Skubitz *et al. Clinical Proteomics* (2026) 23:32
https://doi.org/10.1186/s12014-026-09610-7

Clinical Proteomics

RESEARCH Open Access

Repurposing the liquid-based Pap test for the detection of ovarian cancer protein biomarkers

Amy P. N. Skubitz[1,2,3,9*], Kristin L. M. Boylan[1,3], Limeng Liu[4], Ashley Petersen[4], Yuqian Gao[5], Joshua R. Hansen[5], Jesse B. Trejo[5], Thomas L. Fillmore[6], Yang Wang[7], Melissa A. Geller[2], Peter A. Argenta[2], Samantha Hoffman[2], Karin Rodland[8], Tao Liu[5] and Paul D. Piehowski[6]

Abstract

Background Earlier detection is strongly associated with increased survival for women with ovarian cancer. Unfortunately, current screening strategies, employing serial serum or ultrasound assessments, lack adequate sensitivity and specificity for use in the low-prevalence general population. In contrast, screening for cervical cancer by Pap tests has been routinely performed for over 50 years. Since ovarian cancer cells have been observed in Pap tests, ovarian cancer protein biomarkers may also be present; yet Pap samples have not been rigorously examined for diagnostic proteins. Assessment of cervical effluent, as can be collected in a Pap test, has the potential to differentiate ovarian cancer cases from healthy controls, and thereby demonstrate that intra-abdominal pathology may be detected using this commonly acquired specimen. We hypothesize that proteins shed by ovarian cancer cells can be detected in the SurePath™ liquid-based Pap test fixative using mass spectrometry (MS)-based proteomics, making it possible to distinguish women with ovarian cancer from healthy women.

Methods Candidate ovarian cancer biomarkers were successfully identified in liquid-based Pap test samples from 20 cases of high grade serous ovarian cancer, 10 benign ovarian conditions, and 10 healthy control samples, by performing Tandem Mass Tag™ isobaric labeling, 2D liquid chromatography-MS/MS, and bioinformatics integration. Selected reaction monitoring (SRM) MS-based targeted proteomics was then performed using a panel of candidate biomarkers to quantify their abundance in an expanded patient cohort of 90 liquid-based Pap tests.

Results A multi-protein classifier was developed using the SRM-MS data comprised of 6 proteins and achieving an AUC of 0.880 (95% CI: 0.738–0.989); sensitivity at 90% specificity was 0.800.

Conclusion This pilot study provides evidence that the cervical effluent collected in SurePath™ fixative from Pap tests has the potential to be used as a biospecimen for the detection of ovarian cancer proteins. Our long-term goal is to develop a noninvasive screening test that can be incorporated into a routine Pap test or a self-administered home test, so that women can be screened simultaneously for cervical and ovarian cancer.

Keywords Ovarian cancer detection, Pap test, Mass spectrometry-based proteomics, Biomarker

*Correspondence:
Amy P. N. Skubitz
skubi002@umn.edu

Full list of author information is available at the end of the article

Background

Earlier detection is critical to improved prognosis for women with ovarian cancer, which is the 5th leading cause of cancer deaths among women in the U.S [1]. Given the lack of specific symptoms, most patients are diagnosed at later stages of disease, when the long-term survival rate is < 20% [2]. Serum CA-125 levels and transvaginal ultrasound, even when used in combination, are neither sensitive nor specific enough for general population screening given disease prevalence [3, 4]; highlighting the need for novel "early detection" strategies.

In this study, we set out to develop a paradigm shift that will change the way that women are screened for ovarian cancer by leveraging the standard Papanicolaou (Pap) test. Screening for cervical cancer by Pap tests has been routinely performed for over 50 years [5] and > 30 million are performed annually in the U.S [6]. The liquid-based Pap test consists of collecting cervical cells from the external cervical opening and placing them into a vial containing an alcohol-based fixative [7, 8]. After processing the cells from the vials and staining them, they are examined by a pathologist to identify the presence of premalignant and malignant cells. The fact that ovarian cancer cells have been identified in liquid-based Pap tests [9–13] suggests that ovarian cancer proteins and peptides can also traverse the fallopian tube to the cervical opening. The liquid fixative solution in which the cells are collected for Pap tests is typically discarded after examination of the cells. To our knowledge, this study is the first to analyze cervical effluent in the Pap test fixative by mass spectrometry (MS)-based proteomic techniques to search for ovarian cancer peptide/protein biomarkers.

Cervical effluent present in the residual Pap test fixative provides several important advantages for biomarker development. First, proteins present in Pap test fixative are collected from a site anatomically near the ovarian cancer (i.e. proteins may be secreted or shed from the tumor and flow through the fallopian tube into the uterus and out the cervical opening). Tumor-specific biomarkers have been shown to be present at higher levels in biospecimens proximate to the tumor, such as tumor interstitial fluid, compared to blood [14]. Our group and others showed that ascites fluid contains higher levels of ovarian cancer biomarkers than serum from the same patient by immunoassay [15, 16] and quantitative proteomic analysis [17, 18]. In addition, emerging paradigms have focused on the fallopian tube epithelium as the initial site of at least some high grade serous ovarian cancer (HGSOC) [19], further emphasizing the potential of cervical effluent as a source of ovarian cancer biomarkers. These studies, therefore, would suggest that cervical effluent samples obtained during a routine Pap test may contain higher levels of ovarian cancer proteins than blood.

Second, comprehensive quantitative proteomic analyses of HGSOC tumor tissue by investigators for the Clinical Proteomic Tumor Analysis Consortium (CPTAC) has revealed a proteomic signature associated with chromosomal instability and homologous recombination defects [20]. These genomic alterations are hallmarks of HGSOC and could provide insight into relevant biomarkers discovered in Pap test samples.

More recently, CPTAC reported a comparison of proteins from ovarian cancer tumors to proteins expressed in the fallopian tube [21]. These results strengthen the idea that cervical samples are a rich source of cancer biomarkers. These data identified 509 proteins that were differentially expressed between tumor tissue and normal fallopian tube, which could be used to discern biomarkers with the potential for early detection in Pap test samples.

In an earlier feasibility study, we optimized sample preparation and handling procedures for SurePath™ Pap test fixative from 120 post-menopausal women with normal cervical cytology [22]. Briefly, proteomic techniques were used to separate the proteolytic peptides in the samples by reversed-phase liquid chromatography (RPLC), then detected by MS. More than 700 proteins were identified from a pool of 40 healthy individuals; an average of 427 proteins were identified in Pap tests from each of 5 healthy individuals. We coined the term, "Normal Pap Test Core Proteome" to represent the 153 proteins that were present in at least 4 out of 5 individuals, and quantified the level of protein by calculating a normalized spectral abundance factor for each protein. When we compared our "Normal Pap Test Core Proteome" to proteins identified by others who had directly sampled cervical vaginal fluid [23–26], we found many similarities [27]. We concluded that the Pap test fixative may serve as a proxy for samples derived from the genital tract.

Other studies have examined the Pap test for somatic mutations in DNA for the identification of gynecological cancers [28]. Similar to our study, Kinde et al. used the Pap test sample; however, they examined DNA isolated from the cellular component of the Pap test for mutations that were known to be present in tumors from the same patient. Subsequently, this group used DNA isolated from Pap brush samples to detect mutations in 18 genes commonly mutated in ovarian cancer, combined with a test for aneuploidy [29], however the sensitivity of ovarian cancer detection was low (33%). More recently, two groups in Europe used droplet digital PCR technology to detect TP53 variants identified in tumors in Pap test samples from ovarian cancer patients [30, 31]. Interestingly, both groups were able to detect the TP53 variants in archival Pap test samples ranging from 20 months [30] to 6 years [31] prior to diagnosis. However, another recent analysis of TP53 mutations in Pap test samples

from women with and without HGSOC demonstrated that using a highly sensitive method of sequencing, they were able to detect "abundant, low frequency" TP53 mutations in women without ovarian cancer [32].

HGSOC are the most common subtype of ovarian cancer, and they are typically diagnosed at a late stage when treatments are less effective [5]. Thus, HGSOC was the focus of this study for the discovery and validation phases of cancer biomarkers. In this study, we have used Tandem Mass Tags™ (TMT) isobaric labeling and 2-dimensional liquid chromatography coupled to tandem mass spectrometry (2D LC-MS/MS) to compare the protein profiles of Pap tests collected from women prior to gynecologic oncology surgery versus Pap tests collected from healthy women. Selected proteins that we found to be differentially expressed between these groups were then examined by selected reaction monitoring (SRM) targeted MS on a larger cohort of Pap tests to develop a multi-protein classifier. Our rationale for generating a multi-protein classifier that could distinguish HGSOC from healthy controls is based on our vision of developing an at-home self-sampling test to be used for population-based screening; women with positive results would then require additional tests in the clinic. By incorporating the multi-protein classifier into a routine Pap test, it may be possible for women to be screened simultaneously for cervical and ovarian cancer.

Methods

Collection of "Mock" Pap tests

Under an IRB approved protocol (1112M07362), we developed "Mock" Pap test kits to mimic the sample collection that would occur in a clinical setting, using a Pap test broom and a SurePath™ vial (BD Biosciences). After obtaining signed consent, "Mock" Pap tests were collected prior to surgery from chemo-naïve women with abdominal masses suspected to have ovarian cancer. High grade serous ovarian carcinomas (HGSOC) are the most common subtype of ovarian cancer and are the focus of this study. As samples were collected at the time of surgery, most HGSOC cases were late stage (III or IV) with only a few early stage samples collected. "Mock" Pap tests were also collected from women with benign ovarian conditions as controls for ovarian cancer. As an additional control, "Mock" Pap tests were collected from healthy women over age 40 in the University of Minnesota Women's Clinic. Samples were sent to the lab where they were vortexed and the collection broom was removed from the SurePath™ vial before storage at 4 °C.

Sample preparation

SurePath™ vials were vortexed and 5 mL of the Mock Pap fluid was centrifuged at 400xg to remove cells. Cell-free Pap test fluid samples were precipitated using acetone to remove contaminants and concentrate the proteins, then solubilized in lysis buffer (8 M urea, 75 mM NaCl, 50 mM Tris, pH 8.0, 1 mM EDTA). Protein concentration was determined by BCA assay (Thermo Scientific). The recovered proteins, up to 100 µg, were reduced with 5 mM DDT for 30 min and subsequently alkylated with 10 mM iodoacetamide for 30 min; then diluted 1:3 with 50 mM Tris-HCl (pH 8.0) and subjected to proteolytic digestion with LysC (Wako Chemicals) at 1 mAU:50 µg enzyme-to-substrate ratio for 2 h at room temperature, followed by the addition of sequencing-grade modified trypsin (Promega) at a 1:50 enzyme-to-substrate ratio and overnight incubation at room temperature. The digested proteins were acidified with 50% trifluoroacetic acid (TFA, Sigma) to a pH value of approximately 2.0. Tryptic peptides were desalted on reversed-phase C18 SPE columns (Waters) and dried using a Speed-Vac (Thermo Scientific). Peptide recovery was then determined by BCA assay.

TMT labeling of peptides

Thirty µg of desalted peptides from each sample were labeled with 11-plex TMT reagents according to the manufacturer's instructions (ThermoFisher Scientific) with minimal modification to accommodate reduced sample mass (ThermoFisher Scientific). Mass tags for 40 "Mock" Pap test samples (Additional Table 1) were distributed between the three sample types (20 HGSOC, 10 benign ovarian conditions, and 10 healthy controls) across four TMT 11-plex experiments using a randomized block design. Channel 131 N was used for labeling the internal reference sample (pooled from all samples) throughout the sample analysis. Peptides labeled by different TMT reagents were mixed, dried using a Speed-Vac, reconstituted with 3% acetonitrile, 0.1% formic acid and desalted on tC18 SepPak SPE columns.

2-D LC-MS/MS analysis

The pooled sample of TMT labeled peptide samples was fractionated by high-pH RPLC as described [20, 33, 34]. Approximately 500 µg of 11-plex TMT labeled sample was loaded on the reversed phase Agilent Zorbax 300 Extend-C18 column and 96 fractions were collected. The fractions were then concatenated into 24, by combining 4 fractions that are 24 fractions apart, to create 24 total for subsequent analysis by LC-MS/MS analysis. Fractionated samples were separated using a nanoACQUITY UPLC system (Waters) by RPLC [20, 34, 35] coupled to an Orbitrap Fusion Lumos MS (ThermoFisher Scientific). Orbitrap precursor spectra (AGC 4×10^5) were collected from 350 to 1800 m/z for 110 min at a resolution of 60 K along with data dependent Orbitrap higher-energy collision dissociation (HCD) MS/MS spectra (centroid) at a resolution of 50 K (AGC 1×10^5).

Global proteomics data analysis

MS data sets were searched against protein sequence database using MS-GF+ [36, 37], followed by mzRefinery [38] to characterize and correct for any instrument systematic mass measurement errors. Results were filtered with a false discovery rate (FDR) of < 1%. The intensities of all TMT reporter ions were extracted using MASIC software [39]. Protein abundances were summarized from peptide-level reporter ion intensities and normalized to the plex-specific reference channel by dividing each protein intensity by its corresponding reference intensity. The resulting relative protein abundances were log2-transformed, globally median-normalized across samples to account for loading/labeling differences and then corrected for TMT plex-specific batch effects. Differential expression analysis was performed comparing HGSOC against both healthy controls and benign cases using the limma R package [40] . Significantly differentially expressed proteins were defined as those with p-value < 0.05.

Targeted multiplex SRM-MS assay

Peptide selection

Peptides for targeted assays were selected from proteins identified in the TMT data and based on tryptic peptides ranging from 6 to 24 amino acids in length which produce multiply charged ions and uniquely represent the target proteins. Fully tryptic peptides allowing K/R.P [41] were selected and those peptides with post-translational modifications (PTMs) were excluded [42–44], and in general, methionine-containing peptides were excluded due to its potential for oxidation during storage and sample preparation except when there were no other alternatives. UniProt Knowledgebase was used to check if any PTMs has been reported for a given sequence. The detectability of peptides was determined using online depositories such as PeptideAtlas [45], and our own data from discovery proteomics studies. For peptides which were not in the database or existing data, computational tools such as PREGO [46] were used. The peptide sequence uniqueness was checked with protein coverage summarizer [47], which maps candidate peptides against the whole human proteome and reports the number of proteins matched by each peptide. Only peptides that matched a single protein (uniquely mapping peptides) were selected. Four to five candidate peptides for each target protein were selected and a total of 144 crude heavy peptides were synthesized (Vivitide, Gardner, MA) and used to optimize the LC-SRM-MS assays. For the full list of the 144 peptides used for assay development see Additional Table 2A.

Optimization of SRM-MS assays for target peptides

A crude heavy peptide mixture at 3 pmol/µL for each peptide was created by mixing stock solutions of all heavy peptides (2 nmol/µL in 15% ACN), and analyzed by LC-MS/MS using HCD on a Lumos MS coupled to nanoACQUITY UPLC system to check the peptide quality and to obtain information on peptide fragmentation (i.e., transitions). Skyline was used to build the peptide library [48]. The output file from Skyline was loaded onto a Thermo Scientific TSQ Altis triple quadrupole (QQQ) MS for LC-SRM-MS analysis to obtain optimal transitions and collision energy (CE) for each peptide. This was initially performed in a background of diluted *Shewanella oneidensis* digest, and subsequently in a pooled Pap test protein digest composed of multiple Pap test protein samples from all three types of cases (HGSOC, benign, and healthy) to remove any transitions with endogenous interference. The data quality of endogenous and heavy peptides was monitored based on signal, peak shape and stability in retention time. Seven of the peptides lacked sufficient heavy and endogenous signals and were therefore excluded from the final assay. A minimum of three transitions per peptide was selected. An additional 112 peptides of 56 proteins (6 peptides comprising 5 proteins overlapped with the TMT list) from the CPTAC earlier studies [21, 49] were also included in the final assay (Additional Table 2B). The final SRM-MS assay consisted of 243 peptides comprising 81 proteins (Additional Table 3). The final list of 1470 transitions, including both endogenous and heavy peptides, was assessed using a scheduled LC-SRM-MS method with under 500 concurrent transitions per acquisition cycle, using a 12-minute retention time window.

LC-SRM-MS

For LC-SRM-MS analyses, a heavy internal standard mixture of about 120 fmol/µL for each peptide was prepared, and 2 µL of the heavy standard mixture was added to 18 µL of each patient's peptide sample containing 5 µg of peptide digests. Ninety "Mock" Pap test samples were analyzed in the SRM-MS assay: 40 HGSOC, 20 benign ovarian conditions, and 30 healthy controls (Additional Table 4). The final concentration of the samples was adjusted to 0.2 µg/µL and analyzed on a nanoACQUITY UPLC system interfaced to an TSQ Altis QQQ MS equipped with a nano-ESI source operated in positive ion-mode with the ESI voltage set to 2,100 V and a capillary temperature of 350 °C [50, 51]. The generated raw data was analyzed using Skyline 4 [48]. All LC-SRM data were imported into Skyline software, where peak boundaries were manually inspected to ensure accurate peak assignment and integration. Undetected peptides and peptides that did not correlate well with other peptides within the proteins were excluded, leaving a total of 110 peptides, corresponding to 66 proteins. The total peak area ratios of endogenous light peptides to their corresponding heavy isotope-labeled internal standards

were calculated using Skyline and used for subsequent analysis.

Statistical analysis

The raw intensity values for each peptide were mean-normalized by dividing each peptide intensity by its mean value across samples. In four cases, the heavy peptide standard was not detected, so the peptide could not be quantified. For the small number of peptides with zero values (< 1%: 0.44% at the protein level and 0.64% at the peptide level), values were imputed using one-half of the minimum observed value for the corresponding peptide. Because the normalized peptide intensities exhibited right-skewness, a log2 transformation was applied. To further correct for sample-level variation, median normalization was performed by subtracting the median value of each sample from all peptide measurements within that sample. For protein-level analysis, peptide measurements corresponding to the same protein were averaged after peptide-level mean normalization and prior to log2 transformation and sample-level median normalization. Benign cases were selected as "controls" in this study, based on their close pathology to ovarian cancer. Thus, of the 20 women with benign ovarian conditions, there were 8 cases of ovarian serous cystadenomas, 3 cases of benign ovarian fibromas, 3 cases of ovarian serous cystadenofibroma, 2 cases of fibrothecoma, serous adenofibroma, 2 cases of hemorrhagic cysts, 1 case of benign Brenner tumor with serous cystadenoma, and 1 case of a benign simple cyst. Comparisons between groups were performed using the limma method on the log-transformed intensity values with adjustment for the three known batches. The Benjamini & Hochberg adjusted p-value was used to control the false discovery rate (expected proportion of false discoveries among the rejected hypotheses). The fold change was calculated as the ratio of the geometric means. To plot the individual protein and peptide data by group, the log intensity values were adjusted for the batch effect using removeBatchEffect from the limma package in R [52].

Development of HGSOC-healthy classifiers

The multi-protein classifier comparing the data from 40 HGSOC samples versus 30 healthy samples was developed using L1-penalized logistic regression. The tuning parameter for the regression model was selected via 10-fold cross-validation to be the largest value for which the cross-validation error was within 1 standard error of the minimum cross-validation error. To account for the known batch effect, an unpenalized batch indicator was included in the model. The area under the Receiver Operating Characteristic curve (AUC) and sensitivity at 90% specificity were calculated using the predicted probabilities from the held-out cross-validation folds with confidence intervals based on the bias-corrected bootstrap case cross-validation method of Jiang et al. [53]. Following determination of the proteins included in the multi-protein classifier, stability selection was conducted using 100 bootstrap resamples [54] as a validation step to assess robustness, with proteins selected in more than 50% of resamples considered stable. In a similar manner, the HGSOC-healthy multi-peptide classifier was developed.

Evaluation of the multi-protein classifier

The HGSOC versus healthy multi-protein classifier was applied to HGSOC and healthy sample groups to compare predicted HGSOC risk profiles. For healthy and HGSOC samples, predicted probabilities of HGSOC were obtained from the held-out testing folds of the original 10-fold cross-validation procedure, ensuring out-of-sample evaluation. The same classifier was then applied, without retraining, to an independent set of benign samples to generate predicted probabilities of HGSOC. Predicted probabilities were summarized and compared across 40 HGSOC, 30 healthy, and 20 benign groups to assess whether benign samples exhibited cancer-like or healthy-like prediction patterns under the HGSOC-healthy classifier.

Results

Global proteome analysis of "Mock" Pap test samples

Forty "Mock" Pap test samples were run in four TMT 11-plex experiments for biomarker discovery. Twenty 20 HGSOC, 10 benign ovarian, and 10 healthy cases were selected (Additional Table 1) with a pooled sample as the reference channel. A total of 3105 proteins were identified in the "Mock" Pap test samples in the four TMT experiments. A total of 700 proteins were identified with complete TMT quantification across all 40 Pap test samples and showed potential differential expression when comparing HGSOC against either the healthy or benign cases. When the expression levels of these 700 proteins were compared by Eisen clustering (cosine metric) maintaining samples within their diagnostic groups, group-specific patterns were observed (Fig. 1A), with the majority of the 20 HGSOC samples showing patterns that distinguished them from the 10 benign and 10 healthy samples. When unsupervised Eisen clustering was performed using the average linkage method and cosine metric, the majority of the HGSOC samples grouped together in 3 major clusters, with most of the 10 benign ovarian samples clustering nearby (Fig. 1B). The majority of the 10 healthy samples clustered away from the HGSOC samples. The separation of HGSOC from the healthy samples was further exhibited by three-dimensional principal component analysis (Fig. 1C), where confidence (95%) ovals highlight the clustering

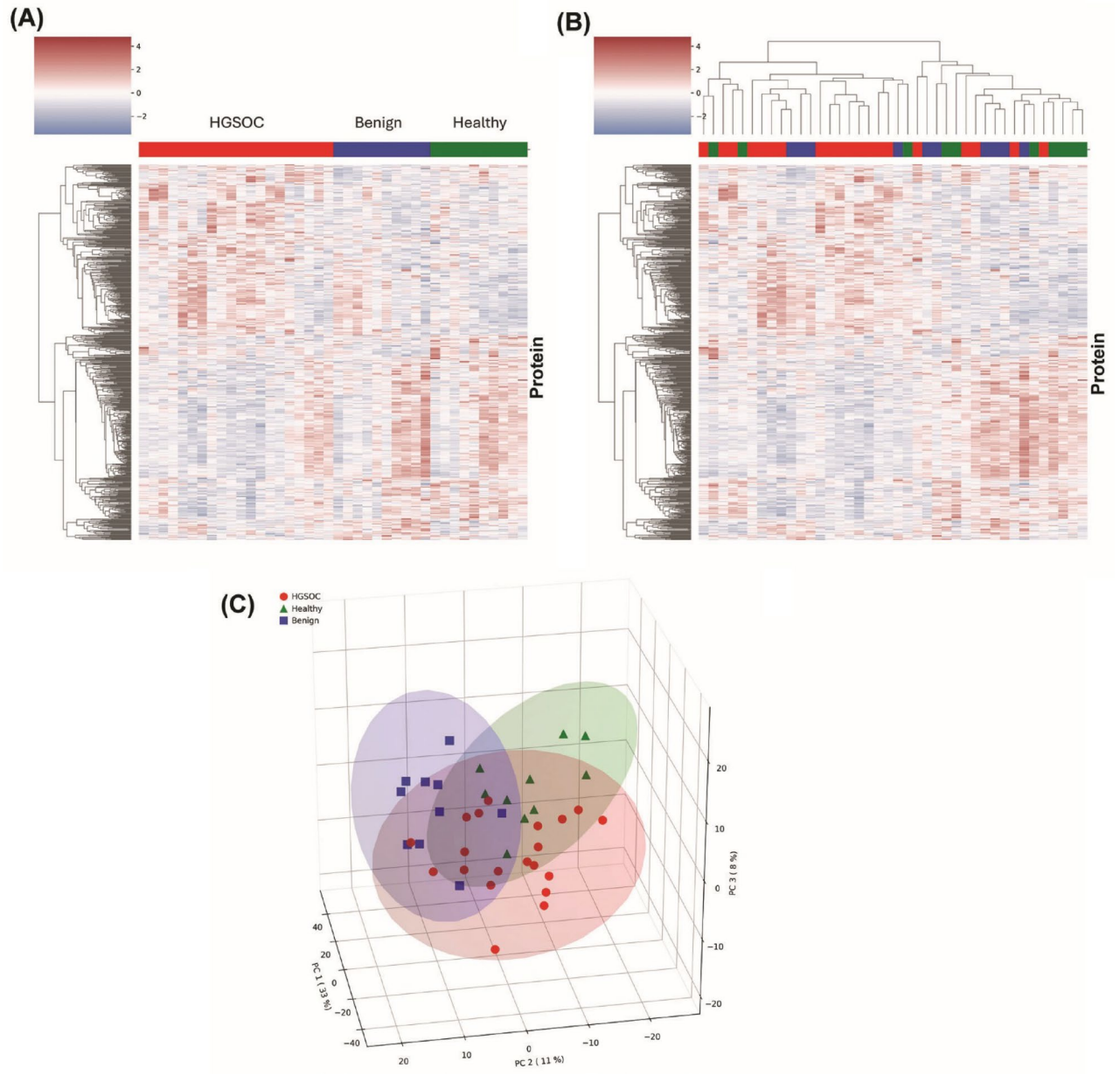

Fig. 1 *Differential expression of proteins based on diagnosis using TMT-labeled quantitative proteomics.* A total of 700 proteins that had complete TMT quantification across all 40 Pap test samples [(20 HGSOC (red), 10 benign ovarian conditions (blue), and 10 healthy (green)] and showed potential differential expression (uncorrected p-value < 0.05) when comparing HGSOC against either the healthy or benign cases. (A and B) Heatmap of protein-level clusters (rows) derived via average-linkage Eisen clustering (cosine metric). (**A**) Samples (columns) are maintained in a predefined order to highlight group-specific patterns without additional column clustering; (**B**) Heat-map of unsupervised Eisen clustering of both proteins (rows) and samples (columns). Clustering was performed using the average linkage method and cosine metric; (**C**) Three-dimensional Principal component analysis (3D-PCA). Total explained variances are shown on the axes. Confidence (95%) ovals highlight the clustering of HGSOC, benign, and healthy cohorts

of HGSOC, benign, and healthy cohorts. Thirty proteins that showed differential expression (p-value < 0.05) across HGSOC, benign, and healthy cases were selected as candidate biomarkers for SRM-MS assay development (Table 1) encompassing both over- and under-expressed proteins in the HGSOC cases relative to the benign and healthy (Fig. 1A). The abundance distributions of these 30 selected proteins across the three diagnostic groups are shown as boxplots in Additional Fig. 1.

Development of a targeted SRM-MS assay

We developed a multiplex SRM-MS assay as a high throughput method for targeted validation of the 30 biomarkers identified in our TMT experiments. We selected a total of 144 target peptide sequences from the 30

Table 1 Proteins from TMT experiments differentially expressed between HGSOC cases and healthy controls and benign conditions

UniProtID	Entry name	Protein name
P07741	APT	Adenine phosphoribosyltransferase
P19652	A1AG2	Alpha-1-acid glycoprotein 2 (Orosomucoid-2)
P04003	C4BPA	C4b-binding protein alpha chain
P08603	CFAH	Complement factor H
Q16531	DDB1	DNA damage-binding protein 1
Q9C0E8	LNP	Endoplasmic reticulum junction formation protein lunapark
Q8N9N8	EIF1A	Eukaryotic translation initiation factor 1 A domain-containing protein
Q96CS3	FAF2	FAS-associated factor 2
P23142	FBLN1	Fibulin-1
P11216	PYGB	Glycogen phosphorylase
P00738	HPT	Haptoglobin (Zonulin)
P00739	HPTR	Haptoglobin-related protein
P07910	HNRPC	Heterogeneous nuclear ribonucleoproteins C1/C2
P26583	HMGB2	High mobility group protein B2
P16402	H13	Histone H1.3
Q12906	ILF3	Interleukin enhancer-binding factor 3
Q9Y608	LRRF2	Leucine-rich repeat flightless-interacting protein 2
P55083	MFAP4	Microfibril-associated glycoprotein 4
O00533	NCHL1	Neural cell adhesion molecule L1-like protein
Q8NDX1	PSD4	PH and Sect. 7 domain-containing protein 4
Q13492	PICAL	Phosphatidylinositol-binding clathrin assembly protein
O95865	DDAH2	Putative hydrolase DDAH2
Q9NWH9	SLTM	SAFB-like transcription modulator
P82979	SARNP	SAP domain-containing ribonucleoprotein
P50452	SPB8	Serpin B8
Q01082	SPTB2	Spectrin beta chain, non-erythrocytic 1
Q15459	SF3A1	Splicing factor 3 A subunit 1
Q16563	SYPL1	Synaptophysin-like protein 1
P48643	TCPE	T-complex protein 1 subunit epsilon
O60635	TSN1	Tetraspanin-1

candidate biomarker proteins to use for assay development (Additional Table 2A). A minimum of three transitions from 3 peptides per protein were retained for all target proteins. The final list of transitions for both endogenous and heavy peptides was assessed using a scheduled LC-SRM-MS method under 500 transitions per event. The multiplexed LC-SRM-MS assay developed from the 30 TMT protein biomarkers was combined with 112 peptides from 56 ovarian cancer candidate proteins developed by the CPTAC (Additional Table 2B), based on differential expression between HGSOC and normal fallopian tube epithelium [21]. Following optimization, the resulting SRM-MS assay was used to quantify a total of 243 peptides from 81 ovarian cancer candidate biomarker proteins (Additional Table 3).

Targeted proteomics measurements using SRM-MS

We used our multiplex SRM-MS assay to quantify the expression of 243 peptides corresponding to 81 proteins in 90 "Mock" Pap test samples (40 HGSOC; 20 benign; 30 healthy) (Additional Table 4). The expression levels of 110 peptides corresponding to 66 proteins were quantified in the 90 Pap test samples relative to the expression of the corresponding heavy isotope peptide.

The SRM-MS data for the 90 Pap test samples was analyzed by limma to determine which peptides/proteins were differentially expressed in HGSOC versus healthy controls. Sixty-one of the 110 peptides were present at significantly (adjusted $p<0.05$) higher levels in the HGSOC samples compared to the healthy controls (Fig. 2A; Table 2). The levels of 8 peptides (corresponding to proteins A1AG2, CRP, and HPT) were 8–12 times higher in the HGSOC samples compared to the healthy controls.

At the protein level, 36 of the 66 proteins were significantly (adjusted $p<0.05$) higher in HGSOC samples compared to the healthy controls (Fig. 2B; Table 3). The levels of the proteins A1AG2, CRP, and HPT were 6–9 times higher in the HGSOC samples than in the healthy controls, confirming similar behavior at the peptide and protein levels. No peptides or proteins were present at significantly (adjusted $p<0.05$) lower levels in the HGSOC samples compared to the healthy controls (Fig. 2A and B). Proteins and peptides identified as upregulated in the SRM analysis showed consistent upregulation in the TMT dataset, supporting the robustness of these findings across platforms.

In contrast, when the peptide levels in the HGSOC samples were compared to the samples from benign cases, none of the 110 peptides in the SRM-MS assay were present at significantly higher levels (adjusted $p<0.05$) in the HGSOC cases. Although the levels of several of the peptides measured in the HGSOC Pap tests were up to 5 times higher than in the benign controls, the adjusted p-values were not significant, potentially due to a smaller number of benign samples. Only one of the 66 proteins (TSN1) was present at significantly (adjusted $p=0.040$) higher levels in the HGSOC samples compared to the benign controls (Additional Fig. 2). The levels of several other proteins present in the HGSOC Pap tests were over 3 times higher than in the benign controls (CFAH, IBP3, FBLN1) and one protein (ILF3) was 2.3 times lower in HGSOC compared to benign Pap tests although none were statistically significant; again this is potentially due to a smaller number of benign samples.

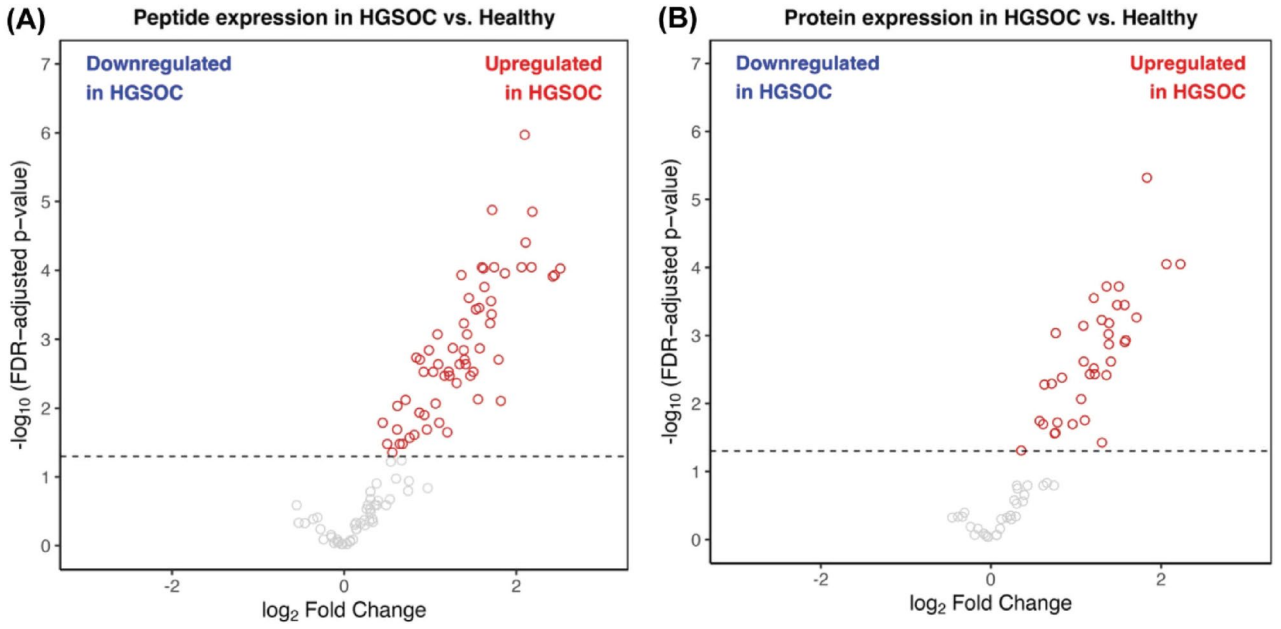

Fig. 2 *Quantification of candidate protein biomarkers by SRM-MS*. Results of the SRM-MS targeted assay for 110 peptides representing 66 proteins are shown as volcano plots of the adjusted p-values vs. the fold change. Points above the dashed line correspond to adjusted p-values < 0.05 (red circles). (**A**) Sixty-one peptides were significantly upregulated in HGSOC vs. healthy control samples. (**B**) 36 proteins were significantly upregulated in HGSOC vs. healthy control samples

The expression levels of each of the peptides and proteins were graphed as boxplots to visualize their expression in HGSOC, benign conditions, and healthy controls. An example of the expression levels of one of the proteins (FBLN1) and the two peptides that were used in the SRM-MS assay is shown in Fig. 3. In this case, the level of expression of the peptides (Fig. 3A and B) and protein (Fig. 3C) showed an upward trend from healthy controls to benign ovarian conditions to HGSOC.

However, the boxplots for the majority of the 61 peptides and 36 proteins showed a trend whereby the level of expression in the HGSOC and benign samples were elevated compared to the healthy cases, as shown for the protein A1AG2 and its peptides (Fig. 4).

Development of a multi-protein classifier

The classification performance distinguishing between healthy controls and HGSOC for each protein individually was summarized using the AUC and sensitivity at 90% specificity with bootstrap confidence intervals. The sensitivity was calculated at 90% specificity vs. a higher specificity due to the limited number of healthy control samples. Sixteen proteins had an estimated AUC of > 0.7 (Table 4).

We also considered whether the classification performance could be improved by integrating results from multiple proteins into a single classifier, which we refer to as a multi-protein classifier.

There were six proteins included in the multi-protein classifier (A1AG2, CRP, HPT, ICAL, PYGB, and C4BPA) with all proteins having positive weights, indicating that higher levels are associated with a higher likelihood of the sample being ovarian cancer (Fig. 5A). The predicted HGSOC risk score is equal to:

$$\begin{aligned} &expit(2.644 + 0.301 \times ICAL + 0.287 \\ &\times A1AG2 + 0.265 \times C4BPA \\ &+ 0.243 \times PYGB + 0.185 \times HPT \\ &+ 0.113 \times CRP) \end{aligned}$$

where $expit(x) = e^x/(1 + e^x)$. While this risk score would typically equal the estimated probability of HGSOC, the intercept estimate is biased given the case-control study design and, thus, we call it more generally a risk score. The AUC for the multi-protein classifier was 0.880 (95% CI: 0.738–0.989) and the sensitivity at 90% specificity was 0.800 (95% CI: 0.470-1.00), higher than any of the individual proteins (Table 4; six proteins in bold).

We evaluated the robustness of protein selection using a bootstrap-based stability analysis, repeating the L1-penalized logistic regression 100 times on resampled datasets. Among the top-performing individual proteins, only the six proteins included in the multi-protein classifier were selected in more than 50% of bootstrap iterations, indicating that the multi-protein classifier preferentially incorporated proteins with

Table 2 Sixty-one peptides with significantly higher levels of expression when comparing HGSOC versus healthy controls

Protein name	Peptide sequence	Fold change	*p*-value	Adjusted *p*-value
1433B	AVTEQGHELSNEER	5.539	< 0.001	< 0.001
A1AG2	EQLGEFYEALDCLCIPR	5.574	< 0.001	< 0.001
A1AG2	EHVAHLLFLR	6.480	< 0.001	< 0.001
A1AG2	NWGLSFYADKPETTK	8.151	< 0.001	< 0.001
A1AG2	SDVMYTDWK	8.904	< 0.001	< 0.001
A1BG	ATWSGAVLAGR	3.689	0.002	0.004
A1BG	CLAPLEGAR	4.334	0.001	0.003
A2GL	TLDLGENQLETLPPDLLR	4.258	< 0.001	< 0.001
A2GL	DLLLPQPDLR	4.955	< 0.001	< 0.001
ANXA1	GTDVNVFNTILTTR	3.307	0.011	0.022
APOE	AATVGSLAGQPLQER	2.886	0.004	0.009
APT	IDYIAGLDSR	1.903	0.018	0.033
APT	DALEPGQR	2.673	< 0.001	0.001
APT	AAIGLLAR	2.806	0.001	0.003
C4BPA	EDVYVVGTVLR	3.899	< 0.001	< 0.001
CD47	STVPTDFSSAK	2.976	0.001	0.002
CERU	ALYLQYTDETFR	4.171	< 0.001	0.001
CERU	DIASGLIGPLIICK	5.031	< 0.001	< 0.001
CFAH	CFEGFGIDGPAIAK	2.252	0.012	0.024
CFAH	LSYTCEGGFR	4.620	< 0.001	< 0.001
CFAH	WQSIPLCVEK	5.705	< 0.001	< 0.001
CO3	TIYTPGSTVLYR	4.478	0.001	0.003
CO3	VHQYFNVELIQPGAVK	5.446	< 0.001	0.001
CO4B	ASSFLGEK	4.829	< 0.001	0.001
CRP	GYSIFSYATK	7.861	< 0.001	< 0.001
CRP	ESDTSYVSLK	8.818	< 0.001	< 0.001
DAG1	VVENGALLSWK	4.104	0.001	0.002
DDAH2	GAEIVADTFR	1.747	0.024	0.044
DDAH2	GGGDLPNSQEALQK	2.513	0.001	0.003
DDB1	YLAIAPPIIK	1.647	0.018	0.033
DDB1	IVVFQYSDGK	2.037	0.003	0.008
EIF1A	TPGNNLHEVETAQGQR	1.853	0.004	0.009
FBLN1	TGYYFDGISR	3.525	< 0.001	0.001
FBLN1	GYHLNEEGTR	3.816	0.001	0.002
FBLN1	EFTRPEEIIFLR	6.012	0.001	0.002
HPT	VVLHPNYSQVDIGLIK	8.243	< 0.001	< 0.001
HPT	DYAEVGR	11.310	< 0.001	< 0.001
HPT	VGYVSGWGR	11.541	< 0.001	< 0.001
HPT	VTSIQDWVQK	12.322	< 0.001	< 0.001
IBP3	ALAQCAPPPAVCAELVR	3.012	0.008	0.016
ICAL	QAEPELDLR	2.405	0.001	0.002
ICAL	AAAPAPVSEAVCR	4.732	0.003	0.007
IL18	SDIIFFQR	4.010	< 0.001	0.001
KNG1	TVGSDTFYSFK	4.795	< 0.001	< 0.001
KNG1	YFIDFVAR	5.505	< 0.001	< 0.001
LEG1	FNAHGDANTIVCNSK	2.392	0.005	0.012
LEG3	IQVLVEPDHFK	2.307	< 0.001	0.002
LEG3	IALDFQR	4.048	0.001	0.002
LG3BP	SDLAVPSELALLK	3.394	0.001	0.003
LUM	FNALQYLR	3.357	0.001	0.003
MSLN	EIDESLIFYK	2.134	0.014	0.027
MUC16	VLQGLLGPIFK	2.610	0.010	0.020

Table 2 (continued)

Protein name	Peptide sequence	Fold change	*p*-value	Adjusted *p*-value
NCHL1	VIAVNEVGR	1.846	0.010	0.020
PIGR	ADEGWYWCGVK	6.172	0.003	0.008
PYGB	LVTSIGDVVNHDPVVGDR	2.951	< 0.001	0.001
S10A4	ELPSFLGK	1.968	0.018	0.033
S10A4	ALDVMVSTFHK	2.533	0.006	0.013
SYPL1	TVTATFGYPFR	3.200	0.001	0.003
TCPE	DVDFELIK	1.562	0.008	0.016
TCPE	IADGYEQAAR	5.098	< 0.001	< 0.001
TSN1	VEGCFNQLLYDIR	4.017	< 0.001	0.001

Names of the proteins with peptides having adjusted p-values <0.05. The fold changes are listed along with the p-values and adjusted p-values

greater selection stability. The AUC values for the 15 peptides that were used in 6 proteins that comprise the multi-protein classifier are shown in Table 5.

Predicted risk of HGSOC

Predicted probabilities from the HGSOC-healthy multi-protein classifier demonstrated clear separation between HGSOC and healthy samples (Fig. 5B). HGSOC samples exhibited consistently high predicted probabilities, with most values well above the 0.5 decision threshold, whereas healthy samples showed substantially lower probabilities concentrated below this threshold. When the classifier was applied to benign samples, 19 of 20 benign samples were classified as HGSOC based on the 0.5 probability threshold, while only one benign sample was classified as healthy (Fig. 5B). Compared with healthy controls, benign samples displayed higher median predicted probabilities and greater overlap with HGSOC, indicating cancer-like prediction profiles under the HGSOC-healthy classifier.

Development of a multi-peptide classifier

In bottom-up proteomics approaches, peptide sequences resulting from trypsin digestion are detected and quantified by the mass spectrometer. These abundances are then used to infer the abundance of the protein using the canonical sequences. Therefore, we hypothesized that individual peptide sequences could provide additional insight. Consequently, we also performed peptide-level classification analysis and determined the AUC values for each of the 110 peptides quantified by SRM based on the 40 HGSOC cases and the 30 healthy cases. In total, 29 peptides (corresponding to 18 proteins) had an estimated AUC of > 0.7; the 15 peptides with the highest AUCs are shown in Additional Table 5. There were four peptides included in a multi-*peptide* classifier, with all peptides having positive weights, indicating that higher levels are associated with a higher likelihood that the sample is ovarian cancer (Fig. 5C). In a similar manner, the predicted HGSOC risk score is equal to:

$$
\begin{aligned}
&expit(1.640 + 0.378 \times NWGLSFYADKPETTK[A1AG2] + 0.070 \\
&\times ESDTSYVSLK[CRP] + 0.070 \\
&\times IADGYEQAAR[TCPE] + 0.024 \\
&\times VVLHPNYSQVDIGLIK[HPT])
\end{aligned}
$$

where $expit\,(x) = e^x/(1 + e^x)$. The AUC for the multi-peptide classifier was 0.825 (95% CI: 0.724–0.913), and the sensitivity at 90% specificity was 0.625 (95% CI: 0.400–0.850). We evaluated the robustness of peptide selection using a bootstrap-based stability analysis, repeating the L1-penalized logistic regression 100 times on resampled datasets. Among the top-performing individual peptides, three of the four peptides included in the multi-peptide classifier were selected in more than 50% of bootstrap iterations. These four peptides were derived from the proteins A1AG2, CRP, HPT, and TCPE; only TCPE was not one of the proteins in the multi-protein classifier above. All four peptides were confidently identified in the SRM assay based on co-elution of endogenous and heavy isotope-labeled internal standards and consistent transition ion-ratio patterns, as shown in Additional Fig. 3.

Discussion

We used SRM-MS, a targeted quantification approach, to quantify the expression of specific proteins in the cervical effluent from Pap tests. By comparing the AUC values of the proteins for their ability to distinguish HGSOC from healthy controls, we developed a multi-protein classifier comprised of six proteins (A1AG2, CRP, HPT, ICAL, PYGB, and C4BPA) with an AUC of 0.880; higher than any of the individual proteins. These data demonstrate proof-of-principle that the multi-protein classifier can differentiate HGSOC from healthy controls. Pap tests are a routine, readily accepted method for cervical cancer screening. Thus, by combining this testing into routine Pap tests, it could prove to be a cost-effective strategy that would be agreeable to patients and clinicians, and thereby may increase uptake.

Three of the six proteins that comprised the multi-protein classifier [A1AG2 (orosomucoid2), CRP (C-reactive

Table 3 Thirty-six proteins with significantly higher levels of expression between HGSOC versus healthy controls

Protein name	Fold change	*p*-value	Adjusted *p*-value
A1AG2	6.261	< 0.001	< 0.001
A1BG	3.891	0.001	0.004
A2GL	4.503	< 0.001	< 0.001
APOE	2.886	0.004	0.009
APT	2.304	0.002	0.004
C4BPA	3.899	< 0.001	< 0.001
CD47	2.976	0.001	0.002
CERU	4.418	< 0.001	< 0.001
CFAH	3.684	< 0.001	0.001
CO3	4.883	< 0.001	0.001
CO4B	4.829	< 0.001	0.001
CRP	7.876	< 0.001	< 0.001
DAG1	4.104	0.001	0.002
DDAH2	2.044	0.002	0.005
DDB1	1.774	0.008	0.018
EIF1A	1.429	0.027	0.049
FBLN1	3.994	< 0.001	0.001
HPT	9.285	< 0.001	< 0.001
IBP3	3.012	0.007	0.018
ICAL	3.354	< 0.001	< 0.001
IL18	4.010	< 0.001	0.001
KNG1	4.820	< 0.001	< 0.001
LEG1	2.114	0.014	0.027
LEG3	2.963	< 0.001	0.001
LG3BP	3.394	0.001	0.004
LUM	3.357	0.001	0.003
MSLN	2.134	0.013	0.027
MUC16	2.610	0.010	0.020
NCHL1	1.846	0.009	0.020
PIGR	3.696	0.020	0.038
PYGB	1.874	0.002	0.005
S10A4	2.185	0.009	0.019
SYPL1	3.200	0.001	0.004
TCPE	2.141	< 0.001	0.001
TSN1*	4.017	< 0.001	0.001
X1433B	5.539	< 0.001	0.001

Proteins with adjusted p-values <0.05. Shown are fold changes, p-values and adjusted p-values. Proteins shown in bold are the six proteins in the multi-protein classifier

*Also significantly different when comparing HGSOC versus benign cases

protein), and HPT (haptoglobin)] had the highest individual AUC values and the highest fold change between HGSOC and healthy controls. Although their levels of expression were significantly different between HGSOC and healthy control, they were not significantly different between HGSOC and benign ovarian conditions. All three proteins are acute phase reactants that are components of blood and considered biomarkers of inflammation. The elevated levels of these proteins in Pap test samples likely reflects the presence of serum-derived proteins in cervical mucus, or possibly they were derived from the ascites fluid, which has a similar composition to serum [17]. We previously identified HPT and A1AG2 as serum biomarkers for ovarian cancer by two other MS-based proteomic techniques, iTRAQ and DIGE [55, 56]. Similarly, serum CRP levels are also known to be elevated in ovarian cancer; high levels of CRP are associated with tumor size and stage [57] and a poor prognosis [reviewed [58]]. In addition, in pre-diagnostic serum, women with elevated CRP levels are at risk for developing breast and ovarian cancer [59–61]. Furthermore, CRP is an approved biomarker for inflammation and cardiovascular disease [62]. Although CRP and other biomarkers might not be ovarian cancer-specific, they may still have a clinical application whereby women would be triaged for further gynecologic work-ups, including potentially CA125 blood tests and transvaginal ultrasounds.

The fourth protein in the multi-protein classifier, C4BPA (Complement component 4 binding protein alpha), is also secreted into the blood. It may play a role in the immune response by negative regulation of complement C4 [63]. In an analysis of data from The Cancer Genome Atlas (TCGA), C4BPA expression in breast cancer was positively correlated with expression of several immune cell marker genes (B-cells, neutrophils, CD8 T-cells) and was a favorable prognostic marker [64]. Similarly, using RNA-seq data from the ovarian cancer TCGA, C4BPA was identified as part of a 7 gene signature to predict prognosis [65]; high C4BPA expression was associated with a favorable prognosis. The overexpression of C4BPA protein in ovarian cancer tissues compared to normal ovary was also validated by immunohistochemistry (IHC) [65]. In pancreatic adenocarcinoma (PDA), C4BPA was identified as a serum biomarker using TMT-labeled serum from matched pre- and post-surgery specimens. Subsequent biomarker validation of C4BPA by ELISA showed a significant increase in C4BPA levels in PDA compared to both pancreatitis and healthy controls [66]. In another study, fully sialylated C4BPA was identified as a serum biomarker in ovarian cancer using glyco-peptide screening [67]. In both studies [66, 67], C4BPA was elevated in early-stage disease, however, the ovarian cancer study focused on ovarian clear cell carcinoma, although they found elevated sialylated C4BPA in serum from HGSOC patients as well [67].

The fifth protein in the multi-protein classifier, PYGB (Glycogen phosphorylase, brain isoform) plays an important role in glycogen metabolism and ATP production. It is over-expressed in the tissues of multiple cancer types (gastric, hepatocellular carcinoma, non-small cell lung, and prostate) including ovarian cancer, and is associated with advanced stage disease and poor prognosis [68] (reviewed by [69]). In colorectal cancer, PYGB expression correlated with increased dysplasia, and was found prior to p53 overexpression, suggesting it may be an early marker of carcinogenesis [70].

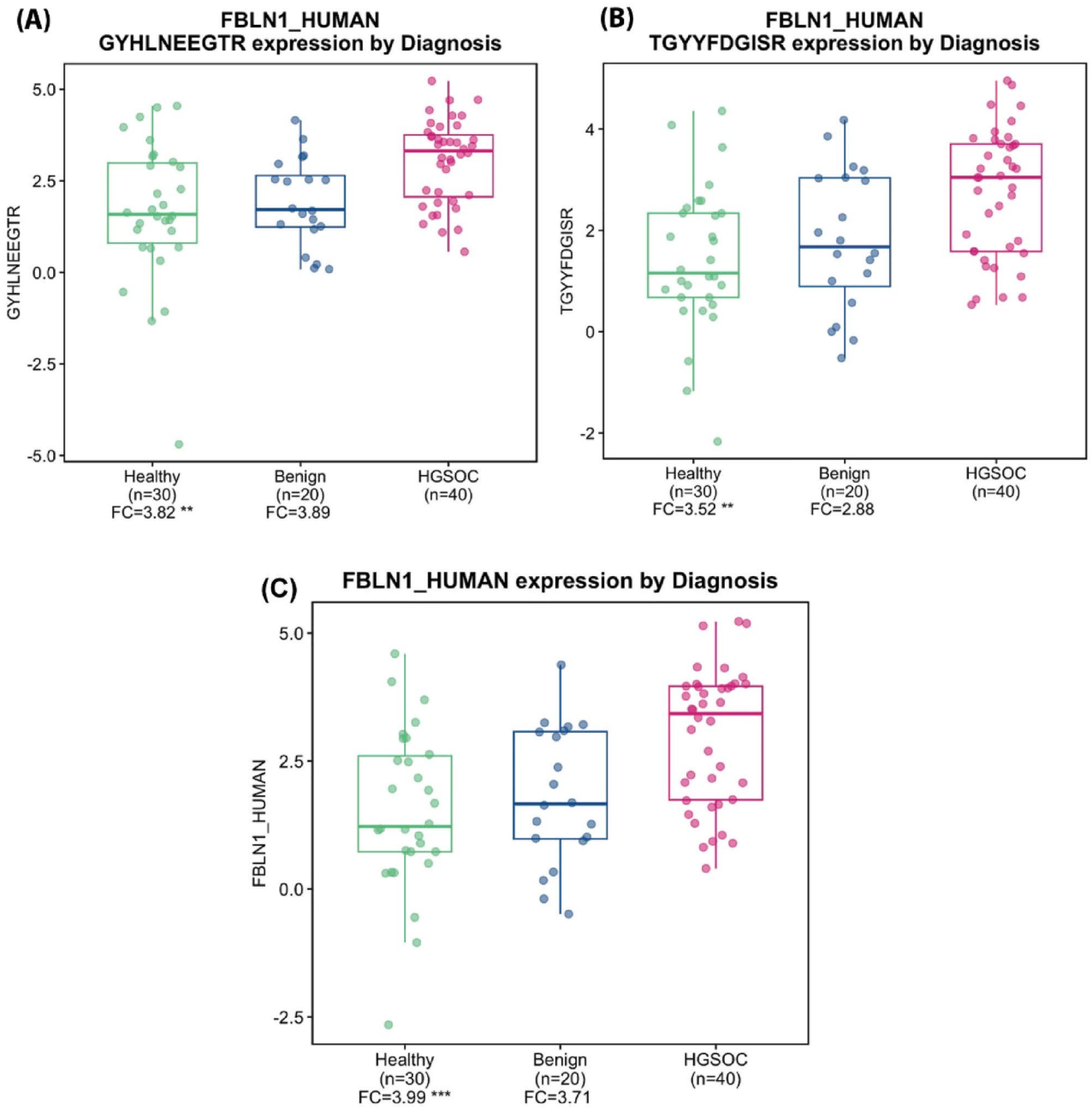

Fig. 3 *Quantification of the candidate ovarian cancer biomarker FBLN1 by SRM-MS in Pap tests*. Normalized values from each sample in the SRM-MS data were plotted with the median and 25th and 75th percentiles for the healthy controls (green), the benign ovarian conditions (blue) and the HGSOC patients (red). The fold change (FC) for comparison between HGSOC and healthy or benign samples is shown below each category, stars correspond to statistical significance (***: adjusted $p < 0.001$, **: adjusted $p < 0.01$). Plots for the protein FBLN1: (**A**) peptide GYHLNEEGTR, (**B**) peptide TGYYFDGISR, and (**C**) protein FBLN1. The expression values for these peptides and protein FBLN1 showed a trend for low levels in healthy controls, with higher levels in benign, and significantly higher levels in HGSOC compared to healthy controls

The final protein in the multi-protein classifier, ICAL [Calpastatin, Calpain-10 (also known as Sperm BS-17 component)], is a specific inhibitor of the proteinase calpain. The calpain-calpastatin system plays a role in cancer progression by regulation of cell adhesion, migration, proliferation and apoptosis [reviewed by [71]]. According to the Human Protein Atlas, calpastatin is a "Cancer Related Protein", although it has not been recognized as a "Candidate Cancer Biomarker" [72]. ICAL is highly expressed in over half of ovarian cancer tumors, and high expression was associated with the HGSOC subtype and better overall survival [73]. A recent proteomic analysis of saliva from patients with newly diagnosed or recurrent glioblastoma found ten proteins, including ICAL,

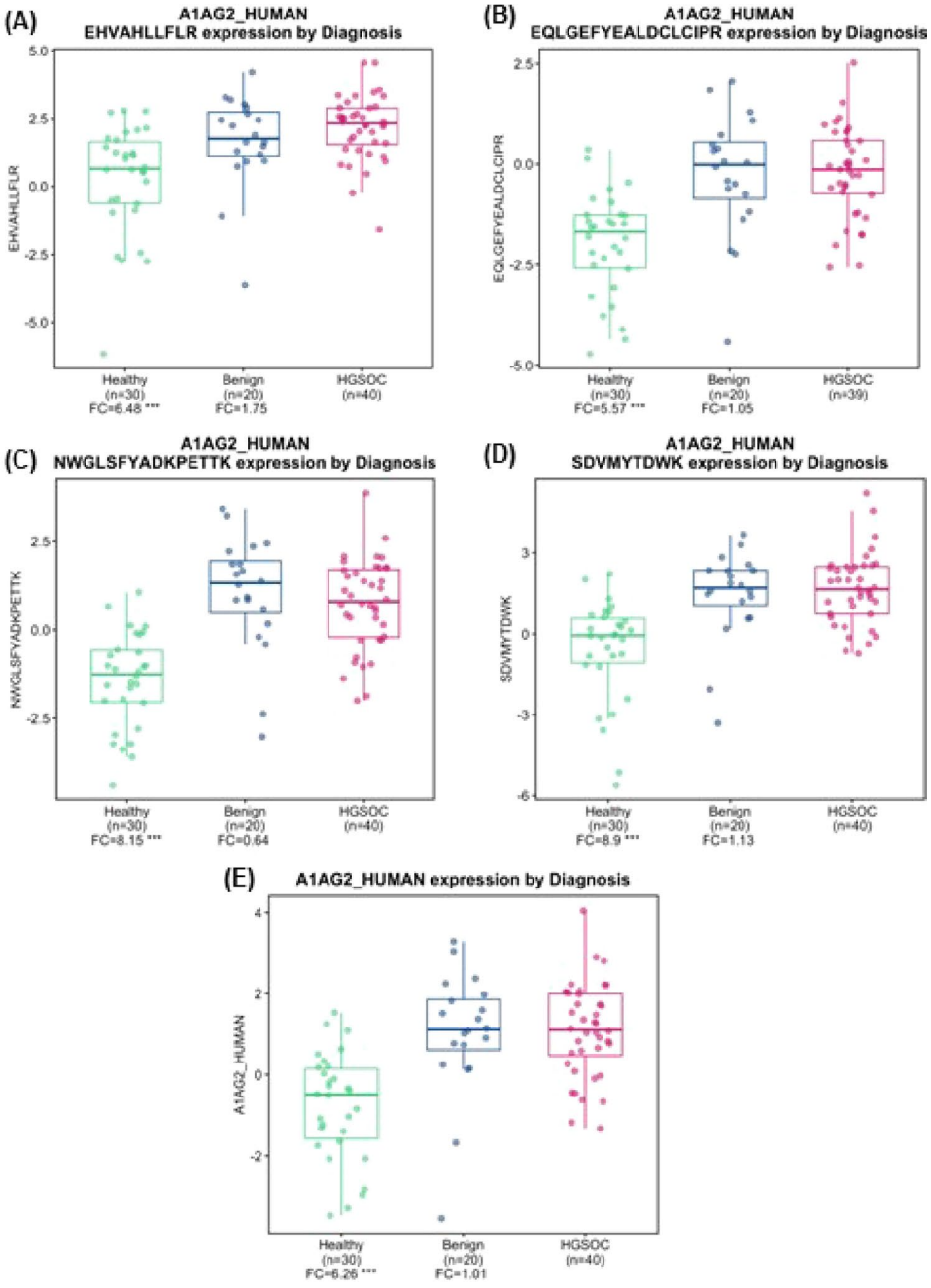

Fig. 4 *Quantification of A1AG2, a candidate ovarian cancer biomarker, levels by SRM-MS in "Mock" Pap tests.* Normalized values from each sample in the SRM-MS data were plotted with the median and 25th and 75th percentiles for the healthy controls (green), the benign ovarian conditions (blue) and the HGSOC patients (red). The fold change (FC) for comparison between HGSOC and healthy or benign samples is shown below each category, stars correspond to statistical significance (***: adjusted $p < 0.001$, **: adjusted $p < 0.01$). Plots for: (**A**) peptide EHVAHLLFLR, (**B**) peptide EQLGEFYEALDCLCIPR, (**C**) peptide NWGLSFYADKPETTK (**D**) peptide SDVMYTDWK and (**E**) protein A1AG2. The expression values for these peptides and protein A1AG2 were significantly higher in HGSOC compared to the healthy controls

that were expressed only in saliva prior to surgery and not in post-treatment saliva samples [74] suggesting that salivary calpastatin levels could be a biomarker for disease recurrence. Moreover, in glioblastoma, high levels of ICAL is associated with a poor prognosis [72, 75].

In our SRM analysis, we found many more proteins with differential expression when comparing the HGSOC samples to healthy controls than when compared to the benign samples. This cannot be attributed to the way that the candidate biomarkers were selected, since we selected proteins from the TMT data that were differentially

Table 4 AUC values for the 16 proteins with AUC > 0.7

Protein name	AUC (95% CI)	AUC rank	Sens. at 90% spec. (95% CI)	Sens. at 90% spec. rank	Proportion of selection (%)
A1AG2_HUMAN	**0.836 (0.737, 0.920)**	**1**	**0.600 (0.225, 0.800)**	**2**	**77**
CRP_HUMAN	**0.821 (0.718, 0.912)**	**2**	**0.525 (0.275, 0.800)**	**4**	**69**
HPT_HUMAN	**0.800 (0.690, 0.894)**	**3**	**0.500 (0.275, 0.800)**	**5**	**84**
X1433B_HUMAN	0.776 (0.660, 0.879)	4	0.625 (0.200, 0.800)	1	22
TCPE_HUMAN	0.761 (0.635, 0.875)	5	0.275 (0.025, 0.800)	20	27
LEG3_HUMAN	0.760 (0.637, 0.873)	6	0.300 (0.100, 0.800)	15	6
A2GL_HUMAN	0.753 (0.632, 0.863)	7	0.375 (0.075, 0.650)	11	27
SYPL1_HUMAN	0.745 (0.616, 0.865)	8	0.250 (0.000, 0.700)	22	9
ICAL_HUMAN	**0.740 (0.617, 0.848)**	**9**	**0.550 (0.100, 0.700)**	**3**	**59**
APT_HUMAN	0.736 (0.610, 0.850)	10	0.450 (0.075, 0.700)	6	2
IL18_HUMAN	0.733 (0.609, 0.845)	11	0.400 (0.025, 0.675)	9	17
KNG1_HUMAN	0.731 (0.612, 0.840)	12	0.400 (0.225, 0.600)	9	0
CERU_HUMAN	0.728 (0.604, 0.838)	13	0.450 (0.200, 0.600)	6	1
PYGB_HUMAN	**0.727 (0.605, 0.840)**	**14**	**0.425 (0.169, 0.675)**	**8**	**55**
C4BPA_HUMAN	**0.716 (0.591, 0.828)**	**15**	**0.325 (0.200, 0.625)**	**13**	**77**
DDB1_HUMAN	0.713 (0.581, 0.836)	16	0.200 (0.000, 0.525)	31	5

The six proteins in the multi-protein classifier from Fig. 5A are in bold

expressed between HGSOC and healthy and benign controls. Many serum biomarkers for ovarian cancer may also be elevated in subjects with benign ovarian conditions [76]. In our SRM analysis only one protein, TSN1 (tetraspanin 1) was significantly elevated in the HGSOC samples compared to benign. Tetraspanins are components of the membrane that form microdomains involved in receptor clustering and signal transduction and participate in cell adhesion, migration, proliferation and extracellular vesicle formation [reviewed by [77]]. TSN1 is expressed in ovarian cancer tissues [78] and overexpression of TSN1 in benign endometriosis promoted cell proliferation and invasion, potentially leading to ovarian clear cell carcinoma [79]. In a study to identify plasma protein biomarkers for ovarian cancer detection, TSN1 was part of a 7-protein signature that could distinguish early-stage ovarian cancer from benign ovarian disease [80], supporting our finding that TSN1 is a biomarker of ovarian cancer.

Many of the proteins that were identified as candidate protein biomarkers for ovarian cancer in the discovery phase of this study (i.e. the TMT experiments), could be classified as highly abundant serum proteins. However, those proteins still exhibited differential levels of expression in the SRM-MS assay when comparing HGSOC to healthy samples. Some of these proteins have been suggested to serve as possible ovarian cancer biomarkers in the literature, e.g. HPT and LRG [55, 56, 81]. Of interest, many of the "classical" ovarian cancer biomarkers [e.g. CA125 (MUC16) and HE4] that had been discovered in tissue or serum were also included in our SRM-MS assay as heavy-labeled peptides from earlier CPTAC studies [21, 49]. However, these proteins were not sufficiently robust to be included in our multi-protein classifier.

Other groups have developed a technique to rinse the inside of the uterus, then analyze the DNA of the cells from the washings for mutations [82]. Recently, a comprehensive proteomic analysis of microvesicles isolated from uterine lavage samples was used to create and validate a multi-protein classifier for ovarian cancer detection with 70% sensitivity and 76.2% specificity [83]. This methodology may prove to be too cumbersome to perform on an annual basis for ovarian cancer screening. Given the relative ease of collecting a Pap test or vaginal swab in comparison to a uterine lavage, the Pap test may be accepted more readily in clinics to screen for both ovarian and cervical cancers.

Because of the inherent variability in the collection of Pap test samples, the amount of cervical effluent present in the SurePath™ vials was determined by measuring the protein concentration following acetone precipitation. On average, each Pap test SurePath™ fixative sample yielded ~ 140 µg of protein, excluding 5 Pap tests which contained > 500 µg of protein, most likely due to blood contamination which was visible in the fixative. To control for variability in sample collection, we used the same amount of protein (50 µg) from each sample when performing the TMT experiments, and the same amount of protein (5 µg) from each sample when performing the SRM-MS assay. In future applications, it may be necessary to include a panel of "housekeeping" proteins that will control for variability in sample collection.

A limitation of this preliminary study was the inability of our multi-protein classifier to distinguish HGSOC from benign ovarian conditions. The two major drivers of this limitation are the size and selection of the clinical cohort. First, our selection of "Mock" Pap tests from 20 women with benign ovarian serous pathology were

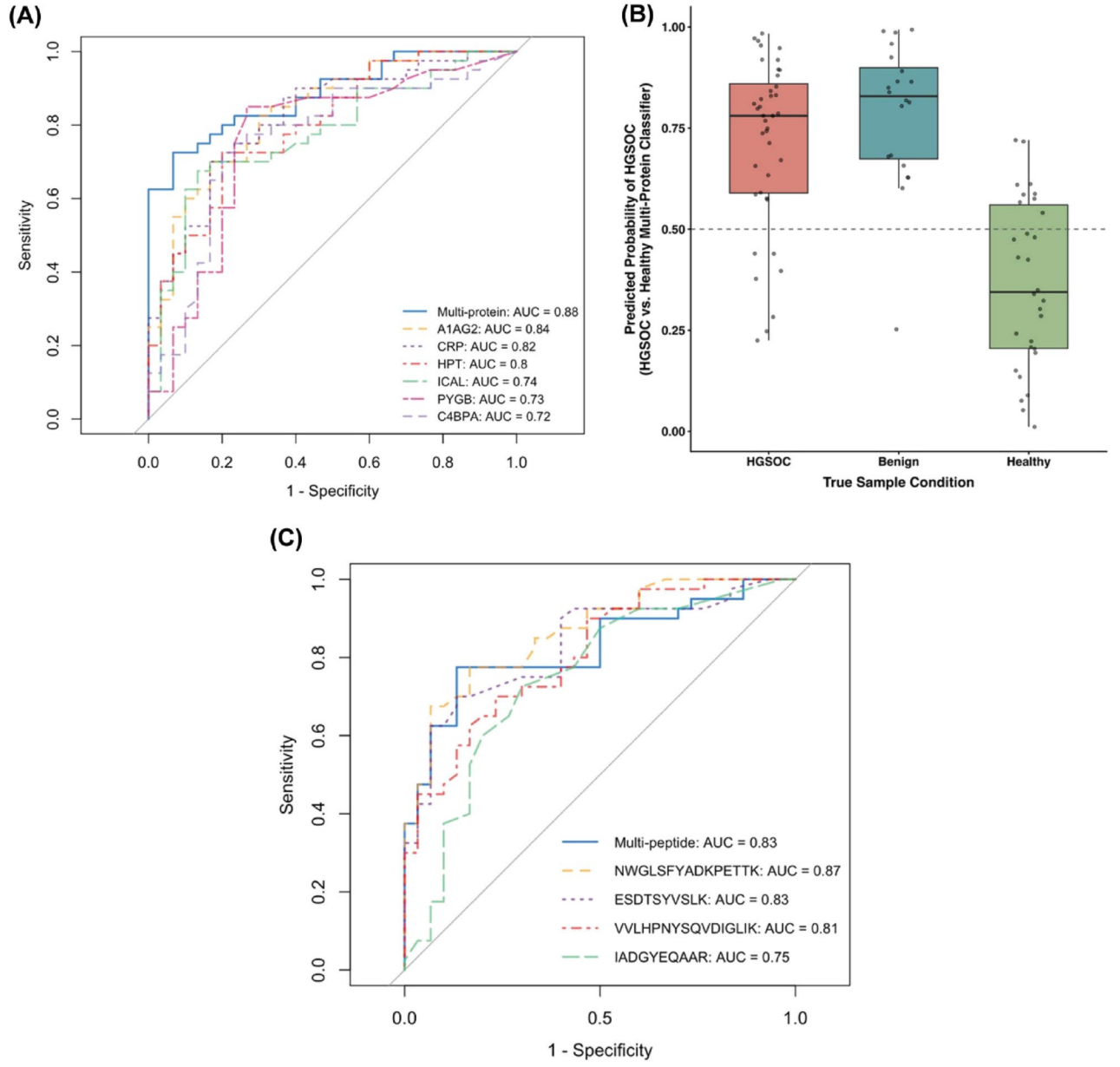

Fig. 5 *Multi-protein and multi-peptide classifiers for HGSOC vs. healthy Pap tests.* (**A**) The ROC curves for the HGSOC-healthy multi-protein classifier and each of the six individual proteins included in the multi-protein classifier (A1AG2, CRP, HPT, ICAL, PYGB, and C4BPA) as described in Table 4. (**B**) Predicted probability of HGSOC from the HGSOC-healthy multi-protein classifier across true sample conditions. Boxplots summarize cross-validated predictions for HGSOC and healthy samples and projected predictions for independent benign samples. The dashed line indicates the 0.5 classification threshold. (**C**) The ROC curves for the multi-peptide classifier and each of the four individual peptides included in the multi-peptide classifier as described in Additional Table 5. Peptide NWGLSFYADKPETTK (protein A1AG2), ESDTSYVSLK (protein CRP), VVLHPNYSQVDIGLIK (protein HPT), and IADGYEQAAR (protein TCPE)

very similar to cases of HGSOC, in that they had all progressed to a palpable abdominal mass that required surgery. Second, practical constraints limited the total number of patients included in our TMT discovery study to 40. The limited number of samples that were used in this pilot study, and the limited clinical information available about the patients, precluded us from studying the effect of possible confounders that may have affected the outcome of our results, e.g. parity, smoking, BRCA, BMI, and menopausal status. Future studies could include these factors and others that were not captured in this study.

In future studies, we intend to use a larger cohort of women for the discovery of additional candidate protein biomarkers and include more of these candidate proteins into our SRM-MS assay. This pilot study was limited to only 30 proteins developed for the targeted SRM-MS assay from among the 700 potential protein candidates. We will also use a larger cohort of women for the targeted

Table 5 AUC values for the 15 peptides that comprised the six proteins in the multi-protein classifier

Protein name	Protein AUC rank	Protein proportion of selection (%)	Peptide sequence	Peptide AUC (95% CI)	Peptide AUC rank	Peptide proportion of selection (%)
A1AG2_HUMAN	1	77	NWGLSFYADKPETTK	0.865 (0.774, 0.939)	1	77
(AUC = 0.836)			EQLGEFYEALDCLCIPR	0.856 (0.761, 0.937)	2	38
			SDVMYTDWK	0.815 (0.708, 0.906)	4	4
			EHVAHLLFLR	0.744 (0.618, 0.857)	17	0
CRP_HUMAN	2	69	ESDTSYVSLK	0.829 (0.726, 0.918)	3	66
(AUC = 0.821)			GYSIFSYATK	0.761 (0.645, 0.869)	13	9
HPT_HUMAN	3	84	VVLHPNYSQVDIGLIK	0.808 (0.698, 0.899)	5	68
(AUC = 0.800)			VTSIQDWVQK	0.802 (0.692, 0.897)	6	15
			DYAEVGR	0.788 (0.674, 0.887)	7	2
			VGYVSGWGR	0.774 (0.658, 0.876)	9	4
ICAL_HUMAN	9	59	AAAPAPVSEAVCR	0.704 (0.577, 0.822)	29	32
(AUC = 0.740)			QAEPELDLR	0.683 (0.554, 0.803)	43	18
PYGB_HUMAN	14	55	LVTSIGDVVNHDPVVGDR	0.738 (0.620, 0.848)	20	34
(AUC = 0.727)			DFYELEPEK	0.626 (0.494, 0.749)	61	31
C4BPA_HUMAN	15	77	EDVYVVGTVLR	0.716 (0.591, 0.828)	27	54
(AUC = 0.716)						

Boxplots showing the level of expression of A1AG2 and its peptides are shown in Fig. 4

SRM-MS assay and the development of a multi-protein classifier, including samples from women with early stage HGSOC. This preliminary study was limited to the use of Pap tests primarily from women with late stage HGSOC; women with other gynecologic malignancies, including endometrial cancer and cervical cancer, were not studied. This, too, will be the focus of future studies to independently confirm the cancer of origin.

The U.S. Preventive Services Task Force (USPSTF) cervical cancer screening guidelines recommend that clinicians stop testing for cervical cancer by use of the Pap test when women reach 65 years of age [84]. In addition, the guidelines recommend the frequency of Pap testing decrease from annually to every 3 to 5 years, depending on the patient's age and other circumstances [84]. Although these current recommendations may initially be seen as a limitation to the utility of the Pap test for ovarian cancer, we envision that once further studies are completed, the Pap test recommendations may be updated to increase the age at which women are tested, since ovarian cancer predominantly is detected in postmenopausal women, with a median age of 63 years [85]. We also envision that this testing will be converted to a self-sampling at-home test using a cervical swab, so that ovarian cancer detection can be monitored in women of all ages on a frequent basis.

Since Pap tests are routinely performed in the clinic and relatively inexpensive, we envision that the multi-protein classifier could be used as a screening tool first for women at high risk of ovarian cancer (e.g. BRCA mutation carriers) and later in a normal risk population. To do so, it may be necessary to perform longitudinal sampling to determine whether levels of some of these proteins increase over time. Women with asymptomatic benign lesions who have a risk of ovarian cancer based on the multi-protein classifier would benefit from being sent to a gynecologic oncologist for a more extensive work-up, which may involve blood tests for CA125/HE4 levels and a transvaginal ultrasound. In the future, state-of-the-art technology for DNA mutations coupled with MS technology for proteins may be used in the clinic for cancer detection. While our results clearly demonstrate the future promise of this approach, finding a clinically acceptable screening strategy will require additional experimentation.

Conclusions

This study demonstrates that proteins are shed/secreted by ovarian cancer cells into the cervical opening and can be captured during a routine Pap test. These promising results provide proof-of-principle that it is possible to quantify ovarian cancer biomarkers from the cervical effluent in the cell-free Pap test fixative. These results provide a new paradigm in which the routine Pap test may have a new application: to detect both cervical cancer and ovarian cancer.

Future studies will be needed for validation of sensitivity and specificity using significantly larger patient cohorts, ideally in a prospective cohort of women prior to diagnosis. The assay could initially be used to test women in high-risk groups, and eventually for screening the general population. Perhaps longitudinal tracking of changes in protein abundance will emerge as the best method to distinguish between benign and malignant lesions in an

early detection context. Such studies have the potential to improve the detection of early stages of ovarian cancer, when the disease is treatable, thereby increasing survival rates of women.

In summary, we have developed a novel method for population-level screening of normal risk women that piggybacks on Pap tests, an existing approved public health screening protocol. Our results have been validated in an independent and clinically relevant population, demonstrating robust discrimination between healthy patients and those with tumors or benign lesions. Although our multi-protein classifier cannot distinguish between tumor and benign, it may still prove useful as a referral tool for a more extensive gynecological exam, including a transvaginal ultrasound, which is currently the best tool we have for early detection of ovarian cancer.

Abbreviations

A1AG2	Orosomucoid2
AUC	Area under the receiver operating characteristic curve
C4BPA	Complement component 4 binding protein alpha
CI	Confidence interval
CPTAC	Clinical proteomic tumor analysis consortium
CRP	C-reactive protein
DIGE	Differential in-gel electrophoresis
ELISA	Enzyme-linked immunosorbent assay
FDR	False discovery rate
HCD	Higher-energy collision dissociation
HGSOC	High grade serous ovarian carcinoma
HPLC	High pressure liquid chromatography
HPT	Haptoglobin
ICAL	Calpastatin, calpain-10, (also known as Sperm BS-17 component)
IHC	Immunohistochemistry
iTRAQ	Isobaric tags for relative and absolute quantitation
LRG	Leucine-rich alpha-2 glycoprotein
MS	Mass spectrometry
PDA	Pancreatic adenocarcinoma
Pap	Papanicolaou
PTM	Post-translational modification
PYGB	Glycogen phosphorylase, brain isoform
ROC	Receiver operating characteristic
RPLC	Reversed phase liquid chromatography
SRM	Selection reaction monitoring
TCGA	The cancer genome atlas
TSN1	Tetraspanin 1
TMT	Tandem mass tag

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12014-026-09610-7.

Supplementary Material 1: Additional Figures 1–3.

Supplementary Material 2: Additional Tables 1–5.

Acknowledgements

We would like to acknowledge the women who agreed to be part of this project by allowing us to collect "Mock" Pap test samples for research purposes only. We also would like to acknowledge the many faculty and staff in the department of Obstetrics, Gynecology, and Women's Health who enrolled and consented the women in this study and contributed to the collection of the "Mock" Pap tests from the clinics and operating room. We thank the pathologists who helped us with this project and Dr. Stefan Pambuccian (deceased) who was a driving force in the initial brainstorming for this project. We are grateful to the staff and personnel who worked on early stages of this project. We thank staff and personnel in the following University of Minnesota core facilities that were involved in many aspects of this study: BioNet Tissue Procurement Facility, Center for Mass Spectrometry and Proteomics, Masonic Cancer Center Analytical Biochemistry Shared Resource, and the Minnesota Supercomputing Institute. We are also grateful to BD Biosciences for providing the SurePathTM vials and brooms used for this project.

Author contributions

APNS conceived of the work, developed the workflow and collaborations for the project, oversaw the IRB approval, procured funding for the project, and was a major contributor in writing the manuscript. KLMB managed the sample collection and curated the samples, contributed to writing the original draft and reviewing and editing the manuscript and figures. LL conducted the statistical analyses for the SRM data, prepared Figs. 2, 3, 4 and 5 and Additional Fig. 2 and contributed to the writing of the manuscript. AP supervised the statistical analyses for the SRM data and figures and contributed to the writing of the manuscript. JRH and JBT prepared samples for TMT and SRM proteomics experiments. TLF collected mass spectrometry data and provided troubleshooting and maintenance of LC-MS/MS instruments. YW processed and contributed statistical analyses for the TMT proteomics dataset. YG selected peptide sequences for targeted measurements, collected and processed SRM data, contributed statistical analyses for the TMT proteomics dataset, prepared Fig. 1 and Additional Figs. 1 and 3, and contributed to the writing of the manuscript. MAG, PAA, and SH contributed to the development of the clinical collection of Pap test samples, collected "Mock" Pap test samples from women in the clinic, and contributed to the writing and review of the manuscript. KR and TL contributed to the design of the experiments, interpretation of the data, and writing of the manuscript. PDP designed the proteomics experiments, oversaw production and interpretation of data, and contributed to the preparation of figure 1 and the writing of the manuscript. All authors read and approved the final manuscript.

Funding

Funding was provided by the University of Minnesota Masonic Cancer Center Brainstorm Award (APNS), the Minnesota Ovarian Cancer Alliance (APNS), the Cancurables Foundation (APNS), the Department of Defense Ovarian Cancer Research Program Pilot Award Number W81XWH-16-1-0070, the Department of Defense Ovarian Cancer Research Program Teal Award Number W81XWH-22-1-0486 (APNS), and the National Institutes of Health National Cancer Institute Number 5R01CA262153 (APNS). Research reported in this publication was supported by NIH grant P30 CA77598 utilizing the Biostatistics Core shared resource of the Masonic Cancer Center, University of Minnesota and by the National Center for Advancing Translational Sciences of the National Institutes of Health Award Number UM1TR004405. Funding for the BioNet Tissue Procurement Facility was made possible through the University of Minnesota's NIH Clinical and Translational Science Award UM1TR004405.

Data availability

The TMT dataset generated and analyzed during the current study is available in the massIVE repository, https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=a78e6827fd4c4f20ae4c8a6a358fed90. The SRM dataset generated and analyzed during the current study is available in the Panorama public repository, https://panoramaweb.org/UMN-PNNL_OvaCancer_Pap.url. Both datasets are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

"Mock" Pap tests were obtained from women in the clinic or preoperatively under approval from the University of Minnesota Institutional Review Board (Protocol #1112M07362) to mimic the sample collection that would occur in a clinical setting, using a Pap test broom and a SurePath™ vial (BD Biosciences).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details
[1]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, United States of America
[2]Department of Obstetrics, Gynecology & Women's Health, University of Minnesota, Minneapolis, MN 55455, United States of America
[3]Ovarian Cancer Early Detection Program, University of Minnesota Medical School, Minneapolis, MN, United States of America
[4]Division of Biostatistics and Health Data Science, School of Public Health, University of Minnesota, Minneapolis, MN, United States of America
[5]Biological Sciences Division, Pacific Northwest National Laboratories, Richland, WA, United States of America
[6]Environmental Molecular Sciences Division, Pacific Northwest National Laboratory, Richland, WA, United States of America
[7]AbbVie, Inc, North Chicago, IL, United States of America
[8]Department of Cell, Developmental, and Cancer Biology, Oregon Health and Science University, Portland, OR, United States of America
[9]University of Minnesota, MMC 395, 420 Delaware St. SE, Minneapolis, MN 55455, United States of America

Received: 3 March 2026 / Accepted: 30 April 2026
Published online: 17 June 2026

References

1. Siegel RL, Kratzer TB, Wagle NS, Sung H, Jemal A. Cancer statistics, 2026. Cancer J Clin. 2026;76(1):e70043.
2. Jelovac D, Armstrong DK. Recent progress in the diagnosis and treatment of ovarian cancer. Cancer J Clin. 2011;61(3):183–203.
3. Jacobs IJ, Menon U, Ryan A, Gentry-Maharaj A, Burnell M, Kalsi JK, et al. Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. Lancet. 2016;387(10022):945–56.
4. Pinsky PF, Yu K, Kramer BS, Black A, Buys SS, Partridge E, et al. Extended mortality results for ovarian cancer screening in the PLCO trial with median 15years follow-up. Gynecol Oncol. 2016;143(2):270–5.
5. Koss LG. The Papanicolaou test for cervical cancer detection. A triumph and a tragedy. JAMA. 1989;261(5):737–43.
6. Solomon D, Breen N, McNeel T. Cervical cancer screening rates in the United States and the potential impact of implementation of screening guidelines. Cancer J Clin. 2007;57(2):105–11.
7. Linder J, Zahniser D. The ThinPrep Pap test. A review of clinical studies. Acta Cytol. 1997;41(1):30–8.
8. Howell LP, Davis RL, Belk TI, Agdigos R, Lowe J. The AutoCyte preparation system for gynecologic cytology. Acta Cytol. 1998;42(1):171–7.
9. Welch JM, Ciesla MC, Selvaggi SM. Serous papillary cystadenocarcinoma of the ovary diagnosed with the ThinPrep Pap Test. Acta Cytol. 2000;44(6):1122–4.
10. Wang H, Chen PC. Primary serous peritoneal carcinoma presenting first on a routine Papanicolaou smear: a case report. Acta Cytol. 2010;54(4):623–6.
11. Mulvany NJ, Mitchell G, Allen DG. Adenocarcinoma cells in Pap smears. Pathology. 2009;41(5):411–8.
12. Gupta D, Balsara G. Extrauterine malignancies. Role of Pap smears in diagnosis and management. Acta Cytol. 1999;43(5):806–13.
13. Risse EK, Ouwerkerk-Noordam E, Meijer-Marres EM, Boon ME. Exploiting the residual of cervical thin layer brush samples through cytohistology in cases with invasive carcinoma with application of antibodies. Acta Cytol. 2010;54(2):175–82.
14. Haslene-Hox H, Oveland E, Berg KC, Kolmannskog O, Woie K, Salvesen HB, et al. A new method for isolation of interstitial fluid from human solid tumors applied to proteomic analysis of ovarian carcinoma tissue. PLoS ONE. 2011;6(4):e19217.
15. Sedlaczek P, Frydecka I, Gabrys M, Van Dalen A, Einarsson R, Harlozinska A. Comparative analysis of CA125, tissue polypeptide specific antigen, and soluble interleukin-2 receptor alpha levels in sera, cyst, and ascitic fluids from patients with ovarian carcinoma. Cancer. 2002;95(9):1886–93.
16. Derycke MS, Pambuccian SE, Gilks CB, Kalloger SE, Ghidouche A, Lopez M, et al. Nectin 4 overexpression in ovarian cancer tissues and serum: potential role as a serum biomarker. Am J Clin Pathol. 2010;134(5):835–45.
17. Amon LM, Law W, Fitzgibbon MP, Gross JA, O'Briant K, Peterson A, et al. Integrative proteomic analysis of serum and peritoneal fluids helps identify proteins that are up-regulated in serum of women with ovarian cancer. PLoS ONE. 2010;5(6):e11137.
18. Elschenbroich S, Ignatchenko V, Clarke B, Kalloger SE, Boutros PC, Gramolini AO, et al. In-depth proteomics of ovarian cancer ascites: combining shotgun proteomics and selected reaction monitoring mass spectrometry. J Proteome Res. 2011;10(5):2286–99.
19. Perets R, Drapkin R. It's totally tubular… Riding the new wave of ovarian cancer research.Cancer Res. 2016;76(1):10–7.
20. Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, et al. Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. Cell. 2016;166(3):755–65.
21. McDermott JE, Arshad OA, Petyuk VA, Fu Y, Gritsenko MA, Clauss TR, et al. Proteogenomic characterization of ovarian HGSC implicates mitotic kinases, replication stress in observed chromosomal instability. Cell Rep Med. 2020;1(1):100004.
22. Boylan KLM, Afiuni-Zadeh S, Geller MA, Hickey K, Griffin TJ, Pambuccian SE, et al. A feasibility study to identify proteins in the residual Pap test fluid of women with normal cytology by mass spectrometry-based proteomics. Clin Proteom. 2014;11(1):30.
23. Dasari S, Pereira L, Reddy AP, Michaels JE, Lu X, Jacob T, et al. Comprehensive proteomic analysis of human cervical-vaginal fluid. J Proteome Res. 2007;6(4):1258–68.
24. Pereira L, Reddy AP, Jacob T, Thomas A, Schneider KA, Dasari S, et al. Identification of novel protein biomarkers of preterm birth in human cervical-vaginal fluid. J Proteome Res. 2007;6(4):1269–76.
25. Shaw JL, Smith CR, Diamandis EP. Proteomic analysis of human cervicovaginal fluid. J Proteome Res. 2007;6(7):2859–65.
26. Zegels G, Van Raemdonck GA, Coen EP, Tjalma WA, Van Ostade XW. Comprehensive proteomic analysis of human cervical-vaginal fluid using colposcopy samples. Proteome Sci. 2009;7:17.
27. Zegels G, Van Raemdonck GA, Tjalma WA, Van Ostade XW. Use of cervicovaginal fluid for the identification of biomarkers for pathologies of the female genital tract. Proteome Sci. 2010;8:63.
28. Kinde I, Bettegowda C, Wang Y, Wu J, Agrawal N, Shih Ie M, et al. Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers. Sci Transl Med. 2013;5(167):167ra4.
29. Wang Y, Li L, Douville C, Cohen JD, Yen TT, Kinde I, et al. Evaluation of liquid from the Papanicolaou test and other liquid biopsies for the detection of endometrial and ovarian cancers. Sci Transl Med. 2018;10(433):eeap8793.
30. Arildsen NS, Martin de la Fuente L, Masback A, Malander S, Forslund O, Kannisto P, et al. Detecting TP53 mutations in diagnostic and archival liquid-based Pap samples from ovarian cancer patients using an ultra-sensitive ddPCR method. Sci Rep. 2019;9(1):15506.
31. Paracchini L, Pesenti C, Delle Marchette M, Beltrame L, Bianchi T, Grassi T, et al. Detection of TP53 Clonal Variants in Papanicolaou Test Samples Collected up to 6 Years Prior to High-Grade Serous Epithelial Ovarian Cancer Diagnosis. JAMA Netw open. 2020;3(7):e207566.
32. Krimmel-Morrison JD, Ghezelayagh TS, Lian S, Zhang Y, Fredrickson J, Nachmanson D, et al. Characterization of TP53 mutations in Pap test DNA of women with and without serous ovarian carcinoma. Gynecol Oncol. 2020;156(2):407–14.
33. Mertins P, Tang LC, Krug K, Clark DJ, Gritsenko MA, Chen L, et al. Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry. Nat Protoc. 2018;13(7):1632–61.
34. Dou Y, Kawaler EA, Cui Zhou D, Gritsenko MA, Huang C, Blumenberg L, et al. Proteogenomic Characterization of Endometrial Carcinoma. Cell. 2020;180(4):729–48. e26.
35. Piehowski PD, Petyuk VA, Sontag RL, Gritsenko MA, Weitz KK, Fillmore TL, et al. Residual tissue repositories as a resource for population-based cancer proteomic studies. Clin Proteom. 2018;15:26.
36. Kim S, Gupta N, Pevzner PA. Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. J Proteome Res. 2008;7(8):3354–63.
37. Kim S, Pevzner PA. MS-GF+ makes progress towards a universal database search tool for proteomics. Nat Commun. 2014;5:5277.
38. Gibbons BC, Chambers MC, Monroe ME, Tabb DL, Payne SH. Correcting systematic bias and instrument measurement drift with mzRefinery. Bioinformatics. 2015;31(23):3838–40.
39. Monroe ME, Shaw JL, Daly DS, Adkins JN, Smith RD. MASIC: a software program for fast quantitation and flexible visualization of chromatographic

profiles from detected LC-MS(/MS) features. Comput Biol Chem. 2008;32(3):215–7.
40. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47–e.
41. Riviere LR, Tempst P. Enzymatic digestion of proteins in solution. Current protocols in protein science. 2001;Chap. 11:Unit 11 1.
42. Carr SA, Abbatiello SE, Ackermann BL, Borchers C, Domon B, Deutsch EW, et al. Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. Mol Cell Proteom. 2014;13(3):907–17.
43. Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Syst Biol. 2008;4:222.
44. Liebler DC, Zimmerman LJ. Targeted quantitation of proteins by mass spectrometry. Biochemistry. 2013;52(22):3797–806.
45. Deutsch EW, Lam H, Aebersold R. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. EMBO Rep. 2008;9(5):429–34.
46. Searle BC, Egertson JD, Bollinger JG, Stergachis AB, MacCoss MJ. Using Data Independent Acquisition (DIA) to Model High-responding Peptides for Targeted Proteomics Experiments. Mol Cell Proteom. 2015;14(9):2331–40.
47. Protein Coverage Summarizer. 2018 [Available from: https://github.com/PNNL-Comp-Mass-Spec/protein-coverage-summarizer
48. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010;26(7):966–8.
49. Song E, Gao Y, Wu C, Shi T, Nie S, Fillmore TL, et al. Targeted proteomic assays for quantitation of proteins identified by proteogenomic analysis of ovarian cancer. Sci data. 2017;4:170091.
50. Shi T, Song E, Nie S, Rodland KD, Liu T, Qian WJ, et al. Advances in targeted proteomics and applications to biomedical research. Proteomics. 2016;16(15–16):2160–82.
51. Gao Y, Wang YT, Chen Y, Wang H, Young D, Shi T, et al. Proteomic tissue-based classifier for early prediction of prostate cancer progression. Cancers. 2020;12(5):1268.
52. Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit R, Irzarry R, Huber W, editors. Bioinformatics and Computational Biology Solutions using R and Bioconductor. New York: Springer; 2005. pp. 397–420.
53. Jiang W, Varma S, Simon R. Calculating confidence intervals for prediction error in microarray classification using resampling. Stat Appl Genet Mol Biol. 2008;7(1):Article8.
54. Meinshausen N, Bühlmann P. Stability Selection. J Royal Stat Soc Ser B: Stat Methodol. 2010;72(4):417–73.
55. Boylan KLM, Andersen JD, Anderson LB, Higgins L, Skubitz APN. Quantitative proteomic analysis by iTRAQ for the identification of candidate biomarkers in ovarian cancer serum. Proteome Sci. 2010;8:31.
56. Andersen JD, Boylan KL, Xue FS, Anderson LB, Witthuhn BA, Markowski TW, et al. Identification of candidate biomarkers in ovarian cancer serum by depletion of highly abundant proteins and differential in-gel electrophoresis. Electrophoresis. 2010;31(4):599–610.
57. Yang D, Li H, Sun X, Yang S, Wang K, Liu Z. Clinical usefulness of high levels of C-reactive protein for diagnosing epithelial ovarian cancer. Sci Rep. 2020;10(1):20056.
58. Zhang W, Zhang Z, Qian L. Prognostic and clinicopathological significance of C-reactive protein in patients with ovarian cancer: a meta-analysis. World J Surg Oncol. 2024;22(1):8.
59. Peres LC, Mallen AR, Townsend MK, Poole EM, Trabert B, Allen NE, et al. High Levels of C-Reactive Protein Are Associated with an Increased Risk of Ovarian Cancer: Results from the Ovarian Cancer Cohort Consortium. Cancer Res. 2019;79(20):5442–51.
60. Trabert B, Pinto L, Hartge P, Kemp T, Black A, Sherman ME, et al. Pre-diagnostic serum levels of inflammation markers and risk of ovarian cancer in the prostate, lung, colorectal and ovarian cancer (PLCO) screening trial. Gynecol Oncol. 2014;135(2):297–304.
61. Wulaningsih W, Holmberg L, Ng T, Rohrmann S, Van Hemelrijck M. Serum leptin, C-reactive protein, and cancer mortality in the NHANES III. Cancer Med. 2016;5(1):120–8.
62. Ridker PM, Moorthy MV, Cook NR, Rifai N, Lee IM, Buring JE, Inflammation. Cholesterol, Lipoprotein(a), and 30-Year Cardiovascular Outcomes in Women. N Engl J Med. 2024;391(22):2087–97.
63. Hair PS, Wagner SM, Friederich PT, Drake RR, Nyalwidhe JO, Cunnion KM. Complement regulator C4BP binds to Staphylococcus aureus and decreases opsonization. Mol Immunol. 2012;50(4):253–61.
64. Zou J, Chen Y, Ji Z, Liu D, Chen X, Chen M, et al. Identification of C4BPA as biomarker associated with immune infiltration and prognosis in breast cancer. Transl Cancer Res. 2024;13(1):25–45.
65. Wang G, Liu X, You Y, Chen S, Chang X, Yang Q. Development and clinical validation of a seven-gene signature based on tumor stem cell-related genes to predict ovarian cancer prognosis. J Ovarian Res. 2024;17(1):58.
66. Sogawa K, Takano S, Iida F, Satoh M, Tsuchida S, Kawashima Y, et al. Identification of a novel serum biomarker for pancreatic cancer, C4b-binding protein alpha-chain (C4BPA) by quantitative proteomic analysis using tandem mass tags. Br J Cancer. 2016;115(8):949–56.
67. Mikami M, Tanabe K, Matsuo K, Miyazaki Y, Miyazawa M, Hayashi M, et al. Fully-sialylated alpha-chain of complement 4-binding protein: Diagnostic utility for ovarian clear cell carcinoma. Gynecol Oncol. 2015;139(3):520–8.
68. Zhou Y, Jin Z, Wang C. Glycogen phosphorylase B promotes ovarian cancer progression via Wnt/beta-catenin signaling and is regulated by miR-133a-3p. Biomed Pharmacother. 2019;120:109449.
69. Yang C, Wang H, Shao M, Chu F, He Y, Chen X, et al. Brain-type glycogen phosphorylase (PYGB) in the pathologies of diseases: a systematic review. Cells. 2024;13(3):289.
70. Tashima S, Shimada S, Yamaguchi K, Tsuruta J, Ogawa M. Expression of brain-type glycogen phosphorylase is a potentially novel early biomarker in the carcinogenesis of human colorectal carcinomas. Am J Gastroenterol. 2000;95(1):255–63.
71. Nian H, Ma B. Calpain-calpastatin system and cancer progression. Biol Rev Camb Philos Soc. 2021;96(3):961–75.
72. The Human Protein Atlas. [Available from: https://www.proteinatlas.org/. Accessed 28 January 2026].
73. Zhang S, Deen S, Storr SJ, Chondrou PS, Nicholls H, Yao A, et al. Calpain system protein expression and activity in ovarian cancer. J Cancer Res Clin Oncol. 2019;145(2):345–61.
74. Muntiu A, Moresi F, Vincenzoni F, Rossetti DV, Iavarone F, Messana I, et al. Proteomic profiling of pre- and post-surgery saliva of glioblastoma patients: a pilot investigation. Int J Mol Sci. 2024;25(23):12984.
75. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. Science. 2017;357:6352.
76. Boylan KLM, Geschwind K, Koopmeiners JS, Geller MA, Starr TK, Skubitz APN. A multiplex platform for the identification of ovarian cancer biomarkers. Clin Proteom. 2017;14:34.
77. Zhou Z, Yang Z, Zhou L, Yang M, He S. The versatile roles of testrapanins in cancer from intracellular signaling to cell-cell communication: cell membrane proteins without ligands. Cell Biosci. 2023;13(1):59.
78. Scholz CJ, Kurzeder C, Koretz K, Windisch J, Kreienberg R, Sauer G, et al. Tspan-1 is a tetraspanin preferentially expressed by mucinous and endometrioid subtypes of human ovarian carcinomas. Cancer Lett. 2009;275(2):198–203.
79. Shin HY, Yang W, Chay DB, Lee EJ, Chung JY, Kim HS, et al. Tetraspanin 1 promotes endometriosis leading to ovarian clear cell carcinoma. Mol Oncol. 2021;15(4):987–1004.
80. Gyllensten U, Hedlund-Lindberg J, Svensson J, Manninen J, Ost T, Ramsell J, et al. Next generation plasma proteomics identifies high-precision biomarker candidates for ovarian cancer. Cancers. 2022;14(7):1757.
81. Andersen JD, Boylan KLM, Jemmerson R, Geller MA, Misemer B, Harrington KM, et al. Leucine-rich alpha-2-glycoprotein-1 is upregulated in sera and tumors of ovarian cancer patients. J Ovarian Res. 2010;3:21.
82. Maritschnegg E, Heitz F, Pecha N, Bouda J, Trillsch F, Grimm C, et al. Uterine and Tubal Lavage for Earlier Cancer Detection Using an Innovative Catheter: A Feasibility and Safety Study. Int J Gynecol Cancer. 2018;28(9):1692–8.
83. Barnabas GD, Bahar-Shany K, Sapoznik S, Helpman L, Kadan Y, Beiner M, et al. Microvesicle Proteomic Profiling of Uterine Liquid Biopsy for Ovarian Cancer Early Detection. Mol Cell Proteom. 2019;18(5):865–75.
84. Curry SJ, Krist AH, Owens DK, Barry MJ, Caughey AB, Davidson KW, et al. Screening for Cervical Cancer: US Preventive Services Task Force Recommendation Statement. JAMA. 2018;320(7):674–86.
85. SEER Cancer Stat Facts: Ovarian Cancer Bethesda, MD: National Cancer Institute. 2025 [Available from: https://seer.cancer.gov/statfacts/html/ovary.html. Accessed 24 April 2026].

Publisher's note