Guidelines for the design, analysis and interpretation of ‘omics’ data: focus on human endometrium

Signe Altmäe1,2,*, Francisco J. Esteban3, Anneli Stavreus-Evers4, Carlos Simón5, Linda Giudice6, Bruce A. Lessey7, Jose A. Horcajadas8,9, Nick S. Macklon10,11, Thomas D’Hooghe12,13, Cristina Campoy2, Bart C. Fauser11, Lois A. Salamonsen14, and Andres Salumets1,15

1Competence Centre on Reproductive Medicine and Biology, Tartu, Estonia 2School of Medicine, Department of Paediatrics, University of Granada, 18012 Granada, Spain 3Department of Experimental Biology, University of Jaén, 23071 Jaén, Spain 4Department of Women’s and Children’s Health, Uppsala University, Akademiska Sjukhuset, 75185 Uppsala, Sweden 5Fundación Instituto Valenciano de Infertilidad (FIVI) and Instituto Universitario IVI/INCLIVA, Valencia University, 46021 Valencia, Spain 6Department of Obstetrics, Gynecology and Reproductive Sciences, University of California San Francisco, San Francisco, CA 94143-0132, USA 7Division of Reproductive Endocrinology, Department of Obstetrics and Gynecology, University Medical Group, Greenville Hospital System, Greenville, South Carolina, SC 29605, USA 8Araid-Hospital Miguel Servet, 50004 Zaragoza, Spain 9Department of Genetics, Universidad Pablo de Olavide, 41013 Sevilla, Spain 10Department of Obstetrics and Gynaecology, Division of Developmental Origins of Adult Disease, University of Southampton, Princess Anne Hospital, SO16 5YA Southampton, UK 11Department of Reproductive Medicine and Gynaecology, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands 12Leuven University Fertility Center, Department of Obstetrics and Gynecology, University Hospital Leuven, Leuven, Belgium 13Department of Development and Regeneration, KU Leuven (Leuven University), 3000 Leuven, Belgium 14Prince Henry’s Institute of Medical Research, Melbourne, Victoria 3168, Australia 15Department of Obstetrics and Gynaecology, University of Tartu, 51014 Tartu, Estonia

*Correspondence address. Competence Centre on Reproductive Medicine and Biology, 50410 Tartu, Estonia. E-mail: signealtmae@ugr.es

Submitted on April 19, 2013; resubmitted on August 4, 2013; accepted on August 19, 2013

TABLE OF CONTENTS

- Introduction
- ‘Omic’ studies applied to human endometrium
  - Genomics in human endometrium
  - Epigenomics in human endometrium
  - Transcriptomics in human endometrium
  - Proteomics in human endometrium
  - Metabolomics in human endometrium
- Study design in endometrial ‘omic’ studies
- Sample collection and processing
- Data processing and analysis
  - Genomics
  - Epigenomics
  - Transcriptomics
  - Proteomics
  - Metabolomics
- Validation of biomarker sets
- Systems biology in integrative endometrial ‘omic’ studies
- Conclusions and future perspectives

BACKGROUND: ‘Omic’ high-throughput analyses, including genomics, epigenomics, transcriptomics, proteomics and metabolomics, are widely applied in human endometrial studies. Analysis of endometrial transcriptome patterns in physiological and pathophysiological conditions...
has been to date the most commonly applied ‘omics’ technique in human endometrium. As the technologies improve, proteomics holds the next big promise for this field. The ‘omics’ technologies have undoubtedly advanced our knowledge of human endometrium in relation to fertility and different diseases. Nevertheless, the challenges arising from the vast amount of data generated and the broad variation of ‘omics’ profiling according to different environments and stimuli make it difficult to assess the validity, reproducibility and interpretation of such ‘omics’ data. With the expansion of ‘omics’ analyses in the study of the endometrium, there is a growing need to develop guidelines for the design of studies, and the analysis and interpretation of ‘omics’ data.

METHODS: Systematic review of the literature in PubMed, and references from relevant articles were investigated up to March 2013.

RESULTS: The current review aims to provide guidelines for future ‘omics’ studies on human endometrium, together with a summary of the status and trends, promise and shortcomings in the high-throughput technologies. In addition, the approaches presented here can be adapted to other areas of high-throughput ‘omics’ studies.

CONCLUSION: A highly rigorous approach to future studies, based on the guidelines provided here, is a prerequisite for obtaining data on biological systems which can be shared among researchers worldwide and will ultimately be of clinical benefit.

Key words: endometrium / epigenomics / genomics / metabolomics / proteomics

Introduction

The completion of the Human Genome Project in 2000 (Venter et al., 2001) triggered a rapid development of several fields in molecular biology that together are described as ‘omics’, also known as the ‘omics revolution’. It was stated then that ‘with 35 000 genes and hundreds of thousands of protein states to identify, correlate and understand, it no longer suffices to rely on studies of one gene, gene product or process at a time. We have entered the ‘omic’ era in biology’ (Weinstein, 2001).

‘Omics’ refer to the application of high-throughput techniques which simultaneously examine changes in the genome (assessment of variability in DNA sequence in the genome, i.e. genomics), epigenome (epigenetic modifications of DNA, i.e. epigenomics), transcriptome (gene expression profiling, assessment of variability in composition and abundance of messenger RNA (mRNA) levels, i.e. transcriptomics), proteome (variability in composition and abundance of the proteins, i.e. proteomics) or metabolome (variability in composition and abundance of metabolites, i.e. metabolomics or metabonomics) in a biological sample. In addition to these well-established ‘omics’ fields, the ‘omics’ are ever expanding with new fields in biological data, such as exomics (exons in a genome, i.e. exome), lipidomics (collection of lipids, i.e. lipidome), secretomics (secreted protein, i.e. secretome), interactomics (interactome, or ‘systems biology’) and others (wikipedia.org/omics.org).

In reproductive medicine, Bellver et al. (2012) defined the ‘omics’ high-throughput analyses as ‘reproducutoomics’. At present, reproducutoxics applications are foreseen as: (i) genomics for whole-genome genetic screening in adults prior to conception, and diseases (male factor, implantation failure, recurrent abortion), comparative genome hybridization (CGH) analyses for PGD and preimplantation genetic screening, and in prenatal diagnosis (Wells and Levy, 2003; Rubio et al., 2007; Bellver et al., 2012; Treff, 2012); (ii) transcriptomics and proteomics to provide robust molecular tools for endometrial evaluation (Boomsma et al., 2009a, b; Diaz-Gimeno et al., 2011; Cheong et al., 2013; Diaz-Gimeno et al., 2013; Garrido-Gomez et al., 2013; Salamonsen et al., 2013), oocyte screening through cumulus cells investigation (Assou et al., 2010; Koks et al., 2010) and sperm selection (Sel et al., 2010; Altmäe and Salumets, 2011) and (iii) proteomics and metabolomics to provide future alternatives for non-invasive methods for embryo selection (Bellver et al., 2012) and for the diagnosis of endometriosis (Fassbender et al., 2012a, b; Fassbender et al., 2012a, b). Nevertheless, ‘omics’ technologies still require improved reproducibility and clinical predictive value based on large sample cohorts (thereby clinical applicability), before they can provide new diagnostic, prognostic and therapeutic tools.

The current review focuses on all ‘omics’ studies applied to human endometrium. While the ‘omics’ technologies have undoubtedly advanced our knowledge of human endometrium in health and different diseases, the challenges arising from the huge amount of data generated and broad variation of ‘omics’ profiling according to different environments and stimuli make it difficult to determine the use, validity, reproducibility and interpretation of ‘omics’ data. The current review provides guidelines for future ‘omics’ studies on human endometrium and summarizes the current status together with the promise and shortcomings of the high-throughput technologies.

‘Omics’ studies applied to human endometrium

The human endometrium is a dynamic tissue that undergoes cyclic growth, differentiation, desquamation and regeneration that are driven by the ovarian steroid hormones oestrogen and progesterone as well as other hormones, cytokines and chemokines (Wilcox et al., 1999; Lessey, 2000; Salamonsen et al., 2009). The main function of the endometrium is to provide precisely timed support to enable embryo implantation and for further fetal growth and maturation. Although the endometrium is non-receptive to embryos for most of the menstrual cycle, it becomes receptive during a spatially and temporally restricted period in the secretory phase known as the ‘window of implantation’ (Harper, 1992; Giudice, 1999). Gaining insight into the complex mechanisms controlling changes within the endometrium is crucial to understand not only implantation but also gynaecological disorders, such as endometriosis, uterine fibroids or endometrial cancer, that can impact endometrial function leading to infertility or pregnancy loss.

From the first histologic dating methods (Noyes et al., 1950; Noyes et al., 1975) to the new ‘omics’ technologies, extensive efforts have been applied to understanding and characterizing receptive endometrium. Despite the common use of the traditional endometrial dating criteria, its accuracy, reproducibility and functional relevance have been
SNPs) have led to a rapid development of genome-wide association studies (GWAS). GWAS are currently the most commonly used technique for searching for SNPs/loci associated with a trait of disease (Day and Loos, 2011). Genomics analyses in human endometrium are still limited; to date, four GWAS in endometriosis (Adachi et al., 2010; Uno et al., 2010; Painter et al., 2011; Nyholt et al., 2012) with summarizing tables of these findings (Rahmioglu et al., 2012; Burney, 2013) and one study in endometrial cancer patients (Ikeda et al., 2012) have been published. The inconsistency in the GWAS and association studies in general has been attributed to small sample sizes, proper controls and heterogeneity within populations (Guo, 2006; Sundqvist et al., 2013). The identification of true genetic association requires a large sample size and replication in different populations. Also a general problem in association studies is that papers with positive results may be published preferentially, thus masking the real situation (Altmae et al., 2011). In the field of endometriosis, big advances in the search for risk loci have been achieved by two recent comprehensive and large-scale GWAS, where seven SNPs were identified and confirmed in different populations (Painter et al., 2011; Nyholt et al., 2012). These recent GWAS of >4600 cases and >9300 controls of Japanese and European ancestry demonstrated that only weakly associated SNPs represent true endometriosis risk loci that can be used for risk prediction and future targeted disease therapy across these different populations (Painter et al., 2011; Nyholt et al., 2012). No SNP-based microarray studies regarding endometrial receptivity or infertility have yet been published.

Although occurring less frequently than SNPs, copy number variations (CNVs) play an important role in genetic variation. CNVs located in the promoter regions of genes can influence gene expression levels and thereby contribute to the development of complex disease traits (Redon et al., 2006). Recent analysis of somatic copy number aberrations and CNVs in patients with endometriosis found no association with disease aetiology (Saare et al., 2012). As a small sample was analysed, the study results are rather preliminary and need to be confirmed in a larger study group. Previous GWAS focusing on genomic alterations in endometria of patients with endometriosis have shown various chromosomal alterations; however, only a few of these alterations have been observed in more than one study (Guo et al., 2004; Wu et al., 2006; Veiga-Castelli et al., 2010). These inconsistencies in findings highlight the need for studies that are powered sufficiently in order to detect (or rule out) the effects of interest, as has been neatly demonstrated on ~15 000 women in a search for endometriosis risk loci (Painter et al., 2011; Nyholt et al., 2012).

Epigenomics in human endometrium

The cyclic changes in gene expression in the human endometrium during the menstrual cycle are believed, in part, to be under epigenetic control (Lessey, 2010; Munro et al., 2010; Guo, 2012). Several genes expressed by the endometrium have already been identified as being epigenetically regulated (Munro et al., 2010), including leukemia inhibitory factor (LIF) (Uchida et al., 2005), glycodelin (Uchida et al., 2007), matrix metalloproteinases (Clark et al., 2007), E-cadherin (Rahnama et al., 2009), mucin I (Yamada et al., 2008) and others. Epigenetic modifications including DNA methylation, histone acetylation and RNA interference are involved in functional changes in endometrium associated with pregnancy (endometrial receptivity, decidualization) and can be altered in diseases affecting the endometrium (endometriosis, cancer, implantation

Figure 1 ‘Omic’s’ publications in human endometrium studies. Y-axis indicates the number of studies, and X-axis denotes the year of publication. The systematic review of the literature in PubMed was conducted up to March 2013. The eligible studies were additionally identified using reference lists of review articles and other relevant studies. Abstracts from conference proceedings were also considered. No language or any other restrictions were applied. Search terms are presented in detail in Supplementary data, Table S1, and the search results in Supplementary data, Figure S1. In short, keywords ‘endometrium’, ‘endometriosis’ and ‘embryo implantation’ were one-by-one searched with each pair term. After excluding duplicates, a total of 2478 manuscripts were identified and following critical selection 269 manuscripts of ‘omic’s’ studies in human endometrium remained (including studies on normal endometrium, endometrial receptivity, implantation and implantation failure, endometriosis and endometrial cancer): 23 of genomics, 164 of transcriptomics, 26 of epigenomics, 54 of proteomics and 2 of metabolomics.

Genomics in human endometrium

Recent advances in genotyping technology together with detailed information of common gene variants (single nucleotide polymorphisms,
failure) (see recent reviews Revel et al., 2011; Estella et al., 2012; Guo, 2012).

For epigenomics analysis of the endometrium, the most applied ‘omics’ platform so far has been microRNA (miRNA) arrays. miRNAs function as posttranscriptional regulators of gene expression and operate through RNA interference, either degrading or translationally repressing target mRNAs (Bhattacharyya et al., 2006; Carthew and Sontheimer, 2009). Array studies in the endometrium have demonstrated that aberrant miRNA expression is associated with benign gynecological conditions, such as endometriosis (Burney et al., 2009; Ha, 2011; Zelenko et al., 2012), gynaecological malignancies (Gilbert-Estelles et al., 2012; Ramon et al., 2012; Torres et al., 2013) and fertility disorders (Li et al., 2011; Revel et al., 2011). The importance of miRNAs in human endometrial receptivity has also been highlighted (Kuokkanen et al., 2010; Li et al., 2011; Revel et al., 2011; Sha et al., 2011; Estella et al., 2012) and, in association with gene expression data, a subset of miRNA target genes, including LIF, that could serve as potential biomarkers for endometrial receptivity has been proposed (Altmäe et al., 2013). To date, only three studies have applied deep sequencing (RNA-seq, in which all of the RNA is sequenced) for identifying miRNA expression profiles in human endometrium related to infertility (Sha et al., 2011) and endometriosis (Creighton et al., 2010; Hawkins et al., 2011). Most recently, miRNAs identified as specific to endometrial exosomes have been described, providing a new paradigm for endometrial-embryo interactions (Ng et al., 2013) and maybe a new opportunity for biomarker discovery.

Methylation of DNA is another level of epigenetic control, which has important implications for diseases including endometriosis (Izawa et al., 2011). A recent whole-genome scanning of methylation status in >25,000 promoters using methylated DNA immunoprecipitation with hybridization to promoter microarrays demonstrated that the overall methylation profile was highly similar between the endometrium and the endometriotic lesions (Borghese et al., 2010).

Clearly, our current knowledge of the endometrial epigenome and its physiological and pathophysiological significance is somewhat limited. Assessing the impact of the epigenome on endometrial physiology and pathophysiology remains a challenge for future studies.

Transcriptomics in human endometrium

Microarray-based gene expression technology that allows simultaneous monitoring of the expression of thousands of genes is the most widely used platform for transcriptome analysis. Our literature search (see Supplementary data, Table S1 and Figure S1) resulted in over 160 relevant manuscripts (61% of all identified ‘omics’ studies) (Figure 1). Endometrial transcriptomics has been applied to many aspects of endometrial physiology and pathophysiology, including the normal menstrual cycle (see reviews Horcajadas et al., 2004; White and Salamonsen, 2005; Giudice, 2006; Sherwin et al., 2006; Horcajadas et al., 2007; Aghajanova et al., 2008a, b; Bellver et al., 2012; Haouzi et al., 2012; Ruiz-Alonso et al., 2012), implantation and implantation failure (see reviews Toth et al., 2011; Koot et al., 2012), infertility including treatment protocols (see reviews Martinez-Conjejero et al., 2007; Ruiz-Alonso et al., 2012), the impact of endometriosis (see reviews Matsuzaki, 2011; Fassbender et al., 2012a, b), endometrial cancer (see reviews Sherwin et al., 2006; Doll et al., 2008) and others (see reviews Horcajadas et al., 2007; Ruiz-Alonso et al., 2012; Garrido-Gomez et al., 2013). While any given study yields numerous candidate genes to explore, the number of genes, which have been identified in more than one study as potential biomarkers in endometrial physiology and pathophysiology, has remained somewhat small. This can be attributed to the differences in experimental design, timing and conditions of endometrial sampling, patient and control selection criteria, array platforms and annotation versions used, applied strategies for data processing and a lack of consistent standards for data presentation and deposition of complete data sets in public repositories (Horcajadas et al., 2007; Altmäe et al., 2010; Ruiz-Alonso et al., 2012; Ulbrich et al., 2013). Together, these constraints have made it nearly impossible to perform a meta-analysis of similar studies on specific stages of endometrial development (Ulbrich et al., 2013). Further, lack of uniformity in the validation of specific diseases such as endometriosis has limited the potential from this cumulative data to be realized.

Importantly, recent studies in which different cellular compartments of endometrium were analysed demonstrated cell type-specific gene expression profiles (Yanaihara et al., 2005; Evans et al., 2012a, b; Ulbrich et al., 2013). This was not surprising given the considerable changes in endometrial cellular composition with cycle stage. However, most transcriptome analyses have used biopsies of entire endometrial tissue containing all cell types: thus measured mRNA abundance reflects an average of all cell types present. In the two published ‘omics’ studies on human endometrium in which stromal and epithelial fractions were isolated by laser capture microdissection, distinct mRNA signatures were related to the day of the cycle (Yanaihara et al., 2005; Evans et al., 2012a, b). In addition, Evans et al., (2012a, b) compared two different microarray platforms, Affymetrix and Agilent, demonstrating concordance in their results. Clearly, one of the future tasks for transcriptome studies will be separate analyses of endometrial compartments, which will in turn provide a better understanding of endometrial physiology, the interactions between different cell types and their regulatory processes. However, the challenge is to obtain sufficient good quality RNA for expression analysis following microdissection. Recently published transcriptomes of endometrial cell constituents isolated by fluorescence-activated cell sorting (FACS) have also demonstrated cell-specific gene expression and identified multiple biological pathways and processes (Spitzer et al., 2012) and provided an alternative approach to laser dissection.

In addition to microarrays, there is the emerging alternative of RNA-seq, in which all RNAs are sequenced and most of the genes being expressed can be revealed. A comparison of the results derived from an Affymetrix microarray study and an RNA-seq study of bovine endometrium revealed a consistent overlap between the results but many more differentially expressed genes for the sequencing data (Ulbrich et al., 2013), as reported elsewhere (Malone and Oliver, 2011). RNA-seq technology detects more exons and alternative splicing events than microarray as it is entirely independent of prior knowledge. Microarrays routinely fail to pick up ~25% of genes with low expression but such low-abundance transcripts are detected in RNA-seq reads (Werner, 2011). Nevertheless, while it is true that RNA-seq is independent of prior knowledge, the biological analysis of the data is not (Werner, 2011). RNA-seq analyses in human endometrium in health and diseased states are required.

The window of implantation is arguably the most relevant time to study gene expression profile as a way to establish biomarkers of a receptive endometrium (see reviews Bellver et al., 2012; Ruiz-Alonso et al.,...
Proteomics in human endometrium

Proteomic research is currently considered a ‘hot topic’ and is increasingly being applied to human endometrium (Salamonsen et al., 2013). Any analysis of the full proteome is a challenging task as the proteome is large and of unknown complexity, being a result of alternative splicing of primary transcripts, the presence of sequence variation and epigenetic and post-translational modifications. This complexity is well reflected by the fact that although mRNA expression precedes protein translation, the correlation between a transcript level and the abundance of the corresponding protein product is poor (Fassbender et al., 2010; Ning et al., 2012), including in the endometrium (Stephens et al., 2010). Nevertheless, with the latest advances in mass spectrometry (MS) instrumentation, proteomics has emerged as a powerful tool for biomarker research.

Our literature search resulted in >50 manuscripts describing endometrial proteomic studies (21% of all identified ‘omics’ studies) (Figure 1). The most studied applications of proteomics include the search for biomarkers of the receptive endometrium (reviews Garrido-Gomez et al., 2010; Berlanga et al., 2011; Hannon et al., 2011a, b; Koot et al., 2012; Edgell et al., 2013; Upadhyay et al., 2013) and endometriosis (reviewed in Meehan et al., 2010; May et al., 2011; Fassbender et al., 2012a, b; Fassbender et al., 2013; Upadhyay et al., 2013). Edgell et al. (2013) presented very recently a list of nine validated proteins from endometrial tissue of relevance to endometrial receptivity, where membrane-associated progesterone receptor component 1 (PGRC1) and annexins (ANXA2 and ANXA4) together with others were validated. A recent systematic review of endometrial biomarkers of endometriosis assessing hormones, cytokines, proteomic factors and histological analysis of endometrial tissue concluded that none of the biomarkers alone or in a panel was 'unequivocally clinically useful' due to low numbers of subjects in discovery and replication studies, cycle- and stage dependence and low sensitivity and specificity (May et al., 2011).

Early proteomic studies focused on analysis of endometrial tissue which has the same limitations of changing cellular composition as those described for transcriptomics (reviewed in Salamonsen et al., 2013). Furthermore, the high abundance of structural proteins is also an issue as these can mask the lower abundance proteins during analysis.

A more promising approach is the analysis of endometrial secretions, which can be accessed in uterine fluid (reviews Cheong et al., 2013; Edgell et al., 2013). Endometrial fluid collection is less invasive than endometrial biopsy. The fluid contains numerous secreted proteins associated with endometrial maturation and receptivity (Boomsma et al., 2009a, b; Casado-Vela et al., 2009; Scotchie et al., 2009; Hannon et al., 2010; Hannon et al., 2011a, b). Furthermore, uterine secretions are less complex than tissue in terms of their protein repertoire and may provide a subset of biomarkers for functional endometrial disturbances.

Although albumin and other plasma proteins make up some 90% of the total protein in uterine fluid, this can be removed (Hannon et al., 2009), enabling the lower abundance proteins to be more readily analysed. Uterine fluid can be harvested by either aspiration (provides ~5 μl fluid for analysis) or lavage, which washes the endometrial surface, probably removing loosely attached proteins. A comparison of these two sampling methods demonstrated that although there were many similarities in protein profiles, the results differed somewhat between methodologies, both having advantages and disadvantages (Hannon et al., 2012). Therefore, consistency in the sampling technology is crucial to enable comparisons between data sets. Another potential non-invasive approach, namely analysis of the proteome of menstrual blood, has also been proposed for assessment of infertility and endometrial pathology (Yang et al., 2012).
hardware (having an achievable ≏50 attomole peptide sensitivity), will lead to further developments in endometrial proteomics, as it will enable detection of the proteome in minute samples of protein. New techniques for pre-fractionation will further enhance sensitivity. Thus, the full potential of proteomics in endometrial research is yet to be achieved.

**Metabolomics in human endometrium**

The study of cellular metabolic products (metabolomics) as potential biomarkers lags behind the other ‘omics’ technologies discussed in this review. Metabolomics is the study of small, low-molecular weight products of metabolism (Nicholson and Lindon, 2008). Metabolomics reflects events well downstream of gene expression and is considered to be closer to the actual phenotype than genomics, transcriptomics or proteomics (Allen et al., 2003). The concentration of a specific metabolite is the cumulative effect of the activity of all enzymes involved in the synthesis and catalysis of a given compound, thereby having the potential to provide integrative information about tissue function within the larger context of the organism (Nielsen and Oliver, 2005). Although metabolomics should remain an integrated approach, its complexity leads to subdivisions, such as lipidomics and glycomics (Lagarde et al., 2003).

Two studies have been published on metabolomics in human endometrium, each novel in its focus (Vouk et al., 2012; Vilella et al., 2013). Vouk et al. (2012) provided the first report of a metabolomics approach to identification of biomarkers for the diagnosis of endometriosis from plasma (Vouk et al., 2012). Their data indicated that elevated levels of sphingomyelins and ether-phospholipids are associated with endometriosis, and eight lipids are presented as novel endometriosis-associated biomarkers (Vouk et al., 2012). Another recent study examined the lipidomics of human endometrium, demonstrating for the first time a significant increase of lipid levels in endometrial fluid at the window of implantation, which could provide a new tool for endometrial receptivity prediction (Vilella et al., 2013).

The high potential for metabolomics to unravel the genetics of human metabolism has been demonstrated (Illig et al., 2010); in contrast to most GWAS with clinically relevant end-points, most of the associations in metabolic traits were linked to genetic variants in genes with a matching metabolic function (Illig et al., 2010). If this is indeed the case, the use of ‘omics’ to understand the broad picture of the physiology needs to begin with an initial survey of the genomics of an individual/phenotype, allowing investigators to place in context any newly discovered metabolomic end-points.

Other ‘omics’ are awaiting elucidation in endometrial research, including microbiome in health and disease, glycome and exposome, to name a few. The integration of this information with the genome, epigenome, transcriptome, proteome and metabolome awaits a major effort of computational biology approaches and phenotypic assessment of patients with a variety of normal and abnormal presentations and hormonal milieu.

Study design in endometrial ‘omics’ studies

Study design is a crucial step in the general conduct of ‘omics’ studies. For the endometrium, the complex dynamic nature of the tissue response to the cyclic hormonal milieu makes the study design particularly important. The endometrium is composed of many cell types (epithelial, stromal fibroblasts or pre-decidual cells, leukocytes and cells of the vasculature), and there is considerable heterogeneity of cellular composition across the continuum of phases of the cycle. This is particularly the case for the epithelial compartments (luminal and glandular) and the numbers and subsets of leukocyte populations. This intrinsic variability needs to be considered in designing and analysing studies on endometrium (Salmonsen et al., 2009; Savaris and Giudice, 2009; Edgell et al., 2013). In addition, there could be inter-cycle variability; however, a recent study demonstrates consistency at the transcriptomic level between different biopsies from the same patient taken with an interval of 2 years (Diaz-Gimeno et al., 2013). Furthermore, pathological states of the endometrium must be considered, including the impact of structural (fibroids) or immune (endometriosis) alterations. Together with concurrent medication or exposures to environmental toxins as well as genetic susceptibility, all of these factors can alter ‘omics’ outcomes (Horcajadas et al., 2007; Savaris and Giudice, 2009). The importance of study design has previously been addressed in several reviews (White and Salmonsen, 2005; Horcajadas et al., 2007; Savaris and Giudice, 2009; Bellver et al., 2012; Ruiz-Alonso et al., 2012; Edgell et al., 2013) but we hereby briefly summarize the critical points.

First and foremost, researchers should be precise and consistent in defining the phenotype of the study population. Heterogeneity resulting from misclassifying participants decreases both sensitivity and power. The phenotypic heterogeneity across studies often makes it difficult to generalize study findings and replicate the results (Wu et al., 2013). While a clear phenotype definition could potentially reduce sample size, it will increase biological homogeneity and thus increase the statistical power (Gracie et al., 2011). For endometriosis, this problem of heterogeneity is an inherent problem. Half of women with this disease are infertile, while the other half are not. Many have pain that is unrelated to the extent of disease. Endometriosis may also not be diagnosed if it is symptom free. Once treated, the eutopic endometrial changes seen in the endometrium may revert to normal, while selection of patients ahead of laparoscopic diagnosis can be equally problematic. Endometriosis appears to present as mild and severe disease, with each having different genetic risk factors (Nyholm et al., 2012). Medications used to treat this disease may alter the progress and therefore the diagnostic biomarkers that ‘omics’ studies are attempting to establish. Further, the phenome of normal patients and those with endometrial disorders is important in interpreting the obtained data. Vigano et al. (2012) have pioneered this approach (Vigano et al., 2012), along with the World Endometriosis Research Foundation, which held a global consensus meeting on the endometriosis phenotype (World Endometriosis Research Foundation: Endometriosis Phenome and Biobanking Harmonisation Project (EPHect), March 2013, http://endometriosisfoundation.org/ephect/).

Although it is important to carefully define and phenotype cases with a trait, disease or susceptibility, it is as important to define adequately ‘normal’ controls (Savaris and Giudice, 2009; Edgell et al., 2013).

Other sources of biological variability can arise through differences in tissue composition of collected samples, phase or day of cycle, frequency of sample collection, medical history and others (Ulbrich et al., 2013). In endometrial tissue analysis, endometrial biopsy sampling using two different techniques, biopsy and curettage hysterectomy, resulted in identical results in transcriptome expression level (Talbi et al., 2006). A biopsy taken in 1 month may affect endometrial receptivity in the next biopsy.
processing, utilization, storage and distribution (Sheldon et al., 2003). Nevertheless, the timing of the biopsy sampling is critical, as its composition varies considerably in a constantly changing endometrium (Giudice, 1999; Mirkin et al., 2005; Haouzi et al., 2009; Fassbender et al., 2010).

Histological dating of the endometrium is another critical aspect of studies on human endometrium. Although the classical histological dating by Noyes criteria has several limitations, it has remained the gold standard for establishing landmarks of endometrial development. Hormonal reference of the endometrium in natural cycles using urinary LH measurement or, even better, serum LH detection that has superior precision, is important (Evans et al., 2012a, b). In endometrial studies, the sample collection conditions, including the location and timing with respect to day of the cycle, should be as uniform as possible with adoption of standard operating procedures for tissue acquisition, processing, utilization, storage and distribution (Sheldon et al., 2011).

Race and ethnicity are also important. So far, ethnicity has been considered mainly in genetic/genomics studies but it applies to all ‘omics’ disciplines. Genetic variation may differ among racial/ethnic groups, which may influence the results at all ‘omics’ levels, a concept known as population stratification (see Figure 2 for the complexity of biological processes). This can lead to excess false-positive results and failure to detect true associations (Wu et al., 2013). Therefore, the studies should focus as far as possible on a single race/ethnic group or apply advanced analytic techniques for adjusting for ethnic genetic differences (Bryc et al., 2010). At the very least, these factors should be included in databases in the hopes that any differences may be identified in subsequent meta-analysis.

Sample size is critical to good study design, as discussed in detail in a prior review (Savaris and Giudice, 2009). Calculation of sample size demonstrates the power of the study, which takes into account the variance of individual measurements, the acceptable false-positive rate and the desired discriminatory power of the used platform (Savaris and Giudice, 2009). Researchers are strongly encouraged to maximize the sample size in their ‘omics’ studies. Collaboration between groups offers the optimal solution for this inherent problem in our field, as the sample collection is in general invasive and thus it can be difficult to motivate participants and clinicians to participate. Clear examples of adequately powered collaborative studies, where a large sample size was analysed and independent replicates performed, have resulted in strong candidate genes for endometriosis (Painter et al., 2011; Nyholt et al., 2012).

Although many significant results have been derived from ‘omics’ studies in the last decade, most would agree that many studies exhibit an unacceptably large degree of variability with low reproducibility. There is a need for better documentation and uniformity of the types of data collected. Groups of scientists have been working to establish standards of minimum information that must be collected and reported, to ensure the interpretability of the experimental results generated using ‘omics’ technologies (Brazma et al., 2001). These reporting standards include MIAME (Minimum Information about a Microarray Experiment), MIAPE (Minimum Information about a Proteomics Experiment), MIGS-MIMS (Minimum Information about a Genome/Metagenome Sequence), MiMIX (Minimum Information about a Molecular Interaction eXperiment), MINIMiSS (Minimal Metagenome Sequence Analysis Standards), MINISEQe (Minimum Information about a high-throughput Nucleotide Sequencing Experiment) and CIMR (Core Information for Metabolomics Reporting), which have been discussed in detail in previous reports (Taylor et al., 2008; Chervitz et al., 2011). Furthermore, consideration of the complexity of human endometrial physiology in health and disease and additional documentation of the data collected are urgently required. Indeed, in a study evaluating the replication of data analyses in 18 articles on microarray-based gene expression profiling published in Nature Genetics in 2005–2006, only two analyses could be fully reproduced and six partially, while ten could not be reproduced (Ioannidis et al., 2009). The conclusion was that the main reason for lack of reproducibility was lack of access to data and discrepancies due to incomplete data annotation or specification of data processing and analysis. Therefore, more strict publication rules enforcing public data availability should be encouraged (Ioannidis et al., 2009). Table I highlights the points to consider for ‘omics’ studies on endometrium, and scientists in the field are encouraged to follow these guidelines when designing, conducting and reporting their projects.

**Sample collection and processing**

Sample collection for ‘omics’ technologies requires separate discussion of the considerations, obstacles and detailed protocols involved. For research in human endometrium, the biological sample collected is determined by the research question of interest, as a variety of sample types can be analysed (see Table II). Quality of the sample is determined at various steps in the process of collection and storage and is one of the most important factors in the overall success or failure of a study. Sampling, processing and preservation of study material should be carefully established and documented by standard operating procedures, a priori; the consistency in tissue handling should be closely followed (Huang et al., 2001; Spruessel et al., 2004; Shayeghi et al., 2005; Micke et al., 2006), to assure the highest-quality samples, representing as closely as possible the physiology of the original *in vivo* tissue.

Sample processing for different ‘omics’ techniques has been previously discussed in detail (White and Salamonsen, 2005; Horgan et al., 2009; Savaris and Giudice, 2009; Gunaratne et al., 2010; Robert, 2010; Pritchard et al., 2012; Slattery et al., 2012; Fassbender et al., 2013; Ulbrich et al., 2013). In terms of study design and sample collection, researchers are encouraged to take careful note of these guidelines. In Table II, we provide an outline of the ‘omics’ techniques and respective sample sources that can be utilized in endometrial studies.

**Data processing and analysis**

Adequate data analysis is always crucial for providing conclusive results, which can guide us to an appropriate interpretation of the data, in terms of both biological sense and clearly defined target clinical applications. Several excellent papers provide both protocols and instructions for data analysis for each ‘omic’ (see detailed references below, and general standards for ‘omic’ data description, exchange, terminology and experimental execution have also been described; Chervitz et al., 2011). The aim of these data standards is not only to improve reproducibility and to avoid discrepancies but also to determine a true clinical application. Thus, researchers are encouraged to read and follow such standards (Chervitz et al., 2011). In this section, an overview of the analysis and interpretation of ‘omics’ data provides some specific quality criteria details for each ‘omic’, with a special focus on human endometrium (see Table III for summary of data processing and analysis). The detailed procedures are thoroughly described in the cited literature.
Obtain and assign the signal intensity data from the scanned array image to targets. In both cases, specialized software, either provided by the manufacturer or from other free or commercially available platforms, is used to obtain and assign the signal intensity data from the scanned array image to each target probe. For data analysis, different methods for normalization, segmentation and calling can be chosen, which may influence the final results (Pinto et al., 2011; Brady and Vermeesch, 2012). The reference baseline library used for SNP arrays is also important, with large in-house reference datasets improving the quality of the results.

Even though different somatic copy number aberrations (SCNAs) in endometriotic foci detected with CGH technology can be found in the literature, SNP genotyping arrays in human endometrial research are scarce. Recently, and as a good example of experimental design and appropriate technology and data analysis, no endometriosis specific de novo SCNAs, or regions of copy-neutral loss of heterozygosity (cn-LOH), were found in eutopic or ectopic endometrium (Saare et al., 2012). This array study applied the advantages of SNP targets, which enable the detection of cn-LOH from the common B-allele frequency measurement obtained with this technology. In addition, given the variability between calling algorithms, two different programs were used to minimize the number of false discoveries, providing robust results.

### Epigenomics

The emerging science of epigenetics and genomics, coined epigenomics, is providing unique opportunities for the detection of heritable changes due to modifications in the DNA or chromatin that does not include alterations in DNA sequence. These modifications include DNA methylation and hydroxymethylation, chromatin remodelling, histone modifications (methylation, acetylation, ubiquitylation, phosphorylation and sumoylation) and gene regulation by non-coding RNAs (including miRNAs) (Weichenhan and Plass, 2013). The most studied epigenetic modifications at ‘omics’ level (miRNAs and DNA methylation) are also based on microarray platforms, although the substrates, pre-processing and data analysis depend on the modification to be studied (Callinan and Feinberg, 2006). A typical sequence for epigenomic array data analysis includes background correction, quality control (checking for positive and negative controls), data preprocessing (data cleaning and transformation), normalization and statistical analyses (exploratory and differential expression tests), together with appropriate further target validation (Deatherage et al., 2009), following similar quality criteria as described for other ‘omics’, but selecting specific algorithms, mainly related to the normalization method (Adriansen et al., 2012) depending on the platform used (Marabita et al., 2013).

Epigenomics also comprises the study of small and large non-coding RNAs. Studies on miRNA expression in both health and disease are widely carried out using microarray technology (Yin et al., 2008). While the general protocol for data analysis is similar to that for mRNA gene expression microarrays, the different nature of mRNA and miRNA expression experiments needs careful selection of the normalization method and quality assessment of data analysis (Sarkar et al., 2009). Such a miRNA analysis has been described for these small regulators in human endometrial receptivity (Altmae et al., 2013).

### Transcriptomics

Global gene expression analysis, functional genomics and transcriptomics are synonyms for describing and quantifying the set of mRNAs present in a given cell population or tissue at any point in time, and are mainly used to compare the global gene expression between different experimental or biomedical situations. Gene expression microarrays enable development of datasets that include the expression levels of all gene products.
the genes of the genome just in one experiment. Once the raw data are obtained, preprocessing, normalization, statistical analysis for obtaining the differential gene expression between situations, multivariate data exploration and gene enrichment functional analyses are included in well-established protocols that have been described comprehensively (Cordero et al., 2007; Mocellin and Rossi, 2007; Weeraratna and Taub, 2007; Durinck, 2008; Suarez et al., 2009; Zhang et al., 2009).

The need for an appropriate annotation of both the experimental approach and data analysis is a critical quality criterion for gene expression microarray data sharing, reanalysis and comparison. A particularly well-conducted initiative is the multidisciplinary EMERALD project that, based on MIAME standards, provides the quality metrics and even tools and platforms at any step of the microarray data, always aspiring to appropriate further model validations and practical clinical applications (Beisvag et al. 2011). Also, in terms of optimization and interpretation of gene expression data from different platforms, recent

---

**Table II**  Sample processing for ‘omics’ studies in human endometrium (adapted from Gracie et al., 2011).

<table>
<thead>
<tr>
<th>'Omnics' platform</th>
<th>Sample</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomics (DNA)</td>
<td>Blood</td>
<td>Stable at room temperature but best if refrigerated or frozen</td>
</tr>
<tr>
<td></td>
<td>Saliva</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Endometrial tissue</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Surgical biopsy (endometriosis, cancer)</td>
<td></td>
</tr>
<tr>
<td>Epigenomics (DNA, RNA)</td>
<td>Blood</td>
<td>Require proper collection, meticulous processing and sample storage at −80°C to ensure sample integrity</td>
</tr>
<tr>
<td></td>
<td>Endometrial tissue</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Surgical biopsy</td>
<td></td>
</tr>
<tr>
<td>Transcriptomics (RNA)</td>
<td>Blood</td>
<td>Require appropriate collection, careful processing and storage at -80°C to ensure RNA integrity and quality for analysis</td>
</tr>
<tr>
<td></td>
<td>Menstrual blood</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Endometrial fluid</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Endometrial tissue</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Surgical biopsy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Urine</td>
<td></td>
</tr>
<tr>
<td>Proteomics (proteins)</td>
<td>Blood</td>
<td>Require rapid sample preparation and preservation (± protease inhibitors) prior to storage at −80°C to prevent non-specific protein degradation</td>
</tr>
<tr>
<td></td>
<td>Menstrual blood</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Endometrial fluid</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Endometrial tissue</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Surgical biopsy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Urine</td>
<td></td>
</tr>
<tr>
<td>Metabolomics (metabolites)</td>
<td>Blood</td>
<td>Require rapid metabolic ‘quenching’ (flash freezing or acid precipitation) to prevent degradation of metabolites</td>
</tr>
<tr>
<td></td>
<td>Menstrual blood</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Endometrial fluid</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Urine</td>
<td></td>
</tr>
</tbody>
</table>

---

**Table III**  Data processing and analysis for ‘omics’ studies.

<table>
<thead>
<tr>
<th>'Omnics' platform</th>
<th>Preprocessing/analysis</th>
<th>Data processing</th>
<th>Statistical analysis</th>
<th>Biological interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomics</td>
<td>Calling; Standard deviation of the intensity ratios (DLR for CGH; MAPD for SNPs)</td>
<td>Normalization (median; lowest)</td>
<td>Exploratory (PCA, clustering); parametric/non-parametric test; multiple hypothesis testing (FDR correction)</td>
<td>Reference baseline library (for SNPs)</td>
</tr>
<tr>
<td>Epigenomics</td>
<td>Background correction; quality control; cleaning and transformation</td>
<td>Normalization (T-quantiles; VSN for miRNAs)</td>
<td></td>
<td>Gene enrichment functional analysis; gene targets (miRNAs); Gene networks; validation</td>
</tr>
<tr>
<td>Transcriptomics</td>
<td>Background correction; quality control; cleaning and transformation</td>
<td>Normalization (quantiles for expression)</td>
<td></td>
<td>Isoform and functional analysis; protein–protein interaction networks; validation</td>
</tr>
<tr>
<td>Proteomics</td>
<td>Extraction/derivation</td>
<td></td>
<td></td>
<td>Metabolic networks; validation</td>
</tr>
<tr>
<td>Metabolomics</td>
<td>Reproducible spectra</td>
<td>Chemometric software</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CGH, comparative genome hybridization; DLR, derivative log ratio; FDR, false discovery rate; MAPD, median absolute pairwise difference; miRNA, microRNA; PCA, principal component analysis; SNPs, single nucleotide polymorphisms; VSN, variance stabilizing normalization.

---
global gene expression analysis (Loven et al., 2012) has demonstrated that common assumptions (such as that cells produce similar levels of RNA per cell) lead to erroneous interpretations which can be solved using the appropriate controls.

The transcriptomics of the human endometrium has been thoroughly reviewed (Ruiz-Alonso et al., 2012). This includes both the endometrial phase-specific transcriptomic gene profiles and common temporal gene expression patterns, also taking into account the necessary quality criteria. In addition, examples of endometrial transcriptomics as a diagnostic tool are provided.

**Proteomics**

The complexity of the proteome, and its functional interpretation, is one of the current challenges in biology (Cox and Mann, 2011; Matthiesen et al., 2011). Proteomic analyses are mainly based on high-throughput MS and protein array technologies and, as for other ‘omics’, the standards for data analysis, sharing and integration need to be carefully considered (Becnel and McKenna, 2012). Since protein array data analysis is adapted from gene expression microarrays (Sundaresh et al., 2006) the quality criteria described above are also valid for protein arrays. However, the uniquely quantitative nature of MS-based proteomics involves characteristic analysis with particular challenges, the most important being the peptide feature detection and quantification from the raw map (Cappadona et al., 2012). Rigorous methods for the assessment of quality of spectral data are also described (Cairns et al., 2008). Raw spectra processing methods and common data analysis strategies have been discussed recently (Matthiesen et al., 2011). Importantly, the ‘International Workshop on Proteomic Data Quality Metrics’ in 2010, identified and addressed issues regarding the development and use of open access proteomics data. The key principles underlying a framework for data quality assessment in MS data were enumerated and included both the need for an evolving list of comprehensive quality metrics and also standards accompanied by software analytics (Kinsinger et al., 2012). A revision of the bioinformatics analysis of qualitative and quantitative proteomic data has also been published (Kumar and Mann, 2009). To support the publication of MS studies, the proteomics identifications database repository includes a curation pipeline acting as a practical data deposition quality control (Csordas et al., 2012; Vizcaino et al., 2013).

**Metabolomics**

Metabolomics, the high-throughput identification of the general profile of metabolites in a system, shares with proteomics the use of MS methodology (Horgan et al., 2009). MS together with H nuclear magnetic resonance spectroscopy (NMR) is the most commonly applied techniques in metabolic profiling, with evaluation of different acquisition schemes generating reproducible spectra (high-quality NRM data) in different sample conditions. Data can be further assessed with automated processing (Aranibar et al., 2006). A comparison of the different metabolomics technologies and a detailed description of the spectral and statistical analysis tools in metabolic profiling studies can be found elsewhere (Wishart, 2010). Metabolites which are detected at different levels between control and disease samples are generally designated as biomarkers but they are not usually validated for practical clinical applications and their potential usefulness in the clinic needs careful consideration (Koulman et al., 2009).

**Validation of biomarker sets**

It is obvious that any set of biomarkers needs considerable validation to ensure that it differentiates between the physiologically or pathologically different groups of interest. Commonly, only very small sample sets are analysed and thus the risk of a type I error is significant. Independent validation of proposed markers in additional sets of clinical material is essential, preferably in a range of laboratories in different countries to allow for ethnic and environmental influences. To date, only a limited number of laboratories have published data on biomarkers in endometrium, most of these being in communities that are predominantly Caucasian. Multi-site collection and validation are essential to prove that any biomarker set is robust. Importantly, the patient groups tested need to be as uniform as possible; for example, in terms of exclusions (steroidal contraception, endometrial disorders not under study) or the stimulation protocols in an IVF cycle. As noted above, collaboration to achieve these requirements is essential. Other aspects of validation are discussed in detail in Edgell et al. (2013) and Dominguez et al. (2009).

Western blotting has been the method of choice for validation of individual proteins identified as potential biomarkers. However, accurate protein quantification by Western blotting presents a challenge, at least in part as it is dependent on the form of the protein recognized by the antibody selected for use. Recently, reliable MS-based methods to quantify proteins, known as selected reaction monitoring or multiple reaction monitoring, have become well established (Aebersold et al., 2013; Editorial, 2013) and are now easily developed for essentially any protein. Since this method outperforms Western blotting in terms of limit of detection, linear dynamic range, ability to multiplex and reproducibility, it is now clear that this must become the ‘gold standard’ for validation.

A recent development for validation of the cellular localization of any protein is matrix assisted laser desorption imaging mass spectrometry profiling which can establish molecular disease signatures in their histological context: essentially removing the need for immunohistochemistry with its dependence on antibodies. This has been applied to examination of mouse implantation sites (Burnum et al., 2008), for cervical cytology specimens (Schwamborn et al., 2011) and for serous ovarian cancer (Longuespee et al., 2013). It is likely also to prove useful in the context of validation of biomarkers in an endometrial setting.

**Systems biology in integrative endometrial ‘omics’ studies**

The integration of ‘omics’ techniques is called ‘systems biology’. This captures information from genomics, epigenomics, transcriptomics, proteomics, metabolomics, etc., and its combination with theoretical models for predicting the behaviour of a cell, tissue or organism (Figure 2).

The global analysis of the genome and the morphofunctional characterization of proteins and metabolites provide a vast amount of information with a very high potential to unravel the complex interaction of molecular networks underlying the function of any organism in health and disease. Thus, systems biology provides an integrative approach to understanding biology, entailing the functional analysis of the structure and dynamics of cells and focusing on complex interactions, rather than the characteristics of the isolated components of biological systems. The challenge of systems biology resides in the compilation of
data derived from very different areas: biology, chemistry, statistics, physics, mathematics and computational engineering. As systems biology attempts to provide a comprehensive interpretation of all this knowledge, the high-throughput ‘omics’ platforms have to be integrated for the analysis, display and recording of information to guarantee compatibility and accessibility to these data sets (Chervitz et al., 2011).

Because systems biology approaches are focused on the global analysis of multiple interactions at different levels, the main strategy usually employs networks as a representation of interacting molecules. Modules are built around each discrete regulated function, with the interactions among modules finally arising as complex networks. Thus, the process begins with a model based on sets of data, and conclusions are obtained when the experimental data and the model are juxtaposed (Weston and Hood, 2004).

Figure 2: The complexity of ‘omics’ fields in biological processes that contribute to the study and understanding of biological systems. There are more than 25,000 genes in the human genome, encoding ~100,000–200,000 transcripts and 1 million proteins, whereas there are as few as 2500–3000 metabolites that make up the human metabolome (Botros et al., 2008). The genome is essentially invariant among cells and tissues, while the epigenome has a low/moderate temporal variance and influences both transcriptome and proteome. The transcriptome has a high temporal variance and is translated into the proteome differentially in tissues and physiological states, affecting the metabolome in a tissue-specific manner. This ‘simple’ model is modulated by multiple factors: (A) differential splicing that can be affected by the proteome; (B) post-translational modification of proteins; (C) transcription factor binding; (D) receptor ligand binding and (E) environmentally induced factors (adapted from Gracie et al. 2011; Bellver et al. 2012).

inspire new analytical approaches to unravel complex networks in human endometrial physiology and pathophysiology. However, the authors acknowledge the limitation of microarray technology, with its focus on a static snapshot analysis of a dynamic process, and its unilateral analysis of either endometrium or in vitro cultured embryo. It is now well recognized that the endometrium is responsive to both the preimplantation and implanting embryo. In a very elegant in vivo study in women (Licht et al., 2001) demonstrated a clear responsiveness of the endometrium to infused hCG, which included changes in both vascular endothelial growth factor and LIF concentrations in the endometrium. This was further verified in primary endometrial epithelial cells, which responded to hCG by increased secretion of six cytokines/chemokines known to increase during the receptive phase and to be important for implantation (Paiva et al., 2011), indicating that embryo-derived factors can enhance endometrial epithelial receptivity. The decidualized endometrial stromal cells are also responsive to an implanting blastocyst, migrating around the embryo rather than simply being invaded by the embryo (Weimar et al., 2012), and have been proposed as ‘sensors’ for embryo quality (Teklenburg et al., 2010). While studies to assess such responsiveness using homologous co-cultures or perfusion studies are useful for analyzing the specific components of the endometrium and their response to the embryo, all models are limited in their ability to represent the in vivo situation (Teklenburg and Macklon, 2009).

Clearly the mass of data generated within ‘omics’ studies is far from being fully utilized. Over 1 million gene expression data sets are publicly available in repositories (Gene Expression Omnibus, GEO and ArrayExpress) but few scientists fully utilize this data to find new information (Baker, 2012); rather they use only a small set of the data to compare with their own findings (Baker, 2012). In human endometrium, only two studies apply a bioinformatic approach for ‘omics’ data mining (Tapia et al., 2011; Zhang et al., 2012). Zhang et al. (2012) analysed 45 microarrays from three independent studies of endometrial receptivity in the GEO database, and identified a series of potential biomarkers of endometrial receptivity. Tapia et al. (2011) used array data from seven different studies, comparing endometrial gene expression profiles from the proliferative versus mid-secretory phase, and from early secretory versus mid-secretory phase, and ected new transcription factors orchestrating human endometrial receptivity. These two studies open a promising new era for biomarker search in human endometrial physiology in health and disease. The challenge for the future will be analyse huge sets of data simultaneously, taking advantage of the existing data in publicly available databases for raising power, credibility and reliability of findings.

Conclusions and future perspectives

‘Omics’ high-throughput analyses have started to revolutionize our understanding of human endometrial physiology and pathophysiological conditions. Nevertheless, our understanding of the complex phenotypes of endometrial physiology, endometrial disorders and fertility complications remains incomplete, inconsistent and without strong clinical application.

This review has summarized the current status and trends in ‘omics’ technologies applied to human endometrium. While significant advances have been made in assessing endometrial receptivity and discovering
potential biomarkers for endometriosis and receptive endometrium, many of the ‘omics’ studies have not been replicated and their practical value has been limited, without translation into clinical practice. In addition to sufficiently powered studies (i.e. larger sample size), there is a growing need for integrated approaches to investigate complex phenotypes across ‘omics’ categories. Most studies to date (i) have analysed a relatively small sample size and (ii) assessed a single level of ‘omics data in isolation, emphasising the need for highly powered studies and integrated ‘omics’ approaches as a future path for research. For that, universal guidelines should be established so that data and sample collections can be merged, compared, validated and replicated.

We provide here a set of guidelines that we encourage are followed for conducting transparent and well-designed studies, and providing ‘good-reporting-practice’ in studies of the endometrium, as well as other biological systems. Central to these studies should be accurate phenotype definition, the choice and quality of the sample and adequate sample size. This necessitates collaboration among multiple research groups and the use of high-quality biological samples which have been collected and using consistent and precise approaches. This will provide international datasets that are transparent (Gracie et al., 2011) and will enable researchers to address unanswered questions and validate their results. We encourage scientists to create and join integrated databases and multicentre international networks, to advance the knowledge and potential biomarker search in health and disease conditions related to human endometrium. The World Endometriosis Research Foundation recent EPHect effort is a good example of phenotyping patients with an endometrial disorder. There are multiple other opportunities for ‘omics’ studies on endometrium in this regard.

Although the ‘omics’ technologies have high potential to deliver, there are still important technical limitations and constraints, including data analysis, that need further development. In research of human endometrium, one important technical consideration is the analysis of very small amounts of sample material. While there are a range of different commercial kits for sample material amplification, where linear amplification is assumed, it is well known that amplification creates errors (Vanneste et al., 2012). Furthermore, the trend towards analysis of endometrial

Figure 3  High-confidence embryo-endometrium interaction network derived from protein–protein interaction data and literature curation. Node colour represents tissue-specific differential gene expression: blue, expressed in embryo; red, expressed in endometrium; grey, expressed in both tissues. The biggest interaction network highlights the importance of cell adhesion molecules, where integrins, collagens and laminins are present. The second largest interaction network represents proteins involved in cytokine–cytokine receptor interactions, where osteopontin, apolipoprotein D, leptin (LEP) and leukaemia inhibitory factor (LIF) pathways intertwine (from Altmäe et al., 2012; published with permission from Molecular Endocrinology).
compartments in isolation for understanding better the function of each compartment, requires prior sample amplification or FACS-isolated endometrial cell populations, or highly sensitive analytical tools such as those emerging for proteomic analyses. Moreover, technical advances are towards high-throughput single-cell analysis, which will better characterize cell populations and provide spatiotemporal resolution in systems biology (Mannello et al., 2012). However, next-generation single-cell analysis technology is still in its infancy, with a plethora of artefacts remaining (Prof. Thierry Voet, personal communication). Thus, it is important to be aware of and acknowledge the limitations of current technologies, while trusting that future developments will overcome these shortcomings.

Thus, the future of the endometrial ‘omics’ field lies in well designed, sufficiently powered studies together with the application of new-generation technologies, complex data analyses and integrated systems biology approaches. Provision of integrated databases and multicentre collaboration will enable new insights, and provide valid and reliable biomarkers. We are on the threshold of realizing the promise of the ‘omics’ technologies in endometrial research.

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

Authors’ roles

Funding
This research was funded by Estonian Ministry of Education and Research (grant SF0180044s09); Enterprise Estonia (grant EU30020), EU-FP7 Eurostars Program (grant NOTED, EU41564) and EU-FP7 IAPP Project (grant SARM, EU324509); Marie Curie post-doctoral fellowship (FP7, no 329812, NutriOmics); Spanish Ministry of Education (Grant no. SB2010-0025); grant from Junta de Andalucía (BIO-302); the NHMRC of Australia (#1002028, #494802), The Monash IVF Education and Research Foundation and the Victorian Government’s Operational Infrastructure Program; and the National Institutes of Health, Eunice Kennedy Shriver National Institute of Child Health and Human Development.

Conflict of Interest
None declared.

References


