

Guidelines for biomarker discovery in endometrium: Correcting for menstrual cycle bias reveals new uterine disorders associated genes

Almudena Devesa-Peiro, M.Sc. & Patricia Sebastian-Leon, Ph.D.; Antonio Pellicer,
Ph.D.; Patricia Diaz-Gimeno, Ph.D.

Supplementary Information

Contents page

Figure S1. Menstrual cycle effect on endometrial gene expression in transcriptomic studies evaluating differences between menstrual cycle phases in women with normal endometrium.

Table SI. Keywords employed at Gene Expression Omnibus (GEO) to search endometrial transcriptomic studies either evaluating uterine disorders or menstrual cycle progression in control women with normal endometrium.

Table SII. Top uterine disorder biomarkers previously masked by the menstrual cycle effect.

Table SIII. Functional annotation for the newly discovered potential uterine disorder biomarkers.

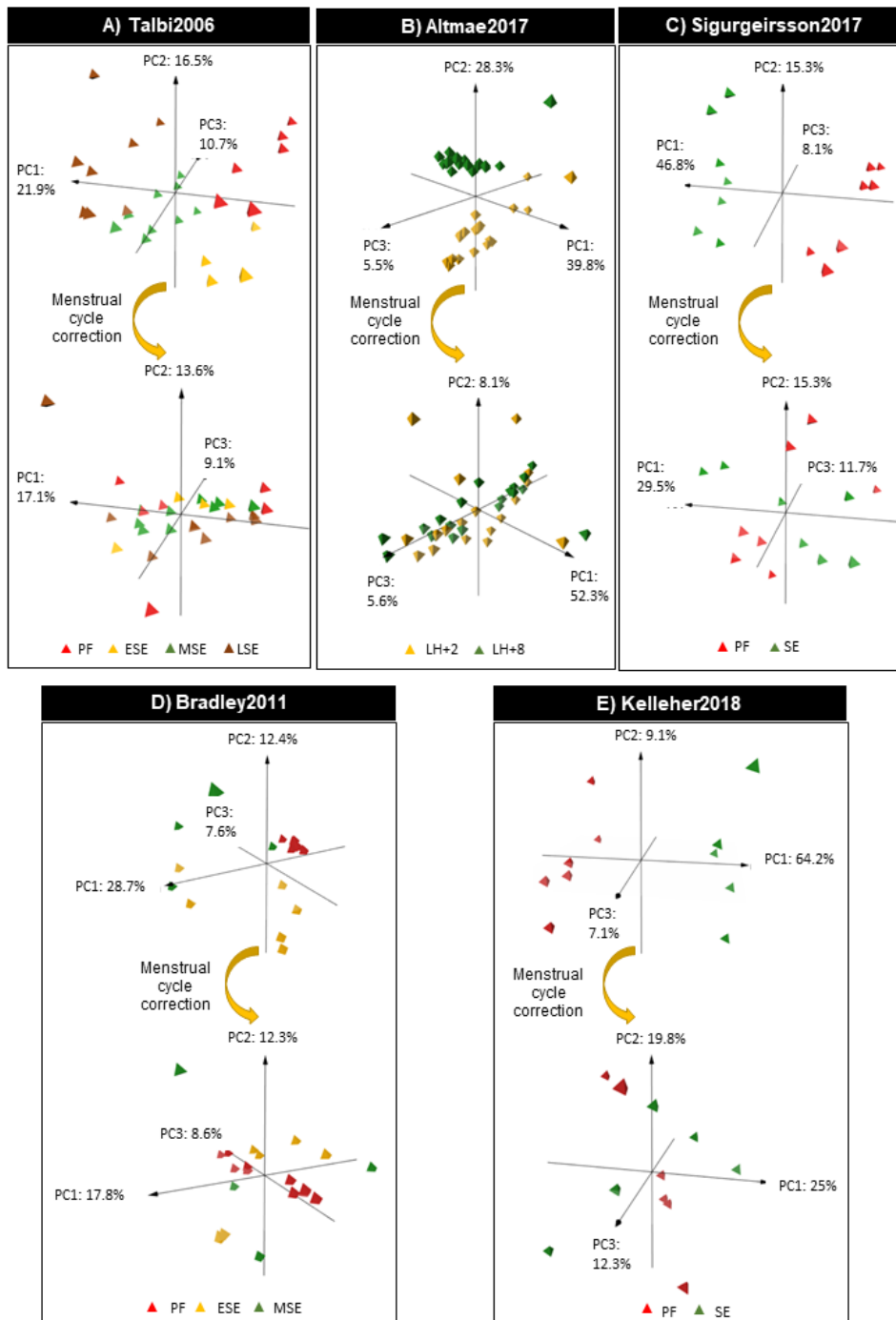


Figure S1. Menstrual cycle effect on endometrial gene expression in transcriptomic studies evaluating differences between menstrual cycle phases in women with normal endometrium. For each study (A-E), the result of the principal component analysis is plotted for the first three components before and after applying the menstrual cycle effect correction method. Based on their gene expression profiles, samples are primarily grouped by the menstrual cycle phase before applying the correction method. However, these groups are no longer present after the menstrual cycle effect has been removed from the data. PC, principal component. PF, proliferative. ESE, early secretory. MSE, mid-secretory. SE, secretory. LH, luteinizing hormone.

Table SI. Keywords employed at Gene Expression Omnibus (GEO) to search endometrial transcriptomic studies either evaluating uterine disorders or menstrual cycle progression in control women with normal endometrium. Study searches were carried out from October 2016 to January 2019 with no restrictions on date or publication language. RIF, Recurrent implantation failure. RPL, recurrent pregnancy loss.

Endometrial pathology	Query
Endometriosis	endometriosis AND "homo sapiens"[Organism]
Uterine myomas	((leiomyo* OR fibroleiomyoma OR leiofibromyoma OR myofibroma OR (uter* AND (fibroid OR myoma OR fibromyoma)) AND "homo sapiens"[Organism]
Adenomyosis	(adenomyo* OR (uter* AND adenomyo*)) AND "homo sapiens"[Organism]
Endometrial leiomyosarcomas	((uter* sarcoma) OR leiomyosarcoma OR adenocarcinoma OR ((uter* OR endometr* OR clear cell OR papilla* OR serous) AND (carcinoma OR adenocarcinoma OR cancer))) AND "homo sapiens"[Organism]
Endometrial adenocarcinomas	((uter* OR endomet*) AND (adenocarcinoma OR carcinoma OR cancer)) AND "homo sapiens"[Organism]
RIF	((recurrent implantation failure) OR RIF OR (implantation failure) OR (unexplained infertility)) AND "homo sapiens"[Organism]
RPL	((recurrent OR habitual) AND (abortion OR miscarriage OR pregnancy loss)) OR RPL) AND "homo sapiens"[Organism]
Endometrium control	(uter* OR endometr*) AND "homo sapiens"[Organism]
Malformations	((uter* OR mülleri*) AND malformations) AND "homo sapiens"[Organism]
Endometrial carcinosarcoma	((uter* OR endometr*) AND carcinosarcoma) OR ((uter* OR endometr*) AND stromal AND (carcinosarcoma OR sarcoma)) OR (malignant mixed Müllerian tumor)) AND "homo sapiens"[Organism]

Table SII. Top uterine disorder biomarkers previously masked by the menstrual cycle effect.

Gene symbol, average expression, t-Student statistic, p-value, false discovery rate (FDR), fold change (FC) and absolute percentage of increased/decreased expression between cases and controls after differential expression analysis with menstrual cycle effect correction are provided for the top 20 (in terms of absolute FC) endometrial biomarkers of A) eutopic endometriosis (Burney2007, n=685), B) ectopic ovarian endometriosis (Hawkins2011, n=302), and C) recurrent implantation failure (Koot2016, n=30) that were previously masked by the menstrual cycle effect. Included new uterine disorder biomarkers (544 for eutopic endometriosis, 158 for ectopic ovarian endometriosis, and 27 for RIF) that were not previously reported in the original articles or in the databases Disgenet v.6 (Piñero *et al.*, 2019), Phenopedia v.6.2.3 (Yu *et al.*, 2010), and/or GeneCards v.4.14.0 (Stelzer *et al.*, 2016) are underlined.

Available at <https://dx.doi.org/10.6084/m9.figshare.13643147>.

Table SIII. Functional annotation for the newly discovered potential uterine disorder biomarkers. For the new potential biomarkers of eutopic endometriosis (n=544) (B), ectopic endometriosis (n=158) (C) and RIF (n=27) (D) not previously reported in the literature, gene names, GeneCards description (Stelzer *et al.*, 2016), annotated GO biological processes, molecular functions and cellular components (Ashburner *et al.*, 2000) and annotated KEGG pathways (Ogata *et al.*, 1999) are indicated. For GO annotation, only experimental-evidenced GO-gene associations were included, a propagated GO version was used for considering the whole GO-tree structure, and annotated GO terms were filtered by those having more than five and less than 500 associated genes. Obtained GO and KEGG annotated terms were then grouped in broader functional categories (A). Table SIIIA shows, for each defined functional group, literature references supporting its role in the pathophysiology of endometriosis and/or RIF; together with the number of associated genes and gene names by study/uterine disorder. Keywords used in PubMed to search for functions altered in endometriosis and RIF patients included “endometriosis”, “RIF”, “recurrent implantation failure”, “function”, “pathway”, “gene ontology” and “KEGG”. GO, Gene Ontology. KEGG, Kyoto Encyclopedia of Genes and Genomes.

Available at <https://dx.doi.org/10.6084/m9.figshare.13643147>.