

A dia-PASEF workflow improves proteome coverage and quantitative performance in repeated-measures analysis of cervicovaginal fluid from a preconception cohort (EARLY-PREG)

Mauricio Hernandez · Pablo Saldivia · Veronica Latapiat · Barbara Antilef · Guillermo Nourdin · Felipe Castro · Cristian Vargas · Elard S. Koch

Received: 20 November 2025 / Accepted: 6 October 2026
© The Author(s) 2026

Abstract

Cervicovaginal fluid (CVF) represents a valuable source of molecular information reflecting time-dependent proteomic changes across the menstrual cycle. However, the large-scale discovery of CVF biomarkers remains challenging because of limitations in the sensitivity, reproducibility, and data completeness of mass spectrometry-based workflows, particularly within repeated-measures longitudinal study designs. Here, we evaluated a standardised data-independent acquisition (DIA)-based proteomics workflow (dia-PASEF) for longitudinal proteomic analysis of serial CVF samples collected across an ovulation-anchored seven-day window (d0-d6) from a prospective preconception cohort (EARLY-PREG). The performance of this dia-PASEF approach was compared with that of a dda-FracOffline strategy. Across the time window, 2,817 quantifiable proteins were detected with dda-FracOffline, whereas 4,229 quantifiable proteins were detected with dia-PASEF, with mean coefficients of variation of $9.37 \pm 1.75\%$ and $5.27 \pm 0.68\%$, respectively. Using day 0 (ovulation) as a reference for day-to-day comparisons, more differentially expressed proteins (DEPs) were identified with dia-PASEF than with dda-FracOffline across pairwise comparisons (952–1,116 DEPs for dia-PASEF versus 161–460 DEPs for dda-FracOffline). Pathway analyses demonstrated broader biological coverage with dia-PASEF. In conclusion, dia-PASEF is a sensitive and reproducible approach for longitudinal repeated-measures proteomic analysis of CVF. This method is adaptable to other biofluids and supports the investigation of time-dependent molecular signatures and associated biological pathways under physiologically relevant conditions in female reproductive system research.

Keywords Cervicovaginal fluid · Proteomics · Dia-PASEF · Menstrual cycle · Biomarkers · Women's health

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

Mauricio Hernandez, Pablo Saldivia and Veronica Latapiat contributed equally to this work.



Introduction

Proteomic analysis of female reproductive tissues is being increasingly recognised as a relevant approach for studying women's fertility and reproductive health. At the molecular level, insights gained from proteomic studies of tissues and biofluids¹ may contribute to our understanding of pathophysiological processes affecting different biological systems. Despite its importance, the proteome and its time-dependent molecular changes in female reproductive fluids and tissues under physiological and pathological conditions remain relatively under-explored^{2–4}. Furthermore, despite the accelerated development of proteomics research, the discovery of reliable biomarkers remains challenging, mainly due to methodological and analytical limitations affecting longitudinal repeated-measures designs⁵.

Cervicovaginal fluid (CVF) is recognised as a relevant biofluid source for biomarker discovery in proteomics research because of its continuity with the female reproductive tract and pivotal role in fertility, pregnancy outcomes, and local immune responses^{6–11,12,13}. As a noninvasive source of biological information, CVF reflects physiological and pathological changes, including the menstrual cycle, pregnancy, and disease progression. Recent studies have linked specific proteins in CVF to endometriosis^{14,15}, pregnancy-related complications^{16,17} and several cancer types, such as ovarian, endometrial, and cervical malignancies^{6–11}.

Recent advances in proteomics techniques have substantially advanced the identification of protein and quantification of their levels. However, most studies rely on data-dependent acquisition (DDA), a time-consuming workflow with limitations in quantitative capabilities, reproducibility, and accuracy, particularly in large-scale longitudinal analyses with repeated measures over time¹⁸. In the search for new strategies for more extensive proteome coverage, data-independent acquisition (DIA)-based methodologies have recently been developed¹⁹. For example, Njoku and collaborators⁷ recently applied a DIA method using sequential window acquisition of all theoretical fragment ion mass spectra (SWATH-MS) to analyse CVF samples, which enabled the quantification of the levels of 2,425 proteins, a greater number of proteins than could be quantitatively analysed via a standard DDA approach.

Addressing the limitations of DDA methods is important for advancing the epidemiological and clinical applications of CVF proteomic analysis, particularly through longitudinal cohort studies^{20,21} that enable the identification of biomarkers and the characterisation of time-dependent pathophysiological changes^{22,23}. In fact, following biomarker variation over time may reveal molecular signatures linked to specific biological processes, offering more comprehensive insights into female reproductive health and disease. Despite the potential of DIA, studies involving the use of spectral library-based DIA to analyse CVF are relatively rare compared with studies involving the use of traditional DDA, which has long been the dominant approach^{24,25}. Furthermore, to date, no proteomic studies have systematically evaluated longitudinal changes in the CVF proteome across consecutive days of the human menstrual cycle.

Although SWATH-MS has been explored as an alternative approach to DIA to assess the CVF proteome, its reliance on fixed isolation windows may introduce spectral interference that potentially limits resolution. With respect to new technologies that address this limitation, in 4D proteomics, enabled by timsTOF mass spectrometry (MS), trapped ion mobility spectrometry (TIMS) is integrated with parallel accumulation–serial fragmentation (PASEF), increasing proteome resolution and coverage^{26–28}. In this context, DIA strategies such as parallel accumulation–serial fragmentation combined with data-independent acquisition (dia-PASEF) have been applied as quantitative research tools²⁹ for complex biofluids such as plasma; in this strategy, ion mobility enables improved separation of ions on the basis of their shape, size, and charge²⁶.

In this study, we aimed to implement and validate an optimised and reproducible dia-PASEF method for time-dependent proteomic analyses of serial CVF samples collected throughout the menstrual cycles of women participating in the EARLY-PREG open longitudinal preconception cohort (NCT07358026)³⁰. The dia-PASEF approach was compared with the traditional DDA approach coupled with offline peptide fractionation (dda-FracOffline). We hypothesised that the dia-PASEF approach would provide greater proteome coverage and reproducibility than

traditional DDA, supporting its potential application for biomarker discovery in repeated-measures longitudinal studies of women's fertility, reproductive health, and pregnancy.

Methods

EARLY-PREG cohort biobank

EARLY-PREG (ClinicalTrials.gov identifier NCT07358026) is a preconception longitudinal open cohort designed to investigate female menstrual cycles, fertility and maternal–embryonic communication at the early stages of pregnancy during the peri-implantation window. The EARLY-PREG biobank is a longitudinal biobank that constitutes a structured and temporally indexed repository of multiple maternal biological samples, including CVF, urine, blood fractions and saliva. In this MS validation study, synchronised and well-documented CVF samples from the EARLY-PREG biobank were used to compare dia-PASEF with the traditional DDA approach coupled with offline peptide fractionation (dda-FracOffline).

Longitudinal CVF samples

At present, the EARLY-PREG biobank includes different biofluids from female volunteers ($n=223$) collected daily during their menstrual cycles from the end of menstruation, which represents the beginning of a new cycle. Ovulation was determined by quantifying the surge in luteinising hormone (LH) levels in urine. Samples were collected from the day after the end of menstruation until the 14th day post-ovulation, as determined by the surge in LH levels during each cycle, implementing a previously described method³¹ used to detect ovulation within ± 1 day (95.7% confidence). LH concentrations were determined through a radioimmunoassay (IRMA) (Barnafi-Krause Laboratory, Santiago, Chile). Through this approach, the day of the surge in LH levels was defined as day 0 (d0), which served as a cycle-anchored time reference for synchronising repeated measures across individuals.

Sample size

This study was designed as a methodological validation of the dia-PASEF workflow to assess its ability to capture longitudinal proteomic changes in cervicovaginal fluid (CVF) during the human menstrual cycle, using offline-fractionated DDA-MS as the reference analytical approach rather than as a population-based biomarker discovery study. Consequently, the primary objective was to maximise internal validity and analytical comparability while minimising biological variability. First, to enable within-individual repeated-measures analyses, we predefined an ovulation-anchored luteal phase window comprising seven consecutive days (d0–d6). Participants were selected exclusively on the basis of the availability of CVF samples for all days across this period. Women with a luteal phase shorter than seven days or for whom stored CVF samples from the selected window were missing were excluded. Second, to further reduce biological heterogeneity and heteroscedasticity in longitudinal analyses, participants were selected according to strict matching criteria, including age, body mass index, menstrual cycle length, bleeding duration, completeness of longitudinal sampling, and other relevant reproductive characteristics. Finally, only women with samples from non-conception cycles available in the EARLY-PREG biobank were selected to avoid pregnancy-related physiological changes during the analysed window; ultimately, the final analysed cohort comprised six participants (42 longitudinal CVF samples). The characteristics of the selected participants are summarised in Table 1.

Table 1 Characteristics of the women whose samples were used in the proteomic experiments.

Factors	Women (n=6)					
	P1	P2	P3	P4	P5	P6
Age (mean±SD)	32.3±3.6					
Weight (kg)	62	57	66	67	60	74
Height (m)	1.52	1.60	1.68	1.61	1.58	1.7
BMI ¹	26.8	22.2	24.4	25.8	24	25.6
Waist circumference (cm)	72	64	66	70	68	82
Age at menarche (mean±SD)	12.7±1.2					
Menstrual cycle length ² (days)	28	30	24	30	29	33
Menstruation length ³ (days)	5	6	6	6	6	6

¹Body mass index (BMI) was calculated as weight (kg) divided by height squared (m²).

²Menstrual cycle length refers to the number of days from the first day of menstruation to the first day of the next cycle.

³Menstruation duration corresponds to the number of days of menstrual bleeding per cycle.

Protein extraction and processing for MS analysis

CVF aliquots (100–500 µL) were thawed on ice and combined with equal volumes of 100 mM NH₄HCO₃ and 100 mM DTT. The mixtures were homogenised by vortexing, incubated at 95 °C for 15 min, and cooled on ice for 5 min. The samples were then subjected to cold ultrasound disruption at 4 °C (10-second on/off pulses at 50% amplitude) and subsequently centrifuged at 19,000 × g for 20 min at 4 °C. The supernatant (soluble phase) was carefully separated, while the pellet (insoluble phase) was retained for further processing.

The soluble phase was concentrated in Vivaspins tubes (10,000 MWCO filter, Sartorius), washed with 400 µL of ultrapure water, and centrifuged at 12,000 × g for 10 min at 4 °C. The filtrate (<10,000 Da) was dried in a SpeedVac and stored at –20 °C, while proteins retained on the filter (>10,000 Da) were recovered by washing with ~600 µL of water, combined, and dried. The insoluble pellet and soluble protein concentrate were resuspended in 8 M urea with 25 mM NH₄HCO₃. The mixture was sonicated on ice and centrifuged at 19,000 × g for 15 min at 4 °C. The protein content was quantified using a Qubit protein assay (Invitrogen, #Q33212).

For MS preparation, 300 µg of total protein from the soluble and insoluble phases was reduced with 20 mM DTT in 25 mM NH₄HCO₃ at 25 °C for 1 h. Alkylation was performed with 20 mM iodoacetamide in 25 mM NH₄HCO₃ in the dark for 1 h. The samples were diluted eightfold with 25 mM NH₄HCO₃ and digested with sequencing-grade trypsin (#V5071; Promega) at a 1:50 protease/protein ratio for 16 h at 37 °C. Digestion was stopped by the addition of 10% formic acid. Peptides were cleaned using Sep-Pak C18 Spin columns (Waters), dried in a rotary concentrator at 40 °C, and stored for subsequent MS analysis.

dda-FracOffline

Peptide fractionation

Offline high-pH reversed-phase fractionation of CVF peptides was performed using a protocol adapted from our previously published tear proteomics workflow³². A total of 200 µg of tryptic peptides from soluble and insoluble CVF phases were separated using an ÄKTA avant 25 system, with refrigerated fraction collection at 6 °C. Peptides were loaded onto a BEH reversed-phase column (2.1 mm × 5 cm, Waters) at a pH of 10 and a flow rate of 0.2 mL/min. Buffer A consisted of 5 mM ammonium formate in water, whereas buffer B consisted of 5 mM ammonium formate in 90% ACN. A binary gradient in which the buffer B concentration increased linearly from 3% to 40% over 30 min and then to 60% over 15 min and to 85% in the final 5 min was applied. We collected a total of 24 fractions (400 µL each), which were vacuum-dried and then reconstituted in water containing 2% ACN and 0.1% FA. These fractions were concatenated into 8 final fractions.

Mass spectrometry

nLC–MS analysis was performed on a nanoElute (Bruker Daltonics) system coupled online with a hybrid TIMS-quadrupole TOF mass spectrometer, timsTOF Pro (Bruker Daltonics, Germany), using a nanoelectrospray ion source (CaptiveSpray, Bruker Daltonics)³³. For long-gradient nLC–MS/MS runs, we followed the standard protocol described by Meier et al., 2021. Approximately 200 ng of peptides were separated on an Aurora series reversed-phase column (25 cm × 75 µm ID, 1.9 µm; IonOpticks, Australia) at a flow rate of 400 nL/min in an oven compartment heated to 50 °C. For the analysis of off-line fractionated CVF samples, we applied a multistep gradient beginning with an increase in the buffer B concentration from 2% to 17% over 60 min, to 25% in 30 min, to 37% in 10 min, and finally to 95% in 10 min (held for 10 min). The column was equilibrated using 4 volumes of solvent A. The mass spectrometer was operated in data-dependent PASEF³⁴ mode, including 1 TIMS-MS survey and 10 PASEF MS/MS scans per acquisition cycle. Mobility values were acquired across a 1/K0 range of 1.6 to 0.6 Vs cm⁻², using equal ion accumulation and a ramp time of 100 ms each for the dual-TIMS analyser.

Data searching

Computational processing of the DDA data was performed using FragPipe v19.1 software. The pipeline integrates the MSFragger (version 3.7)³⁵ search engine for peptide-spectrum matching, Philosopher (version 3.4.13)³⁶ for FDR control and validation, and EasyPQP (<https://github.com/grosenberger/easypqp>; version 0.1.9) for spectral library generation. Raw (.d) files were used as inputs for MSFragger searches. Protein identification was conducted against the reference proteome of *Homo sapiens* (UP000005640) from UniProt/SwissProt (reviewed sequences only; downloaded on Feb. 15, 2021), supplemented with a database of common contaminant proteins in MS experiments available in FragPipe. Decoy sequences were generated using reversed protein sequence entries to enable target-decoy estimation. For MSFragger analysis, the mass tolerance for precursor and fragment ions was set to 20 ppm. The enzyme specificity was defined as ‘strict trypsin’, allowing full tryptic cleavage and a maximum of two missed cleavages. Carbamidomethylation of cysteine was included as a fixed modification, whereas methionine oxidation and N-terminal acetylation were set as variable modifications. The final spectral library was filtered at 1% FDR at both the peptide and protein levels.

dia-PASEF

Mass spectrometry

The timsTOF Pro mass spectrometer was operated in dia-PASEF mode with the following settings: mass range, 100 to 1700 m/z; ion mobility range, 1/K0 start, 0.57 V/cm²; end, 1.47 V/cm²; ramp and accumulation times, 100 ms; capillary voltage, 1600 V; dry gas, 3 L/min; and dry temperature, 160 °C. The dia-PASEF ion mobility parameters were as follows: Each cycle consisted of one MS1 full scan and 64 MS2 windows covering 400–1200 m/z and 0.57–1.46 1/K0. Each window was 25 Th wide by 0.2 V s/cm² high. There was an overlap of 0.05 in 1/K0 and no overlap in m/z. The cycle time was 1.80 s. The CID collision energy ranged from 20 to 52 eV as a function of the inverse mobility of the precursor²⁴. Each fraction was injected into a nanoELUTE nanoliquid chromatography system (Bruker Daltonics). Peptides (200 ng of digest) were separated within 60 min at a flow rate of 400 nL/min on an Aurora Series CSI reversed-phase column (25 cm × 75 µm i.d., C18, 1.6 µm; IonOpticks, Australia) at 50 °C. Mobile phases A and B were water and acetonitrile with 0.1 vol% FA, respectively. The B concentration was linearly increased from 2 to 17% within 37 min, followed by an increase to 25% within 15 min and further to 35% within 8 min, followed by a washing step at 85% B and re-equilibration.

Data search

To construct the DDA spectral library, a pool of tryptic peptides from all the samples was used. For peptide fractionation, mass spectrometry and data searching, the methods described in the DDA section were used (see Methods Sect. 5.1, 5.2 and 5.3). Proteomic data analyses were performed on a high-performance computing server with dual AMD EPYC 7552 48-core processors (2.20 GHz) and 512 GB of RAM. Raw dia-PASEF data were processed via DIA-NN v1.7.15 software³⁷ using the spectral library generated in the previous step. Mass accuracy thresholds were set to a default average of 10 ppm for MS1 and MS2 spectra. Match-between-runs (MBR) was enabled; the quantification mode was set to “any LC (high accuracy)”. The default settings were used for all remaining parameters. Identified proteins were filtered using a precursor-level q value < 1% and a global protein-level q value < 1%.

Bioinformatics and statistical analysis

The intensity matrix reports from MSFragger and DIA-NN were exported and processed in the R statistical environment. The intensity values for each run were log₂-transformed and normalised using quantile normalisation across samples. Missing values were imputed via the missForest algorithm³⁸, applying the criterion that the intensity value should be present in at least 50% of the samples per analysed category³⁹. Absolute and relative numbers (%) are reported for longitudinally differentially expressed proteins (DEPs) along the menstrual cycle, and significant differences were determined using Bayesian t tests for pairwise comparisons between days (e.g., d0 vs. dn, d1 vs. dn), considering the normalised intensity means of biological replicates⁴⁰. Proteins were considered differently expressed at $p < 0.05$. Exploratory analysis and data visualisation were performed in R (v.3.6.0) using EnhancedVolcano (v.1.2.0), ComplexHeatmap (v.2.0.0), and standard base packages. Because the primary comparisons were performed within individuals across time, each participant served as her own control, reducing between-subject variability and increasing sensitivity to detect temporal proteomic changes. The coefficients of variation were determined using the statistically significant differential expression values provided by dia-PASEF and dda-FracOffline. Quantifiable proteins were defined as those that were consistently detected across each condition and presented measurable relative abundance values. Principal component analysis (PCA) of the intensity matrices was performed. For each sample, the total signal intensity was calculated as the sum of the protein intensities and log₁₀-transformed. Pearson correlations between log₁₀-transformed total intensity values and PC1 and PC2 scores were calculated. The relative contributions of total signal intensity, participant, and sampling day were estimated using multiple linear regression. Total signal intensity was used as a sample-level technical proxy and was not considered an independent measure of instrument performance.

Pathway and tissue enrichment analysis

Pathway enrichment analysis was performed using ReactomePA (v.1.42.0), for which the human Reactome database is queried with log₂-fold change (logFC) and p value data obtained for each statistical comparison. For the analysis of enrichment across human systems, tissues and cells, we used the TISSUES 2.0 database⁴¹ and the same statistical parameters used for pathway analysis through topGO (v.2.5.0). Boxplots were generated to show the distribution of protein intensities using ggpubr (v.0.6.0). Statistical comparisons between groups were conducted using t tests and chi-square tests with rstatix (v.0.7.2). Adjusted p values were calculated without additional correction, and significance levels are included directly on the graphs. The z score was calculated on the basis of the logFC of each protein belonging to the designated reactome and tissue categories using the GOplot (v.1.0.2) package. Importantly, this z score is not equivalent to the conventional z score; instead, it provides a directional estimate indicating whether the proteins associated with a biological process, molecular function, or cellular component tended to be upregulated (positive value) or downregulated (negative value). The z score was

determined on the basis of the number of proteins assigned as upregulated ($\log FC > 0$) and those classified as downregulated ($\log FC < 0$), labelled “up” and “down”, respectively, in the dataset.

Panel of selected biomarkers

To assess the efficiency of dia-PASEF in biomarker detection, an FDA-proposed biomarker panel derived from a previously published study⁴² was used to directly compare the ability of dia-PASEF and dda-FracOffline to detect these clinically relevant proteins and quantify their abundance. In addition, we conducted an extensive literature-based curation of candidate proteins associated with physiological and pathological conditions affecting the female reproductive tract. Searches were performed using keywords such as “menstrual cycle”, “ovarian carcinoma”, “endometrial carcinoma”, “cervical carcinoma”, and “endometriosis”, integrating evidence from proteomic, transcriptomic, and genomic studies.

Ethics approval and consent to participate

The EARLY-PREG protocol was approved by the Ethics Committee of the Servicio de Salud de Concepción, Chile (CEC: 17-03-06; NCT Number: NCT07358026). All participants provided written informed consent prior to enrolment. All procedures were conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines.

Results

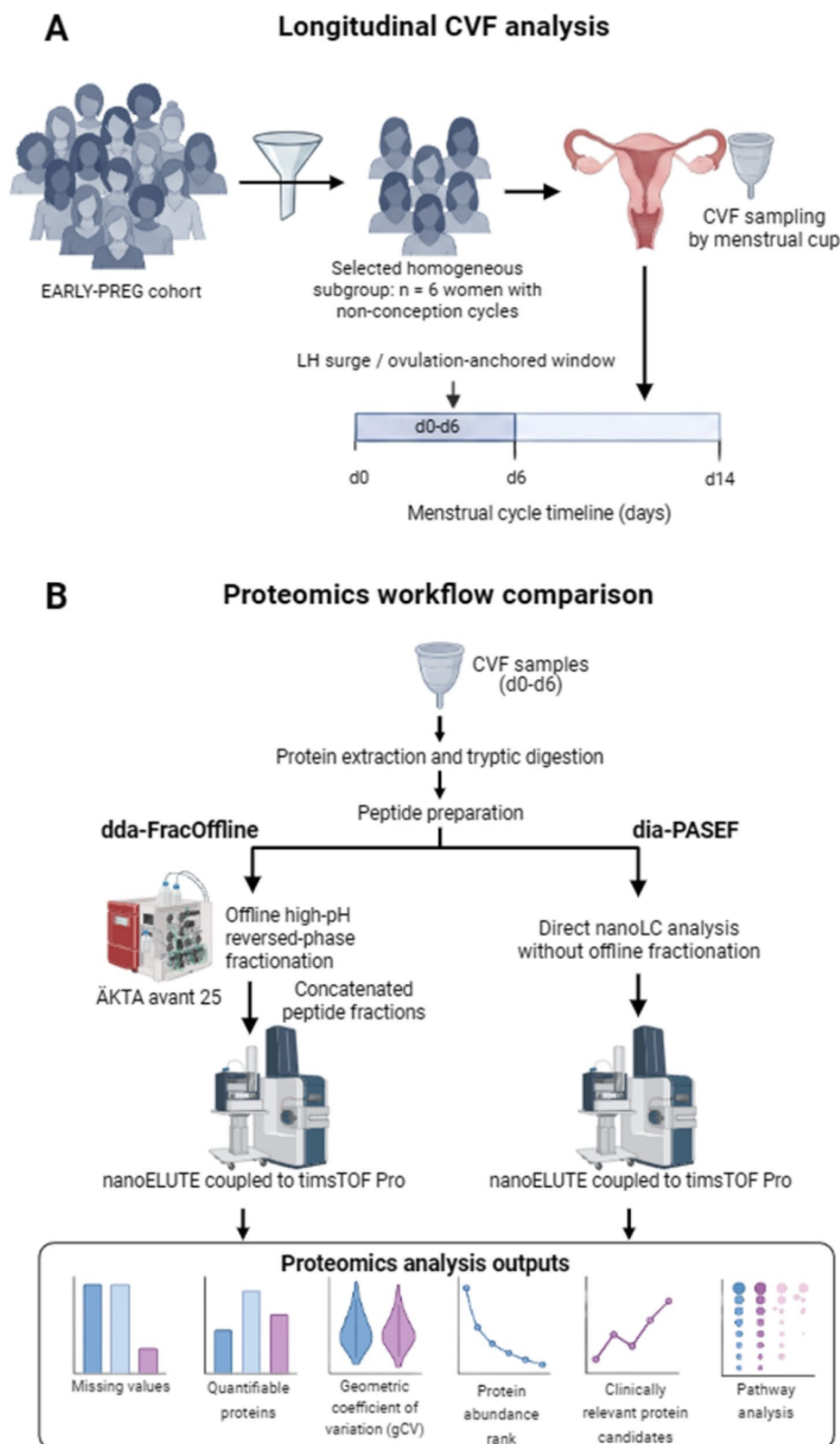
Characterisation of participants and sampling

The detailed characteristics of the participants are provided in Table 1. The length of the menstruation phase of the cycle (5.8 ± 0.4), the total length of the menstrual cycle (29.0 ± 3.0), age (32.3 ± 3.6), and other variables were within comparable ranges (Fig. 1A). Forty-two CVF samples ($n=42$) stored at -80°C were processed for protein extraction and tryptic digestion and then analysed using the two proteomic workflows (Fig. 1B) as described in the Methods section. All 42 samples were analysed on a timsTOF Pro mass spectrometer over a two-week acquisition period. This included 336 chromatographic runs using dda-FracOffline and 42 direct runs using dia-PASEF. Thus, for each protocol, we obtained a 7-day time series ($d0$ to $d6$) of CVF proteomes for each woman. For each analytical workflow, this design yielded a within-individual repeated-measures time series comprising a 7-day ovulation-anchored window ($d0$ – $d6$) of CVF proteomes per participant.

Quantitative characteristics of the dda-FracOffline and dia-PASEF datasets

We conducted a comparative analysis focusing on dataset-level quantitative characteristics, including intersample correlations, the proportion of missing values, and protein identification and quantification performance. First, pairwise correlations between biological samples were assessed as a descriptive measure of dataset consistency. The mean correlation coefficient was approximately 0.70 for dda-FracOffline and 0.90 for dia-PASEF (Figure S2). Second, we evaluated the proportion of missing values across samples. The dia-PASEF workflow yielded a substantially lower proportion of missing values than dda-FracOffline, with some dda-FracOffline samples reaching up to 75% missing values (Fig. 2A). Third, correlation analysis across samples demonstrated lower intersample consistency for dda-FracOffline than for dia-PASEF (Fig. 2B). Next, across the ovulation-anchored time series ($d0$ – $d6$), 1,491 and 7,183 proteins across time points were identified by dda-FracOffline (Fig. 2C), whereas between 3,811 and 5,398 proteins across time points were detected by dia-PASEF (Fig. 2D). Finally, we analysed protein identification and quantification performance. A total of 9,593 proteins were identified by

Fig. 1 Proteomic workflow for the longitudinal repeated-measures analysis of CVF. Schematic representation of the experimental design and strategies used to analyse CVF samples from women in the EARLY-PREG cohort. **(A)** CVF samples were collected across a 7-day ovulation-anchored window (d0–d6) of the menstrual cycle, enabling within-individual repeated-measures analysis. **(B)** Two nLC–MS/MS proteomics approaches were compared: dda-FracOffline and dia-PASEF. Following common protein extraction and tryptic digestion techniques, peptide preparations were split between the two analytical workflows. dda-FracOffline included offline high-pH reversed-phase fractionation, whereas dia-PASEF involved direct analysis without offline fractionation. Both workflows included mass spectrometry analysis using the nanoELUTE system coupled to a timsTOF Pro instrument. The final box provides a comparative evaluation of both methods, including missing values, the number of quantifiable proteins, the geometric coefficient of variation (gCV), the abundance rank distribution, clinically relevant protein candidates, longitudinal proteomic changes, and pathway enrichment analysis. Dataset-level analysis revealed a stronger association between PCA variation and total signal intensity in dda-FracOffline than in dia-PASEF, particularly for PC1 ($r=0.90$ vs. 0.24). Consistently, total signal intensity accounted for 29.6% of the modelled variation for dda-FracOffline compared with 6.5% for dia-PASEF, with additional contributions from participant and sampling day (Figure S1).



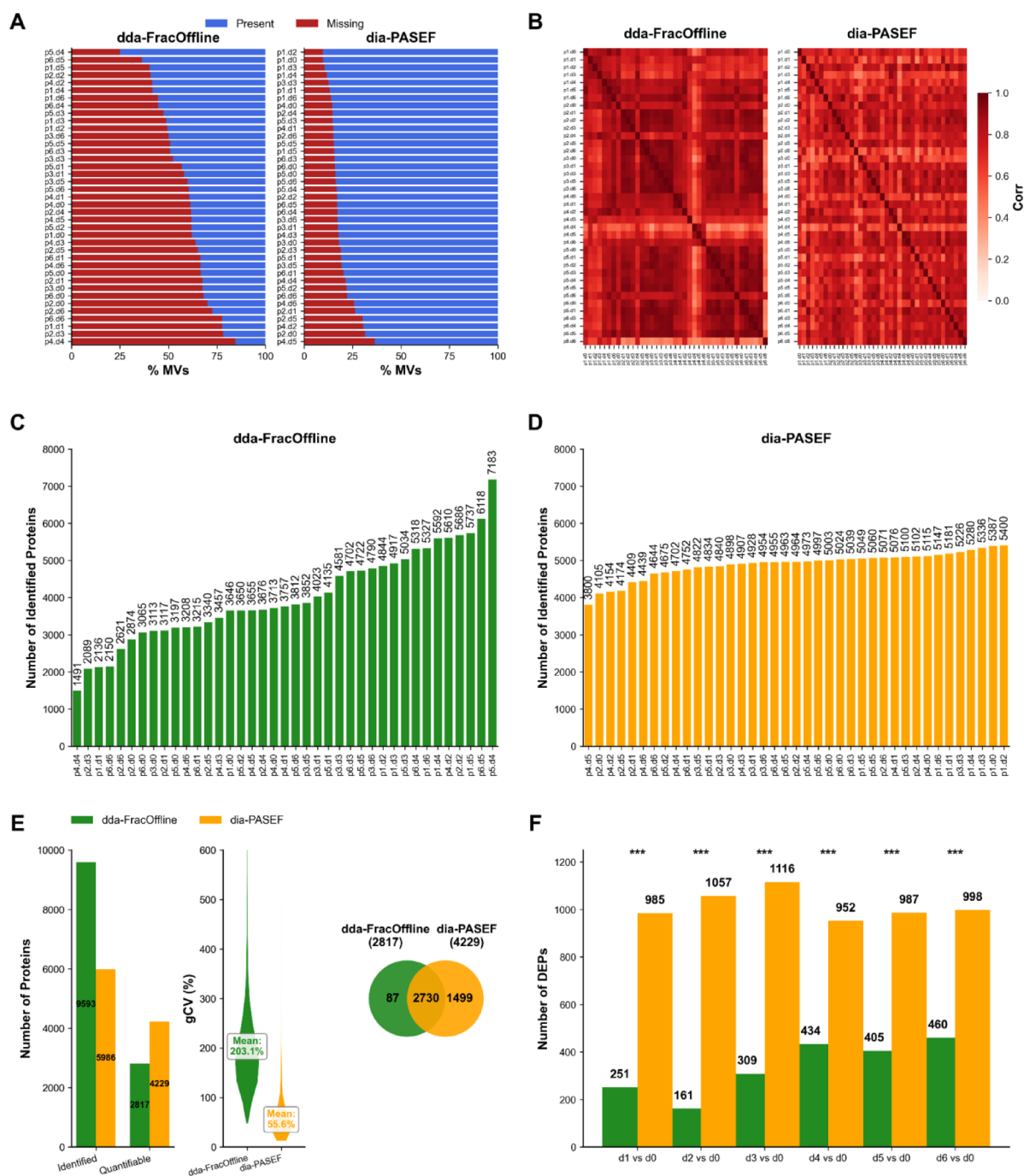


Fig. 2 Quality control of proteomic data obtained by dda-FracOffline and dia-PASEF. **(A)** Bar plots of the proportion of missing values (% MVs) for each sample, showing present values (blue) and missing values (red). **(B)** Pearson correlations between chromatographic runs: positive correlations (red) and no correlations (white). **(C–D)** Number of identified proteins per sample: green bars correspond to dda-FracOffline **(C)**, and orange bars correspond to dia-PASEF **(D)**. **(E)** Overview of identified and quantifiable proteins: bar plots showing the total numbers of identified proteins (left) and quantifiable proteins (centre) for dda-FracOffline (green) and dia-PASEF (orange). The violin plot (right) shows the geometric coefficient of variation (gCV) distributions across quantifiable proteins. The Venn diagram presents the shared and unique proteins identified by each analytical strategy. **(F)** Number of differentially expressed proteins (DEPs) identified by each strategy through longitudinal comparisons at d0. The bar plot displays the total number of DEPs obtained by each strategy, with statistical significance (***, p value < 0.001) indicated above the bars. Statistical significance was determined using chi-square tests for categorical variables. DEPs were identified using dda-FracOffline (green) and dia-PASEF (orange).

dda-FracOffline, whereas 5,986 proteins were identified by dia-PASEF. However, when considering quantifiable proteins (identified and consistently measured across samples), 2,817 proteins were identified by dda-FracOffline, whereas 4,229 proteins were identified by dia-PASEF (Fig. 2E, left), corresponding to a 50% higher quantification yield for dia-PASEF. Overall, both workflows yielded 4,316 unique quantifiable proteins, including 2,730 shared proteins, 87 unique to dda-FracOffline, and 1,499 unique to dia-PASEF (Fig. 2E). The geometric coefficient of variation (gCV) for quantifiable proteins—which was calculated across samples for each protein—was lower for dia-PASEF than for dda-FracOffline (mean 55.6% vs. 203.1%, respectively; Fig. 2E, centre), indicating lower across-sample quantitative variability in the dia-PASEF dataset. Although broader overall protein identification was achieved with dda-FracOffline, dia-PASEF demonstrated greater consistency in protein detection across the ovulation-anchored repeated-measures time series.

Differential protein expression

Following quality control, we assessed the time-dependent changes in differentially expressed proteins (DEPs) across the menstrual cycle to characterise variations in protein abundance in CVF within the ovulation-anchored, repeated-measures 7-day window ($d0$ – $d6$). Comparative analyses revealed significant differences between methodologies in the detection and quantification of DEPs across longitudinal CVF samples (Figure S3). At most time points, compared with dda-FracOffline, dia-PASEF resulted in higher absolute and relative numbers of DEPs, e.g., 985 (23%) vs. 251 (8.9%) between $d0$ and $d1$ and 1,057 (24.9%) vs. 161 (5.7%) between $d0$ and $d2$. From $d0$ to $d6$, dia-PASEF maintained a more stable detection rate (Fig. 2F). These findings indicate that dia-PASEF enables more consistent detection and quantification of differentially expressed proteins in longitudinal repeated-measures designs, facilitating the capture of time-dependent proteomic changes across the menstrual cycle (Fig. 2F and Figure S3).

Longitudinal enrichment analysis of pathways, systems, tissues and cells

To interpret the functional relevance of the quantified proteins, we performed tissue and pathway enrichment analyses. We evaluated tissue representation on the basis of DEPs identified across longitudinal comparisons of CVF collected within the ovulation-anchored repeated-measures window (Figure S4). The analysis revealed enrichment in tissues associated with the female reproductive system, including the uterus and uterine cervix, as well as immune-related tissues. The dia-PASEF workflow demonstrated broader and more consistent tissue coverage, particularly at later time points ($d5$ and $d6$), with elevated enrichment scores ($-\log_{10}$ adjusted p value) and the highest protein densities. We further examined the biological processes associated with the differential protein profiles detected across the 7-day window. Common pathways identified with both strategies between time points $d0$ and $d6$ (Fig. 3) included ‘Formation of the cornified envelope’, ‘Metabolism of proteins’, and ‘Processing of capped intron-containing pre-mRNA’. Additional pathways, such as ‘Signalling by RHO GTPases’ and ‘Mitotic Metaphase/Anaphase Transition’, were related to cell signalling and cell-cycle-related processes (Fig. 3). Compared with dda-FracOffline, dia-PASEF showed increased pathway coverage and enrichment across the time-dependent series.

Identification of physiological and pathological biomarkers

We evaluated the ability of each strategy to identify known biomarkers of physiological and pathological processes and quantify their abundance. We began with a set of disease biomarkers proposed by the U.S. Food and Drug Administration (FDA-proposed biomarkers)⁴², focusing on fluids related to various pathologies and examining literature-curated sets of proteins related to female reproductive-tract pathophysiology (Table S1).

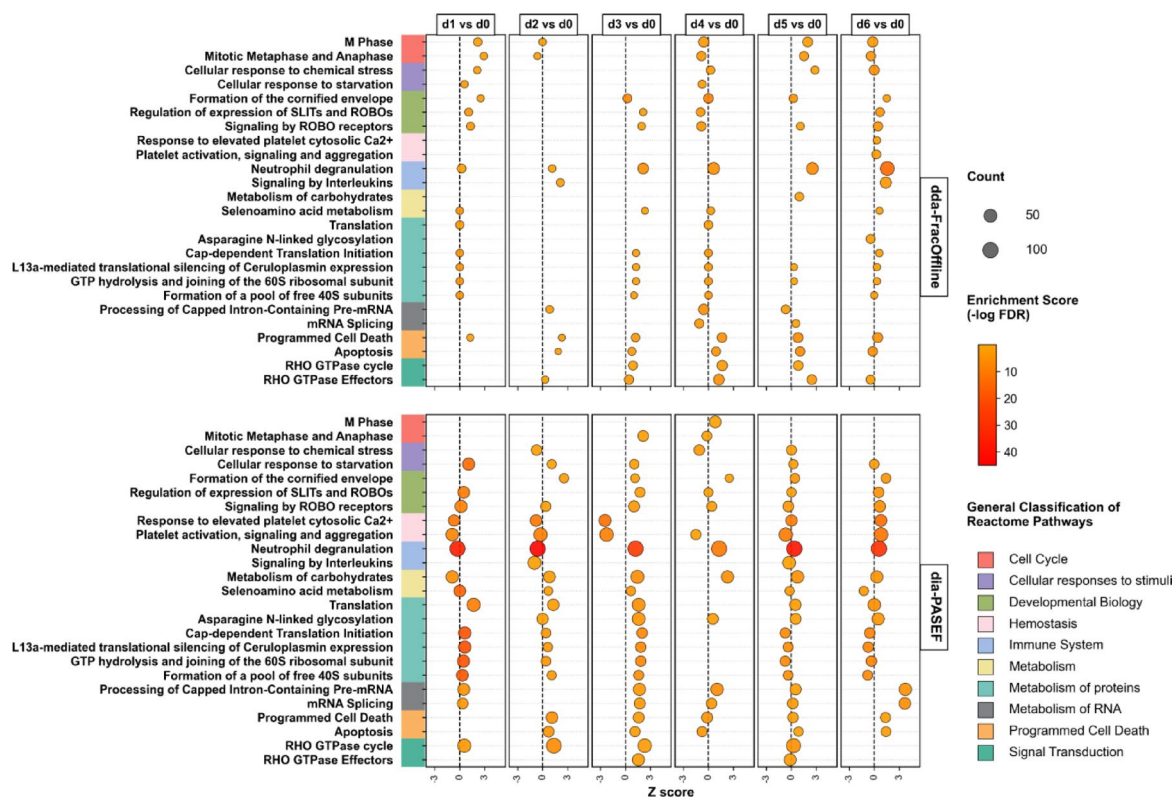


Fig. 3 Differentially expressed proteins and biological pathways identified via longitudinal analysis using dda-FracOffline and dia-PASEF. Reactome enrichment analysis emphasising the most enriched biological pathways. DEPs were identified from longitudinal comparisons for each method. The Y-axis includes Reactome pathways, and the X-axis represents the activation coefficient (Z score) of each pathway.

Comparison of FDA-proposed biomarkers via dda-FracOffline and dia-PASEF

The FDA-proposed biomarker panel comprises 121 proteins detectable in biological fluids and associated with female reproductive pathologies⁴². The ability of each strategy to detect these biomarkers in CVF is compared in Figs. 4 and 92 of the 121 proteins, corresponding to 76% and 64%, were identified by dia-PASEF and dda-FracOffline, respectively; 76 proteins (63%) were identified by both strategies (Fig. 4A). A protein rank-abundance plot demonstrates that dia-PASEF captured proteins across a broader dynamic range than dda-FracOffline (Fig. 4B). Sixteen FDA biomarkers were uniquely identified by dia-PASEF (NT5E, ACE, GLA, C4A, C1QC, C1R, ACP5, PROS1, BRAF, VWF, GALT, MSH2, PKLR, TFRC, TOP2A, and BTB), whereas the protein GFAP was detected only by dda-FracOffline (Fig. 4A). Overall, dia-PASEF showed better quantification consistency and reproducibility than dda-FracOffline for the detection of clinically relevant biomarker candidates in CVF.

Selected biomarkers from literature review

A curated list of biomarkers for pregnancy, cervical carcinomas, ovarian carcinomas, endometrial carcinomas, and endometriosis previously reported in the literature is provided in Table 2. Proteins from this curated list are outlined in black in Fig. 4B. We examined time-dependent changes in the expression of four representative biomarkers from the curated list to illustrate the longitudinal detection capability of the dia-PASEF workflow (Fig. 4C–F; trajectories in Fig. 4G). To illustrate the longitudinal capability of dia-PASEF in detecting clinically relevant proteins over repeated measures, we tracked the expression of insulin-like growth factor-binding protein 4 (IGFBP-4), which is implicated in early pregnancy and foetal growth restriction; MUC1, a mucin contributing

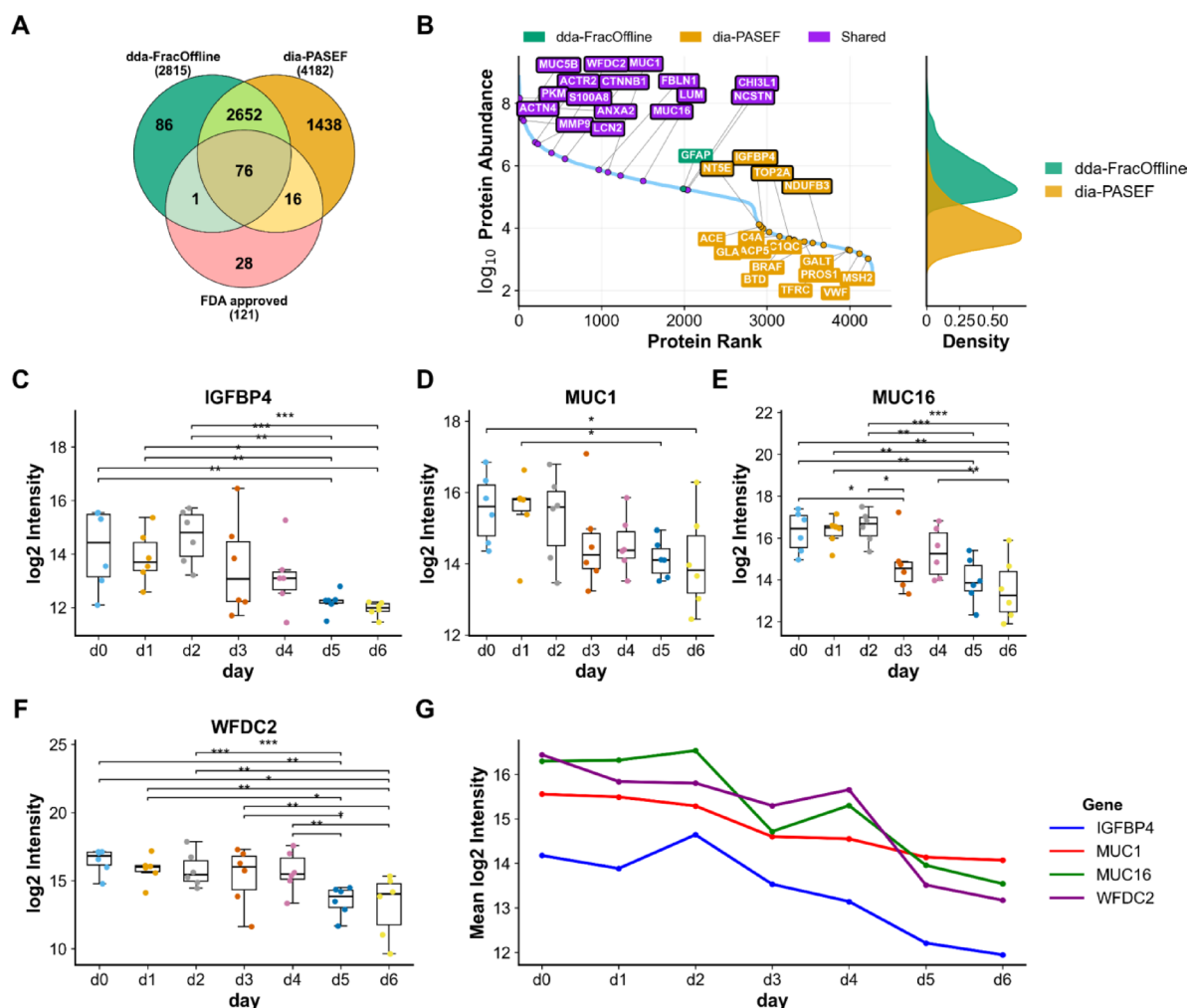


Fig. 4 Biomarker identification and longitudinal protein analysis from *d0* to *d6* by dda-FracOffline and dia-PASEF. **(A)** Venn diagram illustrating the overlap of quantifiable proteins and FDA-proposed biomarkers, highlighting 76 shared proteins and proteins uniquely detected by each strategy. **(B)** Abundance rank plot of proteins whose abundance was quantified in CVF samples via dda-FracOffline alone (green), dia-PASEF alone (orange), and both strategies (purple). FDA-proposed biomarker proteins are highlighted, with the black borders indicating biomarkers included in Table 2. The right panel displays the density distribution of protein abundance. **(C, D)** Box plots showing longitudinal expression patterns of the pregnancy-pathology-associated protein IGFBP4 and the physiological process-related protein MUC1. **(E, F)** Box plots of MUC16 and WFDC2 abundance (***, p value < 0.001; **, p value < 0.01; *, p value < 0.05). **(G)** Trajectory of the longitudinal expression of IGFBP4, MUC1, MUC16 and WFDC2.

to reproductive tract integrity; and the following proteins previously investigated as biomarkers of ovarian cancer: CA-125 (MUC16) and HE4 (WFDC2). Insulin-like growth factor-binding protein 4 (IGFBP4) is associated with early pregnancy and foetal growth restriction (Fig. 4C, Figure S5). MUC1, expressed throughout the female reproductive tract—including the endocervical, ectocervical, and vaginal epithelia—plays a role in maintaining epithelial barrier function (Fig. 4D, Figure S5). Both proteins are physiologically expressed across the menstrual cycle. In addition, the clinically relevant biomarkers CA-125 (Fig. 4E, Figure S5) and HE4 (Fig. 4F, Figure S5), which are encoded by MUC16 and WFDC2, respectively, were consistently detected. Longitudinal analysis across the ovulation-anchored repeated-measures window (*d0*–*d6*) revealed significant day-to-day variation in the expression of all four proteins (Fig. 4C–G; Figure S5), supporting the suitability of dia-PASEF for time-dependent biomarker profiling.

Table 2 Summary of biomarkers associated with female pathologies.

Pathology	Gene Symbol	Reference(s)
Pregnancy pathologies	<i>FBLN1</i>	43,44
	<i>NCSTN</i>	45,46
	<i>ACTR2</i>	45,46
	<i>NDUFB3</i>	47
	<i>IGFBP4</i>	48–50
Cervical carcinoma	<i>MUC5B</i>	51–53
	<i>ANXA2</i>	51–54
	<i>ACTN4</i>	51–54
	<i>PKM</i>	53,54
	<i>MUC16</i>	55–59
Ovarian carcinoma	<i>NT5E</i>	60
	<i>WFDC2</i>	9,58–60
	<i>KLK5</i>	61
	<i>MUC1</i>	62,63
	<i>CHI3L1</i>	56,57
Endometrial carcinoma	<i>MUC16</i>	64–66
	<i>CTNNB1</i>	6,64,67,68
	<i>WFDC2</i>	64,65,68
	<i>MMP9</i>	64,68,69
	<i>TOP2A</i>	70,71
Endometriosis	<i>S100A8</i>	15,72–74
	<i>LCN2</i>	15
	<i>LUM</i>	72,74–76

Discussion

CVF is well established as a valuable biological source for research on the pathophysiology of the female reproductive tract. This biofluid contributes to multiple biological processes, including barrier function, immune responses, microbiota preservation, sperm transport, and menstrual cycle regulation. The emergence of advanced proteomics workflows and MS instrumentation provides opportunities for the large-scale identification of reproductive health biomarkers in CVF⁵. In the present study, the performance of dia-PASEF as a proteomic workflow for analysing CVF in longitudinal repeated-measures designs with time-series sampling was evaluated. To our knowledge, this is among the first applications of dia-PASEF for analysing CVF collected across consecutive days of the menstrual cycle, and the results revealed the improved performance of dia-PASEF relative to dda-FracOffline. Comprehensive comparative analyses revealed that dia-PASEF identified more quantifiable proteins with lower across-sample quantitative variability and greater sensitivity for low-abundance proteins. The reduced variability (gCV=55.6% for dia-PASEF versus 203.1% for dda-FracOffline) and lower proportion of missing values enabled more consistent quantification across samples than dda-FracOffline, improving the robustness of the longitudinal analyses (Fig. 2). This interpretation was further supported by the dataset-level analysis, in which total signal intensity showed a stronger association with the major source of variation for dda-FracOffline than for dia-PASEF. These findings indicate that dia-PASEF supports the reliable tracking of protein abundance over time, facilitating the investigation of time-dependent molecular changes in CVF.

Studies of longitudinal effects are essential for assessing the development of physiological phenomena over time, such as in the earliest stages of pregnancy, which remain poorly understood. These analyses are also relevant for monitoring disease progression and therapeutic responses from a pathophysiological perspective. However, longitudinal analysis of CVF remains technically challenging because of the complexity of this biofluid and the limited availability of well-characterised repeated-measures datasets spanning the menstrual cycle. Thus, the application of advanced acquisition strategies such as dia-PASEF represents a methodological advancement, as

it provides analytical features suited for complex biofluids and improved detection of low-abundance proteins in time-dependent study designs.

The LC–MS/MS acquisition time was substantially shorter for dia-PASEF, for which we used a single 60-min gradient per fraction, than for dda-FracOffline, for which we used a 120-min gradient per fraction. Considering the eight final concatenated fractions, this corresponded to an LC–MS/MS acquisition time of approximately 1 h per sample for dia-PASEF versus 16 h per sample for dda-FracOffline. In addition, the database search time for dia-PASEF was shorter (a few minutes per sample) than that for dda-FracOffline (approximately 1–2 h per sample). These results support the suitability of dia-PASEF for longitudinal repeated-measures proteomic studies.

The higher reproducibility observed for dia-PASEF is consistent with its systematic precursor selection strategy, which enhances proteome depth and sensitivity⁷⁷. In contrast, the sequential precursor selection inherent to dda-FracOffline is associated with greater variability and reduced efficiency in protein sampling, limiting its applicability in large-scale and time-dependent study designs.

The ability of dia-PASEF to capture comprehensive fragmentation data within a single acquisition cycle contributes to improved protein identification, reduced missing values, and broader proteome coverage. The greater quantitative stability observed with dia-PASEF further supports its application in large-scale longitudinal proteomics, in line with previous reports describing the analytical advantages of data-independent acquisition strategies²⁶.

Although both methodologies allow in-depth protein identification, we observed that dia-PASEF enabled more consistent protein quantification than dda-FracOffline, primarily because the latter showed lower quantifiable protein coverage and a higher proportion of missing values²⁹. When dia-PASEF was compared with other DIA approaches applied to CVF, such as SWATH-MS–based methods⁷, a greater number of proteins was identified and quantified. Previous proteomic studies using SWATH-MS for analysing CVF have reported limitations, including reduced sensitivity for low-abundance proteins, peptide cofragmentation effects²⁶, and higher proportions of missing values.

Enrichment analyses enabled exploration of the functional relevance of the identified proteins contributing to CVF pathophysiology (Fig. 3). Both strategies highlighted pathways related to cellular signalling, protein metabolism, immune responses, interleukin signalling, mitotic metaphase–anaphase transition, mRNA splicing, and RHO GTPase signalling, which are involved in reproductive system biology^{78,79}. dia-PASEF showed higher enrichment scores across the time-dependent window, suggesting improved biological coverage. For example, neutrophil gelatinase-associated lipocalin (NGAL; encoded by LCN2), which is involved in neutrophil degranulation¹⁵ and signalling pathways, was consistently detected (Figs. 3 and 4H).

Tissue enrichment analysis confirmed the biological relevance of the identified proteins, highlighting their associations with tissues of the female reproductive system⁸⁰, including the uterus and uterine cervix. These findings are consistent with the ability of dia-PASEF to provide system-specific sensitivity and enable time-dependent biological profiling, particularly for reproductive and immune-related functions in longitudinal CVF proteomic analyses (Figure S4).

We compared proteins identified by dia-PASEF and dda-FracOffline with those in the FDA-proposed biomarker panel⁴². Among the FDA-proposed biomarkers, 76 proteins were detected by both workflows; in addition, dia-PASEF detected 16 unique proteins, whereas dda-FracOffline detected one unique protein (Fig. 4A). The dia-PASEF workflow identified known biomarkers across a broader dynamic range, including both high- and low-abundance proteins (Fig. 4B), supporting its suitability for comprehensive biomarker profiling of CVF. These findings indicate improved detection performance of dia-PASEF for FDA-proposed biomarkers within this longitudinal repeated-measures design.

Low-abundance proteins were linked to biological processes relevant to menstrual cycle regulation and gynaecological diseases. For example, C1QC and VWF are involved in immune responses and vascular regulation. A broader and more uniform dynamic range was observed for dia-PASEF, likely reflecting improved signal-to-noise performance, a feature that is advantageous for longitudinal analyses of complex biofluids³⁴ such as CVF. As an additional strategy for evaluating clinically relevant biomarkers, dia-PASEF facilitated the detection of

low-abundance proteins across the time-dependent repeated-measures window. IGFBP-4 (Fig. 4C), whose levels have been associated with early pregnancy and foetal growth restriction^{48,50}, and MUC1 (Fig. 4D), a mucin expressed throughout the female reproductive tract—including the endocervical, ectocervical, and vaginal epithelia—were consistently detected. MUC1 plays a role in maintaining epithelial integrity and barrier function^{62,63}.

We also detected proteins associated with pathological processes, including alpha-actinin-4 (ACTN4) and annexin A2 (ANXA2), in CVF, which have been reported in previous studies^{51–53} (Table 2). Two proteins clinically relevant to gynaecological carcinomas were identified: cancer antigen 125 (CA-125), which is encoded by the MUC16 gene (Fig. 4E), and human epididymis protein 4 (HE4), which is encoded by the WFDC2 gene (Fig. 4F). These proteins were detected in CVF samples collected across the ovulation-anchored window (*d0–d6*). Both biomarkers have been widely investigated in ovarian^{58,59} and endometrial carcinoma^{64,65}. Longitudinal analysis revealed significant day-to-day variation in the expression of these proteins across the repeated-measures window (Figure S5). These findings suggest that dia-PASEF facilitates time-dependent monitoring of the abundance of clinically relevant proteins in CVF, supporting its potential application in studies of pathologies related to the female reproductive tract.

Proteins present in CVF may also represent promising candidates for longitudinal pathophysiological markers. Because dia-PASEF yielded a reduced proportion of missing values and enabled more consistent time-dependent protein profiling, it generated a more complete quantitative dataset (Fig. 2A). These characteristics support its application for tracking protein dynamics during physiological processes, such as endometrial receptivity, as well as in pathological contexts, over time. A consistent pattern of protein variation was observed across specific days of the cycle (Fig. 4G, Figure S5), providing insight into defined physiological windows within the ovulation-anchored repeated-measures framework. These findings support the longitudinal monitoring of protein abundance in CVF under physiological conditions and suggest potential future applications of dia-PASEF in disease or therapeutic monitoring, which would require validation in appropriate clinical cohorts. No pooled QC or reference samples were included; therefore, instrument performance and batch effects could not be independently assessed. The stronger association with total signal intensity observed for dda-FracOffline may reflect its greater analytical complexity but cannot be attributed specifically to instrument batch effects.

In conclusion, this study provides evidence supporting the suitability of dia-PASEF as an analytical platform for longitudinal repeated-measures proteomic analysis of CVF, addressing limitations inherent to DDA-based workflows and supporting future research on biomarker discovery for fertility, early pregnancy, and gynaecological health and disease. Overall, this approach provides a scalable methodological framework for time-dependent proteomic analyses of complex reproductive biofluids.

Limitations

The principal limitation of this study is the small number of participants, which resulted from the stringent selection criteria applied when selecting individuals from the preexisting EARLY-PREG biobank. However, the study was designed as a methodological validation of the dia-PASEF workflow for longitudinal proteomic analysis of CVF rather than as a population-based biological study. Consequently, the primary objective was to evaluate the analytical performance, reproducibility, proteome coverage of the workflow, and its ability to detect temporal proteomic changes under highly controlled conditions.

Although only six women were included, due to the repeated-measures design, 42 longitudinal CVF samples were collected, with each participant serving as her own control. This design minimised between-subject variability and provided an appropriate framework for comparing the analytical performance of the proteomic workflows. Accordingly, the differentially expressed protein and pathway analyses should be regarded as exploratory and should not be interpreted as population-level biological findings. Larger prospective studies are needed to validate these observations in more diverse cohorts. Nevertheless, this work shows the suitability of dia-PASEF as a robust platform for longitudinal repeated-measures proteomic analysis of CVF and establishes a methodological framework for future biomarker discovery studies in the field of reproductive health.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-75386-3>.

Acknowledgements The authors gratefully acknowledge Dr. Ute Woehlbier of the Center for Integrative Biology at Universidad Mayor for her expert help with the English editing of this paper. We want to thank Estefania Nova-Lamperti from the Molecular and Translational Immunology Laboratory, Department of Clinical Biochemistry and Immunology, Pharmacy Faculty, Universidad de Concepción, Concepción, Chile, for her support in figure design. muestras biotiniladas. I.F.-M. received paid services from Fundación de Investigación San Ramón (FISAR), Concepción, Chile.

Author contributions ESK was the leading author. ESK and MH conceptualised the study. ESK, MH and CV contributed to the study design. MH designed, supervised and performed the proteomics methodology. PS, BA and VL conceived the study. MH, PS and GN developed the experimental system, performed the experiments, and analysed the data. FC and GN supported the experimental work. MH, PS, BA, VL reviewed and commented on the manuscript drafts, and ESK reviewed and commented on drafts and approved the final manuscript. ESK is the guarantor of the study, accepts full responsibility for the research, had access to the data, and controlled the decision to publish.

Funding This work is part of the EARLY-PREG preconception open cohort study and was supported by research grants awarded by Fundación de Investigación San Ramón (FISAR), based in Chile. The pilot study and first recruitment wave were supported by grants #MEL109112011 and #MEL109112011R4 awarded to E.S.K., C.V., and J.F.S. The second recruitment wave was supported by grants #MEL109112011R5 and #MEL131032017R1 awarded to E.S.K. The third wave was supported by grant #MEL205062018 awarded to E.S.K. and M.H. Current funding for the fourth recruitment wave and mass spectrometry research on maternal CVF is supported by grant REH042024-01 awarded to M.H., G.N., and E.S.K.

Data availability Proteomics data can be accessed through the ProteomeXchange Consortium via PRIDE, with the dataset identifier PXD071350.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Beutgen, V. M. & Graumann, J. Advances in aqueous humor proteomics for biomarker discovery and disease mechanisms exploration: a spotlight on primary open angle glaucoma. *Front. Mol. Neurosci.* **17**, 1397461 (2024).
2. Masina, R. & Caldas, C. Precision Cancer Medicine 2.0-Oncology in the postgenomic era. *Mol. Oncol.* **18**, 2065–2069 (2024).
3. Rafelski, S. M. & Theriot, J. A. Establishing a conceptual framework for holistic cell states and state transitions. *Cell* **187**, 2633–2651 (2024).
4. Holmes, C., Carlson, S. M., McDonald, F., Jones, M. & Graham, J. Exploring the post-genomic world: differing explanatory and manipulatory functions of post-genomic sciences. *New. Genet. Soc.* **35**, 49–68 (2016).
5. Kaur Jawanda, I., Soni, T., Kumari, S. & Prabha, V. Deciphering the potential of proteomic-based biomarkers in women's reproductive diseases: empowering precision medicine in gynecology. *Biomarkers* **29**, 7–17 (2024).
6. Martinez-Garcia, E. et al. Cervical fluids are a source of protein biomarkers for early, non-invasive endometrial cancer diagnosis. *Cancers (Basel)*. **15**, 911 (2023).
7. Njoku, K. et al. Comprehensive library generation for identification and quantification of endometrial cancer protein biomarkers in cervico-vaginal fluid. *Cancers (Basel)*. **13**, 3804 (2021).

8. Njoku, K. et al. Detection of endometrial cancer in cervico-vaginal fluid and blood plasma: leveraging proteomics and machine learning for biomarker discovery. *EBioMedicine* **102**, 105064 (2024).
9. Hedlund Lindberg, J. et al. Toward ovarian cancer screening with protein biomarkers using dried, self-sampled cervico-vaginal fluid. *iScience* **27**, 109001 (2024).
10. Köse, B. et al. Identifying biomarkers for cervical neoplasia: A label-free proteomic analysis of cervicovaginal fluid. *Eur. Arch. Med. Res.* **40**, 36–43 (2024).
11. Gutiérrez, A. L. et al. Identification of candidate protein biomarkers for CIN2+ lesions from self-sampled, dried cervico-vaginal fluid using LC-MS/MS. *Cancers (Basel)*. **13**, 2592 (2021).
12. Lacroix, G., Gouyer, V., Gottrand, F. & Desseyn, J. L. The cervicovaginal mucus barrier. *Int. J. Mol. Sci.* **21**, 8266 (2020).
13. Zegels, G., Van Raemdonck, G. A., Tjalma, W. A. & Van Ostade, X. W. Use of cervicovaginal fluid for the identification of biomarkers for pathologies of the female genital tract. *Proteome Sci.* **8**, 63 (2010).
14. Fernandez-Hermida, Y., Grande, G., Menarguez, M., Astorri, A. L. & Azagra, R. Proteomic markers in cervical mucus. *Protein Pept. Lett.* **25**, 463–471 (2018).
15. Grande, G. et al. Cervical mucus proteome in endometriosis. *Clin. Proteom.* **14**, 7 (2017).
16. Zhao, Z. et al. Noninvasive cervicovaginal proteomic biomarkers for spontaneous preterm birth. (2024).
17. Ueda, Y. et al. Hyposecretion of cervical MUC5B is related to preterm birth in pregnant women after cervical excisional surgery. *Am. J. Reprod. Immunol.* **91**, e13832 (2024).
18. Birhanu, A. G. Mass spectrometry-based proteomics as an emerging tool in clinical laboratories. *Clin. Proteom.* **20**, 32 (2023).
19. Fang, H. & Greening, D. W. An optimized data-independent acquisition strategy for comprehensive analysis of human plasma proteome. *Methods Mol. Biol.* **2628**, 93–107 (2023).
20. Stefanakis, K., Samiotaki, M. & Mantzoros, C. S. 7562 longitudinal untargeted proteomics of Leptin administration in lean humans under short- and long-term energy deficiency: Novel physiological insights and pleiotropic immunomodulatory effects from two randomized, placebo-controlled clinical trials. *J Endocr. Soc* **8**, (2024).
21. Viode, A. et al. Longitudinal plasma proteomic analysis of 1117 hospitalized patients with COVID-19 identifies features associated with severity and outcomes. *Sci. Adv.* **10**, ead15762 (2024).
22. Alesi, S. et al. Exploring causal associations between serum inflammatory markers and female reproductive disorders: A Mendelian randomisation study. *Biomolecules* **14**, 1544 (2024).
23. Bruinvels, G., Hackney, A. C. & Pedlar, C. R. Menstrual cycle: The importance of both the phases and the transitions between phases on training and performance. *Sports Med.* **52**, 1457–1460 (2022).
24. Rice, S. J. & Belani, C. P. Optimizing data-independent acquisition (DIA) spectral library workflows for plasma proteomics studies. *Proteomics* **22**, e2200125 (2022).
25. Staes, A. et al. Benefit of in silico predicted spectral libraries in data-independent acquisition data analysis workflows. *J. Proteome Res.* **23**, 2078–2089 (2024).
26. Meier, F. et al. diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition. *Nat. Methods.* **17**, 1229–1236 (2020).
27. Skowronek, P. & Meier, F. High-throughput mass spectrometry-based proteomics with dia-PASEF. *Methods Mol. Biol.* **2456**, 15–27 (2022).
28. Mun, D. G. et al. Four-dimensional proteomics analysis of human cerebrospinal fluid with trapped ion mobility spectrometry using PASEF. *Proteomics* **23**, e2200507 (2023).
29. Bader, J. M., Albrecht, V. & Mann, M. MS-Based Proteomics of Body Fluids: The End of the Beginning. *Mol. Cell. Proteomics: MCP.* **22**, 100577 (2023).
30. Koch, E. S. et al. EARLY-PREG: A preconception longitudinal cohort of women seeking pregnancy to investigate maternal-embryonic molecular signalling during the peri-implantation window. *medRxiv* <https://doi.org/10.1101/2025.05.12.25326372> (2026).
31. Godbert, S., Miro, F., Shreeves, C., Gnoth, C. & Johnson, S. Comparison between the different methods developed for determining the onset of the LH surge in urine during the human menstrual cycle. *Arch. Gynecol. Obstet.* **292**, 1153–1161 (2015).
32. Vera-Montecinos, A. et al. High throughput tear proteomics with data independent acquisition enables biomarker discovery in allergic conditions. *Sci. Rep.* **15**, 31181 (2025).
33. Meier, F. et al. Online parallel accumulation-serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol. Cell. Proteom.* **17**, 2534–2545 (2018).
34. Meier, F. et al. Parallel accumulation-serial fragmentation (PASEF): Multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. *J. Proteome Res.* **14**, 5378–5387 (2015).
35. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods.* **14**, 513–520 (2017).
36. da Leprevost, V. Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nat. Methods.* **17**, 869–870 (2020).
37. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods.* **17**, 41–44 (2019).

38. Stekhoven, D. J. & Bühlmann, P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**, 112–118 (2012).
39. Lazar, C., Gatto, L., Ferro, M., Bruley, C. & Burger, T. Accounting for the multiple natures of missing values in label-free quantitative proteomics data sets to compare imputation strategies. *J. Proteome Res.* **15**, 1116–1125 (2016).
40. Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D. & Iverson, G. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon Bull. Rev.* **16**, 225–237 (2009).
41. Palasca, O., Santos, A., Stolte, C., Gorodkin, J. & Jensen, L. J. TISSUES 2.0: an integrative web resource on mammalian tissue expression. *Database (Oxford)*. **2018**, (2018).
42. Whelan, S. A. et al. Assessment of a 60-biomarker health surveillance panel (HSP) on whole blood from remote sampling devices by targeted LC/MRM-MS and discovery DIA-MS analysis. *Anal. Chem.* **95**, 11007–11018 (2023).
43. Beernink, R. H. J., Schuitemaker, J. H. N., Zwertbroek, E. F., Scherjon, S. A. & Cremers, T. I. F. H. Early pregnancy biomarker discovery study for spontaneous preterm birth. *Placenta* **139**, 112–119 (2023).
44. McElrath, T. F. et al. Circulating microparticle proteins obtained in the late first trimester predict spontaneous preterm birth at less than 35 weeks' gestation: a panel validation with specific characterization by parity. *Am. J. Obstet. Gynecol.* **220**, 488e1–488e11 (2019).
45. Wei, C. et al. Identification and verification of diagnostic biomarkers in recurrent pregnancy loss via machine learning algorithm and WGCNA. *Front. Immunol.* **14**, 1241816 (2023).
46. Wei, P. et al. Identification and validation of a signature based on macrophage cell marker genes to predict recurrent miscarriage by integrated analysis of single-cell and bulk RNA-sequencing. *Front. Immunol.* **13**, 1053819 (2022).
47. Yin, X. J., Hong, W., Tian, F. J. & Li, X. C. Proteomic analysis of decidua in patients with recurrent pregnancy loss (RPL) reveals mitochondrial oxidative stress dysfunction. *Clin. Proteom.* **18**, 9 (2021).
48. Qiu, Q. et al. Significance of IGFBP-4 in the development of fetal growth restriction. *J. Clin. Endocrinol. Metab.* **97**, E1429–E1439 (2012).
49. Zhang, T., Du, T., Li, W., Yang, S. & Liang, W. Sex hormone-binding globulin levels during the first trimester may predict gestational diabetes mellitus development. *Biomark. Med.* **12**, 239–244 (2018).
50. Hedderson, M. M. et al. Prepregnancy SHBG concentrations and risk for subsequently developing gestational diabetes mellitus. *Diabetes Care.* **37**, 1296–1303 (2014).
51. Van Ostade, X., Dom, M., Tjalma, W. & Van Raemdonck, G. Candidate biomarkers in the cervical vaginal fluid for the (self-)diagnosis of cervical precancer. *Arch. Gynecol. Obstet.* **297**, 295–311 (2018).
52. Starodubtseva, N. L. et al. Label-free cervicovaginal fluid proteome profiling reflects the cervix neoplastic transformation. *J. Mass. Spectrom.* **54**, 693–703 (2019).
53. Mukherjee, A. et al. Insights on proteomics-driven body fluid-based biomarkers of cervical cancer. *Proteomes* **10**, 13 (2022).
54. Van Raemdonck, G. A. A., Tjalma, W. A. A., Coen, E. P., Depuydt, C. E. & Van Ostade, X. W. M. Identification of protein biomarkers for cervical cancer using human cervicovaginal fluid. *PLoS One.* **9**, e106488 (2014).
55. Olalekan, S., Xie, B., Back, R., Eckart, H. & Basu, A. Characterizing the tumor microenvironment of metastatic ovarian cancer by single-cell transcriptomics. *Cell. Rep.* **35**, 109165 (2021).
56. Qian, L. et al. Proteomic landscape of epithelial ovarian cancer. *Nat. Commun.* **15**, 6462 (2024).
57. Whitwell, H. J. et al. Improved early detection of ovarian cancer using longitudinal multimarker models. *Br. J. Cancer.* **122**, 847–856 (2020).
58. Radu, M. R. et al. Ovarian cancer: Biomarkers and targeted therapy. *Biomedicines* **9**, 693 (2021).
59. Atallah, G. A., Abd Aziz, N. H., Teik, C. K., Shafiee, M. N. & Kampan, N. C. New predictive biomarkers for ovarian cancer. *Diagnostics (Basel)*. **11**, 465 (2021).
60. Lupia, M. et al. CD73 regulates stemness and epithelial-mesenchymal transition in ovarian cancer-initiating cells. *Stem Cell. Rep.* **10**, 1412–1425 (2018).
61. Kim, H. et al. Human kallikrein gene 5 (KLK5) expression is an indicator of poor prognosis in ovarian cancer. *Br. J. Cancer.* **84**, 643–650 (2001).
62. Gipson, I. K. et al. Mucin genes expressed by human female reproductive tract epithelia. *Biol. Reprod.* **56**, 999–1011 (1997).
63. Moncla, B. J., Chappell, C. A., Debo, B. M. & Meyn, L. A. The effects of hormones and vaginal microflora on the glycome of the female genital tract: Cervical-vaginal fluid. *PLoS One.* **11**, e0158687 (2016).
64. Njoku, K., Chiasserini, D., Whetton, A. D. & Crosbie, E. J. Proteomic biomarkers for the detection of endometrial cancer. *Cancers (Basel)*. **11**, 1572 (2019).
65. la Coll-de, E. et al. Prognostic biomarkers in endometrial cancer: A systematic review and meta-analysis. *J. Clin. Med.* **9**, 1900 (2020).
66. Degez, M. et al. Endometrial cancer: A systematic review of HE4, REM and REM-B. *Clin. Chim. Acta.* **515**, 27–36 (2021).
67. Romano, A. et al. Endometrial cancer diagnostic and prognostic algorithms based on proteomics, metabolomics, and clinical data: a systematic review. *Front. Oncol.* **13**, 1120178 (2023).

68. Martinez-Garcia, E. et al. Targeted proteomics identifies proteomic signatures in liquid biopsies of the endometrium to diagnose endometrial cancer and assist in the prediction of the optimal surgical treatment. *Clin. Cancer Res.* **23**, 6458–6467 (2017).
69. Njoku, K. et al. Quantitative SWATH-based proteomic profiling of urine for the identification of endometrial cancer biomarkers in symptomatic women. *Br. J. Cancer.* **128**, 1723–1732 (2023).
70. Mirza, Z. & Abdel-Dayem, U. A. Uncovering potential roles of differentially expressed genes, upstream regulators, and canonical pathways in endometriosis using an in silico genomics approach. *Diagnostics (Basel)*. **10**, 416 (2020).
71. Yang, H., Kang, K., Cheng, C., Mamillapalli, R. & Taylor, H. S. Integrative analysis reveals regulatory programs in endometriosis. *Reprod. Sci.* **22**, 1060–1072 (2015).
72. Azeze, G. G. et al. Proteomics approach to discovering non-invasive diagnostic biomarkers and understanding the pathogenesis of endometriosis: a systematic review and meta-analysis. *J. Transl Med.* **22**, 685 (2024).
73. Méar, L. et al. The eutopic endometrium proteome in endometriosis reveals candidate markers and molecular mechanisms of physiopathology. *Diagnostics (Basel)*. **12**, 419 (2022).
74. Janša, V. et al. Proteomic analysis of peritoneal fluid identified COMP and TGFBI as new candidate biomarkers for endometriosis. *Sci. Rep.* **11**, 20870 (2021).
75. Anastasiu, C. V. et al. Biomarkers for the noninvasive diagnosis of endometriosis: State of the art and future perspectives. *Int. J. Mol. Sci.* **21**, 1750 (2020).
76. Irungu, S. et al. Discovery of non-invasive biomarkers for the diagnosis of endometriosis. *Clin. Proteom.* **16**, 14 (2019).
77. Ludwig, C. et al. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **14**, e8126 (2018).
78. Gérard, N., Caillaud, M., Martoriati, A., Goudet, G. & Lalmanach, A. C. The interleukin-1 system and female reproduction. *J. Endocrinol.* **180**, 203–212 (2004).
79. Zubor, P. et al. Rho GTPases in gynecologic cancers: In-depth analysis toward the paradigm change from reactive to predictive, preventive, and personalized medical approach benefiting the patient and healthcare. *Cancers (Basel)*. **12**, 1292 (2020).
80. Kanaka, V., Drakakis, P., Loutradis, D. & Tsangaris, G. T. Proteomics in the study of female fertility: an update. *Expert Rev. Proteom.* **20**, 319–330 (2023).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Mauricio Hernandez^{1,3} · Pablo Saldivia^{1,2} · Veronica Latapiat¹ · Barbara Antilef¹ · Guillermo Nourdin¹ · Felipe Castro¹ · Cristian Vargas¹ · Elard S. Koch¹

✉ Elard S. Koch
ekoch@melisainstitute.org

¹ MELISA Institute, Dalcáhué, San Pedro de la Paz, Concepción 1120, 4133515, Chile

² Faculty of Biological Sciences, University of Concepción, Concepción, Chile

³ Faculties of Dentistry and Pharmacy, University of Concepción, Concepción, Chile