiling of endometrium from women with endometriosis reveals candidate genes for disease-based implantation failure and infertility. *Endocrinology* 2003;**144**:2870–2881.

Kennedy S, Bergqvist A, Chapron C, D'Hooghe T, Dunselman G, Greb R, Hummelshoj L, Prentice A, Saridogan E. ESHRE guideline for the diagnosis and treatment of endometriosis. *Hum Reprod* 2005;**20**:2698–2704.

Kyama CM, T'Jampens D, Mihalyi A, Simsa P, Debrock S, Waelkens E, Landuyt B, Meuleman C, Fulop V, Mwenda JM *et al.* ProteinChip technology is a useful method in the pathogenesis and diagnosis of endometriosis: a preliminary study. *Fertil Steril* 2006;**86**:203–209.

Kyama CM, Mihalyi A, Gevaert O, Waelkens E, Simsa P, Van de Plas R, Meuleman C, De Moor B, D'Hooghe TM. Evaluation of endometrial biomarkers for semi-invasive diagnosis of endometriosis. *Fertil Steril* 2011;**95**:1338–1343.

Matsuzaki S, Canis M, Vaurs-Barriere C, Boespflug-Tanguy O, Dastugue B, Mage G. DNA microarray analysis of gene expression in eutopic endometrium from patients with deep endometriosis using laser capture microdissection. *Fertil Steril* 2005;**84**(Suppl. 2):1180–1190.

May KE, Conduit-Hulbert SA, Villar J, Kirtley S, Kennedy SH, Becker CM. Peripheral biomarkers of endometriosis: a systematic review. *Hum Reprod Update* 2010;**00**:1–24.

Ponnampalam AP, Weston GC, Trajstman AC, Susil B, Rogers PA. Molecular classification of human endometrial cycle stages by transcriptional profiling. *Mol Hum Reprod* 2004;**10**:879–893.

Poon TC. Opportunities and limitations of SELDI-TOF-MS in biomedical research: practical advices. *Expert Rev Proteomics* 2007; **4**:51–65.

Punyadeera C, Dassen H, Klomp J, Dunselman G, Kamps R, Dijcks F, Ederveen A, de Goeij A, Groothuis P. Oestrogen-modulated gene expression in the human endometrium. *Cell Mol Life Sci* 2005; **62**:239–250.

Robin X, Turck N, Hainard A, Lisacek F, Sanchez JC, Müller M. Bioinformatics for protein biomarker panel classification: what is needed to bring biomarker panels into *in vitro* diagnostics? *Expert Rev Proteomics* 2009;**6**:675–689.

Sensky TE, Liu DT. Endometriosis: associations with menorrhagia, infertility and oral contraceptives. *Int J Gynaecol Obstet* 1980; **17**:573–576.

Sherwin JR, Sharkey AM, Mihalyi A, Simsa P, Catalano RD, D'Hooghe TM. Global gene analysis of late secretory phase, eutopic endometrium does not provide the basis for a minimally invasive test of endometriosis. *Hum Reprod* 2008;**23**:1063–1068.

Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004;**3** (Article 3).

Smyth G. *Limma: Linear Models for Microarray Data*. New York: Springer, 2005.

Stratowa. XPS, a novel framework for distributed storage and analysis of microarray data in the terabyte range: an alternative to bioConductor. In: *Proceedings of the Third International Workshop on Distributed Statistical Computing*, 2003.

Suykens J, Vandewalle J, Moor BD. Optimal control by least squares support vector machines. *Neural Netw* 2011;**14**:23–25.

Talbi S, Hamilton AE, Vo KC, Tulac S, Overgaard MT, Dosiou C, Le Shay N, Nezhat CN, Kempson R, Lessey BA *et al.* Molecular phenotyping of human endometrium distinguishes menstrual cycle phases and underlying biological processes in normo-ovulatory women. *Endocrinology* 2006;**147**:1097–1121.

Wang L, Zheng W, Ding XY, Yu JK, Jiang WZ, Zhang SZ. Identification biomarkers of eutopic endometrium in endometriosis using artificial neural networks and protein fingerprinting. *Fertil Steril* 2010; **93**:2460–2462.

Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;**3**:25.

Xua Q-S, Liang Y-Z. Monte Carlo cross validation. *Chemometrics Intelligent Lab Syst* 2001;**56**:1–11.