High-throughput dia-PASEF workflow improves proteome coverage and quantitative performance in repeated-measures proteomics of cervicovaginal fluid within a preconception cohort

Mauricio Hernandez[1,3]†, Pablo Saldivia[1,2]†, Veronica Latapiat[1]†, Barbara Antilef[1], Guillermo Nourdin[1], Felipe Castro[1], Cristian Vargas[1], Elard S. Koch[1]*.

1 MELISA Institute, Dalcahue 1120, San Pedro de la Paz, 4133515, Concepción, Chile.
2 PhD Program in Molecular Biotechnology, Faculty of Biological Sciences, University of Concepción, Concepción, Chile.
3 PhD Program in Applied Clinical Sciences, Faculties of Dentistry and Pharmacy, University of Concepción

†The authors contributed equally to the development of the manuscript.

Corresponding Author:
*Elard S. Koch
MELISA Institute, Dalcahue 1120, San Pedro de la Paz, 4133515, Concepción, Chile
Telephone +56 412467242 E-mail: ekoch@melisainstitute.org

Abstract

Proteomic studies of the female reproductive system have the potential to advance women's reproductive health and improve understanding of fertility-related biological processes. Cervicovaginal fluid (CVF), a complex mixture of uterine, cervical, and vaginal secretions, represents a valuable source of molecular information reflecting time-dependent molecular changes across the menstrual cycle. However, large-scale biomarker discovery in CVF remains challenging due to limitations in sensitivity, reproducibility, and data completeness in mass spectrometry–based workflows, particularly within repeated-measures longitudinal study designs.

Here, we developed and evaluated a standardized high-throughput dia-PASEF–based proteomics workflow for longitudinal repeated-measures analysis of CVF samples. These samples were collected day-by-day across an ovulation-anchored 7-day window (day 0) and the subsequent six post-ovulatory days within a prospective preconception cohort (EARLY-PREG). This data-independent acquisition approach was directly compared with a conventional DDA-FracOffline strategy.

Across a 7-day ovulation anchored window, DDA-FracOffline detected 2,817 quantifiable proteins, whereas dia-PASEF identified 4,229 quantifiable proteins (approximately 50% increase), with mean coefficients of variation 9.37 ± 1.75% and 5.27 ± 0.68%, respectively. Using day 0 (ovulation day) as the time reference for day-to-day comparisons, dia-PASEF identified higher numbers of differentially expressed proteins (DEPs) than did DDA-FracOffline across pairwise day comparisons (952–1,116 DEPs for dia-PASEF versus 161–460 DEPs for DDA-FracOffline). Pathway enrichment analyses demonstrated broader biological coverage using dia-PASEF. Furthermore, when evaluated against a literature-reported panel of clinical biomarker candidates relevant to female reproductive tract pathophysiology, the dia-PASEF workflow showed improved quantitative consistency and enhanced dynamic range sensitivity.

These findings indicate that the implemented dia-PASEF approach provides a sensitive, reproducible, and scalable workflow for longitudinal repeated-measures proteomics of CVF. The method is adaptable to other biofluids and supports robust investigation of time-dependent molecular signatures associated with key biological pathways under physiologically relevant conditions in female reproductive system research.

Word count: 295

Keywords: Cervicovaginal fluid, High-throughput proteomics, dia-PASEF, Menstrual Cycle, Biomarkers, Women's health.

Introduction

Proteomic studies of the female reproductive system are increasingly recognised as relevant for advancing women's reproductive health and fertility research. At the molecular level, insights gained from proteomic studies of tissues and biofluids[1] may contribute to our understanding of the pathophysiology of 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 underexplored [2–4]. Furthermore, despite the accelerated development of proteomics research, the discovery of reliable biomarkers remains challenging, mainly due to methodological and analytical limitations [5].

Cervicovaginal fluid (CVF) is recognised as a relevant biofluid source for biomarker discovery in proteomics because of its continuity with the female reproductive tract and pivotal role in fertility, pregnancy outcomes, and the local immune responses [6–1112,13]. 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].

High-throughput proteomics has substantially advanced protein identification and quantification. However, most studies rely on the data-dependent acquisition (DDA) of proteins, a time-consuming workflow with limitations in quantitative capabilities, reproducibility, and accuracy, particularly in large-scale longitudinal analyses with repeated measures over time[18]. In the search for new strategies to obtain a more complete proteomic analysis, data-independent acquisition data-independent acquisition (DIA)-based methodologies have recently been developed[19]. For example, Njoku and collaborators [7] 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 a greater number of proteins compared with a standard DDA approach.

Addressing DDA limitations is important for improving clinical applications of CVF proteomics, 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 additional insight into the course of female reproductive health and disease. Despite its potential, studies applying spectral library-based DIA to CVF are relatively rare compared with studies applying the 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 menstrual cycle.

Although SWATH-MS has been explored as an alternative DIA approach to assess the CVF proteome, its reliance on fixed isolation windows may introduce spectral interferences that potentially limit resolution. With respect to new technologies that address this limitation, 4D proteomics, enabled by timsTOF mass spectrometry, integrates trapped ion mobility spectrometry (TIMS) with MS 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 tools [29] for complex biofluids such as plasma, where ionic mobility enables improved separation of ions based on their shape, size, and charge [26].

In this study, we aimed to implement and validate an optimised and reproducible dia-PASEF method for high-throughput time-dependent proteomic analyses of CVF samples collected from menstrual cycles of women participating in the EARLY-PREG longitudinal preconception open cohort [30]. The new dia-PASEF approach was compared with the traditional DDA coupled with the offline peptide fractionation (DDA-FracOffline) method. We hypothesised that the dia-PASEF approach would provide greater proteome coverage and reproducibility compared with traditional DDA, supporting its application for biomarker discovery in longitudinal studies of women's reproductive health, fertility, and early pregnancy.

Methods

1. EARLY-PREG Cohort

EARLY-PREG (ClinicalTrials.gov identifier NCT07358026) is a preconception longitudinal open cohort designed to investigate maternal–embryonic communication at the ultra-early stages of conception through the serial and standardised collection of multiple biofluids, including CVF. Within this cohort, synchronised and well-documented samples were used to compare dia-PASEF with the traditional DDA approach coupled with offline peptide fractionation (DDA-FracOffline).

2. CVF longitudinal samples

The CVF samples for this proteomics MS implementation and validation study were obtained from the biorepository of the EARLY-PREG preconception cohort. This biorepository includes different biofluids from 223 female volunteers collected daily during their menstrual cycles from the end of the menstrual period, which represents the beginning of a new cycle. Ovulation was determined by quantifying the luteinising hormone (LH) surge in urine. The samples were collected from the day after the end of menstruation until the 14th day post ovulation, as determined by the LH surge in each cycle, using a previously described method [31] which detects ovulation within ±1 day (95.7% confidence). LH concentrations were determined through radioimmunoassay (IRMA) (Barnafi-Krause Laboratory, Santiago, Chile). This approach allowed LH to be defined as a day 0 (*d0*), which served as a cycle-anchored time reference for synchronising repeated measures across individuals.

Participants were grouped into three cycle categories: conception, non-conception, and abstinence cycles. Due to the methodological focus of this validation study, homogeneity was prioritised over sample size. A subgroup of CVF samples from six women (n=6) belonging to non-conception cycles was selected according to criteria that controlled for factors such as body mass index, age, menstrual cycle length, and reproductive history.

To enable a within-individual repeated-measures longitudinal analysis, we defined an ovulation-anchored luteal phase window comprising seven consecutive days (*d0–d6*). Women were selected if at least seven consecutive post-ovulatory CVF samples were available in the

155 biorepository. To minimise intra- and inter-individual variability and reduce heteroscedasticity
156 in repeated-measures analyses, we further prioritised homogeneity within selected individuals
157 by considering menstrual cycle length, menstruation duration, age, sampling completeness
158 across days, and other relevant variables within comparable ranges. Detailed characteristics of
159 the selected participants are provided in Table 1.
160
161 **3. Protein extraction and processing for MS analysis**
162
163 CVF aliquots (100–500 µL) were thawed on ice and combined with equal volumes of 100 mM
164 NH4HCO3 and 100 mM DTT. The mixture was homogenised by vortexing, incubated at 95
165 °C for 15 minutes, and cooled on ice for 5 minutes. The samples were then subjected to cold
166 ultrasound disruption at 4 °C (10-second on/off pulses at 50% amplitude) and subsequently
167 centrifuged at 19,000 × g for 20 minutes at 4 °C. The supernatant (soluble phase) was carefully
168 separated, while the pellet (insoluble phase) was retained for further processing.
169
170 The soluble phase was concentrated in Vivaspin tubes (10,000 MWCO filter, Sartorius),
171 washed with 400 µL of ultrapure water, and centrifuged at 12,000 × g for 10 minutes at 4 °C.
172 The filtrate (<10,000 Da) was dried in a SpeedVac and stored at -20 °C, while proteins retained
173 on the filter (>10,000 Da) were recovered by washing with ~600 µL of water, combined, and
174 dried. The insoluble pellet and soluble protein concentrate were resuspended in 8M urea with
175 25 mM NH_4HCO_3. The mixture was sonicated on ice, and centrifuged at 19,000 × g for 15
176 minutes at 4 °C. The protein content was quantified using a Qubit protein assay (Invitrogen,
177 #Q33212).
178
179 For MS preparation, 300 µg of total protein from the soluble and insoluble phases was reduced
180 with 20 mM DTT in 25 mM NH_4HCO_3 at 25 °C for 1 hour. Alkylation was performed with 20
181 mM iodoacetamide in 25 mM NH_4HCO_3 in the dark for 1 hour. The samples were diluted
182 eightfold with 25 mM NH_4HCO_3 and digested with sequencing-grade trypsin (#V5071,
183 Promega) at a 1:50 protease/protein ratio for 16 hours at 37 °C. Digestion was stopped by
184 adding 10% formic acid. Peptides were cleaned using Sep-Pak C18 Spin columns (Waters),
185 dried in a rotary concentrator at 40 °C, and stored for subsequent MS analysis.
186
187 **4. DDA-FracOffline**
188

189 **4.1. Peptide Fractionation**

190

191 Offline high-pH reversed-phase fractionation of CVF peptides was performed using a protocol
192 adapted from our previously published tear proteomics workflow [32]. A total of 200 µg of tryptic
193 peptides from soluble and insoluble CVF phases were separated using an ÄKTA avant 25
194 system with refrigerated fraction collection at 6° C. Peptides were loaded onto a BHE reversed-
195 phase column (2.1 cm × 5 cm, Waters) at pH 10 and a flow rate of 0.2 mL/min. The binary
196 gradient increased linearly from 3% to 40% buffer B (90% ACN in 5 mM NH_4HCO_2) over 30
197 minutes, followed by increases to 60% over 15 minutes, and to 85% in the final 5 minutes. We
198 collected a total of 24 fractions (400 µL each), which were vacuum-dried and then reconstituted
199 in water containing 2% ACN and 0.1% FA. These fractions were concatenated into 8 final
200 fractions.

201

202 **4.2. Mass Spectrometry**

203

204 nLC–MS was performed on a NanoElute (Bruker Daltonics) system coupled online with a
205 hybrid TIMS-quadrupole TOF mass spectrometer, timsTOF Pro (Bruker Daltonics, Germany),
206 via a nanoelectrospray ion source (CaptiveSpray, Bruker Daltonics) [33]. For long gradient nLC–
207 MS/MS runs, we followed the standard protocol described by Meier et al., 2021.
208 Approximately 200 ng of peptides were separated onto an Aurora series reversed-phase
209 column (25 cm × 75 µm ID, 1.9 µm, IonOpticks, Australia) at a flow rate of 400 nL/min in an
210 oven compartment heated to 50 °C. For the analysis of CVF off-line fractionated samples, we
211 applied a multi-step gradient beginning with an increase from 2% to 17% B over 60 minutes,
212 progressing to 25% B in 30 minutes, then to 37% B in 10 minutes, and finally to 95% B in 10
213 min, which was held for 10 min. The column was equilibrated using 4 volumes of solvent A.
214 The mass spectrometer was operated in data-dependent PASEF [34] mode, consisting of 1 TIMS-
215 MS survey and 10 PASEF MS/MS scans per acquisition cycle. Mobility values were acquired
216 across a 1/K0 range of 1.6 to 0.6 Vs cm-2, using equal ion accumulation and ramp time of 100
217 ms each in the dual-TIMS analyser.

218

219 **4.3. Data searching**

220

221 Computational processing of DDA data was performed using the FragPipe v19.1 software. The
222 pipeline integrates MSFragger (version 3.7)[35] search engine, for peptide-spectrum matching,

223 Philosopher (version 3.4.13)[36] for FDR control and validation, and EasyPQP
224 (https://github.com/grosenberger/easypqp; version 0.1.9) for spectral library generation. Raw
225 (.d) files were used as input for MSFragger searches. Protein identification was conducted
226 against reference proteome of *Homo sapiens* (UP000005640) from UniProt/SwissProt
227 (reviewed sequences only; downloaded on Feb. 15, 2021), supplemented with database of
228 common contaminant proteins in MS experiments available in FragPipe. Decoy sequences
229 were generated by reversed protein sequence entries to enable target-decoy estimation. For
230 MSFragger analysis, mass tolerance for precursor and (initial) fragments were set to 20 ppm.
231 The enzyme specificity was defined as 'strict trypsin', allowing fully tryptic cut and a
232 maximum of two missed cleavages. Carbamidomethylation of cysteine was included as a fixed
233 modification, whereas methionine oxidation and N-terminus acetylation were set as variable
234 modifications. The final spectral library was filtered to 1% protein- and peptide FDR levels.
235
236 **5. dia-PASEF**
237
238 **5.1 Mass spectrometry**
239
240 The timsTOF Pro was operated in dia-PASEF mode with the following settings: mass range,
241 100 to 1700 m/z; ion mobility range, 1/k0 start, 0.57 V/cm2, end, 1.47 V/cm2; ramp and
242 accumulation times, 100 ms; capillary voltage, 1600 V; dry gas, 3 L/min; and dry temperature,
243 160 °C. The dia-PASEF Ion Mobility Parameters were as follows: Each cycle consisted of a
244 one MS1 full scan and 64x MS2 windows covering 400–1200 m/z and 0.57–1.46 1/K0. Each
245 window was 25 Th wide by 0.2 V s/cm2 high. There was an overlap of 0.05 in 1/K0 and no
246 overlap in m/z. The cycle time was 1.80 seconds. The CID collision energy was 20 to 52 eV as
247 a function of the inverse mobility of the precursor [24].
248
249 **5.2 Data search**
250
251 For the construction of the DDA spectral library, a pooled of all samples of tryptic-digested
252 peptides was used. For peptide fractionation, mass spectrometry and data searching, the method
253 described in the DDA section was used (see methods sections 3.1, 3.2 and 3.3). Proteomic data
254 analyses were performed on a high-performance computing server with dual AMD EPYC 7552
255 48-core processors (2.20 GHz) and 512 GB of RAM. Raw dia-PASEF data were processed via
256 DIA-NN v1.7.15 software [37], 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; quantification mode was set to "any LC (high accuracy)". All remaining settings were left default. Identifications were filtered using a precursor-level q value < 1% and a global protein-level q value < 1%.

6. Bioinformatics and statistical analysis

The intensity matrix reports from MSFragger and DIANN were exported and processed in the R statistical environment. The intensity values for each run were normalised by adjusting the medians. Missing values were imputed via the missForest algorithm [38], applying the criterion that the intensity value should be present in at least 50% of the samples per analysed category [39]. Longitudinal differentially expressed proteins (DEPs) along the menstrual cycles were represented both as absolute and relative numbers (%), and the significant difference was determined by using a using Bayesian t-tests for pairwise day comparisons (e.g., d0 vs. dn, d1 vs. dn), considering the normalised intensity means of biological replicates [40]. AProteins were considered statistically significant 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. The coefficients of variation were determined using the statistically significant differential expression values of each methodology for both dia-PASEF and DDA-FracOffline. Quantifiable proteins were defined as those that were consistently detected across each condition and presented measurable relative abundance values.

7. Pathway and tissue enrichment analysis

Pathway enrichment analysis was developed using ReactomePA (v.1.42.0), which queries the human database with log2 fold change (logFC) and p value data obtained for each statistical comparison. For the analysis of enrichment by human systems, tissues and cells, we used the TISSUES 2.0 database [41], considering the same statistical parameters used for the pathway analysis through topGO (v.2.5.0). Boxplots were generated to represent 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 were directly annotated on the graphs. The z score was calculated based on 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 a biological process, molecular function, or cellular component tends to be upregulated (positive value) or downregulated (negative value). The z score was determined based on the number of proteins assigned as upregulated (logFC > 0) and those classified as downregulated (log FC < 0), labelled as "up" and "down" in the dataset.

8. Biomarker selection

To assess the efficiency of dia-PASEF in biomarker detection, an FDA-proposed biomarker panel derived from a previously published study [42], was used to directly compare the ability of dia-PASEF and DDA-FracOffline to detect and quantify these clinically relevant proteins. In addition, we conducted an extensive literature-based curation of candidate proteins associated with physiological and pathological conditions of the female reproductive tract. Searches were performed using keywords including “menstrual cycle”, “ovarian carcinoma”, “endometrial carcinoma”, “cervical carcinoma”, and “endometriosis”, integrating evidence from proteomic, transcriptomic, and genomic studies.

9. Data availability

Proteomics data can be accessed through the ProteomeXchange Consortium via PRIDE, with the dataset identifier PXD071350 and PXD071584.

10. 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). 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.

321 **RESULTS**

322

323 **1. Patient Characterization and Sampling**

324

325 After screening all non-conception fertile cycles of women with complete records in the
326 EARLY-PREG cohort database, we selected six women with homogeneity in terms of age,
327 weight, height, BMI, waist circumference, age at menarche, menstrual cycle length, and
328 menstruation length. The length of the menstrual phase of the cycle (5.8 ± 0.4), the length of
329 the fertile cycle (29.0 ± 3.0), age (32.3 ± 3.6), and other variables were within comparable
330 ranges. Detailed participant characteristics are provided in Table 1. Forty-two CVF samples
331 (n=42) stored at -80 °C were processed for protein extraction and tryptic digestion and then
332 analysed with the two mass spectrometry methodologies (Figure 1A) as described in the
333 Methods. To minimise batch effects, all 42 samples were analysed on the timsTOF Pro mass
334 spectrometer within the same calibration batch. This included 336 chromatographic runs using
335 the DDA approach and 42 direct runs using dia-PASEF over a two-week period. Thus, with
336 each protocol, we obtained a 7-day time series (*d0* to *d6*) of CVF proteomes for each woman.
337 For each analytical workflow, this design yielded a within-individual repeated-measures time
338 series comprising a 7-day ovulation-anchored window (d0–d6) of CVF proteomes per
339 participant.

340

341 **Table 1**: Information on the patients 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

342 [1]Body mass index (BMI) was calculated as weight (kg) divided by height squared (m^2).
343 [2]Menstrual cycle length refers to the number of days from the first day of menstruation to the first day of the next
344 cycle.
345 [3]Menstruation duration corresponds to the number of days of menstrual bleeding per cycle.

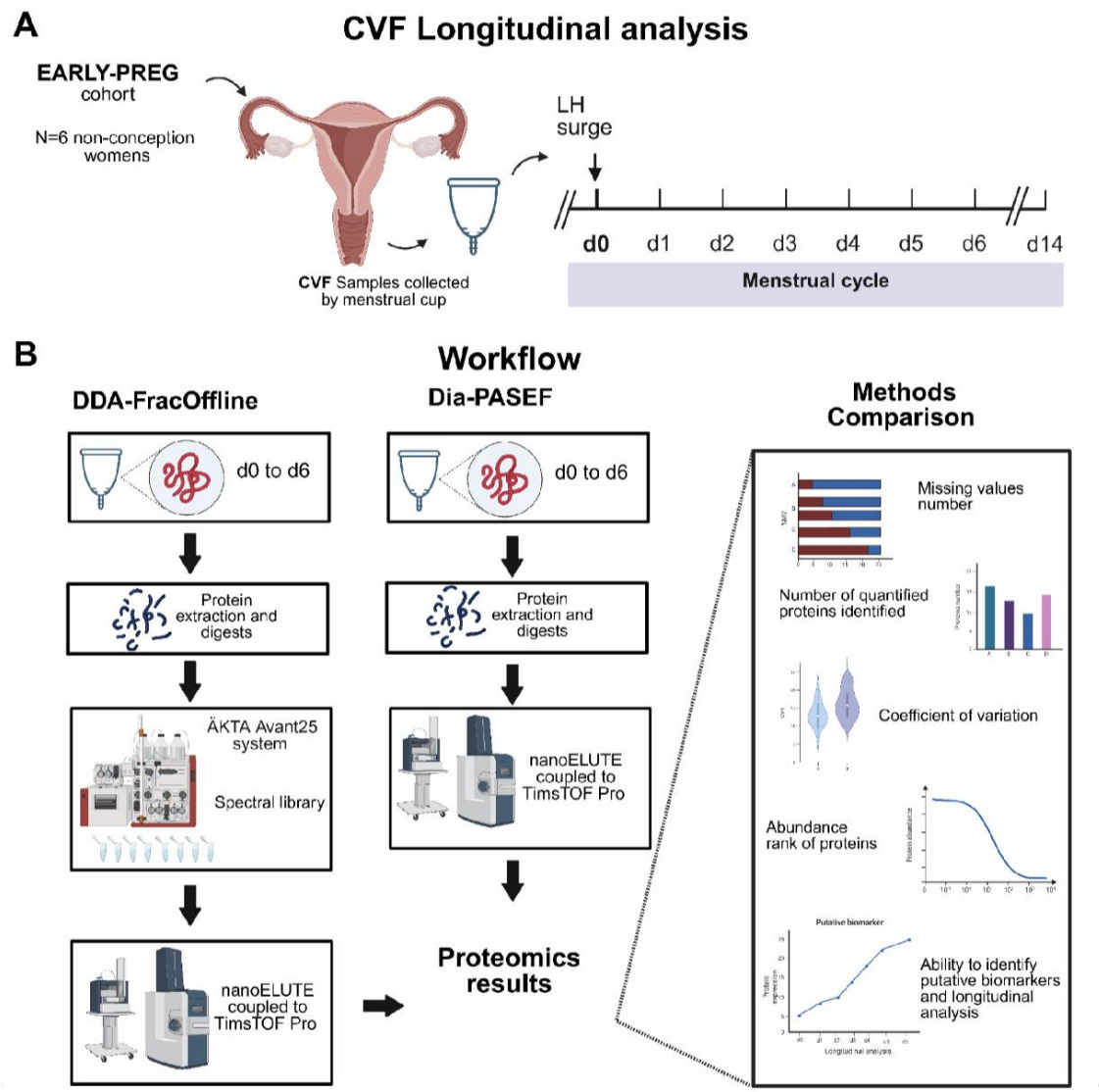

Figure 1. Proteomics workflow for the longitudinal repeated-measures analysis of CVF. Schematic representation of the experimental design and strategies used to analyse CVF samples from women in the EARLY-PREG cohort. (A) CVF samples were collected across a 7-day ovulation-anchored window (d0–d6) of the menstrual cycle, enabling within-individual repeated-measures analysis. (B) Two nLC-MS/MS proteomics approaches were compared: DDA-FracOffline and dia-PASEF. Both workflows included protein extraction, digestion, and mass spectrometry analysis using the nanoElute system coupled to a timsTOF Pro instrument. The right panel summarises the comparative evaluation of both methods, including missing values, number of quantified proteins, coefficient of variation, abundance rank distribution, and their ability to identify putative biomarkers and capture time-dependent longitudinal proteomic changes.

358

359 **2. Quality Control Factors analysed to compare DDA-FracOffline and the dia-PASEF**

360

361 We conducted a comparative analysis focusing on key quality control parameters, including
362 reproducibility, the proportion of missing values, and protein identification and quantification
363 performance. First, reproducibility was assessed using correlation coefficients between
364 biological replicates. The mean correlation coefficient was approximately 0.70 for DDA-
365 FracOffline and 0.90 for dia-PASEF (Figure S1). Second, we evaluated the proportion of
366 missing values across samples. The dia-PASEF workflow showed a substantially lower
367 proportion of missing values compared with DDA-FracOffline, which reached up to 75% at its
368 lowest detection point (Figure 2B). Third, correlation analysis across samples demonstrated
369 lower inter-sample consistency for DDA-FracOffline relative to dia-PASEF (Figure 2A). Next,
370 we analysed protein identification and quantification performance. DDA-FracOffline
371 identified a total of 9,593 proteins, whereas dia-PASEF identified 5,986 proteins. However,
372 when considering quantifiable proteins (identified and consistently measured across samples),
373 DDA-FracOffline quantified 2,817 proteins, whereas dia-PASEF quantified 4,229 proteins
374 (Figure 2E, left), corresponding to a 50% higher quantification yield for dia-PASEF. The mean
375 coefficient of variation (CV) for quantifiable proteins—calculated as the average of individual
376 protein CVs across all samples—was 5.27 ± 0.68% for dia-PASEF and 9.37 ± 1.75% for DDA-
377 FracOffline (Figure 2E, centre) indicating improved quantitative consistency in the dia-PASEF
378 workflow. Finally, across the ovulation-anchored time series (d0–d6), DDA-FracOffline
379 identified between 1,491 and 7,183 proteins across time points (Figure 2C), whereas dia-
380 PASEF detected between 3,811 and 5,398 proteins across time points (Figure 2D). Although
381 DDA-FracOffline showed broader overall protein identification, dia-PASEF demonstrated
382 greater consistency in protein detection across the ovulation-anchored repeated-measures time
383 series.

384

385

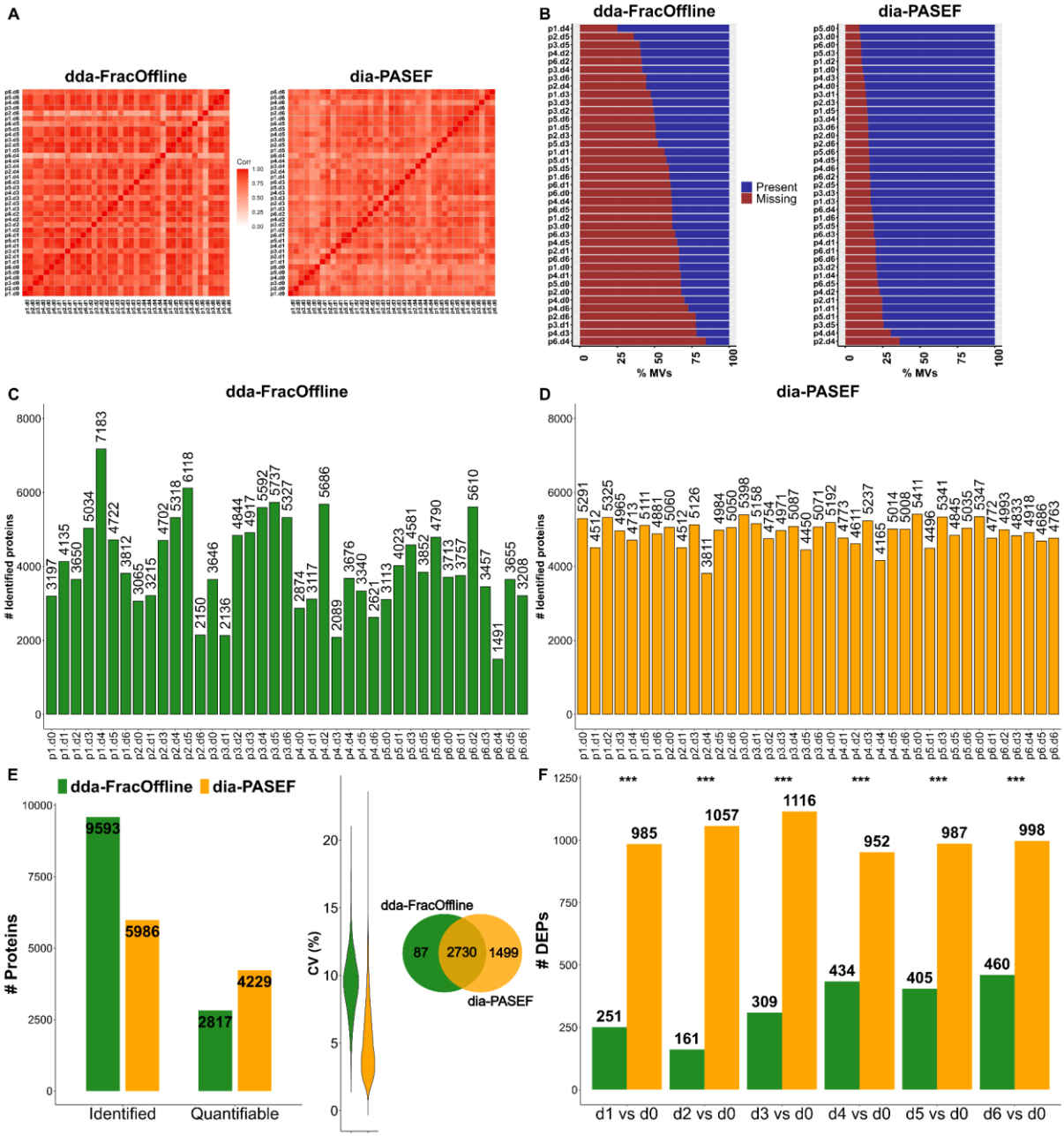

Figure 2. Quality control of proteomic data obtained by DDA-FracOffline and dia-PASEF. (A) Pearson correlations between chromatographic runs, positive correlations (red) and no correlations (white). (B) Bar plots of the proportion of missing values (%MVs) for each sample, showing present values (blue) and missing values (red). (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) represents the coefficient of variation (%CV) for each method. The Venn diagram illustrates the shared and unique proteins identified by each analytical strategy. (F) Number of differentially expressed

397 proteins (DEPs) identified by each strategy through longitudinal comparisons at day d0. The
398 bar plot displays the total number of DEPs obtained by each strategy, indicating the statistical
399 significance (***, p value < 0.001) above the bars. The statistical significance was determined
400 using chi-square tests for categorical variables. DEPs were identified using DDA-FracOffline
401 (green) and dia-PASEF (orange).
402
403 **2.1. Differential Protein Expression**
404
405 Following quality control, we assessed the time-dependent dynamics of differentially
406 expressed proteins (DEPs) across the menstrual cycle to characterise variations in protein
407 abundance in CVF within the ovulation-anchored, repeated-measures 7-day window (*d0–d6*).
408 Comparative analyses revealed significant differences between methodologies in the detection
409 and quantification of DEPs across longitudinal CVF samples (Figure S2). At most time points,
410 dia-PASEF identified higher absolute and relative numbers of DEPs than did DDA-
411 FracOffline, e.g. 985 (23%) vs. 251 (8.9%) between *d0* and *d1* and 1,057 (24.9%) vs 161 (5.7%)
412 between *d0* and *d2*. From d0 to d6, dia-PASEF maintained a more stable detection rate (Figure
413 2F). These findings indicate that dia-PASEF enables more consistent detection and
414 quantification of differentially expressed proteins in longitudinal repeated-measures designs,
415 facilitating the capture of time-dependent proteomic changes across the cycle (Figure 2F and
416 Figure S2).
417
418 **3. Longitudinal enrichment analysis of pathways, systems, tissues and cells**
419
420 To interpret the functional relevance of the quantified proteins, we performed tissue and
421 pathway enrichment analyses. We evaluated tissue representation based on DEPs identified
422 across longitudinal CVF comparisons within the ovulation-anchored repeated-measures
423 window (Figure S3). The analysis revealed enrichment in tissues associated with the female
424 reproductive system, including the uterus and uterine cervix, as well as immune-related tissues.
425 The dia-PASEF workflow demonstrated broader and more consistent tissue coverage,
426 particularly at later time points (*d5* and *d6*), with elevated enrichment scores (-log10 adjusted
427 p value) and the highest protein densities. We further examined the biological processes
428 associated with the differential protein profiles detected across the 7-day window. Common
429 pathways identified with both strategies between time points *d0* to *d6* (Figure 3) included
430 '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 consistent with cell signalling and cell-cycle–related processes (Figure 3). Compared with DDA-FracOffline, dia-PASEF showed increased pathway coverage and enrichment across the time-dependent series.

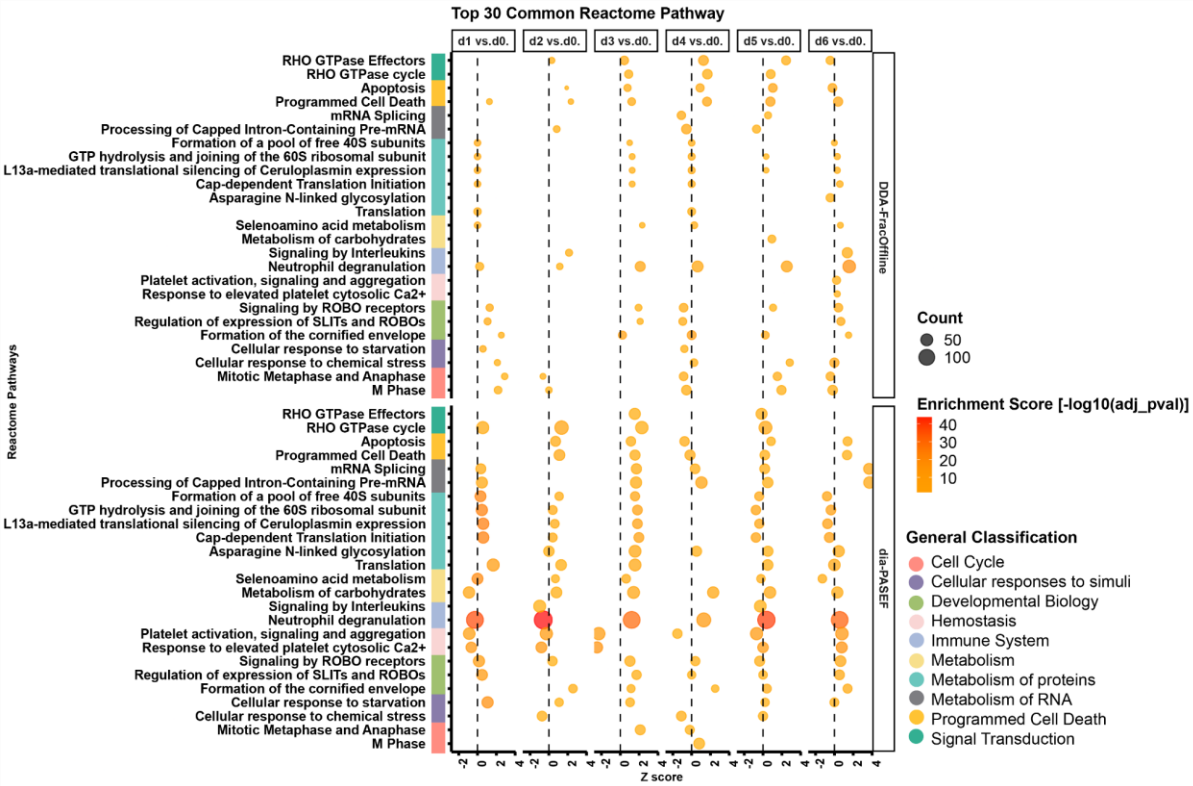

Figure 3. Differentially expressed proteins and biological pathways identified via longitudinal analysis using DDA-FracOffline and dia-PASEF. Reactome enrichment analysis, emphasising the most abundant biological pathways. DEPs were derived from longitudinal comparisons within each method. The Y-axis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

4. Identification of Physiological and pathological biomarkers

We evaluated the ability of each strategy to identify and quantify known biomarkers of physiological and pathological processes. 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).

451 **4.1 Comparison of FDA-proposed biomarkers via DDA-FracOffline and dia-PASEF**
452
453 The FDA-proposed biomarker panel comprised 121 proteins detectable in biological fluids and
454 associated with female reproductive pathologies [42]. Figure 4 compares the ability of each
455 strategy to detect these biomarkers in CVF. Dia-PASEF and DDA-FracOffline identified 91
456 and 76 of the 121 proteins, corresponding to 75% and 63%, respectively; 92 proteins (76%)
457 were identified by both strategies (Figure 4A). A protein rank-abundance plot demonstrated
458 that dia-PASEF captured proteins across a broader dynamic range compared with DDA-
459 FracOffline (Figure 4B). Fifteen FDA biomarkers were unique to dia-PASEF (NT5E, ACE,
460 GLA, C4A, C1QC, ACP5, PROS1, BRAF, VWF, GALT, MSH2, TFRC, TOP2A, and BTD),
461 whereas the protein GFAP was detected only by DDA-FracOffline (Figure 4B). Overall, dia-
462 PASEF showed improved quantification consistency and reproducibility relative to DDA-
463 FracOffline for the detection of clinically relevant biomarker candidates in CVF. 4.2 Selected
464 biomarkers from literature review
465
466 A literature-based curated list of biomarkers previously reported in pregnancy, cervical,
467 ovarian, and endometrial carcinomas, as well as endometriosis, is provided in Table 2. We
468 examined the time-dependent behaviour of four representative biomarkers selected from the
469 curated list to illustrate the longitudinal detection capability of the dia-PASEF workflow
470 (Figure 4C–G). To illustrate the longitudinal repeated-measures capability of dia-PASEF in
471 detecting clinically relevant proteins, we tracked insulin-like growth factor-binding protein 4
472 (IGFBP-4), implicated in early pregnancy and foetal growth restriction; MUC1, a mucin
473 contributing to reproductive tract integrity; and the ovarian cancer biomarkers CA-125
474 (MUC16) and HE4 (WFDC2). Insulin-like growth factor-binding protein 4 (IGFBP4) has been
475 associated with early pregnancy and foetal growth restriction (Figure 4C, Figure S4). MUC1,
476 expressed throughout the female reproductive tract—including endocervical, ectocervical, and
477 vaginal epithelia—plays a role in maintaining epithelial barrier function (Figure 4D, Figure
478 S4). Both proteins are physiologically expressed across the menstrual cycle. In addition,
479 clinically relevant biomarkers CA-125 (Figure 4E, Figure S4) and HE4 (Figure 4F, Figure S4),
480 encoded by MUC16 and WFDC2 respectively, were consistently detected. Longitudinal
481 analysis across the ovulation-anchored repeated-measures window (*d0–d6*) revealed
482 significant day-to-day variation in all four proteins (Figure 4C–G; Figure S4), supporting the
483 suitability of dia-PASEF for time-dependent biomarker profiling.
484

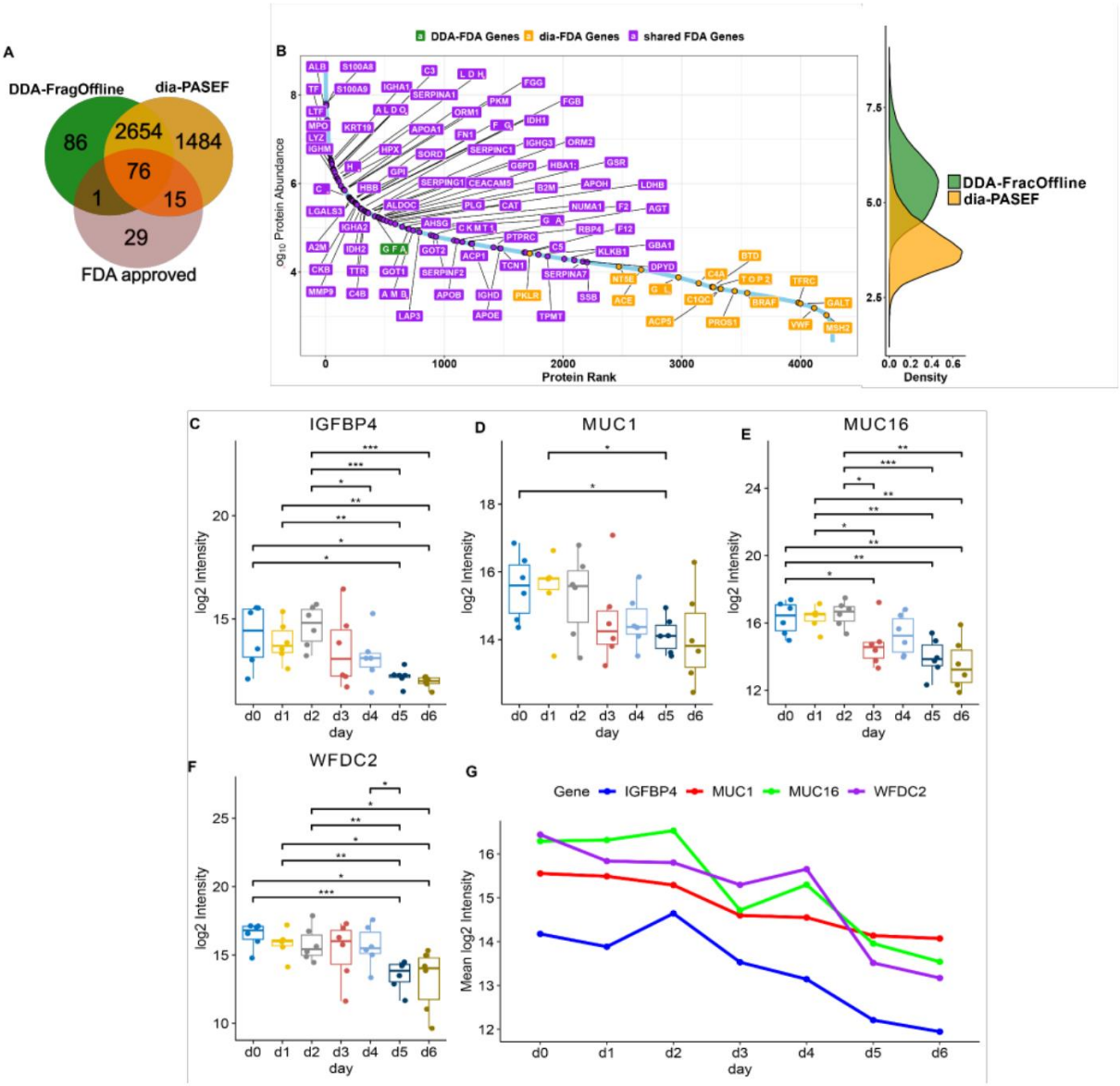

Figure 4. Biomarker identification and longitudinal protein analysis from *d0* to *d6* by DDA-FracOffline and dia-PASEF. (A) Venn diagram illustrating the overlap of proteins identified and FDA-approved biomarkers, highlighting 76 shared proteins and unique identifications for each strategy. (B) Abundance rank plot of proteins identified in CVF samples via DDA-FracOffline alone (green), dia-PASEF alone (orange), and both strategies (purple). FDA approved biomarker proteins are highlighted. The right panel displays the density distribution of protein abundances. (C, D) Box plots show longitudinal expression patterns of the pregnancy-pathology-associated gene IGFBP4 and the physiological protein MUC-1. (E, F) Box plots of MUC16 and WFDC2. (***, p value < 0.001; **, p value < 0.01; *, p value < 0.05) (G) Trajectory of the longitudinal expression of IGFBP4, MUC1, MUC16 and WFDC2.

498

Table 2: Summary of biomarkers associated with female pathologies.

Pathology	Gene Symbol	References
Pregnancy pathologies	*FBLN1*	43,44
	NCSTN	45,46
	ACTR2	45,46
	NDUFB3	47
	IGFBP4	48–50
Cervical carcinoma	*MUC5B*	51–53
	ANXA2	51–54
	ACTN4	51–54
	PKM	53,54
Ovarian carcinoma	*MUC16*	55–59
	NT5E	60
	WFDC2	9,58–60
	KLK5	61
	MUC1	62,63
	CHI3L1	56,57
Endometrial carcinoma	*MUC16*	64–66
	CTTNB1	6,64,67,68
	WFDC2	64,65,68
	MMP9	64,68,69
Endometriosis	*TOP2A*	70,71
	S100A8	15,72–74
	LCN2	15
	LUM	72,74–76

499

500 **Discussion**

501

502 CVF is well established as a valuable biological source for research on the pathophysiology of
503 the female reproductive tract. This biofluid contributes to multiple biological processes,
504 including barrier function, immune responses, microbiota preservation, sperm transport, and
505 menstrual cycle regulation. The emergence of advanced proteomics workflows and MS
506 instrumentation provides opportunities for large-scale identification of reproductive health
507 biomarkers in CVF [5]. The present study evaluated the performance of dia-PASEF as a high-
508 throughput proteomic workflow for analysing CVF in longitudinal repeated-measures designs
509 with time-series sampling. To our knowledge, this is among the first applications of dia-PASEF
510 to CVF collected across consecutive days of the menstrual cycle, showing improved
511 performance relative to DDA-FracOffline. We observed greater quantification depth,
512 reproducibility, and sensitivity for low-abundance proteins following comprehensive
513 comparative analyses. The reduced variability (CV = 5.27 ± 0.68%) and lower proportion of
514 missing values enabled more consistent quantification across samples, improving the
515 robustness of longitudinal analyses compared with DDA-FracOffline (Figure 2). These
516 findings indicate that dia-PASEF supports reliable tracking of protein abundance over time,
517 facilitating the investigation of time-dependent molecular changes in CVF.

518

519 Studies of longitudinal effects are essential for assessing the development of physiological
520 phenomena over time, such as ultra-early pregnancy, during which the earliest stages remain
521 poorly understood. These analyses are also relevant for monitoring disease progression and
522 therapeutic responses from a pathophysiological perspective. However, longitudinal analysis
523 of CVF remains technically challenging due to the intrinsic complexity of this biofluid and the
524 limited availability of well-characterised repeated-measures datasets spanning the menstrual
525 cycle. In this context, the application of advanced acquisition strategies such as dia-PASEF
526 represents a methodological advancement, as it provides analytical features suited for complex
527 biofluids and improved detection of low-abundance proteins in time-dependent study designs.

528

529 We observed a shorter data processing time for dia-PASEF (a few minutes per sample) than
530 for DDA-FracOffline (approximately 1–2 hours per sample), supporting its suitability for high-
531 throughput longitudinal repeated-measures proteomic studies. The higher reproducibility
532 observed with dia-PASEF is consistent with its systematic precursor selection strategy, which
533 enhances proteome depth and sensitivity [77]. In contrast, the sequential precursor selection

534 inherent to DDA-FracOffline is associated with greater variability and reduced efficiency in
535 protein sampling, limiting its applicability in large-scale and time-dependent study designs.
536
537 The ability of dia-PASEF to capture comprehensive fragmentation data within a single
538 acquisition cycle contributes to improved protein identification, reduced missing values, and
539 broader proteome coverage. The greater quantitative stability observed with dia-PASEF further
540 supports its application in large-scale longitudinal proteomics, in line with previous reports
541 describing the analytical advantages of data-independent acquisition strategies [26].
542
543 Although both methodologies allow in-depth protein identification, we observed that dia-
544 PASEF enabled more consistent protein quantification than DDA-FracOffline, primarily
545 because the latter showed lower quantifiable protein coverage and a higher proportion of
546 missing peptides [29]. When comparing dia-PASEF with other DIA approaches applied to CVF,
547 such as SWATH-MS–based methods [7], dia-PASEF identified and quantified a substantially
548 higher number of proteins. Previous proteomic studies using SWATH-MS in CVF have
549 reported limitations including reduced sensitivity for low-abundance proteins, peptide co-
550 fragmentation effects [26], and higher proportions of missing values.
551
552 Enrichment analyses enabled exploration of the functional relevance of identified proteins
553 contributing to CVF pathophysiology (Figure 3). Both strategies highlighted pathways related
554 to cellular signalling, protein metabolism, immune responses, interleukin signalling, mitotic
555 metaphase–anaphase transition, mRNA splicing, and RHO GTPase signalling, which are
556 implicated in reproductive system biology [78,79]. dia-PASEF showed higher enrichment scores
557 across the time-dependent window, suggesting improved biological coverage. For example,
558 neutrophil gelatinase-associated lipocalin (NGAL; encoded by LCN2), involved in neutrophil
559 degranulation [15] signalling pathways, was consistently detected (Figure 3, Figure 4H).
560
561 Tissue enrichment analysis confirmed the biological relevance of the identified proteins,
562 highlighting associations with tissues of the female reproductive system [80], including the uterus
563 and uterine cervix. These findings are consistent with the ability of dia-PASEF to provide
564 system-specific sensitivity and time-dependent biological profiling, particularly for
565 reproductive and immune-related functions in longitudinal CVF proteomics.
566

567 We compared proteins identified by dia-PASEF and DDA-FracOffline with the FDA-proposed
568 biomarker panel [42] . Using the dia-PASEF workflow, we identified 15 unique quantifiable
569 FDA-proposed biomarkers, compared with one unique quantifiable biomarker detected using
570 DDA-FracOffline (Figure 4A). Dia-PASEF identified biomarkers across a broader dynamic
571 range, including both high- and low-abundance proteins (Figure 4B), supporting its suitability
572 for comprehensive biomarker profiling in CVF. These findings indicate improved detection
573 performance of dia-PASEF for FDA-proposed biomarkers within this longitudinal repeated-
574 measures design.
575
576 Proteins detected at low abundance were linked to biological processes relevant to menstrual
577 cycle regulation and gynaecological diseases. For example, C1QC and VWF are involved in
578 immune responses and vascular regulation. A broader and more uniform dynamic range was
579 observed with dia-PASEF, likely reflecting improved signal-to-noise performance, a feature
580 that is advantageous for longitudinal analyses in complex biofluids[34] such as CVF. As an
581 additional strategy for evaluating clinically relevant biomarkers, dia-PASEF facilitated the
582 detection of low-abundance proteins across the time-dependent repeated-measures window.
583 IGFBP-4 (Figure 4C), whose levels have been associated with early pregnancy and foetal
584 growth restriction [48,50] and MUC1 (Figure 4D), a mucin expressed throughout the female
585 reproductive tract—including endocervical, ectocervical, and vaginal epithelia—were
586 consistently detected. MUC1 plays a role in maintaining epithelial integrity and barrier function
587 [62,63] .
588
589 We also detected CVF proteins associated with pathological processes, including alpha-
590 actinin-4 (ACTN4) and annexin A2 (ANXA2), which have been reported in previous studies
591 [51–53] (Table 2). In the context of gynaecological carcinomas, two clinically relevant proteins
592 were identified: cancer antigen 125 (CA-125), encoded by the MUC16 gene (Figure 4E), and
593 human epididymis protein 4 (HE4), encoded by the WFDC2 gene (Figure 4F). These proteins
594 were detected in CVF samples collected across the ovulation-anchored window (*d0–d6*). Both
595 biomarkers have been widely investigated in ovarian [58,59] and endometrial [64,65] carcinoma.
596 Longitudinal analysis revealed significant day-to-day variation across the repeated-measures
597 window (Figure S4). These findings suggest that dia-PASEF facilitates time-dependent
598 monitoring of clinically relevant proteins in CVF, supporting its potential application in studies
599 of pathologies related to the female reproductive tract.
600

601 Proteins present in CVF may also represent promising candidates for identifying longitudinal
602 pathophysiological markers. Owing to its reduced proportion of missing values and more
603 consistent time-dependent protein profiling, dia-PASEF generated a more complete
604 quantitative dataset (Figure 2B). This characteristic supports its application for tracking protein
605 dynamics over time in physiological processes, such as endometrial receptivity, as well as in
606 pathological contexts. A consistent pattern of protein variation was observed across specific
607 days of the cycle (Figure 4G, Figure S4), providing insight into defined physiological windows
608 within the ovulation-anchored repeated-measures framework. These findings support the
609 feasibility of longitudinal biomarker monitoring in CVF, including potential applications in
610 therapeutic follow-up studies.
611
612 This study is focused primarily on achieving methodological robustness to enable a reliable
613 comparison between strategies and to highlight the advantages of the dia-PASEF approach.
614 Despite the limited sample size, our parsimonious approach[81] minimises variability by
615 selecting a homogeneous subgroup of volunteers based on exclusion criteria that control for
616 factors such as body mass index, age, menstrual cycle length, and reproductive history.
617
618 The participants were also synchronised according to the surge in luteinising hormone from *d0*
619 to *d6* (peri-ovulatory to early luteal phase of menstruation), which minimised interindividual
620 variability. This approach enabled the selection of volunteers with complete repeated-measures
621 data for implementation and validation of the proteomic workflows, thereby minimising
622 technical and biological noise in biomarker identification. By using a within-individual
623 longitudinal framework with a semi-counterfactual structure, this design further reduced the
624 influence of population heterogeneity. Owing to the methodological scope and limited sample
625 size of the study, the findings are not directly generalisable to pathological conditions.
626 However, future expansion of the cohort to include participants with defined pathologies will
627 be necessary to evaluate the potential of dia-PASEF in identifying disease-specific biomarkers.
628 In addition, functional validation of candidate biomarkers will be essential to clarify how these
629 proteins are involved in specific physiological or pathological processes and to assess their
630 potential clinical relevance
631
632 This study demonstrates the suitability of dia-PASEF as a robust analytical platform for
633 longitudinal repeated-measures CVF proteomics, addressing limitations inherent to DDA-
634 based workflows and supporting biomarker discovery in fertility, early pregnancy, and

gynaecological research. Overall, this approach provides a scalable methodological framework for time-dependent proteomic investigations in complex reproductive biofluids.

Author's roles

E.S.K. led the study. M.H. and E.S.K. drafted the manuscript and conceptualised the study. E.S.K., M.H., and C.V. contributed to the study design. M.H. designed, supervised, and performed the proteomics methodology. P.S., B.A., and V.L. contributed to study conception. M.H., P.S., and G.N. developed the experimental system, performed the experiments, and analysed the data. F.C. and G.N. supported the experimental work. M.H., P.S., B.A., and V.L. reviewed and commented on manuscript drafts. E.S.K. revised the manuscript and approved the final version. E.S.K. is the guarantor of the study and accepts full responsibility for the integrity of the work.

Acknowledgements

The authors thank Dr. Ute Woehlbier (Center for Integrative Biology, Universidad Mayor) for English language editing. We also acknowledge Estefanía Nova-Lamperti (Molecular and Translational Immunology Laboratory, Department of Clinical Biochemistry and Immunology, Faculty of Pharmacy, Universidad de Concepción, Chile) for assistance with figure design. I.F.-M. received professional service fees from Fundación de Investigación San Ramón (FISAR), Concepción, Chile.

Funding

This work forms part of the EARLY-PREG preconception open cohort and has been 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.

Conflict of interest

E.S.K. serves in an unpaid honorary capacity as a research advisor and reviewer for Fundación de Investigación San Ramón (FISAR). No other competing interests are declared.

669

670 **References**

671 1. Beutgen, V. M. & Graumann, J. Advances in aqueous humor proteomics for biomarker
672 discovery and disease mechanisms exploration: a spotlight on primary open angle glaucoma.
673 *Front. Mol. Neurosci.* **17**, 1397461 (2024).
674 2. Masina, R. & Caldas, C. Precision Cancer Medicine 2.0-Oncology in the postgenomic era. *Mol.*
675 *Oncol.* **18**, 2065–2069 (2024).
676 3. Rafelski, S. M. & Theriot, J. A. Establishing a conceptual framework for holistic cell states and
677 state transitions. *Cell* **187**, 2633–2651 (2024).
678 4. Holmes, C., Carlson, S. M., McDonald, F., Jones, M. & Graham, J. Exploring the post-genomic
679 world: differing explanatory and manipulatory functions of post-genomic sciences. *New Genet.*
680 *Soc.* **35**, 49–68 (2016).
681 5. Kaur Jawanda, I., Soni, T., Kumari, S. & Prabha, V. Deciphering the potential of proteomic-
682 based biomarkers in women's reproductive diseases: empowering precision medicine in
683 gynecology. *Biomarkers* **29**, 7–17 (2024).
684 6. Martinez-Garcia, E. *et al.* Cervical fluids are a source of protein biomarkers for early, non-
685 invasive endometrial cancer diagnosis. *Cancers (Basel)* **15**, 911 (2023).
686 7. Njoku, K. *et al.* Comprehensive library generation for identification and quantification of
687 endometrial cancer protein biomarkers in cervico-vaginal fluid. *Cancers (Basel)* **13**, 3804 (2021).
688 8. Njoku, K. *et al.* Detection of endometrial cancer in cervico-vaginal fluid and blood plasma:
689 leveraging proteomics and machine learning for biomarker discovery. *EBioMedicine* **102**,
690 105064 (2024).
691 9. Hedlund Lindberg, J. *et al.* Toward ovarian cancer screening with protein biomarkers using
692 dried, self-sampled cervico-vaginal fluid. *iScience* **27**, 109001 (2024).
693 10. Köse, B. *et al.* Identifying biomarkers for cervical neoplasia: A label-free proteomic analysis of
694 cervicovaginal fluid. *Eur. Arch. Med. Res.* **40**, 36–43 (2024).
695 11. Gutiérrez, A. L. *et al.* Identification of candidate protein biomarkers for CIN2+ lesions from self-
696 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. Proteomics* **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. Proteomics* **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**, eadl5762 (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. *Molecular & Cellular 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* 2025.05.12.25326372 (2026) doi:10.1101/2025.05.12.25326372.

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. Proteomics* **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. *Nature Methods* **14**, 513–520 (2017).

753 36. da Veiga Leprevost, F. *et al.* Philosopher: a versatile toolkit for shotgun proteomics data
754 analysis. *Nature Methods* **17**, 869–870 (2020).
755 37. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural
756 networks and interference correction enable deep proteome coverage in high throughput. *Nature*
757 *Methods* **17**, 41–44 (2019).
758 38. Stekhoven, D. J. & Bühlmann, P. MissForest--non-parametric missing value imputation for
759 mixed-type data. *Bioinformatics* **28**, 112–118 (2012).
760 39. Lazar, C., Gatto, L., Ferro, M., Bruley, C. & Burger, T. Accounting for the multiple natures of
761 missing values in label-free quantitative proteomics data sets to compare imputation strategies. *J.*
762 *Proteome Res.* **15**, 1116–1125 (2016).
763 40. Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D. & Iverson, G. Bayesian t tests for
764 accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* **16**, 225–237 (2009).
765 41. Palasca, O., Santos, A., Stolte, C., Gorodkin, J. & Jensen, L. J. TISSUES 2.0: an integrative web
766 resource on mammalian tissue expression. *Database (Oxford)* **2018**, (2018).
767 42. Whelan, S. A. *et al.* Assessment of a 60-biomarker health surveillance panel (HSP) on whole
768 blood from remote sampling devices by targeted LC/MRM-MS and discovery DIA-MS analysis.
769 *Anal. Chem.* **95**, 11007–11018 (2023).
770 43. Beernink, R. H. J., Schuitemaker, J. H. N., Zwertbroek, E. F., Scherjon, S. A. & Cremers, T. I. F.
771 H. Early pregnancy biomarker discovery study for spontaneous preterm birth. *Placenta* **139**,
772 112–119 (2023).
773 44. McElrath, T. F. *et al.* Circulating microparticle proteins obtained in the late first trimester predict
774 spontaneous preterm birth at less than 35 weeks' gestation: a panel validation with specific
775 characterization by parity. *Am. J. Obstet. Gynecol.* **220**, 488.e1–488.e11 (2019).
776 45. Wei, C. *et al.* Identification and verification of diagnostic biomarkers in recurrent pregnancy loss
777 via machine learning algorithm and WGCNA. *Front. Immunol.* **14**, 1241816 (2023).
778 46. Wei, P. *et al.* Identification and validation of a signature based on macrophage cell marker genes
779 to predict recurrent miscarriage by integrated analysis of single-cell and bulk RNA-sequencing.
780 *Front. Immunol.* **13**, 1053819 (2022).

781 47. Yin, X.-J., Hong, W., Tian, F.-J. & Li, X.-C. Proteomic analysis of decidua in patients with
782 recurrent pregnancy loss (RPL) reveals mitochondrial oxidative stress dysfunction. *Clin.*
783 *Proteomics* **18**, 9 (2021).
784 48. Qiu, Q. *et al.* Significance of IGFBP-4 in the development of fetal growth restriction. *J. Clin.*
785 *Endocrinol. Metab.* **97**, E1429–39 (2012).
786 49. Zhang, T., Du, T., Li, W., Yang, S. & Liang, W. Sex hormone-binding globulin levels during the
787 first trimester may predict gestational diabetes mellitus development. *Biomark. Med.* **12**, 239–
788 244 (2018).
789 50. Hedderson, M. M. *et al.* Prepregnancy SHBG concentrations and risk for subsequently
790 developing gestational diabetes mellitus. *Diabetes Care* **37**, 1296–1303 (2014).
791 51. Van Ostade, X., Dom, M., Tjalma, W. & Van Raemdonck, G. Candidate biomarkers in the
792 cervical vaginal fluid for the (self-)diagnosis of cervical precancer. *Arch. Gynecol. Obstet.* **297**,
793 295–311 (2018).
794 52. Starodubtseva, N. L. *et al.* Label-free cervicovaginal fluid proteome profiling reflects the cervix
795 neoplastic transformation. *J. Mass Spectrom.* **54**, 693–703 (2019).
796 53. Mukherjee, A. *et al.* Insights on proteomics-driven body fluid-based biomarkers of cervical
797 cancer. *Proteomes* **10**, 13 (2022).
798 54. Van Raemdonck, G. A. A., Tjalma, W. A. A., Coen, E. P., Depuydt, C. E. & Van Ostade, X. W.
799 M. Identification of protein biomarkers for cervical cancer using human cervicovaginal fluid.
800 *PLoS One* **9**, e106488 (2014).
801 55. Olalekan, S., Xie, B., Back, R., Eckart, H. & Basu, A. Characterizing the tumor
802 microenvironment of metastatic ovarian cancer by single-cell transcriptomics. *Cell Rep.* **35**,
803 109165 (2021).
804 56. Qian, L. *et al.* Proteomic landscape of epithelial ovarian cancer. *Nat. Commun.* **15**, 6462 (2024).
805 57. Whitwell, H. J. *et al.* Improved early detection of ovarian cancer using longitudinal multimarker
806 models. *Br. J. Cancer* **122**, 847–856 (2020).
807 58. Radu, M. R. *et al.* Ovarian cancer: Biomarkers and targeted therapy. *Biomedicines* **9**, 693 (2021).
808 59. Atallah, G. A., Abd Aziz, N. H., Teik, C. K., Shafiee, M. N. & Kampan, N. C. New predictive

809 biomarkers for ovarian cancer. *Diagnostics (Basel)* **11**, 465 (2021).

810 60. Lupia, M. *et al.* CD73 regulates stemness and epithelial-mesenchymal transition in ovarian
811 cancer-initiating cells. *Stem Cell Reports* **10**, 1412–1425 (2018).

812 61. Kim, H. *et al.* Human kallikrein gene 5 (KLK5) expression is an indicator of poor prognosis in
813 ovarian cancer. *Br. J. Cancer* **84**, 643–650 (2001).

814 62. Gipson, I. K. *et al.* Mucin genes expressed by human female reproductive tract epithelia. *Biol.*
815 *Reprod.* **56**, 999–1011 (1997).

816 63. Moncla, B. J., Chappell, C. A., Debo, B. M. & Meyn, L. A. The effects of hormones and vaginal
817 microflora on the glycome of the female genital tract: Cervical-vaginal fluid. *PLoS One* **11**,
818 e0158687 (2016).

819 64. Njoku, K., Chiasserini, D., Whetton, A. D. & Crosbie, E. J. Proteomic biomarkers for the
820 detection of endometrial cancer. *Cancers (Basel)* **11**, 1572 (2019).

821 65. Coll-de la Rubia, E. *et al.* Prognostic biomarkers in endometrial cancer: A systematic review and
822 meta-analysis. *J. Clin. Med.* **9**, 1900 (2020).

823 66. Degez, M. *et al.* Endometrial cancer: A systematic review of HE4, REM and REM-B. *Clin.*
824 *Chim. Acta* **515**, 27–36 (2021).

825 67. Romano, A. *et al.* Endometrial cancer diagnostic and prognostic algorithms based on proteomics,
826 metabolomics, and clinical data: a systematic review. *Front. Oncol.* **13**, 1120178 (2023).

827 68. Martinez-Garcia, E. *et al.* Targeted proteomics identifies proteomic signatures in liquid biopsies
828 of the endometrium to diagnose endometrial cancer and assist in the prediction of the optimal
829 surgical treatment. *Clin. Cancer Res.* **23**, 6458–6467 (2017).

830 69. Njoku, K. *et al.* Quantitative SWATH-based proteomic profiling of urine for the identification of
831 endometrial cancer biomarkers in symptomatic women. *Br. J. Cancer* **128**, 1723–1732 (2023).

832 70. Mirza, Z. & Abdel-Dayem, U. A. Uncovering potential roles of differentially expressed genes,
833 upstream regulators, and canonical pathways in endometriosis using an in silico genomics
834 approach. *Diagnostics (Basel)* **10**, 416 (2020).

835 71. Yang, H., Kang, K., Cheng, C., Mamillapalli, R. & Taylor, H. S. Integrative analysis reveals
836 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. Proteomics* **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. Proteomics* **20**, 319–330 (2023).

81. Raykov, T. & Marcoulides, G. A. On desirability of parsimony in structural equation model selection. *Struct. Equ. Modeling* **6**, 292–300 (1999).