MS analysis protocol

Following data acquisition, raw files (.d format) were processed and visualized using XCMS online software (version 3.7.1). These files were then converted to a standardized format (.mzML) with the ProteoWizard MS Convert tool. Subsequently, control and UI group data (.mzML format) were uploaded as individual data sets to XCMS Online for pairwise comparison. Metabolites with a log2fold change (|log2FC|) greater than or equal to 1 and a statistically significant p-value ($p \leq 0.05$) were considered differentially expressed. To identify affected metabolic pathways, differentially expressed metabolites (increased or decreased in the FF) were obtained from XCMS and formatted according to MetaboAnalyst's (version 6.0) "MS Peaks to Pathways" function. This ensured accurate identification of their origin. The data was uploaded and analyzed with a mass accuracy of 10 ppm. The Mummichog algorithm within MetaboAnalyst identified enriched pathways associated with the identified metabolites using a p-value threshold of 0.001. The KEGG database for Homo sapiens (human) was used as the reference for metabolite annotation and pathway assignment.

Statistical analysis

Comprehensive statistical analysis was employed to determine potential biomarkers linked to group variations in metabolite profiles. XCMS software facilitated the initial analysis, generating a cloud plot visualizing changes (increased/decreased) in metabolite levels between the control and UI group. This plot combined fold-change and p-value for a more comprehensive perspective. MetaboAnalyst software subsequently provided a deeper evaluation using advanced techniques like Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA), and Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA). Further analysis of the OPLS-DA model specifically focused on identifying potential biomarkers. Here, Variable Importance in Projection (VIP) scores assessed each metabolite's contribution to group separation, with scores exceeding 2 indicating potential relevance. Additionally, cross-validation ensured the models weren't overfitted. Consequently, metabolites with a VIP score greater than 2, log2fold change (|log2FC|) greater than or equal to 1 and a statistically significant p-value ($p < 0.05$) were designated as differential metabolites, indicating substantial changes

between the groups and holding promise as potential biomarkers. To gain deeper insights into the biological significance of these significantly different metabolites, pathway analysis was conducted on metabolites exhibiting significant changes. Enrichment analysis aimed to uncover potentially impacted metabolic pathways associated with the observed variations. Metabolites with statistically significant p-values were prioritized for further investigation. Additionally, pathway topology analysis was employed to assess the relative importance of each metabolite within the identified pathways, aiding in selecting potential target metabolites for a more focused exploration. Following the selection of potential target metabolites, a comprehensive literature review using PubMed was conducted, employing search terms like "metabolite name and its role in fertility" to gain a deeper understanding of their biological functions