

RESEARCH ARTICLE OPEN ACCESS

Proteomic Analysis of Cervicovaginal Fluid for Diagnostic Endometriosis Biomarker Discovery

Emily S. J. Paterson¹  | Simon Scheck^{2,3} | Torsten Kleffmann⁴ | Roopi Gill^{2,3} | Nicholas Bedford³ | Simon McDowell³ | Jane E. Girling⁵ | Claire E. Henry¹

¹Department of Surgery and Anaesthesia, University of Otago, Ōtākou Whakaihu Waka, Wellington, Aotearoa, New Zealand | ²Department of Obstetrics, Gynaecology and Women's Health, University of Otago, Ōtākou Whakaihu Waka, Wellington, Aotearoa, New Zealand | ³Department of Obstetrics and Gynaecology, Wellington Hospital, Te Whatu Ora - Capital, Coast and Hutt Valley, Wellington, Aotearoa, New Zealand | ⁴Department of Biochemistry, University of Otago, Ōtākou Whakaihu Waka, Dunedin, Aotearoa, New Zealand | ⁵Department of Anatomy, University of Otago, Ōtākou Whakaihu Waka, Dunedin, Aotearoa, New Zealand

Correspondence: Claire E. Henry (claire.henry@otago.ac.nz)

Received: 27 July 2025 | **Revised:** 29 January 2026 | **Accepted:** 14 March 2026

Keywords: biomarkers | cervicovaginal fluid | endometriosis | vaginal swab

ABSTRACT

Purpose: Endometriosis is a prevalent inflammatory condition characterised by the presence of endometrial-like tissue outside the uterus and is associated with significant challenges, including diagnostic delays and continued reliance on laparoscopy. Although extensive research has investigated potential biomarkers in various biofluids, none have been validated for clinical use.

Experimental design: This pilot study explored the proteome of cervicovaginal fluid to identify biomarkers of endometriosis. SWATH-MS was performed over two experiments on cervicovaginal fluid sampled via a low vaginal swab, from people with ($n = 20$) and without ($n = 19$) surgically confirmed endometriosis. STRING, OPLS-DA modelling and ingenuity pathway analysis were used to interrogate the data. ELISA was performed to validate SWATH-MS findings.

Results: There were 29 proteins in experiment one and 47 proteins in experiment two identified as differentially abundant between cases and controls. No proteins were identified as differentially abundant in both experiments. Legumain (LGMN) measured via ELISA was significantly increased in the cervicovaginal fluid of people with endometriosis.

Conclusions and clinical relevance: Cervicovaginal fluid proteins sampled via vaginal swab may have limited biomarker potential for endometriosis. A larger and more diverse cohort would be required to confirm the promise of cervicovaginal fluid LGMN as a candidate biomarker of endometriosis.

1 | Introduction

Endometriosis is an inflammatory condition defined by the presence of endometrial-like tissue outside the uterus, commonly

cited to affect one in ten people assigned female at birth [1, 2]. Endometriosis is renowned for an extensive diagnostic delay ranging internationally from 3.3 years up to over 10 years [3]. Transvaginal ultrasound and magnetic resonance imaging (MRI)

Abbreviations: ACN, acetonitrile; AUC, area under the receiver operator characteristic curve; BCA, bicinchoninic acid assay; COP, combined oral contraceptives; DDA, data-dependent acquisition; DIA, data-independent acquisition; ELISA, enzyme-linked immunosorbent assay; FA, FDR, false discovery rate; formic acid; GLM, generalised linear model; GnRH, gonadotropin-releasing hormone; GO, gene ontology; IPA, ingenuity pathway analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; LGMN, legumain; MCL, Markov clustering; MELAA, Middle Eastern, Latin American, African; NAMPT, nicotinamide phosphoribosyltransferase; OLFM4, olfactomedin 4; OPLS-DA, orthogonal-partial least squares discriminant analysis; rASRM, revised American Society for Reproductive Medicine; QC, quality control; ROC, receiver operator characteristic; SPARC1, SPARC-like protein 1; SWATH-MS, sequential window acquisition of all theoretical mass spectra; TEAB, tri-ethyl ammonium bicarbonate; VIP, variables influence on projection.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2026 The Author(s). *Proteomics – Clinical Applications* published by Wiley-VCH GmbH

can detect endometriomas and deep infiltrating endometriosis (DIE) with high sensitivity and specificity when performed by experienced operators [4]. However, accuracy is highly operator dependent, creating inconsistencies in diagnostic reliability. Subsequently, despite laparoscopy no longer being required for an endometriosis diagnosis, surgery remains the most common method of diagnosis endometriosis globally [5]. Negative laparoscopy, where no evidence of endometriosis or alternative pelvic pathology is found, is reported in up to 42% of cases with a preoperative diagnosis of suspected endometriosis [6].

An effective biomarker could significantly transform the endometriosis diagnostic landscape, reducing delays in accessing specialist care and eliminating the reliance on expensive and invasive procedures to confirm endometriosis. Furthermore, a diagnostic biomarker may significantly reduce the number of unnecessary surgeries, minimising patient harm and optimising resource allocation. Extensive research has been conducted investigating endometriosis biomarkers [7–10] and remains a priority for both clinicians and patients [11, 12], yet, no single or combination of biomarkers has been validated for clinical use; although several are commercially available [13].

Most biomarker research to date has investigated either accessible biofluids like peripheral blood, or more biologically relevant sources such as endometrial tissue or peritoneal fluid. A recent systematic review by Brulport et al. demonstrated inconsistent performance of candidate markers across studies [10]. Other accessible biofluids in proximity to the uterus such as cervical mucus and cervicovaginal fluid remain relatively understudied [10]. Cervicovaginal fluid, comprised of cervical, vaginal and uterine secretions [14], offers notable advantages as a non-invasive and easily accessed biofluid for biomarker research. The proximity to the reproductive tract suggests cervicovaginal fluid may provide valuable insights into localised inflammatory and molecular processes associated with endometriosis. Previously, using cervical mucus aspirated via catheter, Grande et al. [15] identified 15 dysregulated proteins between people with endometriosis and fertile women, demonstrating the potential of cervicovaginal fluid biomarkers [15]. The catheter-based approach may be more suitable for identifying proteins specific to cervical mucus; however, the sampling procedure can cause discomfort, reducing accessibility [16]. A clinically standard sampling procedure, such as vaginal swabs which can be self-sampled, would be ideal to ensure the accessibility and clinical translation of candidate biomarkers.

In this study, we aimed to characterise protein differences in the cervicovaginal fluid sampled by vaginal swab between people with and without endometriosis in a pilot case-control study.

2 | Methods

2.1 | Ethical Approval

This study was approved by the Central Health and Disability Ethics Committee (HDEC; 2022 EXP 12616), received endorsement from the Research Advisory Group—Māori (RAG-M) at

Te Whatu Ora—Capital and Coast (#937) and local approval from the Te Whatu Ora—Capital and Coast research and audit committee.

2.2 | Participant Recruitment and Cervicovaginal Fluid Collection

Participants were recruited between June 2022 and December 2024. Eligible participants were having surgery for suspected endometriosis at Te Whatu Ora—Capital and Coast (Wellington and Kenepuru Hospitals) and aged between 16 and 45 years. People who had current known cervical abnormalities or were menstruating were excluded.

Low vaginal swabs were taken by the operating surgeon after the administration of anaesthesia, prior to surgical preparation. Vaginal swabs were collected first by inserting a FLOQ swab (Copan Diagnostics, #5E089N) into the vagina and turning for approximately 30 s, then placed into a tube with 1 mL of phosphate buffered saline (PBS; Thermo Fisher, #70011044).

Samples were visually assessed, and samples with any blood contamination were excluded. This included any samples with blood visible on the collection device, or if the sample fluid was orange or red. The FLOQ swab was removed, and the sample was briefly vortexed and aliquoted in 200 µL volumes before storing at -80°C . Batched cervicovaginal fluid samples were transported on ice to the Centre for Protein Research at the University of Otago, Dunedin, New Zealand.

To represent the real-world heterogeneity in which a diagnostic biomarker would be applied clinically, there were no exclusion criteria relating to menstrual phase. Menstrual cycle phase was estimated based on participant reported data collected via survey on the day of surgery. Participants were assigned to one of the following groups: combined oral contraceptives (COP), progestin medication, gonadotropin-releasing hormone (GnRH) analogue, proliferative phase, secretory phase, irregular periods, or unknown. Irregular periods were defined as being ± 20 days from the expected date. Participants expecting their period in less than 14 days were grouped into the secretory phase. Participants expecting their period in 14 days or longer were grouped into the proliferative phase. Those who did not report the date of their last menstrual period, or reported they had no periods in the last three months and were not taking hormonal medications, were grouped into the unknown group. If endometrial curettings were taken during the surgery, the phase listed in the histological report was used instead of the estimated phase.

For participants with endometriosis at laparoscopy, disease extent was assessed by the revised American Society for Reproductive Medicine (rASRM) scoring system [17]. Two (2/20) endometriosis samples were not staged at surgery and were retrospectively staged by SS. Participants were classified as not having endometriosis and allocated to the control group when all biopsies were histologically negative for endometriosis, or when there was no visible evidence of endometriosis and no biopsy.

2.3 | Swath-MS

A total of 39 samples, 20 from people with endometriosis and 19 from people without endometriosis were prepared for SWATH-MS analysis. To enable the assessment of the reproducibility of dysregulated protein signatures, sample preparation and SWATH-MS was performed over two separate experimental runs (experiment one: endometriosis $n = 11$, no endometriosis $n = 9$, experiment two: endometriosis $n = 9$, no endometriosis $n = 10$).

Proteins were extracted and processed using the S-Trap micro spin columns (Protifi, #C02-micro) according to manufacturer's instructions. Each 200 μL aliquot of vaginal swab cervicovaginal fluid was lysed with 200 μL of lysis buffer (100 mM triethyl ammonium bicarbonate [TEAB] and 10% sodium dodecyl sulphate [SDS] in water) and sonication. Magnesium chloride (MgCl_2) was added to yield a final concentration of 2 mM, followed by the addition of 200 units of benzonase. Samples were then incubated at 37°C for 10 min, followed by centrifugation for 15 min at 30,000 $\times g$. The supernatant was retained.

To estimate the total protein in each sample, a BCA assay was performed according to manufacturer's instructions (Thermo Fisher Scientific, #23225).

Samples were reduced and alkylated by adding tris(2-carboxyethyl) phosphine (TCEP) and iodoacetamide to a final concentration of 5 and 10 mM, respectively and incubating for 10 min at room temperature in the dark. A sample volume containing 100 μg of protein was then processed using the S-trap micro protocol for tryptic digestion. The digested protein samples were further purified by solid phase extraction on centrifugal peptide desalting columns (Thermo Scientific) following the manufacturer's instructions and dried in a centrifugal vacuum concentrator. For mass spectrometry analysis samples were reconstituted in 5% acetonitrile (ACN), 0.05% formic acid (FA) in water at a concentration corresponding to 1.5 μg of digested protein per μL sample. For each experiment an aliquot from each sample was pooled into a reference sample to (i) generate quality control (QC) samples to monitor instrument performance over the experiments and (ii) perform a peptide pre-fractionation of a representative pooled sample to generate a high-quality spectral library by data-dependent acquisition (DDA) mass spectrometry. The pooled peptide samples from each experiment were pre-fractionated into 11 fractions using the high pH reversed-phase peptide fractionation kit (Thermo Scientific) according to the manufacturer's instructions.

2.4 | Protein Identification and Quantification

Samples were analysed on a 5600 + triple time-of-flight (TOF) mass spectrometer coupled to an Eksper NanoLC 415 system (Eksigent, AB Sciex). A volume of 5 μL of peptide sample was separated on a 75 μm ID and 20 cm long silica emitter tip column that was in-house packed with Luna (Phenomenex) C18 bead material (3.2 μm , 100Å) through an ACN gradient from 5% ACN, 0.05% FA in water to 90% ACN, 0.05% FA in water over a total gradient length of 115 min.

For protein identification and generating a spectral library, each of the high pH reversed phase fractions was analysed by data-dependent acquisition with one MS survey scan in the mass range of 400–1300 m/z followed by 25 data-dependent high energy collision-induced MS/MS spectra per cycle. Each fraction was analysed in two technical replicates. For protein quantification individual samples were analysed in technical triplicates by SWATH-MS with the mass spectrometer operated in data-independent mode. Each DIA cycle consisted of one MS survey scan followed by HCD MS/MS scans of 34 mass windows over the mass range of 400–1250 m/z .

2.5 | Data Analysis

Protein identification was performed by searching the DDA raw data against the complete UniProt Human amino acid sequence database (204294 sequence entries; downloaded 20-11-2023) using ProteinPilot software version 4.5 (AB Sciex). The spectral library was built using the SWATH application (version 2.0) embedded into PeakView software (version 2.2, AB Sciex). The SWATH-MS raw data of individual samples was aligned to the spectral library at a false discovery rate of 1% and confidence $\geq 99\%$ for peak identification in at least one sample. Aligned peak areas of at least 6 fragment ions per peptide signal were exported to MarkerView software (version 1.2, AB Sciex) for further data processing and basic statistics such as unsupervised multivariate analysis using principal component analysis (PCA). For peak area normalisation, high abundance plasma and structural proteins (including keratins and haemoglobins) were removed to improve normalisation and quantification of lower-abundance proteins of potential biological relevance. The proteins removed can be found in (Table S1). The resulting data was then scaled using the median peak intensity of all the samples in the experiment.

The median peak value of each sample triplicate was used for analysis. Intra-assay coefficients of variance (CV%) were calculated for QC samples for each protein across technical replicates. For each protein, the median values for each study group were averaged, followed by calculating the fold change between the endometriosis and no endometriosis groups. Two-tailed unpaired t-tests were performed on peak values. Fold change and p values were log transformed. As this was an exploratory study, a low-stringency threshold was applied without corrections for multiple testing. Proteins were considered differentially regulated when the median peak values in the endometriosis group were 1.5-fold different from the no endometriosis group ($\log_2(\text{fold change}) \geq 0.58$ and ≤ -0.58), with $p < 0.05$ ($-\log_{10} p \leq 1.3$). The list of differentially regulated proteins from the two experiments were analysed separately for all subsequent analyses.

2.6 | Bioinformatic Analysis

2.6.1 | STRING Analysis

The STRING database tool (Version 12.0; [18]) was used to provide insight into the functional associations of the differentially regulated proteins. The minimum required interaction score was set at 0.4 (medium confidence). Markov clustering (MCL) with

TABLE 1 | Characteristics of participant cohort in SWATH-MS analysis.

	Endometriosis (<i>n</i> = 20)	No endometriosis (<i>n</i> = 19)	<i>p</i> value
Age, mean ± standard deviation	32.4 ± 6.8	30.7 ± 6.4	0.42*
Ethnicity, <i>n</i> (%)			
Māori	3 (15%)	2 (11%)	0.56†
Pacific	0 (0%)	0 (0%)	
NZ European	10 (50%)	13 (68%)	
Asian	6 (30%)	2 (11%)	
Other European	1 (5%)	2 (11%)	
MELAA	0 (0%)	0 (0%)	
Symptoms, <i>n</i> (%)			
Dysmenorrhoea	18 (90%)	17 (89%)	0.87‡
Pelvic pain unrelated to periods	13 (65%)	16 (84%)	
Dyspareunia	14 (70%)	12 (63%)	
Dyschesia	12 (60%)	14 (74%)	
Other bowel problems (bloating, diarrhoea, constipation)	15 (75%)	17 (89%)	
Difficulty falling pregnant	9 (45%)	4 (25%)	
Abnormal menstrual bleeding	13 (65%)	15 (79%)	
Other period related symptoms	8 (40%)	12 (63%)	
Approximate menstrual cycle phase, <i>n</i> (%)			
Proliferative	8 (40%)	1 (5%)	0.23†
Secretory	3 (15%)	5 (26%)	
Progestin medication	4 (15%)	5 (26%)	
Combined oral contraceptive	1 (5%)	1 (5%)	
GnRH analogue	0 (0%)	1 (5%)	
Irregular	3 (15%)	4 (21%)	
Unknown	1 (5%)	2 (11%)	
Gravidity, median (Q1-Q3)	0 (0-[1])	0 (0-[3])	0.70†
Parity, median (Q1-Q3)	0 (0-[1])	0 (0-[1])	>0.99†
Taking hormonal medications, <i>n</i> (%)	5 (25%)	7 (37%)	0.50†
Menstruated in the last three months, <i>n</i> (%)	17 (85%)	11 (58%)	0.16†
Previous laparoscopy for endometriosis, <i>n</i> (%)	3 (15%)	5 (26%)	0.45†
Other pathologies identified, <i>n</i> (%)	5 (10%)	9 (47%)	0.19†
Adhesions	3 (15%)	2 (11%)	
Uterine fibroids	2 (10%)	1 (5%)	
Functional ovarian cyst	2 (10%)	1 (5%)	
Adenomyosis	0 (0%)	1 (5%)	
Paratubal or parafimbrial cyst	0 (0%)	4 (25%)	
Adenomatoid tumour	0 (0%)	1 (5%)	
Teratoma	0 (0%)	1 (5%)	
Benign endometrial polyp	0 (0%)	1 (5%)	
rASRM stage, <i>n</i> (%)			
I	6 (30%)	N/A	
II	7 (35%)		

(Continues)

TABLE 1 | (Continued)

	Endometriosis (<i>n</i> = 20)	No endometriosis (<i>n</i> = 19)	<i>p</i> value
III	4 (20%)		
IV	3 (15%)		
Endometriosis phenotypes, <i>n</i> (%) [§]			
Superficial	17 (85%)		
Ovarian	5 (25%)		
Deep infiltrating	7 (35%)		

Abberivation: GnRH—gonadotropin-releasing hormone, MELAA—Middle Eastern, Latin American, African, rASRM—revised American Society for Reproductive Medicine.

*Unpaired t-test.

†Fisher's exact test.

‡Chi-square.

§Multiple endometriosis phenotypes per participant possible, sum does not equal *n* = 20.

a 1.5 inflation parameter was performed to identify groups of interacting proteins. Functional enrichments were significant when the false discovery rate (FDR) < 0.05 and strength > 0.01 with a minimum of two proteins in the network.

2.6.2 | OPLS-DA Modelling

Data preprocessing and visualisation of supervised multivariate analysis was performed in R (version 4.4.1) and R Studio (version 2024.04.2). Data was log-transformed and scaled. Orthogonal-partial least squares discriminant analysis (OPLS-DA) was performed using the R package metabom8 (version 1.0.0, available from github.com/tkimhofer/metabom8, accessed July 11, 2024). The variables extracted and plotted were predicted weights (variables influence on projection [VIP]), predictive component (t_pred_cv) and orthogonal component (t_orth_cv). A VIP score of 1.0 was considered significant [19].

2.6.3 | Ingenuity Pathway Analysis

Ingenuity pathway analysis (IPA, version 111725566) was performed to identify upstream molecules that may be responsible for the observed differences in protein abundances in each dataset. For differentially regulated proteins, the log fold change (log₂(fold change)) and log *p*-value (−log₁₀(*p*)) were input into the IPA core analysis tool. The core analysis from both datasets were input into the comparison analyses tool to identify results common to both datasets. A *z*-score of >1 or <−1 and *p* value of <0.05 were requirements for significance. Where applicable, *p* values and *z*-scores adjusted for multiple comparisons or bias were used.

2.7 | ELISA Validation

Enzyme-linked immunosorbent assays (ELISAs) were performed on cervicovaginal fluid samples to investigate the protein levels of human LGMN (Invitrogen, #EH299RB), NAMPT/Visfatin (Invitrogen, #EH482RB), OLFM4 (Abcam, #ab267805) and SPARCL1

(Abcam, #ab272478). LGMN, SPARCL1, NAMPT, and OLFM4 were chosen based on abundance, interaction with other proteins and predicted molecular functions, in addition to availability of commercial ELISA kits.

Five of the 28 samples included in the ELISA experiments had an ASRM stage reported by the operating surgeon with the remaining 23 samples retrospectively staged by SS. Cervicovaginal fluid sample aliquots were thawed on ice, pooled, then centrifuged at 2,000 × *g* for 10 min at 4°C. The supernatant was retained and the pellet discarded.

To estimate the total protein in each sample, a BCA assay was performed as previously described. ELISAs were then performed according to manufacturer's instructions. All incubations were performed at room temperature with shaking.

LGMN and NAMPT was detected through solid-phase sandwich ELISA. LGMN standards ranged from 28.67 to 7,000 pg/mL and NAMPT standards ranged from 1.229 to 300 ng/mL. Cervicovaginal fluid swab supernatant were diluted 1:1 in assay diluent, plated in duplicate and incubated for 2.5 h prior to biotin conjugate, streptavidin-HRP and TMB steps. OLFM4 and SPARCL1 were detected through quantitative sandwich ELISA. OLFM4 standards ranged from 75,000 to 1,172 pg/mL. SPARCL1 standards ranged from 39.06 to 2,500 pg/mL. To each well, 50 µL of cervicovaginal fluid supernatant, standard and an assay diluent blank were plated in duplicate, followed by the kit antibody cocktail and TMB solution.

All ELISA absorbances were read at 450 nm on a Multiskan microplate spectrophotometer (Thermo Scientific, #51119000). Four parameter logistic curves (4PL) were used to interpolate the unknown sample concentrations. When the absorbance exceeded the standard curve, the concentration of the first standard was assigned. When the absorbance was below the standard curve, the lower limit of detection of each assay was assigned. The interpolated concentrations for NAMPT and LGMN were then adjusted for dilution factor. The amount of LGMN, NAMPT, OLFM4, and SPARCL1 per µg of protein in the sample was calculated using the total protein obtained from the BCA assay.

The thresholds for a replacement test were set at 0.94 sensitivity and a 0.79 specificity [9]. A triage test ruling out endometriosis (SnOUT) required 0.95 sensitivity and 0.50 specificity, while a rule-in test (SpIN) required the inverse thresholds.

First, univariate analysis was performed to evaluate the predictive potential of each individual protein. The pROC package in R was used to generate receiver operator characteristic (ROC) curves for the concentrations and normalised concentrations of LGMN, NAMPT, OLFM4 and SPARCL1. The optimal cut-off threshold was identified using the Youden index, and sensitivity and specificity values were extracted from the ROC curve. Post-hoc analysis examined normalised LGMN across the menstrual cycle using a Kruskal-Wallis test.

To explore multivariate analysis with a simplified approach to the interpretation of biomarker patterns, a biomarker scoring system was developed [20]. Normalised LGMN concentrations, and the concentrations of NAMPT, OLFM4 and SPARCL1 were split into quartiles. For NAMPT and SPARCL1, scores were: 0 for the lower quartile, up to 3 for the upper quartile. As LGMN and OLFM4 showed opposite abundance directionality, scores were: 3 for the lower quartile, down to 0 for the upper quartile. Scores were summed to determine the biomarker score, with a theoretical minimum score of 0 and maximum score of 12. Biomarker scores between the endometriosis and no endometriosis group were compared using Fisher's exact test.

A multivariate logistic regression model was developed using a generalised linear model (GLM) in R studio to provide a more nuanced analysis of the predictive ability of the combined markers. Two models were produced, one using the concentrations of LGMN, NAMPT, OLFM4, and SPARCL1, and the other using the concentrations relative to total protein of the same four proteins. The AUC, sensitivity and specificity values were obtained as in the univariate analysis.

3 | Results

3.1 | Swath-Ms

SWATH-MS was performed on 39 samples over two separate experiments, due to instrument availability. Experiment one comprised 11 endometriosis samples and 9 no endometriosis samples. Experiment two comprised 9 endometriosis and 10 no endometriosis samples. Of those who did not have endometriosis, almost half (9/19) had negative histological results, and the remaining ten participants had no areas suspicious of endometriosis (no histology).

No significant differences in clinical characteristics were found between the endometriosis and no endometriosis groups, whether analysed together as a whole (Table 1) or when the two experimental runs were compared separately (Table S2).

There were 3,740 and 3,655 proteins identified of which 2,154 and 2,400 proteins were quantified in experiment one and two, respectively. A total of 1,741 proteins appeared in both lists of quantified proteins.

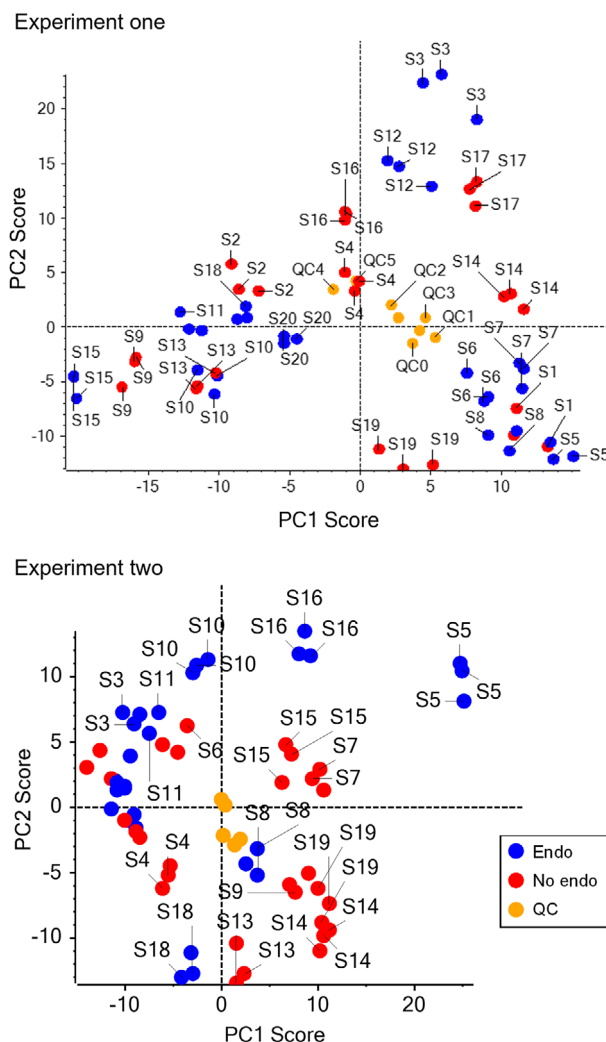


FIGURE 1 | Principal component analysis of datasets from SWATH-MS experiments one and two. The score plots show three technical replicates of each sample and the pooled quality control (QC) samples. In experiment one, principal component 1 (PC1) and PC2 explained 8.8% and 6.6% of the variance, respectively, with a cumulative variance of 15.4%. In experiment two, PC1 and PC2 explained 9.9% and 5.4%, respectively, with a cumulative variance of 15.3%. The plots show sample distribution based on these two principal components. QC—quality control, SWATH-MS—sequential window acquisition of all theoretical mass spectra.

The intra-assay CV% for the top 500 most abundant proteins was 21.9% for experiment one and 18.5% for experiment two, indicating acceptable technical reproducibility. The raw intensity data and CV% over the range of protein abundances are in Supplementary Figure 1.

There were 29 proteins in experiment one and 47 proteins in experiment two that were identified as differentially regulated with a minimum fold change of 1.5 and $p < 0.05$. In experiment one, 12 proteins were upregulated and 17 were downregulated, while in experiment two 36 proteins had increased abundance and 11 had decreased abundance. There were no differentially regulated proteins that appeared in both protein lists. When the significance threshold was increased to $p < 0.1$, OLFM4 was the

TABLE 2 | Top ten differentially regulated vaginal swab proteins identified in experiment one and two, according to *p*-value.

Protein	Protein abbreviation	Uniprot identifier	log2(FC)	-log10(p)
Experiment one				
Small ribosomal subunit protein eS17	RS17	P08708	0.66	2.71
Plasminogen	PLMN	P00747	−1.17	2.60
Gap junction beta-2 protein	CXB2	P29033	0.93	2.14
SH3 domain-binding glutamic acid-rich-like protein 2	SH3L2	Q9UJC5	0.74	2.09
Immunoglobulin heavy variable 1–2	HV102	P23083	−1.31	1.94
Copine-1	CPNE1	Q99829	0.67	1.92
Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3	HACD3	Q9P035	0.78	1.91
Isoform 6 of WASH complex subunit 2C	WAC2C	Q9Y4E1	−1.30	1.90
Isoform 2 of SPARC-like protein 1	SPRL1	Q14515	−0.77	1.73
Isoform 2 of ATP-binding cassette sub-family F member 1	ABCF1	Q8NE71	−0.63	1.73
Experiment two				
Nicotinamide phosphoribosyltransferase	NAMPT	P43490	−0.63	3.13
Proteasome subunit beta type-7	PSB7	Q99436	0.67	2.68
Voltage-gated monoatomic cation channel TMEM109	TM109	Q9BVC6	0.91	2.46
Transmembrane 9 superfamily member 3	TM9S3	Q9HD45	0.60	2.42
Ras-related protein Rab-13	RAB13	P51153	0.83	2.21
ADP-ribosylation factor 4	ARF4	P18085	0.69	2.17
Charged multivesicular body protein 4c	CHM4C	Q96CF2	1.53	2.14
Probable non-functional immunoglobulin heavy variable 3–38	HV338	A0A0C4DH36	1.00	2.04
Eukaryotic initiation factor 4A-III	IF4A3	P38919	0.82	2.03
Isoform 2 of Glucosidase 2 subunit beta	GLU2B	P14314-2	−0.67	1.95

only protein that appeared in both protein lists with a fold change >1.5 with the same directionality. The top ten proteins according to *p*-value from each experiment are presented in Table 2, with the full list of differentially regulated proteins in Table S3.

PCA was performed to evaluate the variance distribution and underlying structure of the datasets (Figure 1). In experiment one, PC1 accounted for 8.8% of the total variance and PC2 explained an additional 6.6%. Similarly, in experiment two, PC1 and PC2 explained 9.9% and 5.4% of the variance, respectively. The cumulative variance for the first two PCs was nearly identical across experiments, capturing 15.4% in experiment one and 15.3% in experiment two. All QC samples clustered in the centre of all samples indicating consistent measurements throughout each experiment.

3.2 | Bioinformatic Analysis

STRING functional protein association networks were created from the differentially regulated protein lists from each experi-

ment. Neither network had significantly more interactions than expected in either experiment (experiment one: 7 observed, 9 expected, $p = 0.77$; experiment two: 18 observed, 12 expected, $p = 0.055$).

MCL analysis explored potential functional modules within the protein networks (Figure 2a,b). In the network from experiment one, there were three clusters containing two or more proteins. These were related to the ribosome and ribonucleoprotein complex KEGG pathway and GO cellular components. There were six clusters identified in experiment two, with enriched GO and KEGG pathways for these clusters relating to the innate immune system, the complement cascade, and the Cul3-RING ubiquitin ligase complex. The remaining nodes in experiment one and two were not enriched for any GO or KEGG pathways. Table 4 contains a tabulated summary of MCL analyses.

A discriminant model was produced for experiment one, with a cross-validated area under the receiver operator characteristic curve (CV_AUROC) of 0.79 suggesting that the model had a good ability to discriminate between the endometriosis and no endometriosis groups (Figure 2c). The model had an R^2 of 0.05,

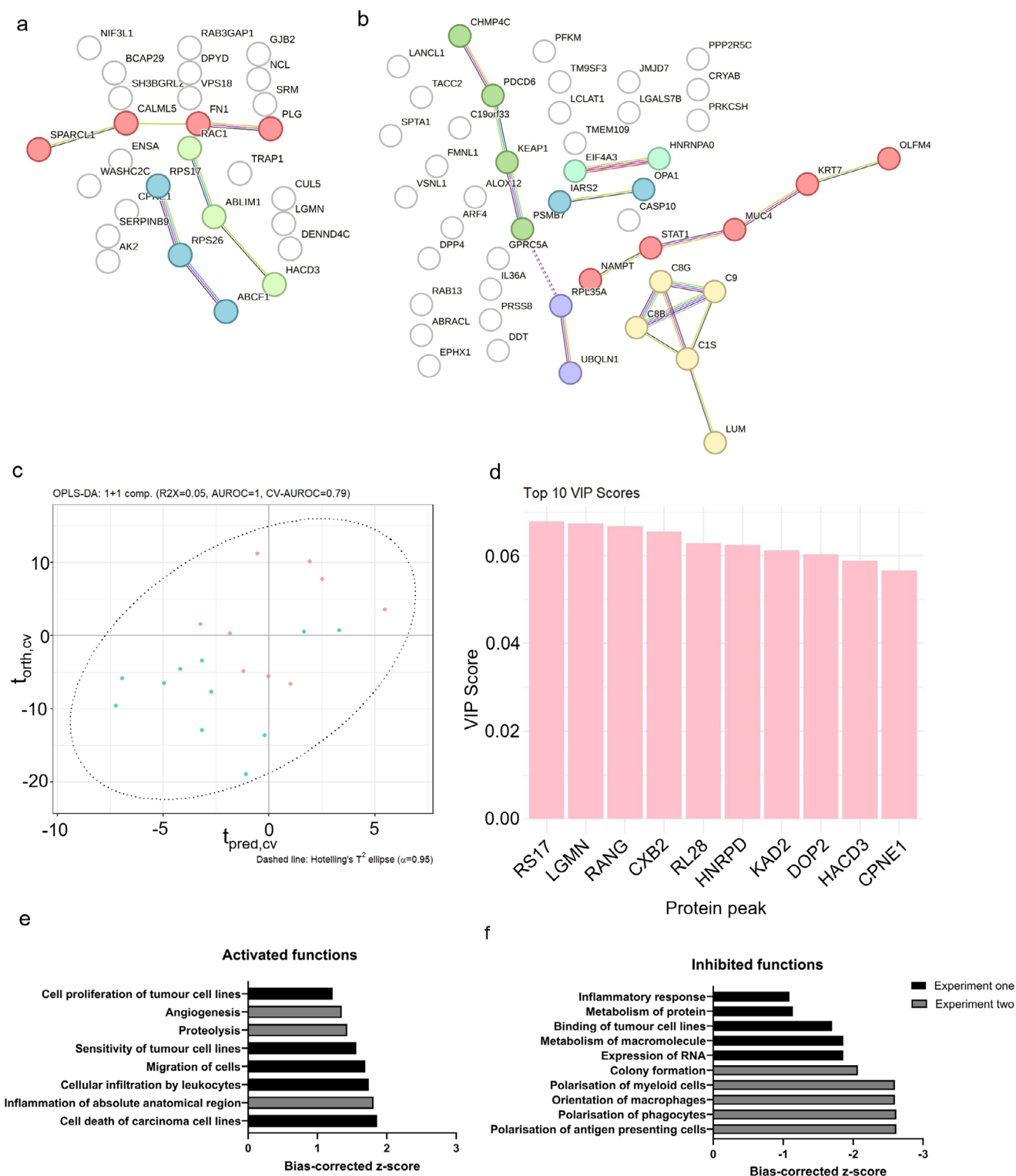


FIGURE 2 | Bioinformatic analysis differentially regulated cervicovaginal fluid proteins identified in SWATH-MS. (a)–(b) STRING functional protein association network of differentially regulated proteins for experiment one (a) and experiment two (b). Proteins with a minimum of 1.5-fold change and $p < 0.05$ were included, with the entire list of proteins identified in the respective experiment used as statistical background. Coloured nodes show groups of interacting proteins, identified through Markov clustering (MCL) with a 1.5 inflation parameter. Neither network contained significantly more interactions than expected ($p > 0.05$). (c) Orthogonal-partial least squares discriminant analysis (OPLS-DA) modelling of experiment one, demonstrating the predictive and orthogonal components of the model. (d) The ten proteins with highest variable influence on projection (VIP) score, indicating the proteins with the most predictive weight in the model. (e)–(f) Predicted molecular and biological functions activated (e) and inhibited (f) by differentially regulated cervicovaginal fluid proteins. Pathways with a Benjamini–Hochberg corrected p value < 0.05 and bias-corrected z -score of > 1 (for activated pathways) or < -1 (for inhibited pathways) were considered significant.

indicating that 95% of the variance is explained by factors outside the model. The individual proteins that contributed the most to the separation of the groups are presented in Figure 2d. There were no significant components identified in the experiment two data (CV_AUROC = 0.59).

Canonical pathway analysis predicts pathways most likely to be affected by the change in protein expression. S100 signalling was inhibited in experiment one ($p = 0.013$, Z-score = -1.00), and ferroptosis signalling ($p = 1.2 \times 10^{-4}$, Z-score = 2.00) and the complement cascade ($p = 1.4 \times 10^{-4}$, Z-score = 2.00) were activated in experiment two. No significant canonical pathways were identified in both experiments.

Upstream regulator analysis identified molecules that may be responsible for the observed changes in protein abundance. Predicted activated upstream regulators were methylprednisolone in experiment one (adj $p = 0.047$, corrected Z-score = 1.22), and hepatocyte nuclear factor-1 alpha (HNF1A) in experiment two (adj $p = 0.021$, corrected Z-score = 1.07). Interleukin-1 beta (IL1B) was predicted to be inhibited in experiment two (adj $p = 0.021$, corrected Z-score = -1.85).

Causal network analysis goes beyond identifying single upstream potential regulators, instead aiming to identify regulatory networks that may explain the observed changes in protein abundance. Master regulator analysis identifies key molecules that influence large sections of the regulatory network. CD55 was one of the top five master regulators in experiment one, and comparison analysis identified CD55 as a significant master regulator in experiment two also, although not in the top five (network bias-corrected p -value = 0.043, z-score = 1.00). Comparison analysis also identified claudin 7 (CLDN7) and density regulated re-initiation and release factor (DENR) as significant regulators with the same directionality in both experiments (CLDN7 experiment one: $p = 0.039$, $z = -1.34$, experiment two: $p = 0.023$, $z = -1.13$. DENR experiment one: $p = 0.0001$, $z = -1.41$, experiment two: $p = 0.025$, $z = -1.00$). The top five master regulators in each experiment by z-score can be found in Table S5 presents the top five master regulators in each experiment by z-score.

Biological and molecular function analysis identifies the processes that are likely to be impacted by the differential regulated proteins. There were no overlapping functions between experiments. Figure 2e,f presents the top five activated and inhibited functions identified in experiment one and two. In experiment two, there were only three activated biological functions meeting the Benjamini-Hochberg corrected p value and bias-corrected Z-score criteria. P values ranged from 0.020 to 0.047 for activated pathways and 0.012 to 0.029 for inhibited pathways.

3.3 | ELISA Validation

Concentrations of LGMN, NAMPT, SPARCL1, and OLFM4 in cervicovaginal fluid were measured by ELISA. The two groups were unequal in size, comprising all remaining samples from individuals with no endometriosis ($n = 11$), and a randomly selected subset of samples for the endometriosis cohort ($n = 28$).

The groups were similar for all reported parameters (Table S6). One participant in the no endometriosis cohort was included in the unknown menstrual phase group despite taking hormonal medications, as it was unrelated to endometriosis treatment (testosterone).

3.4 | Univariate ELISA Analysis

Three samples had insufficient volume remaining to perform the BCA assay to quantify total protein, so were excluded from the normalised protein analysis. LGMN relative to total protein was significantly elevated in people with endometriosis with a median concentration of 27.22 (15.93-111.41) pg/ μ g protein compared to 6.61 (2.43-19.18) pg/ μ g protein in the no endometriosis group (adjusted $p = 0.012$; Figure 3a). The amount of the remaining proteins NAMPT, OLFM4 and SPARCL1 relative to total protein did not differ between groups (Figure 3a). There were no significant differences in the concentrations of LGMN, SPARCL1, NAMPT, or OLFM4 between groups (Figure 2).

LGMN relative to total protein had the highest AUC of 0.826 (95% CI 0.640-1.000, $p = 0.0021$; Figure 3b). When applying a 11.47 pg/ μ g total protein cut-off, the sensitivity was 0.960 and specificity was 0.727. Post-hoc analysis determined that LGMN concentrations did not differ by approximate menstrual phase (Figure S3). The AUC and the sensitivity and specificity at cut-offs determined by Youden's index for each protein are in Table S7.

3.5 | Multivariate ELISA Analysis

To assess whether combining markers could provide greater discriminatory power than individual proteins, multivariate analysis was performed. Firstly, a biomarker score was derived from the quartile concentrations of the four ELISA proteins. The biomarker scores were significantly lower in the endometriosis group compared to controls ($p = 0.048$; Figure 3c), however the AUC was 0.699 (Table S7). Next, logistic regression models were produced to determine whether increasing the complexity and nuance of the model improved discriminatory power. The models showed similar performances, with AUCs of 0.735 and 0.731 for concentration and concentration relative to total protein, respectively (Figure 3d). Neither multivariate model met the pre-defined sensitivity and specificity thresholds.

4 | Discussion

In this study, data-independent proteomic analysis across two experiments were performed on cervicovaginal fluid sampled with vaginal swabs. There was no common signature of protein dysregulation across the two experiments, and bioinformatic analysis showed equally discordant results. Cervicovaginal fluid proteins were measured using ELISA, identifying normalised LGMN concentration as significantly elevated in the endometriosis group.

The two separate proteomics runs aimed to provide a form of validation for each experiment's findings. However, each experiment

identified distinct sets of differentially abundant proteins and at $p < 0.05$ there were none common to both experiments. Unsupervised (PCA) and supervised (OPLS-DA) analysis revealed the data had complex structure rather than one clear source of variation, with individual proteins contributing minimally to group separation. These findings align with the perspective that individual markers are unlikely to effectively discriminate between endometriosis and non-endometriosis cases [21, 22]. Furthermore, validating biomarkers in external cohorts remains elusive.

Predicted dysregulated pathways in endometriosis also showed little consistency between the two experiments. Importantly, lack of reproducibility for both individual proteins and pathways may suggest that cervicovaginal fluid does not exhibit consistent changes associated with endometriosis. Furthermore, none of the proteins identified in this study overlap with results from the one other study to perform proteomic analysis of cervicovaginal fluid for endometriosis biomarkers [15]. However, there were notable methodological differences compared to the current study, such as more inclusion criteria for the endometriosis group, use of an asymptomatic control group, different approaches to sampling cervicovaginal fluids and data-dependent proteomic analysis. What remains unclear is whether this inconsistency stems from the heterogeneity of endometriosis coupled with small cohort sizes, dysregulation patterns being masked by menstrual cyclicity, or if there are genuinely limited changes associated with endometriosis detectable in cervicovaginal fluid.

As the data displayed a complex structure in the SWATH-MS results, multivariate analysis was performed to assess whether coordinated changes together were more predictive. While the biomarker score was significantly different between the two groups, neither this score nor the multivariate logistic regression models outperformed LGMN as a single marker.

LGMN demonstrated strong individual biomarker performance in the ELISA validation. Identified in the OPLS-DA model as one of the top proteins contributing to group separation, LGMN concentrations were significantly increased in individuals with endometriosis and sensitivity (0.960) and specificity (0.727) values meeting the thresholds for a clinical useful rule-out test [9]. Improvement of diagnostic accuracy may be achieved through incorporating other cervicovaginal fluid markers outside those investigated in this study. Interestingly, elevated legumain pseudogene 1 (LGMNP1), which protects LGMN expression by reducing miR-495-3p-mediated silencing [23], in serum extracellular vesicles has also been reported as a candidate marker of predicting endometriosis recurrence [24]. Given the lack of validation observed often in endometriosis biomarker studies and in the current study, a larger and more diverse cohort would be required to confirm the promise of cervicovaginal fluid LGMN as a candidate biomarker of endometriosis. Investigating cervicovaginal LGMN in the context of recurrence may also be important.

LGMN is an asparaginyl-specific cysteine endopeptidase located within lysosomes and expressed primarily by macrophages, particularly M2-like macrophages [25, 26]. LGMN secreted by M2-like macrophages contributes to extracellular matrix degradation through the activation of MMP2 and MMP9 [25] and in vitro

studies suggest LGMN promotes angiogenesis [26]. Macrophages are the dominant innate immune cell in the lower genital tract [27] and are generally associated with an M2-like phenotype in endometriosis [28, 29]. This provides a plausible explanation for the observed increase in cervicovaginal fluid LGMN. However, these hypotheses remain speculative as limited research has explored LGMN in the context of endometriosis, and macrophage polarisation in the vagina and cervix is poorly characterised [30]. Beyond potential utility as a biomarker, increased LGMN levels in endometriosis may reflect altered macrophage phenotypes in cervicovaginal fluid, warranting further investigation.

Bioinformatic analysis predicted three pathways most likely to be affected by the observed differences in protein abundances that may warrant further investigation in the context of endometriosis pathophysiology. Ferroptosis, a regulated form of iron-dependent cell death, was suggested to be activated based on protein abundance changes, contrasting previous reports of inhibition in endometriotic tissue [31, 32]. Iron-induced ferroptosis may contribute to fibrosis and inflammation via increasing proportions of myofibroblasts, [31], congruent with the high iron levels observed in the peritoneal cavity [32]. Endometriosis-associated dysregulation of S100 signalling pathways and complement proteins has been previously reported in both menstrual blood [33] and cervical mucus [15]. Our study validates these findings, suggesting the systemic inflammation and immune dysregulation in endometriosis may extend to the cervicovaginal environment. Dysregulated complement activity has been linked with adverse reproductive outcomes [34], suggesting a potential mechanistic link to endometriosis-associated subfertility. Therefore, this study has identified novel pathways of endometriosis pathophysiology for further research including investigation into targeted therapy.

This study had several limitations. *P* values were not adjusted for multiple comparisons in the SWATH-MS analysis, increasing the likelihood of false positives. However, for the bioinformatic analysis, pathway enrichment requires sets of dysregulated proteins, making entire false-positive pathways unlikely. The two-step study design enabled the comparison of trends identified under these lower stringency thresholds across both experimental datasets, anticipating the emergence of modest yet pathophysiologically relevant changes. However, few conserved alterations were observed, suggesting limited consistent proteomic changes in cervicovaginal fluid in relation to endometriosis. High-abundance keratin, haemoglobin and plasma proteins and keratins were excluded to enhance assessment of lower-abundance proteins. We opted for removal during data analysis as the presence of these proteins were not consistent between samples, and depletion procedures during sample preparation targeting high abundance plasma proteins may have changed the protein complement of the original samples unpredictably. Subsequently, if these proteins were informative these insights will have been missed.

Additional limitations include the lack of screening for vaginal infections or recent sexual activity, and that menstrual cycle phase was not controlled for. While no significant differences in LGMN levels were observed across the cycle, confidence in this finding is constrained by small subgroup sizes and reliance on self-reported information for menstrual phase estimation.

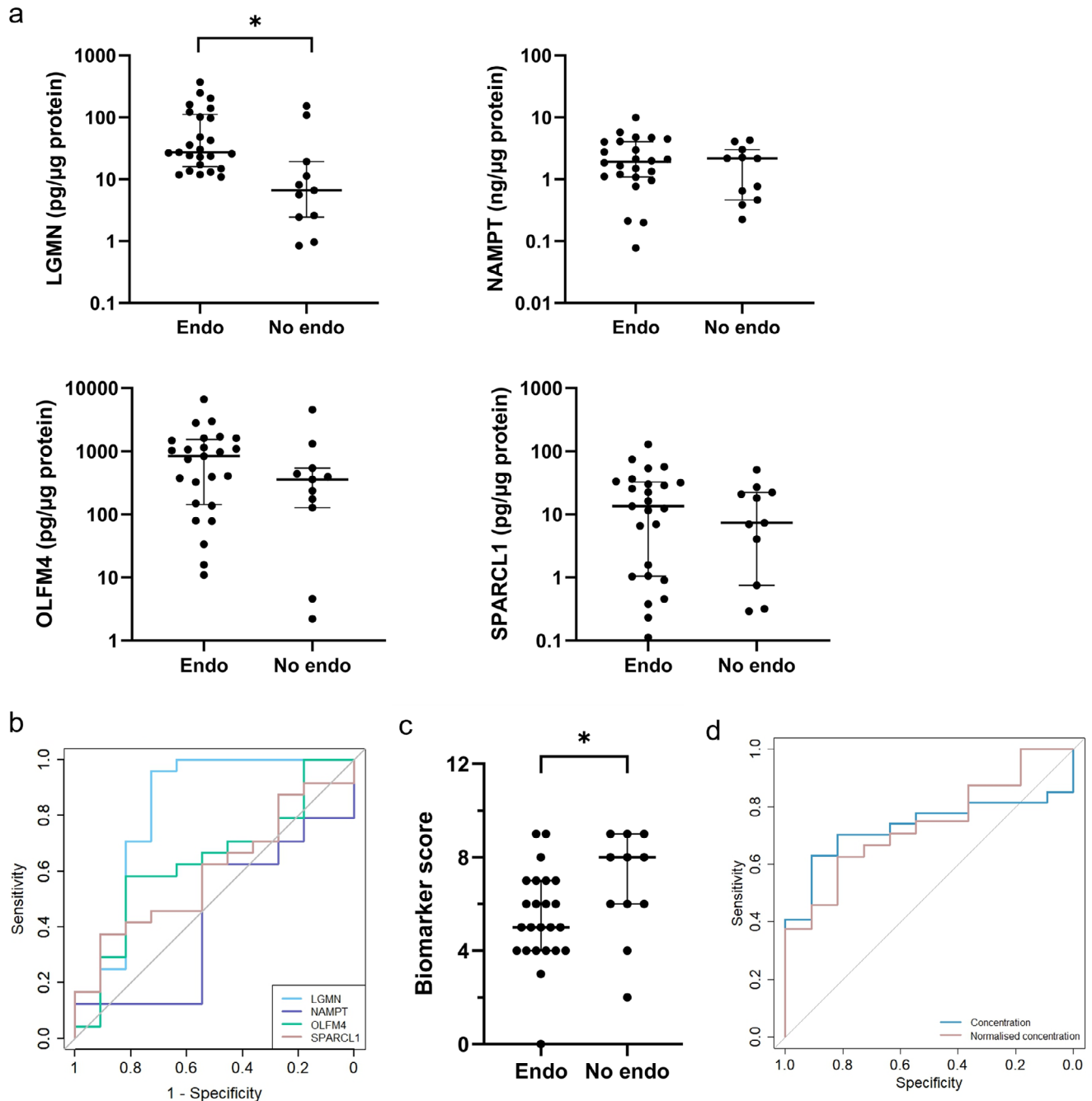


FIGURE 3 | Cervicovaginal fluid proteins measured by ELISA. a) Cervicovaginal fluid LGMN, NAMPT, OLFM4, and SPARCL1 concentrations normalised to total protein (endometriosis $n = 25$, no endometriosis $n = 11$). (b) Receiver operator characteristic (ROC) curves of normalised cervicovaginal fluid protein concentrations. (c) A biomarker score system was developed by dividing the normalised concentration of LGMN and concentrations of NAMPT, OLFM4, and SPARCL1 into quartiles, then assigning a score between 0–3. Scores for individual proteins were summed with a theoretical maximum of 12. (d) ROC curves of multivariate logistic regression models of the concentrations and normalised concentrations of LGMN, NAMPT, OLFM4 and SPARCL1. Reference line (random classification) in gray. Data is presented as the median and interquartile range. * Adjusted $p < 0.05$, assessed by Mann–Whitney tests with Holm–Šidák correction for multiple comparisons (a) or Fisher’s exact test (c). ELISA—enzyme-linked immunosorbent assay, LGMN—legumain, NAMPT—nicotinamide phosphoribosyltransferase, OLFM4 – olfactomedin 4, SPARCL1 – SPARC-like protein 1.

Because this study aimed to evaluate biomarkers in conditions reflective of their intended clinical use, our menstrual cycle estimates were a pragmatic balance between accuracy and clinical feasibility. Finally, the ELISA kits used were not validated by the manufacturers for cervicovaginal fluid, which may have affected assay performance and result reproducibility.

To maintain the translatability of this pre-clinical research, we prioritised a participant cohort representative of the populations the test is clinically intended for, by applying limited exclusion criteria and including control participants that were pre-operatively suspected to have endometriosis. Consequently, the cohorts were highly heterogenous regarding menstrual cycle phase, hormonal

medication use, concurrent pathologies and disease phenotype which may have affected our power to detect differences between groups. We believe this ultimately prevents wasting resources attempting to validate biomarkers that do not perform accurately outside highly selected participant cohorts.

The lack of concordance between the two proteomics runs was a notable result. These findings suggest that cervicovaginal fluid proteins sampled via vaginal swab may have limited biomarker potential for endometriosis. Given the lack of reproducible and validated markers despite thousands of reported candidates [10], this result may be unsurprising—reporting these negative findings is important to prevent redundant work in this space. Exploring subtype specific diagnostic markers or incorporating symptom-based phenotyping in studies with larger cohorts presents an opportunity to reduce heterogeneity, with biomarkers panels showing improved performance at endometriosis disease extremes [35]. However, non-invasive imaging techniques perform well for more severe endometriosis subtypes (endometrioma and DIE) [4]. Thus, diagnostic endometriosis biomarkers with poorer performance in diagnosing superficial endometriosis may have limited clinical utility. Future patient stratification enabled by deeper understanding of endometriosis molecular biology, may enhance biomarker discovery efforts.

5 | Conclusion

This study performed two proteomics experiments on cervicovaginal fluid samples, finding dysregulated cervicovaginal fluid proteins do not show a consistent signature in endometriosis. Bioinformatic analysis predicted the disruption of pathways previously linked to endometriosis, indicating systematic inflammation and altered immune function observed in endometriosis extends to the cervicovaginal environment that warrant further research. Cervicovaginal fluid normalised LGMN concentrations discriminated between endometriosis and controls with accuracy suitable for a clinical triage test, but a much larger cohort would be required to confirm the promise of cervicovaginal fluid LGMN as a candidate biomarker of endometriosis. Overall, these findings suggest that cervicovaginal fluid proteins sampled via vaginal swab may have limited biomarker potential for endometriosis, highlighting areas for additional investigation.

Acknowledgments

This work was supported by the Royal Australian and New Zealand College of Obstetricians and Gynaecologists (RANZCOG; grant number 22-2). Open access publishing facilitated by University of Otago, as part of the Wiley—University of Otago agreement via the Council of Australian University Librarians. The authors extend our deepest gratitude to all participants and their whānau who contributed to this research.

Open access publishing facilitated by University of Otago, as part of the Wiley – University of Otago agreement via the Council of Australasian University Librarians

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [36] partner repository with the dataset identifier PXD065951.

Reviewer access details

Log in to the PRIDE website using the following details:

Project accession: PXD065951

Token: V7IIToU76vai

Alternatively, reviewer can access the dataset by logging in to the PRIDE website using the following account details:

Username: reviewer_pxd065951@ebi.ac.uk

Password: yznvPVAALnwF

References

1. International Working Group of AAGL, ESGE, ESHRE and WES., C. Tomassetti, N. P. Johnson, J. Petrozza, et al., “An international terminology for endometriosis, 2021,” *Human Reproduction Open* 2021, no. 4 (2021), <https://doi.org/10.1093/hropen/hoab029>.
2. Endometriosis New Zealand. Endo Information [Internet]. Christchurch New Zealand: Endometriosis NZ. Available from: 2023 [Accessed 2023 July 25]. <https://nzendo.org.nz/endo-information/>.
3. K. E. Nnoaham, L. Hummelshoj, P. Webster, et al., “Impact of endometriosis on quality of life and work productivity: A multicenter study Across ten countries,” *Fertility and Sterility* 96, no. 2 (2011): 366–373.e8, <https://doi.org/10.1016/j.fertnstert.2011.05.090>, e8.
4. V. Nisenblat, P. M. M. Bossuyt, C. Farquhar, N. Johnson, and M. L. Hull, “Imaging modalities for the non-invasive diagnosis of endometriosis,” *Cochrane Database of Systematic Reviews* no. 2 (2016), <https://doi.org/10.1002/14651858.CD009591.pub2>.
5. C. Hanly, M. Leonardi, A. Eathorne, and M. Armour, “The impact of diagnostic method on sense of control and powerlessness and social support in endometriosis patients-A retrospective cohort study,” *Acta Obstetrica Et Gynecologica Scandinavica* 102, no. 10 (2023): 1390–1395, <https://doi.org/10.1111/aogs.14488>.
6. E. L. Mowers, C. S. Lim, B. Skinner, et al., “Prevalence of Endometriosis During Abdominal or Laparoscopic Hysterectomy for Chronic Pelvic Pain,” *Obstetrics & Gynecology* 127, no. 6 (2016): 1045–1053, <https://doi.org/10.1097/AOG.0000000000001422>.
7. E. Liu, V. Nisenblat, C. Farquhar, et al., “Urinary biomarkers for the non-invasive diagnosis of endometriosis,” *Cochrane Database of Systematic Reviews* no. 12 (2015), <https://doi.org/10.1002/14651858.CD012019>.
8. D. Gupta, M. L. Hull, I. Fraser, et al., “Endometrial biomarkers for the non-invasive diagnosis of endometriosis,” *Cochrane Database of Systematic Reviews* no. 4 (2016), <https://doi.org/10.1002/14651858.CD012165>.
9. V. Nisenblat, P. M. M. Bossuyt, R. Shaikh, et al., “Blood biomarkers for the non-invasive diagnosis of endometriosis,” *Cochrane Database of Systematic Reviews* no. 5 (2016), <https://doi.org/10.1002/14651858.CD012179>.
10. A. Brulport, M. Bourdon, D. Vaiman, et al., “An integrated multi-tissue approach for endometriosis candidate biomarkers: A systematic review,” *Reproductive Biology and Endocrinology* 22, no. 1 (2024): 21, <https://doi.org/10.1186/s12958-023-01181-8>.
11. M. Armour, D. Ciccia, A. Yazdani, et al., “Endometriosis research priorities in Australia,” *Australian and New Zealand Journal of Obstetrics and Gynaecology* 63, no. 4 (2023): 594–598, <https://doi.org/10.1111/ajo.13699>.

12. P. C. Brady, A. W. Horne, P. T. K. Saunders, A. M. Thomas, S. A. Missmer, and L. V. Farland, "Research priorities for endometriosis differ Among patients, clinicians, and researchers," *American Journal of Obstetrics and Gynecology* 222, no. 6 (2020): 630–632, <https://doi.org/10.1016/j.ajog.2020.02.047>.
13. S. M. Scheck, C. Henry, N. Bedford, et al., "Non-invasive tests for endometriosis are here; how reliable are they, and what should we do With the results?," *Australian and New Zealand Journal of Obstetrics and Gynaecology* 64, no. 2 (2024): 168–170, <https://doi.org/10.1111/ajo.13765>.
14. G. Zegels, G. A. A. Van Raemdonck, W. A. A. Tjalma, and X. W. M. Van Ostade, "Use of cervicovaginal fluid for the identification of biomarkers for pathologies of the female genital tract," *Proteome Science* 8, no. 1 (2010): 63, <https://doi.org/10.1186/1477-5956-8-63>.
15. G. Grande, F. Vincenzoni, D. Milardi, et al., "Cervical mucus proteome in endometriosis," *Clin Proteomics* 14 (2017): 7, <https://doi.org/10.1186/s12014-017-9142-4>.
16. B. Gidaszewski, M. Khajehei, and T. McGee, "Outpatient cervical ripening: Discomfort/pain During speculum and Foley catheter insertion," *Midwifery* 67 (2018): 57–63, <https://doi.org/10.1016/j.midw.2018.09.012>.
17. American Society for R. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. *Fertility and Sterility* 1997;67(5):817–821, [https://doi.org/10.1016/S0015-0282\(97\)81391-X](https://doi.org/10.1016/S0015-0282(97)81391-X).
18. D. Szklarczyk, R. Kirsch, M. Koutrouli, et al., "The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest," *Nucleic Acids Research* 51, no. D1 (2022): D638–D646, <https://doi.org/10.1093/nar/gkac1000>.
19. H. J. Kim, E. Y. Seong, W. Lee, et al., "Comparative analysis of therapeutic effects Between medium cut-off and high flux dialyzers using metabolomics and proteomics: Exploratory, prospective study in hemodialysis," *Scientific Reports* 11, no. 1 (2021): 17335, <https://doi.org/10.1038/s41598-021-96974-5>.
20. G. A. Kristono, *Inflammatory Profiles in Acute Coronary Syndromes* PhD thesis.: (University of Otago, 2021).
21. S. H. Ahn, V. Singh, and C. Tayade, "Biomarkers in endometriosis: Challenges and opportunities," *Fertility and Sterility* 107, no. 3 (2017): 523–532, <https://doi.org/10.1016/j.fertnstert.2017.01.009>.
22. K. E. May, S. A. Conduit-Hulbert, J. Villar, S. Kirtley, S. H. Kennedy, and C. M. Becker, "Peripheral biomarkers of endometriosis: A systematic review," *Human Reproduction Update* 16, no. 6 (2010): 651–674, <https://doi.org/10.1093/humupd/dmq009>.
23. K. Liao, Z. Qian, S. Zhang, et al., "The LGMN pseudogene promotes tumor progression by acting as a miR-495-3p sponge in glioblastoma," *Cancer Letters* 490 (2020): 111–123, <https://doi.org/10.1016/j.canlet.2020.07.012>.
24. S. G. Sun, J. J. Guo, X. Y. Qu, et al., "The extracellular vesicular pseudogene LGMNP1 induces M2-Like macrophage polarization by upregulating LGMN and serves as a novel promising predictive biomarker for ovarian endometriosis recurrence," *Human Reproduction* 37, no. 3 (2022): 447–465, <https://doi.org/10.1093/humrep/deab266>.
25. J.-M. Chen, M. Fortunato, R. A. E. Stevens, and A. J. Barrett, "Activation of Progelatinase A by Mammalian Legumain, a Recently Discovered Cysteine Proteinase," *Biochemistry* 382, no. 5 (2001): 777–784, <https://doi.org/10.1515/bchm.2001.382.5.777>.
26. L. Shen, H. Li, Y. Shi, et al., "M2 tumour-associated macrophages contribute to tumour progression via legumain remodelling the extracellular matrix in diffuse large B cell lymphoma," *Scientific Reports* 6, no. 1 (2016): 30347, <https://doi.org/10.1038/srep30347>.
27. J. Pudney, A. J. Quayle, and D. J. Anderson, "Immunological Microenvironments in the Human Vagina and Cervix: Mediators of Cellular Immunity Are Concentrated in the Cervical Transformation Zone1," *Biology of Reproduction* 73, no. 6 (2005): 1253–1263, <https://doi.org/10.1095/biolreprod.105.043133>.
28. M. Bacci, A. Capobianco, A. Monno, et al., "Macrophages Are Alternatively Activated in Patients With Endometriosis and Required for Growth and Vascularization of Lesions in a Mouse Model of Disease," *The American Journal of Pathology* 175, no. 2 (2009): 547–556, <https://doi.org/10.2353/ajpath.2009.081011>.
29. Y. Gou, X. Li, P. Li, et al., "Estrogen receptor β upregulates CCL2 via NF- κ B signaling in endometriotic stromal cells and recruits macrophages to promote the pathogenesis of endometriosis," *Human Reproduction* 34, no. 4 (2019): 646–658, <https://doi.org/10.1093/humrep/dez019>.
30. M. K. Collins, C. R. McCutcheon, and M. G. Petroff, "Impact of Estrogen and Progesterone on Immune Cells and Host–Pathogen Interactions in the Lower Female Reproductive Tract," *The Journal of Immunology* 209, no. 8 (2022): 1437–1449, <https://doi.org/10.4049/jimmunol.2200454>.
31. Y. Zhang, X. Liu, M. Deng, et al., "Ferroptosis induced by iron overload promotes fibrosis in ovarian endometriosis and is related to subpopulations of endometrial stromal cells," *Frontiers in pharmacology* 13 (2022): 930614, <https://doi.org/10.3389/fphar.2022.930614>.
32. A. Van Langendonck, F. Casanas-Roux, and J. Donnez, "Iron overload in the peritoneal cavity of women With pelvic endometriosis," *Fertility and Sterility* 78, no. 4 (2002): 712–718.
33. S. Ji, Y. Liu, L. Yan, et al., "DIA-based analysis of the menstrual blood proteome identifies association Between CXCL5 and IL1RN and endometriosis," *Journal of Proteomics* 289 (2023): 104995, <https://doi.org/10.1016/j.jprot.2023.104995>.
34. M. Sugiura-Ogasawara, K. Nozawa, T. Nakanishi, Y. Hattori, and Y. Ozaki, "Complement as a predictor of further miscarriage in couples With recurrent miscarriages," *Human Reproduction* 21, no. 10 (2006): 2711–2714, <https://doi.org/10.1093/humrep/del229>.
35. E. M. Schoeman, Bringans, K. Peters, et al., "Identification of plasma protein biomarkers for endometriosis and the development of statistical models for disease diagnosis," *Human Reproduction* (2024).
36. Y. Perez-Riverol, C. Bandla, D. J. Kundu, et al., "The PRIDE database at 20 years: 2025 update," *Nucleic Acids Research* 53, no. D1 (2025): D543–D553, <https://doi.org/10.1093/nar/gkae1011>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File: prca70044-sup-0001-SuppMat.docx.