

Supplementary Data File S1

Supplementary Material and methods Transcriptomic endometrial dating (TED) model

Leveraging an artificial intelligence (AI) machine learning model along with a 73-gene signature, the TED model (Diaz-Gimeno et al., 2024b) classifies endometrial biopsies obtained during the mid-secretory phase into four distinct profiles (C1–C4). While profiles C2 and C3 correspond with the early and late mid-secretory phases (EMSE; LMSE) of the menstrual cycle, profiles C1 (early secretory; ESE) and C4 (late secretory; LSE) are indicative of a displaced mid-secretory phase.

Gene signature discovery

Genes were listed in decreasing order, based on an informativity score, using CorrelationAttributeEval (Witten et al., 2016) in Weka (Frank et al., 2010). To study the predictive performance of different sets of ordered genes, five-fold cross-validation processes with 100 iterations were performed independently with support vector machine (SVM) (Noble, 2006), k-nearest neighbors (kNN) (Zhang, 2016) and random forest (RF) algorithms (Breiman, 2001) using RWeka (Hornik et al., 2009). Among all the outputs, the signature with the highest accuracy and most genes was selected.

Optimization of the supervised balanced probabilistic model

Due to the imbalanced nature of our cohort, reflected by a noticeably higher proportion of good prognosis samples compared to poor prognosis samples, and the immensity and complexity of the entire endometrial transcriptomic gene set, the majority class (good prognosis) had to be undersampled. This step is required to prevent potential biases that may lead to a poor performance of the AI algorithm (i.e. incorrectly identifying the poor prognosis patients or minor class) (Blagus and Lusa, 2015).

Following selection of the endometrial disruption signature in the training set, a balanced probabilistic model was developed through undersampling, which consists of selecting randomly from the largest group (good prognosis) the same number of samples from the smallest group (poor prognosis). For this purpose, the training set was randomly split into 100 balanced subgroups, each containing an equal number of poor and random good prognosis samples. Three algorithms were applied individually [i.e. support vector machine (SVM), k-Nearest neighbors (kNN) and Random forest (RF)], or in pairs (i.e. SVM+kNN, SVM+RF and kNN+RF) (Witten et al., 2016), using the selected signature. This balanced probabilistic model was evaluated in the test set 100 times, to generate a prediction-probability of endometrial disruption for each sample. A mean prediction-probability threshold of 0.5 was then used to reclassify samples as having poor (≥ 0.5) or good (< 0.5) prognosis. To externally validate the optimal prediction model, the accuracy (proportion of true results in the population being tested), sensitivity (proportion of correctly identified positives) and specificity (proportion of correctly identified negatives) were calculated independently in the test set. The algorithm with the best overall performance considering the accuracy, specificity and sensitivity was selected as the optimal prediction model for endometrial disruption.

Rates for reproductive outcomes

The pregnancy rate (PR) was evaluated as the proportion of clinically confirmed pregnancies. The ongoing pregnancy rate (OPR) was determined as the number of ongoing pregnancies divided by the total number of clinically confirmed pregnancies. The clinical miscarriage rate (CMR) was defined as the losses of ultrasound-detected pregnancies prior to 20–24 weeks of gestation over the total number of pregnancies in which a gestational sac was visualized. The biochemical miscarriage rate (BMR) was defined as the number of biochemical pregnancies (detected by positive serum β -hCG values but without visualization of the gestational sac within the first 10 weeks of gestation) over the total number of pregnancies. Finally, the cumulative pregnancy rate (cumulative PR), which represents the total number of embryo transfers required to achieve a successful pregnancy, was calculated considering the pregnancy outcomes of all embryo transfers.

Selection of potential biomarkers for qPCR validation

Genes used for qPCR validation were selected according to the results of differential expression analysis. We selected those differentially expressed genes (DEGs) in each comparison with a high value of fold change and a high value of average expression in endometrial tissue.

Supplementary Results

Clinical and transcriptomic selection

After excluding samples with insufficient tissue ($n = 15$), low RNA quality ($n = 41$) and dropouts ($n = 40$), 195 samples were sequenced. Out of 195 selected samples, two outliers were detected and removed during the data quality control step (Supplementary Figure S1A). Moreover, 62 patients with insufficient embryo transfers for clinical classification were removed leaving a total of 131 samples eligible for stratification.

Batch effect corrections

To ensure the transcriptomic differences were related to endometrial disruption, all significant batch effects were corrected. In particular, batch effects due to the sequencing run (Supplementary Figure S1B) and endometrial timing were detected and corrected (Supplementary Figure S1C). No other batch effects owing to experimental variables were detected (Supplementary Figure S2).

Gene signature associated with poor prognosis endometrium

The optimal gene signature was selected after testing the performance of three AI algorithms (SVM, kNN and RF). While SVM and RF algorithms worked best with gene signatures composed of 76 and 15 genes, respectively, the kNN algorithm resulted in a larger gene signature of 236 genes. Notwithstanding, the accuracy of predictions was 88.1% for SVM, 82.7% for RF and 79.2% for kNN (Supplementary Figure S3).

Selection of the best AI probabilistic model

Data were randomly split into training ($n = 105$) and test ($n = 26$) sets, which were homogeneous in terms of baseline characteristics (Supplementary Table S3). The optimal balanced probabilistic

model was selected after analyzing the performances of each algorithm (kNN, SVM, RF) and their combinations (SVM+kNN, SVM+RF, kNN+RF). While the majority of AI models showed suitable accuracies and specificities (65–81% or 70–95% respectively), the SVM and kNN combination was more sensitive (67%), and the only model to exceed 50% accuracy ([Supplementary Table S4](#)).

References

- Blagus R, Lusa L. Joint use of over- and under-sampling techniques and cross-validation for the development and assessment of prediction models. *BMC Bioinformatics* 2015;**16**:363.
- Breiman L. Random forests. *Mach Learn* 2001;**45**:5–32.
- Frank E, Hall M, Holmes G, Kirkby R, Pfahringer B, Witten IH, Trigg L. Weka—a machine learning workbench for data mining. In: Maimon O, Rokach L (eds). *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer US, 2010, 1269–1277.
- Hornik K, Buchta C, Zeileis A. Open-source machine learning: R meets Weka. *Comput Stat* 2009;**24**:225–232.
- Noble WS. What is a support vector machine? *Nat Biotechnol* 2006;**24**:1565–1567.
- Witten IH, Frank E, Hall MA, Pal CJ. *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington, MA: Morgan Kaufmann, 2016.
- Zhang Z. Introduction to machine learning: K-nearest neighbors. *Ann Transl Med* 2016;**4**:218.