High-throughput longitudinal proteomics of cervicovaginal fluid during consecutive days of women's menstrual cycles: a new robust, sensitive, and efficient method

Mauricio Hernandez [1]†, Pablo Saldivia[1,2]†, Verónica Latapiat[1]†, Bárbara 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, Concepcion, Chile.
[2] Programa de Doctorado en Biotecnología Molecular, Facultad de Ciencias Biológicas, Universidad de Concepción, Concepción, Chile.

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

Author email addresses

Pablo Fernando Saldivia: psaldivia@melisainstitute.org

Barbara Evelyn Antilef: bantilef@melisainstitute.org

Verónica Xiomar Latapiat: vlatapiat@melisainstitute.org

Felipe Ignacio Castro: fcastro@melisainstitute.org

Cristian Andres Vargas: cvargas@melisainstitute.org

Mauricio Alejandro Hernández: mhernandez@melisainstitute.org

Guillermo Alejandro Nourdin: gnourdin@melisainstitute.org

Corresponding Author:

* Elard S. Koch

MELISA Institute, Dalcahue 1120, San Pedro de la Paz, 4133515, Concepcion, Chile.
Telephone +56 412467242 E-mail: ekoch@melisainstitute.org

Abstract

Proteomic studies of the female reproductive system are expected to significantly impact women's reproductive health and the ability to address fertility problems. Cervicovaginal fluid (CVF), a complex combination of uterine, cervical, and vaginal secretions, represents an important source of marker molecules to understand different pathophysiological processes of the reproductive tract. However, reliable biomarker discovery on this fluid remains challenging due to limitations in the sensitivity and reproducibility of current mass spectrometry techniques for large-scale longitudinal analyses. To address these challenges, this study compared dia-PASEF with the conventional DDA-FracOffline approach in a longitudinal proteomic study of CVF samples from the EARLY-PREG preconception open cohort biorepository. The study included 7-day menstrual cycles with day-by-day CVF samples since ovulation. Regarding the number of identified and quantified proteins over the entire period, DDA-FracOffline detected 2,817 quantifiable proteins versus 4,229 for dia-PASEF. The dia-PASEF approach performed 33% better than DDA for detecting quantifiable proteins in CVF with a coefficient of variation as low as 5%. The data-independent approach overcomes the limitations of DDA for longitudinal prospective studies with repeated measurements, enabling the quantification of dynamic changes over time in a panel of biomarkers associated with key biological pathways and pathophysiological processes of the female reproductive system. Thus, the dia-PASEF method emerges as a robust, sensitive, and efficient method for longitudinal proteomic studies of CVF in women's health research and diagnosis.

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

Introduction

Proteomic studies of the female reproductive system are expected to significantly impact women's health monitoring and the ability to address fertility problems. 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 dynamic changes in female reproductive 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 methodologic limitations[5].

Cervicovaginal fluid (CVF) has emerged as an important 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 response[6,7]. 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[8,9], pregnancy-related complications[10,11] and several cancer types, such as ovarian, endometrial, and cervical malignancies[12–17].

High-throughput proteomics has revolutionized 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 (DIA)-based methodologies with the potential to outperform DDA-based methods have recently emerged[19]. For example, Njoku and collaborators[13] recently applied a DIA method using sequential window acquisition of all theoretical fragment ion mass spectra (SWATH-MS) to analyse CVF samples, which had an increased capacity to quantify a greater number of proteins than did a DDA standard approach.

Addressing DDA limitations is critical for advancing clinical applications of CVF proteomics, particularly through longitudinal cohort studies[20,21] that provide a unique opportunity to identify potential biomarkers and capture dynamic pathophysiological changes[22,23]. In fact, following dynamic biomarkers over time may reveal molecular signatures linked to specific

biological processes, offering deeper insights 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, no proteomic studies have been conducted to date that evaluate the "evolving proteome" of CVF over time series of 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 may overcome this latter drawback, 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 emerged as a powerful quantitative research tool[29] offering a more robust solution 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 dynamic proteomic analyses of CVF samples collected from menstrual cycles of women participating in the EARLY-PREG longitudinal preconception open cohort[80]. The new dia-PASEF approach is contrasted with the traditional DDA coupled with the offline peptide fractionation (DDA-FracOffline) method. It is hypothesised that the dia-PASEF approach would outperform traditional DDA in both depth of proteome coverage and reproducibility, providing a robust platform for biomarker discovery in longitudinal studies of women's reproductive health, fertility, and early pregnancy.

Results

Patient Characterization and Sampling

After screening all non-conception menstrual cycles of women with complete records in the EARLY-PREG open cohort database[80], we selected six women based on criteria of homogeneity for variables such as age (32.3 ± 3.6), weight, height, BMI, waist circumference, age at menarche (12.7 ± 1.2), menstrual cycle length, and menstruation length. The information for each woman is detailed in Table 1. The length of the menstrual phase of the cycle (5.8 ± 0.4), the total length of the cycle (29.0 ± 3.0), and other variables were within a comparable

range. Forty-two CVF samples (n=42) stored at -80 °C were processed for protein extraction and tryptic digestion and then analysed with the two mass spectrometry methodologies (Figure 1A) as described in the methods section. To avoid batch effect bias, we ran all 42 samples on timsTOF Pro mass spectrometer consecutively in the same calibration batch: 336 chromatographic runs via the DDA method and 42 direct runs via the DIA method over a time frame of 2 weeks. Thus, with each protocol, we obtained a 7-day time series (*d0* to *d6*) of CVF proteomes for cycles of each woman, all synchronized by the ovulation day as *d0*, which is the first day of the series.

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

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

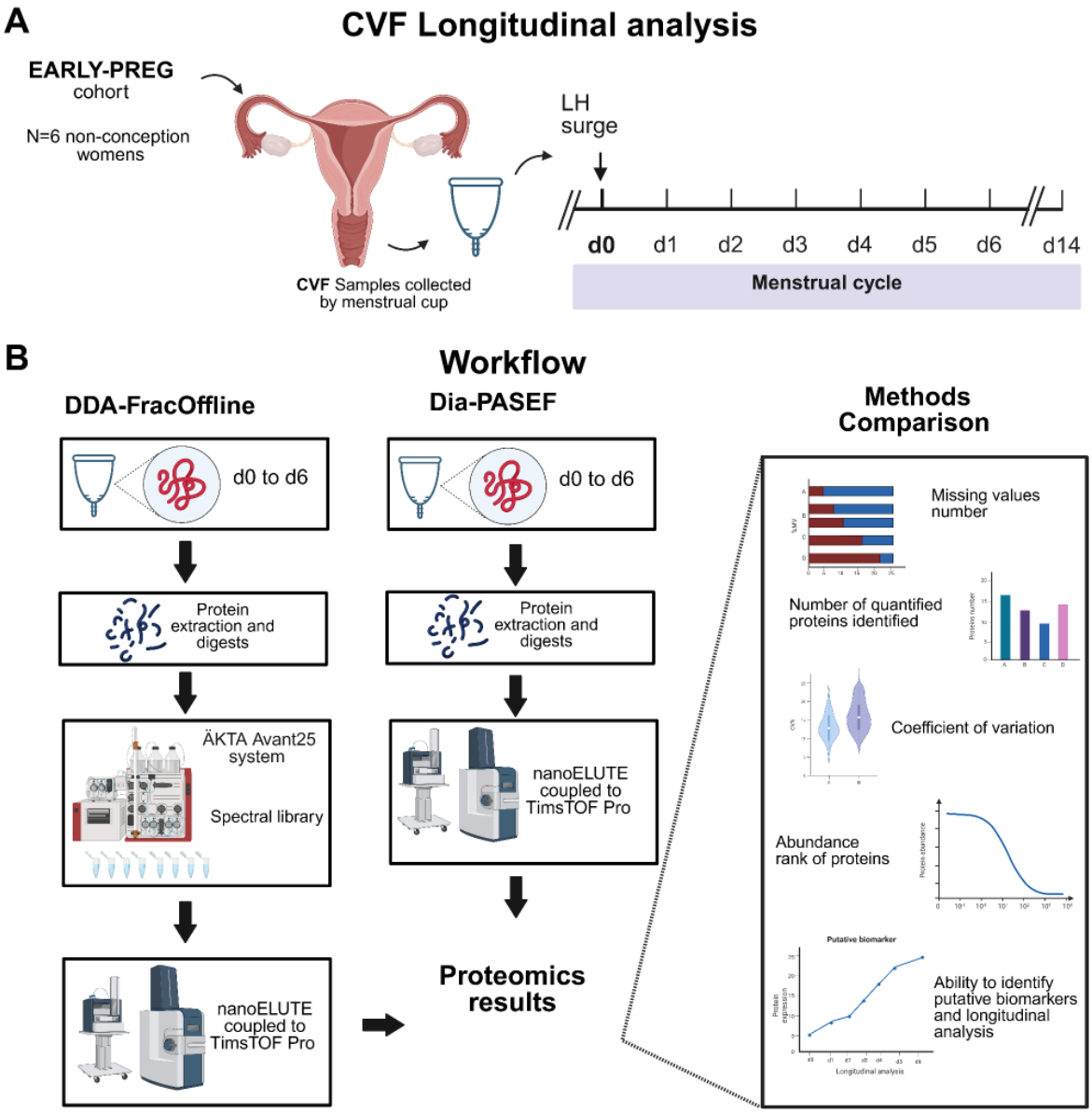

Figure 1: Proteomics workflow for the longitudinal analysis of CVF. Schematic representation of the experimental design and strategies used to analyse CVF samples from women in the EARLY-PREG cohort. (A) CVF samples were collected daily from d0 to d6 of the menstrual cycle. (B) Two nLC-MS/MS proteomics 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 shows the method comparison, evaluating missing values, the number of quantified proteins, the coefficient of variation, the abundance rank distribution, and the ability to identify putative biomarkers and track longitudinal proteomic changes.

Comparative Quality Control Factors of the DDA-FracOffline and dia-PASEF Strategies

We developed an analysis focusing on different quality control factors, such as reproducibility, the proportion of missing values, protein identification and quantification performance. First, we assessed the reproducibility of the two methods considering the correlation coefficients between biological replicates, with coefficients of approximately 0.70 for DDA-FracOffline

and 0.90 for dia-PASEF **(Figure S1)**. Second, we evaluated the proportion of missing values with each methodology. The samples processed with dia-PASEF presented a lower proportion of missing values than did those processed with DDA-FracOffline (reaching up to 75% at its lowest point) **(Figure 2B)**. Third, through correlation analysis of both methodologies, we detected a lower intergroup correlation for DDA-FracOffline than for dia-PASEF **(Figure 2A)**. Next, we analysed the protein identification performance of each method on CVF serial samples. DDA-FracOffline identified a total of 9,593 proteins, whereas dia-PASEF identified 5,986 proteins. However, considering the number of identified and measured proteins (quantifiable proteins), DDA-FracOffline quantified 2,817 proteins, whereas dia-PASEF quantified 4,229 **(Figure 2E, left).** This indicated that dia-PASEF has a higher yield of 33% than does DDA-FracOffline. Additionally, the average coefficient of variation (CV) for quantifiable proteins—calculated as the mean of individual CVs for each protein across all samples—was 5.27 ± 0.68% for dia-PASEF, whereas it was 9.37 ± 1.75% for DDA-FracOffline **(Figure 2E, centre)**. We detected overlap among the quantifiable proteins identified via the two methods, and we observed that the majority were present in the dia-PASEF dataset **(Figure 2E, right)**. Finally, by employing DDA-FracOffline, we identified 1,491 to 7,183 proteins **(Figure 2C)**, in contrast to employing dia-PASEF, which identified 3,811 to 5,398 proteins **(Figure 2D)**. Although DDA-FracOffline demonstrated broader overall proteomic coverage, it was less consistent with protein detection across individual samples of CVF over time.

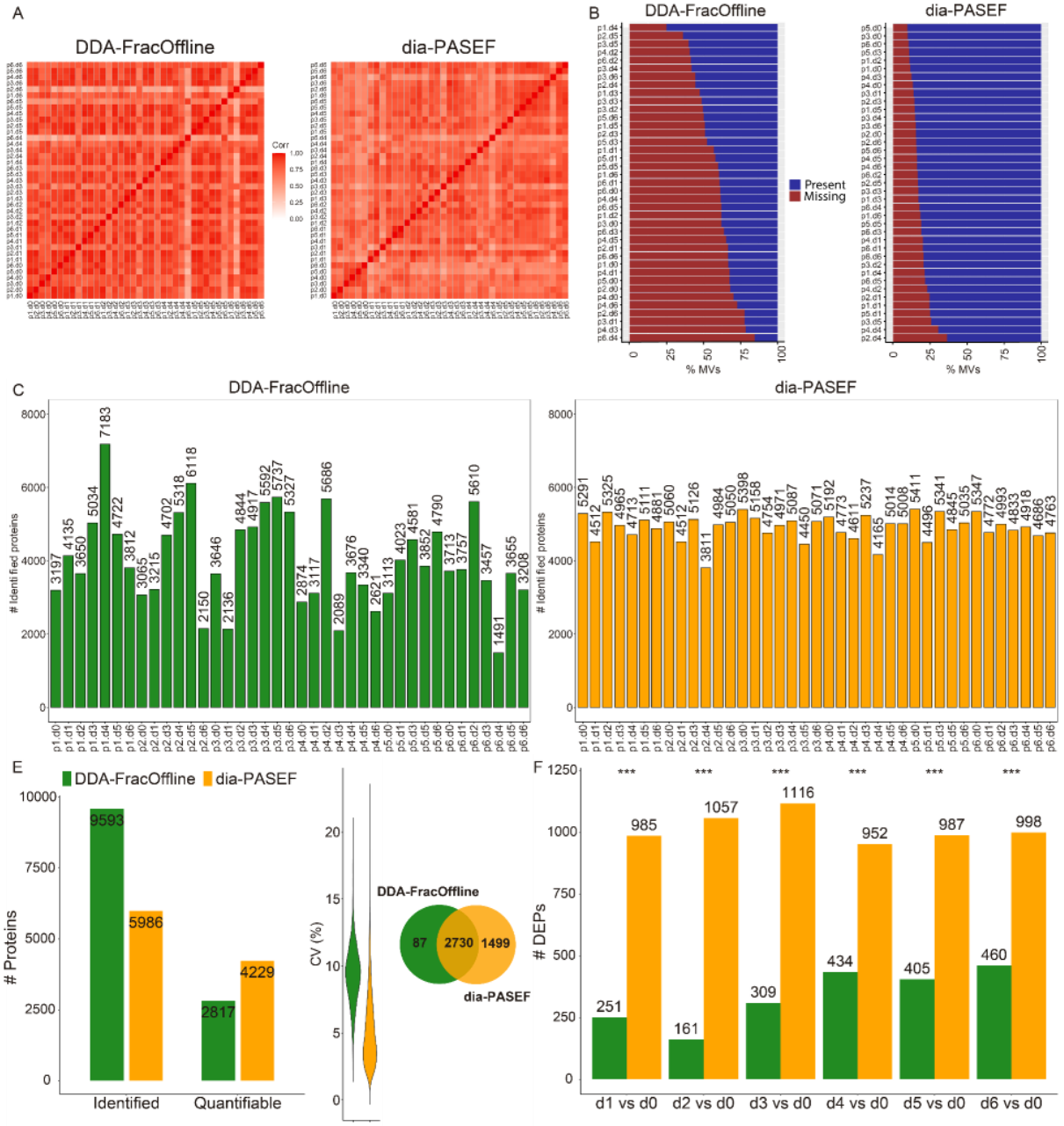

Figure 2. Quality control of proteomic data obtained by DDA-FracOffline and dia-PASEF. (A) Pearson correlations between chromatographic runs, positive correlations (red) and no correlations (white). (B) Bar plots of the proportion of missing values (%MVs) for each sample, showing present values (blue) 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 show the total numbers of identified proteins (left) and quantifiable proteins (centre) for DDA-FracOffline (green) and dia-PASEF (orange). The violin plot (right) presents the coefficient of variation (%CV) for each method. The Venn diagram illustrates the shared and unique proteins identified by each analytical strategy. (F) Comparison of the number of differentially expressed proteins (DEPs) identified through longitudinal comparisons against d0 using each method. The bar plot displays the total number of DEPs for each comparison, 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).

Differential Protein Expression Over Time

After quality control, we analysed the longitudinal dynamics of the differential protein expression profiles across the time window of the cycle to determine the temporal variations in protein expression on CVF samples. Multiple comparison tests revealed a significant difference in the detection of DEPs across longitudinal CVF samples between methodologies **(Figure S2)**. At most time points, dia-PASEF identified a higher absolute and relative numbers of DEPs than did DDA-FracOffline, for example, 985 (23%) vs. 251 (8.9%) between *d0* and *d1* and 1,057 (24.9%) vs. 161 (5.7%) between *d0* and *d2*. In addition, from d0 to d6, dia-PASEF maintained a more stable detection rate (**Figure 2F**). These results highlight that dia-PASEF can detect and quantify a greater number of expressed proteins in longitudinal experimental designs with repeated sampling, allowing the capture of putative changes in the proteome over time **(Figure 2F, Figure S2)**.

Longitudinal enrichment analysis of pathways, systems, tissues and cells

To interpret the functional relevance of the quantified proteins detected, we expanded our analysis to include tissue and pathway enrichment. We evaluated the representation of human tissues based on DEPs across longitudinal CVF comparisons **(Figure S3)**. 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 method demonstrated broader and more consistent tissue coverage, particularly at later time points (*d5* and *d6*), with elevated enrichment scores (-log10 adjusted p value) and the highest protein densities. Furthermore, we explored the biological processes associated with the differential protein profiles detected in CVF across the 7-day time window of the cycle for both strategies. We found representative and common pathways by comparing DDA-FracOffline and dia-PASEF between *d0* to *d6* **(Figure 3)**. Common pathways identified included the 'Formation of the cornified envelope', 'Metabolism of proteins', and 'Processing of Capped Intron-Containing Pre-mRNA' pathways. Additionally, we detected 'Signalling by RHO GTPases' and 'Mitotic Metaphase/Anaphase Transition', reflecting involvement in signalling and the cell cycle **(Figure 3)**. Compared with DDA-FracOffline, dia-PASEF had increased coverage and enrichment on the different days of the studied cycle.

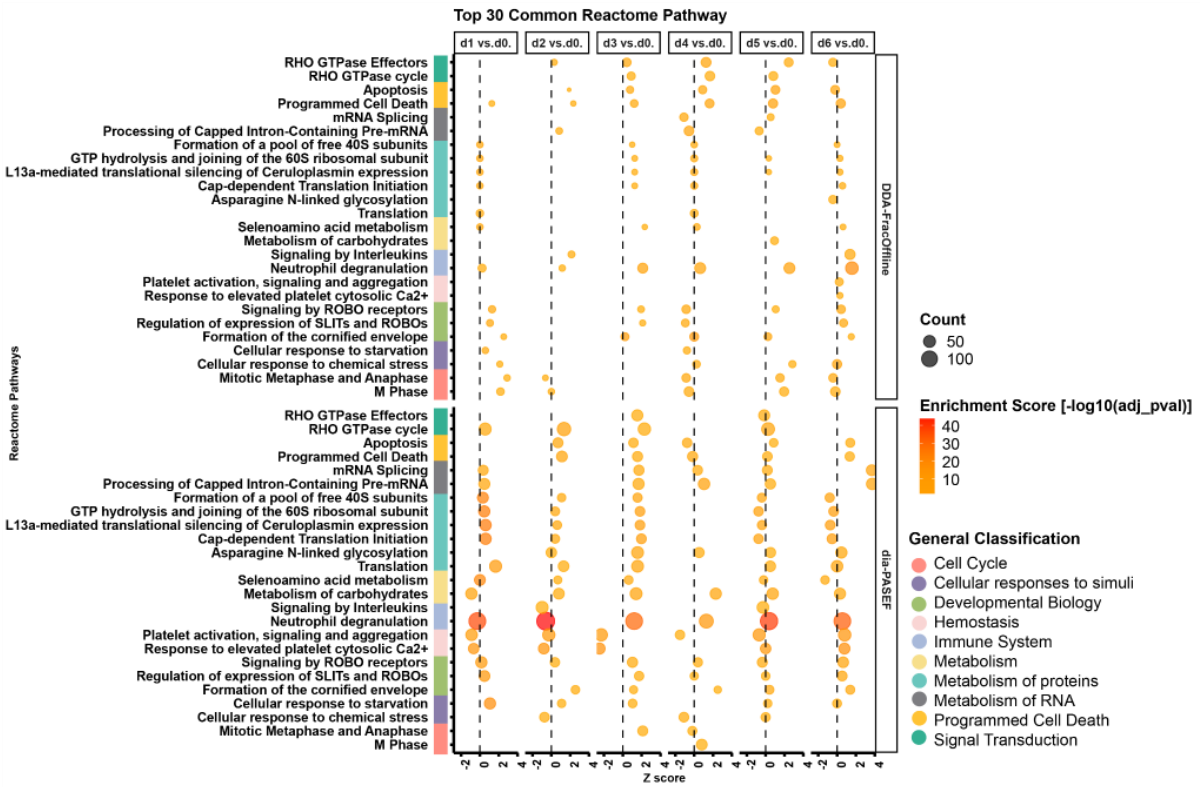

Figure 3. Differentially expressed proteins and biological pathways identified via longitudinal analysis using the DDA-FracOffline and dia-PASEF methods. Reactome enrichment analysis, emphasizing the most abundant biological pathways. DEPs were derived from longitudinal comparisons within each method.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

Performance for Detecting Physiological and pathological biomarkers

Finally, we evaluated the ability of each method to detect and quantify known biomarkers of physiological and pathological processes. First, we used a set of biomarkers proposed by the Food and Drug Administration (FDA) for different diseases[30]. Second, we examined another set of proteins of interest specifically related to the pathophysiology of the female reproductive tract based on an in-house review of the literature (**Table S1**).

The set of biomarkers proposed by the FDA is focused on 121 proteins present in different fluids and is related to multiple pathologies[30]**.** Figure 4 compares the ability of each method to detect FDA biomarkers in CVF samples. A total of 91 out of the 121 proteins between both methods were detected, among which 75% were identified by dia-PASEF and 63% were identified by DDA-FracOffline, as shown in Figure 4A. Additionally, we graphed the detected markers to observe their abundance within the CVF. Our findings indicate that dia-PASEF can detect proteins across a broader dynamic range than can DDA-FracOffline **(Figure 4B)**. In the figure, the FDA biomarkers identified by both methods are purple, whereas the 15 unique

biomarkers detected exclusively by dia-PASEF (*NT5E, ACE, GLA, C4A, C1QC, ACP5, PROS1, BRAF, VWF, GALT, MSH2, TFRC, TOP2A,* and *BTD*) are yellow; *GFAP*, the only FDA-approved biomarker detected exclusively by DDA-FracOffline, is represented in green **(Figure 4B)**. Considering the ability to detect the set of FDA biomarkers in CVF, dia-PASEF could quantify changes in proteins over time and performs better in terms of quantification, annotation, and reproducibility, making it the preferred method for subsequent analysis on this fluid.

In order to assess the ability of the new dia-PASEF method to longitudinally detect other biomarkers of interest for female reproductive health, it was subsequently developed a table of reported biomarkers for processes associated with pregnancy, carcinomas (such as cervical, ovarian, and endometrial), and endometriosis **(Table 2).** The dynamic change of four biomarkers is exemplified in **Figure 4 (C to G)**. First, we focused on insulin-like growth factor binding protein-4 (encoded by the *IGFBP4* gene), which may play a role in early pregnancy and foetal growth restriction **(Figure 4C, Figure S4)**[39,41]. Second, we investigated *MUC1*, a mucin expressed throughout the female reproductive tract, particularly in the endocervical, ectocervical, and vaginal epithelia **(Figure 4D, Figure S4)**. Its widespread expression underscores its essential role in maintaining the integrity and function of the reproductive tract[32,33]. Both genes encode a physiological protein associated with the menstrual cycle. In addition, we analysed proteins that are clinically relevant for diagnosis, such as *CA-125*, which is encoded by the *MUC16* gene **(Figure 4E, Figure S4),** and HE4, which is encoded by the *WFDC2* gene **(Figure 4F, Figure S4)**. Longitudinal analysis of these proteins demonstrated significant changes in protein levels between *d0* and *d6* of the menstrual cycle **(Figure 4G, Figure S4).**

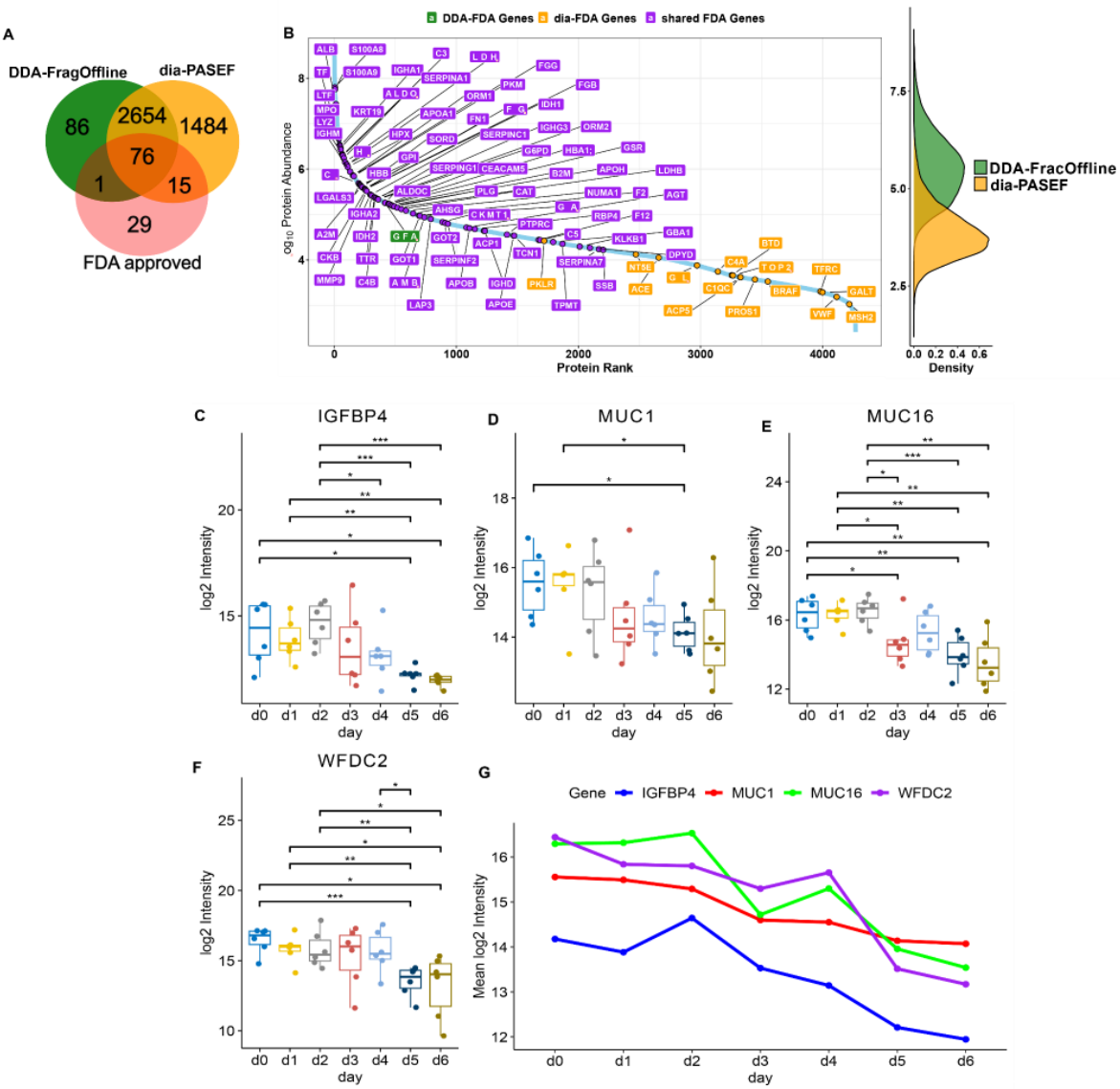

Figure 4. Biomarker identification and longitudinal protein analysis from d0 to d6 by DDA-FracOffline and dia-PASEF. (A) Venn diagram illustrating the overlap of proteins identified and FDA-approved biomarkers, highlighting 76 shared proteins and unique identifications for each method. **(B)** Abundance rank plot of proteins identified in CVF samples via DDA-FracOffline alone (green), dia-PASEF alone (orange), and both methods (purple). Proteins approved as biomarkers by the FDA are highlighted. The right panel displays the density distribution of protein abundances. **(C, D)** Box plots showing 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*.

Table 2: Summary of biomarkers associated with female pathologies.

Pathology	Gene Symbol	References
Pregnancy pathologies	*FBLN1*	34,35
	NCSTN	36,37
	ACTR2	36,37
	NDUFB3	38
	IGFBP4	39–41
Cervical carcinoma	*MUC5B*	42–44
	ANXA2	42–45
	ACTN4	42–45
	PKM	44,45
Ovarian carcinoma	*MUC16*	46–50
	NT5E	51
	WFDC2	15,49–51
	CHI3L1	47,48
	KLK5	52
	MUC1	32,33
Endometrial carcinoma	*MUC16*	53–55
	CTTNB1	12,53,56,57
	WFDC2	53,54,57
	MMP9	53,57,58
Endometriosis	*TOP2A*	59,60
	S100A8	9,61–63
	LCN2	9
	LUM	61,63–65

Discussion

CVF is well known as a valuable biological source for research on the pathophysiology of the female reproductive tract. CVF in fact contributes to various biological processes, including barrier function, immune responses, microbiota preservation, sperm transport, and menstrual cycle regulation. The emergence of new proteomics techniques, methods, and mass spectrometry instruments provides a remarkable opportunity to massively identify reproductive health biomarkers in this biofluid[6]. The present study highlights the performance of dia-PASEF as a high-throughput proteomic tool for analysing CVF in longitudinal cohorts with repeated time series sampling. In fact, this is the first study to use this strategy with CVF for consecutive days of female menstrual cycles, outperforming the gold-standard DDA-FracOffline approach. After carried out extensive comparative analyses, we observed greater depth in protein identification, reproducibility, and sensitivity, particularly for low-abundance proteins. The reduced variability (CV = 5.27 ± 0.68%) and fewer missing values enabled consistent quantification across CVF samples, improving the reliability of the new method for longitudinal analyses compared with the DDA-FracOffline approach (Figure 2). Our results demonstrate that dia-PASEF has the capacity to track protein expression over time, detect biomarkers and monitor complex biological processes particularly for understanding sequential biological events.

Comparison of dia-PASEF with Existing Proteomic Approaches

Longitudinal analyses are essential for evaluating the progression of biological events over time in physiological phenomena, such as ultra-early pregnancy, where early stages remain poorly understood. Furthermore, from a pathophysiological perspective, these analyses potentially enable the monitoring of disease progression and therapeutic response. Longitudinal analysis using CVF samples remains challenging due to the complexity of this biofluids and the limited number of datasets with individuals' data across the menstrual cycle. In this context, the application of advanced acquisition strategies such as dia-PASEF represents a major innovation because it offers features that are advantageous for analysing complex and low-abundance samples.

We observed a shorter data processing time for dia-PASEF (a few minutes per sample) than for DDA-FracOffline (approximately 1-2 hours per sample), highlighting the methodological

advantages of dia-PASEF over DDA-FracOffline for high-throughput and longitudinal proteomic studies. The higher reproducibility observed with dia-PASEF underscores its robustness, likely due to its systematic precursor selection, which enhances proteome depth and sensitivity[66]. In contrast, the sequential precursor selection of DDA-FracOffline leads to greater variability and reduced efficiency in protein sampling[29], limiting the applicability of this method in large-scale studies. The ability of dia-PASEF to capture more data in a single acquisition cycle is a factor that optimizes protein identification, minimizes missing values, provides increased depth, and detects a broader panel of proteins. The increased stability and consistency of dia-PASEF in protein quantification further reinforce the suitability of this method for large-scale proteomic studies, aligning with previous findings on the advantages of data-independent acquisition strategies[26].

Although both methodologies allow in-depth protein identification, we observed that dia-PASEF enabled quantification better than DDA-FracOffline, mainly, because the latter technique has lower protein coverage and a higher proportion of missing peptides[29]. On the other hand, if we compare dia-PASEF with other DIA methods applied to CVF samples such as SWATH-based methods[13] dia-PASEF approach can detect and quantify approximately 50% more proteins. In particular, proteomic studies using SWATH-MS strategies with CVF samples[5] face limitations such as low sensitivity for detecting low-abundance proteins, issues with counting peptides and cofragmentation[26], generating a greater proportion of missing values.

Enriched Biological Pathways and Biomarker Detection

Enrichment analysis allowed for the exploration of the functional implications of identified proteins associated with the pathophysiology of CVF **(Figure 3)**. Both methods consistently highlighted pathways related to cellular signalling, protein metabolism, and the immune response, such as signalling by interleukins[67], mitotic metaphase and anaphase, mRNA splicing, and RHO GTPase, which are crucial in the pathophysiological of the female reproductive system[68]. Nevertheless, dia-PASEF shows superior enrichment scores, indicating its potential in reproductive health studies. For example, proteins such as neutrophil gelatinase-associated lipocalin (NGAL, encoded by the *LCN2* gene) are associated with neutrophil degranulation signalling pathways[9] **(Figure 3, Figure 4H)**.

Tissue enrichment analysis confirmed the biological relevance of the identified proteins, highlighting their associations with the female reproductive system[69], uterus, and uterine cervix. This demonstrates the ability of dia-PASEF to achieve system-specific sensitivity and biological dynamic profiling over time, particularly for reproductive and immune functions in longitudinal CVF proteomics.

Ability to detect biomarkers

We compared the proteins identified by the dia-PASEF and DDA-FracOffline with the biomarkers proposed by FDA and identified by Whelan and collaborators (2023). Using the dia-PASEF method, we identified 15 unique quantifiable proteins, compared to only one quantifiable protein with the DDA-FracOffline method (Figure 4A). The dia-PASEF method significantly outperformed DDA-FracOffline in capturing biomarkers across a broader dynamic range, capturing high- and low-abundance proteins, demonstrating high efficacy in identifying these types of markers (Figure 4B). These findings underscore the superior ability of the dia-PASEF method to identify FDA-proposed biomarkers.

Proteins with low abundance were linked to essential processes, such as menstrual cycle regulation and gynaecological diseases. For example, *C1QC* and VWF play vital roles in the immune response and vascular health. Notably, a broader and more uniform dynamic range was achieved with dia-PASEF, with high sensitivity due to an increase in the signal-to-noise ratio[70], a crucial feature for longitudinal and diagnostic studies in complex biofluids such as CVF.

Ability to detect Biomarkers of interest for reproductive health

As a novel approach to identifying another set of biomarkers of particular interest for reproductive health, the dia-PASEF method enabled the detection of proteins with very low abundance. We observed that the dia-PASEF approach has greater depth in terms of identification and quantification of biomarkers implied in pathologies, for this reason, we selected this method for further analysis. Regarding physiological reproductive health processes, we identified several biomarkers of interest, among which we highlight: IGFBP-4 **(Figure 4C),** whose levels are elevated in early pregnancy and may play a role in foetal growth restriction[39,41]; and MUC-1 **(Figure 4D)**, a key mucin expressed throughout the female reproductive tract, particularly in the endocervical, ectocervical, and vaginal epithelia. Its

widespread expression suggests an essential role in maintaining the integrity and function of the reproductive tract[32,33].

Meanwhile, we also detected CVF proteins associated with pathological processes, such as alpha-actinin-4 (ACTN4) and annexin A2 (ANX2), which have been included in various studies[42–44] **(Table 2)**. Focusing on gynaecological carcinomas, we identified two specific proteins: cancer antigen 125 (CA-125), which is encoded by the *mucin-16* gene **(Figure 4E)**, and human epididymis secretory protein E4 (HE4), which is encoded by the *WFDC-2* gene **(Figure 4F)**. These glycoproteins are widely evaluated, as their high expression levels are indicative of tumours such as ovarian[49,50] and endometrial[53,54] carcinoma.
Using the dia-PASEF method, we identified these proteins in CVF samples collected from *d0* to *d6* of the menstrual cycle. This approach enabled us to perform a longitudinal analysis, which revealed significant changes between phases **(Figure S4)**. These findings demonstrate the efficiency of dia-PASEF in addressing diagnostic and prognostic problems related to female pathologies.

Longitudinal Changes and Group-Specific Variations

Proteins present in CVF may be also promising candidates for identifying longitudinal pathophysiological markers. The ability of dia-PASEF to obtain a larger amount of data is due to its ability to reduce the proportion of missing values and provide consistent protein temporal profiling **(Figure 2B)**. This makes an ideal tool for tracking changes over time, for example, by studying physiological processes such as endometrial receptivity or potential pathologies. Analyses revealed consistent protein changes associated with specific days of the cycle, providing insights into windows **(Figure 4G, Figure S4)**. These findings demonstrate the potential of monitoring disease markers longitudinally, providing a foundation for future applications in therapeutic follow-up.

Strengths, Limitations, and Future Perspectives

This study is focused primarily on achieving methodological robustness to enable a reliable comparison between strategies and to highlight the advantages of the dia-PASEF approach. Despite the limited sample size, our parsimonious approach[71] minimizes variability by selecting a homogeneous subgroup of volunteers based on exclusion criteria that control for factors such as body mass index, age, menstrual cycle length, and reproductive history.

Moreover, participants were synchronized according to the luteinizing hormone surge from d0 to d6 (peri-ovulatory to early luteal phase of the menstrual cycle), minimizing interindividual hormonal variability. This strategy allowed us to select volunteers with complete data for decision-making, implementation, and validation of proteomic approaches, minimizing technical and biological noise, to identify a panel of biomarkers. Additionally, longitudinal studies offer the advantage that each individual acts as their own control via a semi-contrafactual approach, reducing the influence of population heterogeneity[29]. In addition, the findings are not directly generalizable to pathological states because the study was limited to nonpregnant healthy volunteers. However, the cohort should be expanded to include patients with pathologies and validate the potential of dia-PASEF in identifying disease-specific biomarkers. In addition, functional validation of candidate biomarkers will be essential to understand how these proteins are involved in specific physiological or disease processes and to determine whether they could be useful in clinical practice.

Conclusions

In summary, this study confirms the potential of dia-PASEF to address the limitations of DDA for longitudinal studies, enabling the quantification of dynamic changes over time in a panel of biomarkers associated with key biological pathways of the pathophysiology of the female reproductive system. Thus, the dia-PASEF method emerges as a robust, sensitive, and efficient tool for longitudinal proteomic studies of CVF in women's reproductive health research and diagnosis.

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 of Molecular and Translational Immunology Laboratory, Department of Clinical Biochemistry and Immunology, Pharmacy Faculty, Universidad de Concepción, Concepción, Chile, for her support in figure design.

Funding

This work was supported by FISAR, Human Reproduction Grant “Embryo-Mother Crosstalk” REH042024-01, and Grant MEL205062018 from FISAR Foundation. https://www.fisarchile.org/.

Competing interests:

As a senior scientist, ESK has served as an honorary research advisor and/or reviewer on research applications for the FISAR Foundation since 2015. No other conflicts of interest are reported.

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 drafted the manuscript, 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.

Methods

Ethics

The protocol was approved by the Ethics Committee of the Servicio de Salud de Concepción in Chile (CEC: 17-03-06). All patients provided written informed consent before entering the study, and all methods were performed in accordance with the Helsinki Declaration and Good Clinical Practice.

CVF longitudinal samples

The CVF samples for this proteomics mass spectrometry 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[80]. The day of ovulation was determined by quantifying and calculating the luteinizing hormone (LH) surge in the 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 the method proposed by [72], with a confidence of 95.7% for detection of ovulation within ± 1 day. LH concentrations were determined through radioimmunoassay (IRMA) (Barnafi-Krause Laboratory, Santiago, Chile). This method allows the LH surge to be determined as day zero (*d0*) to synchronize the starting point of each cycle.

The volunteers who participated in this cohort were grouped into three conditions: conception, non-conception, and abstinence cycles. Due to the methodological nature of this study, we decided to prioritize homogeneity over sample size. We used a subgroup of CVF samples from six women (n=6) belonging to non-conception cycles selected according to criteria that controlled for factors such as body mass index, age, menstrual cycle length, and reproductive history. First, to enable a comparative longitudinal analysis, we defined a luteal phase window with seven consecutive days (*d0* to *d6*) of sample collection. Thus, we selected women in the EARLY-PREG cohort with a minimum of seven consecutive days of CVF samples since ovulation available in the biorepository. Second, because natural intra- and interindividual variability may lead to substantial heteroscedasticity and potential bias in the quantitative proteomics analysis of repeated measurements over time, we focused on achieving greater homogeneity within the selected individuals by considering factors such as menstrual cycle length, menstruation length, age, sampling completeness across days, and other relevant

variables within comparable ranges. Detailed information about the selected individuals is provided in Table 1.

Protein extraction and processing for MS analysis

CVF aliquots (100–500 µL) were thawed on ice and combined with equal volumes of 100 mM NH_4HCO_3 and 100 mM DTT. The mixture was homogenized by vortexing, incubated at 95 °C for 15 minutes, and cooled on ice for 5 minutes. 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 minutes 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 Vivaspin tubes (10,000 MWCO filter, Sartorius), washed with 400 µL of type 1 water, and centrifuged at 12,000 × g for 10 minutes 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 8M urea with 25 mM NH_4HCO_3, sonicated on ice, and centrifuged at 19,000 × g for 15 minutes at 4 °C. The protein content was quantified via 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 NH_4HCO_3 at 25 °C for 1 hour. Alkylation was performed with 20 mM iodoacetamide in 25 mM NH_4HCO_3 in the dark for 1 hour. The samples were diluted eightfold with 25 mM NH_4HCO_3 and digested with sequencing-grade trypsin (#V5071, Promega) at a 1:50 protease/protein ratio for 16 hours at 37 °C. Digestion was stopped with 10% formic acid. Peptides were cleaned using Sep-Pak C_{18} Spin columns (Waters), dried in a rotary concentrator at 40 °C, and stored for subsequent MS analysis.

DDA-FracOffline

Peptide Fractionation

High-pH reversed-phase fractionation was performed using an ÄKTA avant 25 system (Cytiva) with refrigerated fraction collection. Purified peptides (200 µg) from soluble and insoluble phases were separated on a BHE reversed-phase column (2.1 cm × 5 cm, Waters) at pH 10 and a flow rate of 0.2 mL/min. The binary gradient increased linearly from 3% to 40% buffer B (90% ACN in 5 mM NH_4HCO_2) over 30 minutes, then to 60% over 15 minutes, and finally to 85% within 5 minutes. A total of 24 fractions (400 µL each) were collected, vacuum-

dried, reconstituted in water with 2% ACN and 0.1% FA, and concatenated into 8 final fractions.

Mass Spectrometry

LC–MS was performed on a NanoElute (Bruker Daltonics) system coupled online with a hybrid TIMS-quadrupole TOF mass spectrometer, timsTOF Pro (Bruker Daltonics, Germany), via a nanoelectrospray ion source (CaptiveSpray, Bruker Daltonics)[73]. For long gradient runs (2 hours total run), approximately 200 ng of peptides were separated on a reversed-phase Aurora series 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. To analyse samples from whole-proteome digests, we used a gradient starting with a linear increase from 2% B to 17% B over 60 minutes, followed by further linear increases to 25% B in 30 minutes and to 37% B in 10 minutes and finally to 95% B in 10 min, which was held constant for 10 min. The column was equilibrated using 4 volumes of solvent A. The mass spectrometer was operated in data-dependent PASEF [70] mode with 1 survey TIMS-MS and 10 PASEF MS/MS scans per acquisition cycle. We analysed the ion mobility range from 1/K0 = 1.6 to 0.6 Vs cm-2 using equal ion accumulation and ramp time in a dual-TIMS analyser with runs of 100 ms each. Suitable precursor ions for MS/MS analysis were isolated in a window of 2 Th for m/z < 700 and 3 Th for m/z > 700 by rapidly switching the quadrupole position in sync with the elution of precursors from the TIMS device. The collision energy was lowered stepwise as a function of increasing ion mobility, starting from 20 eV for 1/K0 = 0.6 Vs cm-2 and 59 eV for 1/K0 = 1.6 Vs cm-2. We used m/z and ion mobility information to exclude singly charged precursor ions with a polygon filter mask and further used ‘dynamic exclusion’ to avoid resequencing of precursors that reached a ‘target value’ of 20,000 a.u. The ion mobility dimension was calibrated linearly using three ions from the Agilent ESI LC/MS tuning mix (m/z, 1/K0: 622.0289, 0.9848 Vs cm-2; 922.0097, 1.1895 Vs cm-2; and 1221.9906, 1.3820 Vs cm-2).

Data searching

For protein identification, we used FragPipe v19.1 software with the search engines MSFragger (version 3.7)[74], Philosopher (version 3.4.13)[75], and EasyPQP (https://github.com/grosenberger/easypqp; version 0.1.9). Peptide identification was performed using the MSFragger search engine, with raw (.d) files used as input. For protein searches, we used the reference proteome of *H. sapiens* (UP000005640) from

UniProt/SwissProt (reviewed sequences only; downloaded on Feb. 15, 2021) and the database of common contaminant proteins in mass spectrometry experiments available in FragPipe. Reversed protein sequences were appended to the original databases as decoys. For MSFragger analysis, both the precursor and (initial) fragment mass tolerances were set to 20 ppm. The enzyme specificity was set to 'stricttrypsin', and fully enzymatic peptides were allowed. Up to two missed cleavages were allowed. Carbamidomethylation of cysteine was set as a fixed modification. Oxidation of methionine and acetylation of the N-terminus were set as variable modifications. The final spectral library was filtered to 1% protein- and peptide-level FDRs.

dia-PASEF

Mass spectrometry

The timsTOF was operated in dia-PASEF mode with the following settings: mass range, 100 to 1700 m/z; 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, 200 °C.

The dia-PASEF settings were as follows: Each cycle consisted of a 1x MS1 full scan and 64x 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 seconds. The CID collision energy was 20 to 52 eV as a function of the inverse mobility of the precursor.

Data searching

For the construction of the DDA spectral library, a pool of all samples of tryptic-digested peptides was used. For peptide fractionation, mass spectrometry and data searching, the method described in the DDA section was used (see methods sections 3.1, 3.2 and 3.3).

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.

The processing of dia-PASEF data was performed with DIA-NN v1.7.15 software[76], using the spectral library previously generated. The parameters of maximum mass accuracy tolerances were set to a default average of 10 ppm for both the MS1 and MS2 spectra. The samples were analysed with match-between-runs (MBR) enabled; quantification mode was set to "any LC (high accuracy)". All other settings were left default. The DIA-NN output was filtered at a precursor q value < 1% and a global protein q value < 1%.

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 normalized by adjusting the medians. Missing values were imputed for each condition using the missForest algorithm, applying the criterion that the intensity value should be present in at least 50% of the samples per analysed category[77]. 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 Bayesian t test for multiple comparisons between days (e.g., *d0* vs. *dn*, *d1* vs. *dn*, *d2* vs. *dn*…, etc.), considering the normalized intensity means of biological replicates[78]. Any protein associated with a p value < 0.05 was considered significant. Exploratory data analysis and visualization were created using R (v.3.6.0) with EnhancedVolcano (v.1.2.0), ComplexHeatmap (v.2.0.0), and 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.

Pathway and tissue enrichment analysis

Pathway enrichment analysis was conducted using ReactomePA (v.1.42.0), which queries the human database with logarithm base two of the fold change (logFC) and p value data for each statistical comparison. For the analysis of enrichment by human systems, tissues and cells, we used the TISSUES 2.0 database[79], 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 does not refer to the standard score typically used in statistics but rather a simple calculation intended to indicate whether the biological process, molecular function, or cellular component is more likely to decrease (resulting in a negative value) or increase (resulting in a

positive value). The z score was determined based on the number of proteins assigned as upregulated (log FC > 0) or downregulated (log FC < 0), denoted as "up" and "down,"respectively, in the dataset.

Biomarker selection

To assess the efficiency of dia-PASEF in biomarker detection, we used the (Food and Drug Administration) FDA-biomarker set list obtained from the study by Whelan and collaborators[30], and we compared the efficiency of identifying FDA-approved biomarkers between the dia-PASEF and DDA-FracOffline methods. In addition, we performed extensive bibliography research involving physiological and pathological proteins using keywords such as “menstrual cycle”, “ovarian carcinoma”, “endometrial carcinoma”, “cervical carcinoma” and “endometriosis”, considering a variety of studies with proteomic, transcriptomic and genomic focuses.

Availability of data

Mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier XXXXX. Created in BioRender. Nova Lamperti, E. (2025) https://BioRender.com/7ir5wi5 .

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. Lacroix, G., Gouyer, V., Gottrand, F. & Desseyn, J.-L. The cervicovaginal mucus barrier. *Int. J. Mol. Sci.* 21, 8266 (2020).

7. 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).

8. Fernandez-Hermida, Y., Grande, G., Menarguez, M., Astorri, A. L. & Azagra, R. Proteomic markers in cervical mucus. *Protein Pept. Lett.* 25, 463–471 (2018).

9. Grande, G. *et al.* Cervical mucus proteome in endometriosis. *Clin. Proteomics* 14, 7 (2017).

10. Zhao, Z. *et al.* Noninvasive cervicovaginal proteomic biomarkers for spontaneous preterm birth. (2024).

11. 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).

12. 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).

13. 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).

14. 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).

15. Hedlund Lindberg, J. *et al.* Toward ovarian cancer screening with protein biomarkers using dried, self-sampled cervico-vaginal fluid. *iScience* 27, 109001 (2024).
16. 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).
17. 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).
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. 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).
31. Shaw, J. L. V., Petraki, C., Watson, C., Bocking, A. & Diamandis, E. P. Role of tissue kallikrein-related peptidases in cervical mucus remodeling and host defense. *Biol. Chem.* 389, 1513–1522 (2008).
32. Gipson, I. K. *et al.* Mucin genes expressed by human female reproductive tract epithelia. *Biol. Reprod.* 56, 999–1011 (1997).
33. 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).
34. 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).

35. 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, 488.e1–488.e11 (2019).
36. 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).
37. 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).
38. 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. Proteomics* 18, 9 (2021).
39. Qiu, Q. *et al*. Significance of IGFBP-4 in the development of fetal growth restriction. *J. Clin. Endocrinol. Metab.* 97, E1429–39 (2012).
40. 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).
41. Hedderson, M. M. *et al*. Prepregnancy SHBG concentrations and risk for subsequently developing gestational diabetes mellitus. *Diabetes Care* 37, 1296–1303 (2014).
42. 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).
43. Starodubtseva, N. L. *et al*. Label-free cervicovaginal fluid proteome profiling reflects the cervix neoplastic transformation. *J. Mass Spectrom.* 54, 693–703 (2019).

44. Mukherjee, A. *et al*. Insights on proteomics-driven body fluid-based biomarkers of cervical cancer. *Proteomes* 10, 13 (2022).

45. 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).

46. 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).

47. Qian, L. *et al*. Proteomic landscape of epithelial ovarian cancer. *Nat. Commun.* 15, 6462 (2024).

48. Whitwell, H. J. *et al*. Improved early detection of ovarian cancer using longitudinal multimarker models. *Br. J. Cancer* 122, 847–856 (2020).

49. Radu, M. R. *et al*. Ovarian cancer: Biomarkers and targeted therapy. *Biomedicines* 9, 693 (2021).

50. 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).

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

52. 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).

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

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

55. Degez, M. *et al*. Endometrial cancer: A systematic review of HE4, REM and REM-B.

Clin. Chim. Acta 515, 27–36 (2021).

56. 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).

57. 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).

58. 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).

59. 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).

60. Yang, H., Kang, K., Cheng, C., Mamillapalli, R. & Taylor, H. S. Integrative analysis reveals regulatory programs in endometriosis. *Reprod. Sci.* 22, 1060–1072 (2015).

61. 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).

62. Méar, L. *et al.* The eutopic endometrium proteome in endometriosis reveals candidate markers and molecular mechanisms of physiopathology. *Diagnostics (Basel)* 12, 419 (2022).

63. Janša, V. *et al.* Proteomic analysis of peritoneal fluid identified COMP and TGFBI as new candidate biomarkers for endometriosis. *Sci. Rep.* 11, 20870 (2021).

64. 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).

65. Irungu, S. *et al*. Discovery of non-invasive biomarkers for the diagnosis of endometriosis. *Clin. Proteomics* 16, 14 (2019).

66. Ludwig, C. *et al*. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* 14, e8126 (2018).

67. Gérard, N., Caillaud, M., Martoriati, A., Goudet, G. & Lalmanach, A.-C. The interleukin-1 system and female reproduction. *J. Endocrinol.* 180, 203–212 (2004).

68. 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).

69. 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).

70. 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).

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

72. 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).

73. 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).

74. 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).

75. da Veiga Leprevost, F. *et al*. Philosopher: a versatile toolkit for shotgun proteomics data

analysis. *Nature Methods* 17, 869–870 (2020).

76. 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. *Nature Methods* 17, 41–44 (2019).

77. 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).

78. Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D. & Iverson, G. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review* 16, 225–237 (2009).

79. 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).

80. Avila, D., Latapiat, V., Nourdin, G., Antilef, B., Vargas, C., Koch, E. EARLY-PREG: Preconception, longitudinal, bidirectional, and counterfactual cohort of women trying to conceive for the characterization of maternal–embryonic molecular crosstalk during the first two weeks after conception. *Medrxiv* (2025).