RESEARCH ARTICLE

Proteomics
Clinical Applications

Discovery and validation of serum glycoprotein biomarkers for high grade serous ovarian cancer

Mriga Dutt[1] | Gunter Hartel[1] | Renee S. Richards[1] | Alok K. Shah[1] | Ahmed Mohamed[1] | Sophia Apostolidou[2] | Aleksandra Gentry-Maharaj[2] | Australian Ovarian Cancer Study Group | John D. Hooper[3] | Lewis C. Perrin[3,4] | Usha Menon[2] | Michelle M. Hill[1,5]

[1]QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

[2]MRC Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College London, London, UK

[3]Mater Research Institute – The University of Queensland, Translational Research Institute, Woolloongabba, QLD, Australia

[4]Mater Adult Hospital, South Brisbane, QLD, Australia

[5]UQ Centre for Clinical Research, Faculty of Medicine, The University of Queensland, Brisbane, Australia

Correspondence
Usha Menon, MRC Clinical Trials Unit at UCL,
90 High Holborn, 2nd Floor,
London WC1V 6LJ, UK.
Email: u.menon@ucl.ac.uk

Michelle Hill, QIMR Berghofer Medical
Research Institute, Herston, Brisbane, QLD
4006, Australia.
Email: m.hill2@uq.edu.au

Abstract

Purpose: This study aimed to identify serum glycoprotein biomarkers for early detection of high-grade serous ovarian cancer (HGSOC), the most common and aggressive histotype of ovarian cancer.

Experimental design: The glycoproteomics pipeline lectin magnetic bead array (LeMBA)-mass spectrometry (MS) was used in age-matched case-control serum samples. Clinical samples collected at diagnosis were divided into discovery ($n = 30$) and validation ($n = 98$) sets. We also analysed a set of preclinical sera ($n = 30$) collected prior to HGSOC diagnosis in the UK Collaborative Trial of Ovarian Cancer Screening.

Results: A 7-lectin LeMBA-MS/MS discovery screen shortlisted 59 candidate proteins and three lectins. Validation analysis using 3-lectin LeMBA-multiple reaction monitoring (MRM) confirmed elevated A1AT, AACT, CO9, HPT and ITIH3 and reduced A2MG, ALS, IBP3 and PON1 glycoforms in HGSOC. The best performing multimarker signature had 87.7% area under the receiver operating curve, 90.7% specificity and 70.4% sensitivity for distinguishing HGSOC from benign and healthy groups. In the preclinical set, CO9, ITIH3 and A2MG glycoforms were altered in samples collected 11.1 ± 5.1 months prior to HGSOC diagnosis, suggesting potential for early detection.

Conclusions and clinical relevance: Our findings provide evidence of candidate early HGSOC serum glycoprotein biomarkers, laying the foundation for further study in larger cohorts.

KEYWORDS
high grade serous ovarian cancer, lectin magnetic bead array (LeMBA), mass spectrometry, ovarian cancer screening, serum glycoprotein biomarker

1 | INTRODUCTION

While ovarian cancer is only the third most common gynaecological cancer worldwide, it is the leading cause of gynaecological cancer mortality. One of the main factors for this high mortality rate is diagnosis at an advanced stage due to non-specific symptoms and the lack of effective predictive or diagnostic blood biomarkers. Late diagnosis is associated with high mortality. Only 29% of women with distant metastases survive 5 years, compared to 92% with localized disease [1]. However, the existing ovarian cancer diagnostic tests in clinical use, transvaginal ultrasound and serum CA125, do not have the sensitivity required for detecting the disease in early stage. Indeed, two large ovarian cancer screening trials using a combination of these modalities found no evidence of a reduction in disease-specific mortality on long-term follow-up [2, 3]. Furthermore, neither test is specific to cancer and both trials reported unnecessary surgery in women without cancer [4]. This has led to a concerted effort to discover biomarkers for early detection of ovarian cancer [5].

Cancer is associated with alterations in the glycosylation machinery and glycan structures on circulating proteins [6]. Several studies have shown that specific glycoforms of cancer biomarkers can improve specificity. For ovarian cancer, glycosylated forms of CA125 measured by microarray [7], lectin immunoassay [8] or glycosylation-specific antibodies [9] can significantly improve differential diagnosis. This suggests that a glycoform-specific glycoprotein biomarker panel may achieve the high specificity and sensitivity required for ovarian cancer screening. Glycomic and glycoproteomics studies on ovarian cancer serum and tissues have revealed differential abundance of several types of N-glycans in ovarian cancer, including fucose, sialic acid, high mannose types [10–15]. However, these potential biomarkers are yet to be clinically validated.

Here, we report on a study using lectin magnetic bead array (LeMBA)-coupled mass spectrometry (MS) platform [16] for ovarian cancer serum glycoprotein biomarker discovery and validation. LeMBA is a one-pot, high throughput glycoproteomics method with no need for abundant serum protein depletion, potentially increasing robustness of the biomarker development process as we previously reported for oesophageal adenocarcinoma [17, 18] and canine haemangiosarcoma [19]. It has not been previously used for glycoprotein biomarker studies in ovarian cancer. We have focused on the most common and aggressive histotype, high grade serous ovarian cancers (HGSOC), which accounts for ~70% of ovarian cancers and most of the disease-specific mortality [5]. Furthermore, for discovery of biomarkers with the potential for early detection, in addition to using samples collected at clinical diagnosis as is the norm, we also evaluated samples collected prior to ovarian cancer diagnosis from the multicentre randomised controlled trial, the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS).

2 | MATERIALS AND METHODS

2.1 | Study design

Case control studies were undertaken using serum samples from women with HGSOC patients (cases) and two age-matched groups (controls)—women with a benign ovarian neoplasm and healthy women using sample sets from three independent cohorts - (1) a clinical set of 30 serum samples from the United Kingdom Ovarian Population Study (UKOPS) collected from women at diagnosis of HGSOC ($n = 10$), benign ovarian neoplasms ($n = 10$) and healthy controls ($n = 10$). (Table S1); (2) a pre-clinical set of 30 serum samples from the UKCTOCS trial [20] collected from women at a mean interval of 11.1 ± 5.1 months prior to diagnosis of HGSOC ($n = 10$), benign ovarian neoplasms ($n = 10$) and healthy controls ($n = 10$). (Table S1); and (3) a clinical set of 95 serum samples from the Australian Ovarian Cancer Study (AOCS) collected from women at diagnosis of HGSOC ($n = 39$), benign ovarian neoplasms ($n = 28$) and healthy controls ($n = 28$) (Table S2).

The discovery phase included the UKOPS clinical set and the UKCTOCS pre-clinical set. A shortlist of candidate proteins and lectins from the discovery phase was then validated using the independent clinical set from the AOCS (Figure 1). This study was approved by the QIMR Berghofer Medical Research Institute Research Ethics committee, and the East Midlands—Derby Research Ethics Committee in the UK. AOCS was approved by the Human Research Ethics Committees at the Peter MacCallum Cancer Centre, QIMR Berghofer Medical Research Institute, University of Melbourne and all participating hospitals.

2.2 | Biomarker discovery phase

LeMBA-MS was used for both discovery and validation phases (Figure 1). For discovery, seven lectins were selected from the literature [10, 11, 13, 15]: *Aleuria aurantia* (AAL), Concanavalin-A (Con-A),

Erythrina cristagalli (ECA), *Phaseolus vulgaris* Leucoagglutinin (L-PHA), *Sambucus nigra* (SNA), *Solanum tuberosum* (STL) and *Wisteria floribunda* (WFA), which preferentially target glycoproteins with fucose (α1-3, α1-4, α1-6 linked), mannose (oligomannose and hybrid-type), galactose (β1-4-linked terminal), 2,6-branched tri-, tetraantennary complex-type N-glycan, sialic acid (α2-6-linked and Tn antigen), N-acetylglucosamine ((GlcNAcβ1-4)n (Chitin), oligosaccharide containing GlcNAc and MurNAc), and N-acetylgalactosamine (terminal), respectively. All chemicals and reagents were purchased from Sigma-Aldrich, USA unless stated otherwise.

2.2.1 | Serum denaturation

Thawed serum samples were centrifuged at 16,000 g and 4°C for 15 min to remove cellular debris and the supernatant protein concentration was determined by BCA protein assay (Pierce, Thermo Fischer). To minimise batch effects, bulk serum denaturation was performed for the entire project. An aliquot of each serum sample containing 800 μg of protein was diluted to 10 μg/μL in denaturation buffer (20 mM Tris-HCl pH 7.4, 1% v/v sodium dodecyl sulphate (SDS) and 5% v/v Triton X-100. The internal standard protein chicken ovalbumin was added to each serum sample at 10 pmol. Protein disulphide bonds were reduced by adding 20 mM dithiothreitol (Thermo Fisher, USA) to the samples and incubating at 37°C for 30 min. Following this, 100 mM iodoacetamide (Thermo Fisher, USA) was added to each sample and incubated at room temperature for 30 min in the dark to alkylate free thiol groups. The denatured serum samples were diluted 20 times in LeMBA binding buffer (20 mM Tris-HCl pH 7.4, 300 mM NaCl, 1 mM CaCl$_2$, 1 mM MnCl$_2$, 1% Triton, 1 unit protease inhibitor cocktail) to yield a final protein concentration of 0.5 μg/μL. Aliquots of 100 μL were transferred to microplates in a randomized layout in preparation for LeMBA. Prepared plates were sealed and frozen at −80°C until use.

2.2.2 | Lectin magnetic bead pulldown and on-bead trypsin digest

LeMBA-MS with the selected seven lectins was performed as previously described [16, 18]. First, individual lectins were conjugated to MyOne tosyl-activated Dynabeads (Invitrogen, Australia) by incubating 50 μg of selected lectin (Vector Laboratories, USA) with 100 μL of Dynabeads at 37°C for 24 h. The resulting lectin-bead conjugate was treated with 2% w/v glycine solution to reduce nonspecific binding, and further incubated at 37°C for 16 h. The blocked beads were washed and diluted in lectin storage buffer (20 mM Tris pH 7.4, 150 mM NaCl, 1 mM CaCl$_2$, 1 mM MnCl$_2$, 0.5% Triton-X 100, 1 unit protease inhibitor cocktail).

Pulldown using the prepared lectin magnetic beads was performed on the AssayMAP Bravo liquid handler workstation (Agilent Technologies, USA) using one lectin per microplate. Briefly, 50 μL conjugated beads and 100 μL denatured serum was added to each well of a 96-

Statement of Clinical Relevance

Ovarian cancer continues to be associated with high disease mortality. Much of this is due to the diagnosis at an advanced stage of the most common and aggressive histotype—high grade serous ovarian cancer (HGSOC). This has led to significant efforts to detect the disease earlier when treatment is more effective. However, to date there is no effective screening test. We describe discovery and validation of a novel blood glycoprotein signature using lectin magnetic bead array (LeMBA)-coupled mass spectrometry for HGSOC using a case control study of clinical samples collected at diagnosis. Three biomarker candidates were altered 4–18 months prior to cancer diagnosis in a pilot case control study using pre-clinical samples from the UK Collaborative Trial of Ovarian Cancer Screening. Our findings suggest that serum glycoproteins could be novel biomarkers for earlier detection of HGSOC and lay the foundation for further study in larger cohorts.

well microtiter plate (Greiner, USA), and glycoprotein capture was performed at 4°C for 1 h under gentle shaking. Post incubation, the conjugated beads with the captured glycoproteins were washed seven times in 50 mM ammonium bicarbonate buffer with two microplate changes to minimise trace detergent, reducing and alkylating agent concentrations. The captured glycoproteins were digested with trypsin at 37°C for 18 h after adding 50 mM ammonium bicarbonate buffer and 1 μg of sequencing grade porcine trypsin (Promega, Australia) to each well. The plate was sealed for enzyme digestion. The trypsin was inactivated with 1% v/v formic acid (FA; Merck, USA) and the digested peptides were collected, dried down in a vacuum concentrator, sealed and stored at −80°C until use.

2.2.3 | Data dependent acquisition mass spectrometry

Shotgun proteomics using data-dependent acquisition was performed on a SCIEX 5600 TripleTOF 5600+ mass spectrometer (SCIEX, USA) coupled to a Shimadzu LC-20AD Prominence nano liquid chromatography system (Shimadzu, Japan). All solvents and reagents were of MS grade (Thermo Fisher, USA). The mass spectrometer was controlled using Analyst 1.7 software (SCIEX, USA). Digested peptides were resuspended in 0.1% v/v FA and injected onto a Protocol C18 analytical column (200 Å, 3 μm, 150 mm × 150 μm, Trajan Scientific, Australia) connected to a Protocol guard column (Polar 120 Å, 3 μm, 10 mm × 300 μm, Trojan Scientific, Australia) and the sample injection order was randomised in the worklist. Column compartment was maintained at 45°C. The peptides were eluted using mobile phase A (0.1% v/v FA) over the specified gradient of mobile phase B (95%
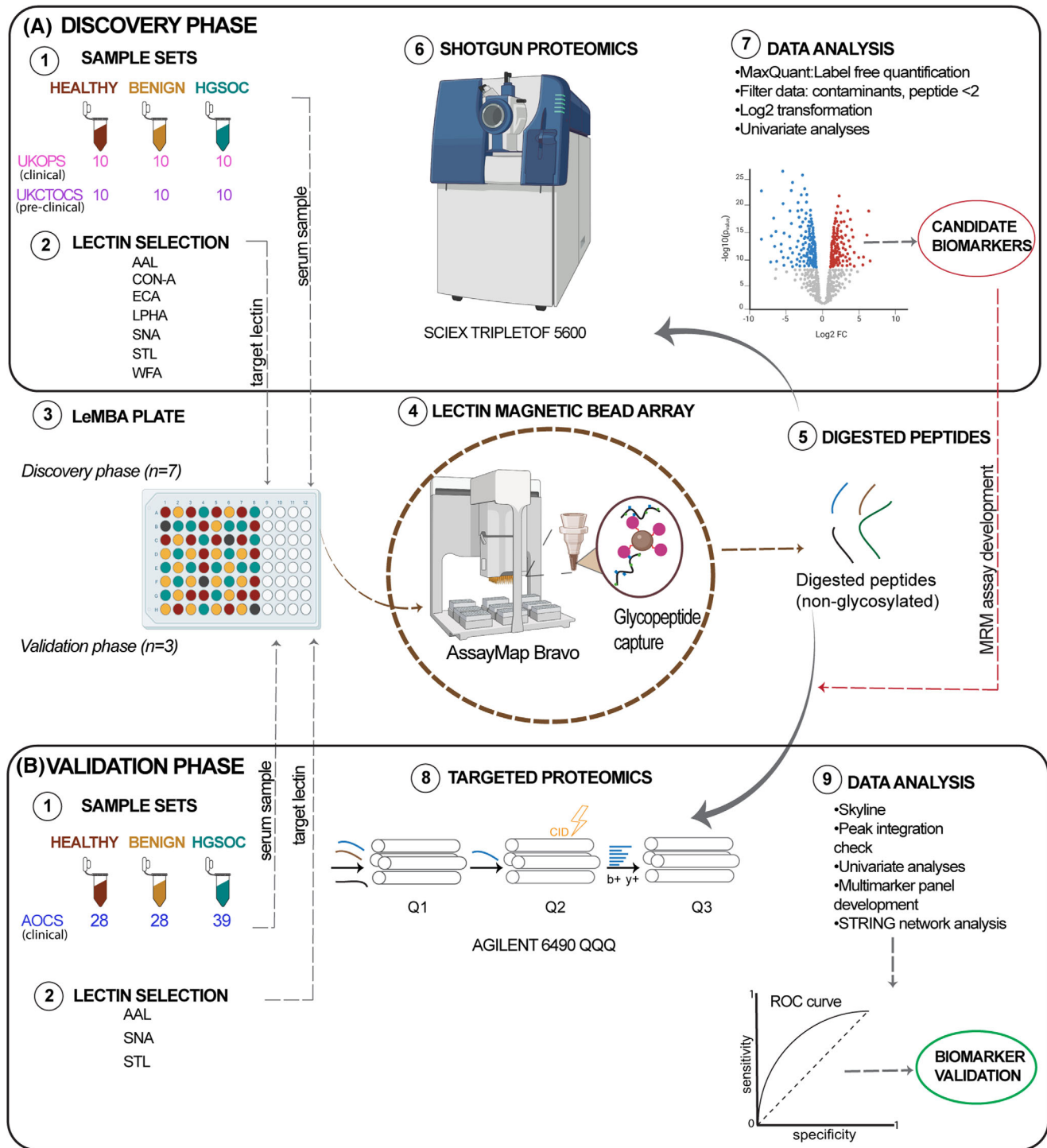
FIGURE 1 Biomarker study design. Discovery and validation of HGSOC biomarkers was conducted in two phases, starting from separate clinical cohorts (1) and sera collection (2). Lectin selection (3) was based on literature for discovery phase and the discovery results for the validation phase. Both phases use LeMBA (4), liquid handler-assisted pulldown (5) and on-bead digestion (6). Shotgun mass spectrometry was conducted for discovery phase (7) followed by discovery of candidates (8) for development of a targeted mass spectrometry assay (9) for validation phase. Both univariate and multivariate analyses were conducted for biomarker validation (10).

acetonitrile, 5% v/v water, 0.1% v/v FA) for 60 min at a flow rate of 1.2 µL/min (5% B at 3 min; 30% B at 37 min; 50% B at 45 min; 100% B at 47 min; 100% B at 51 min; 5% B at 53 min until end of run). The nanospray ion source was set as follows: ion source gas 1 = 35 psi, curtain gas = 30 psi, ion spray floating voltage = 2400 V and interface heater temperature = 180°C. The ion optics parameters were set as declustering potential = 100 V and collision energy (survey scan) = 10 V. Data acquisition was performed using the information dependent acquisition (IDA) and the top 30 precursors from each survey scan were selected for fragmentation. The MS1 spectra was acquired in positive polarity within the mass range = m/z 350–1250 Da, with the accumulation time of 250 ms. The precursor selection mass window in the quadrupole was set to unit resolution (m/z 0.7 window). The MS/MS spectra were acquired using collision induced dissociation (CID) within the mass range = m/z 100–1500 Da with the following parameters: charge states +2 to +5, accumulation time = 100 ms, fragmentation threshold = 150 cps, dynamic exclusion = 15 s and collision energy voltage was set as rolling collision energy with a collision energy spread of 3.

The acquired raw ion spectra for each lectin batch were searched against the reviewed UniProt human proteome database (20,365 proteins, accession date 1 January 2020) using MaxQuant software, v.1.6.6.0 [21]. The MaxQuant contaminant database (247 entries) was also searched to identify contaminants such as keratin. MaxQuant parameters were set as follows: Digestion = trypsin, with two missed cleavages; fixed modification was set to cysteine carbamidomethylation; variable modifications were set as methionine oxidation and N-terminal acetylation; label free quantification (LFQ) was enabled with minimum ratio count set to 2; unique and razor peptides were used for protein identification; match between runs was set as TRUE; and false discovery rate (FDR) for protein and peptide identification was set at 0.01. AB Sciex Q-TOF was set as instrument type using default settings. First search peptide tolerance was set to 0.07 Da and 0.06 Da for the main search. MS/MS tolerance was set at 40 ppm. The search results were imported into R software v1.4.1103 (www.R-project.org) for further data processing and statistical analyses.

2.2.4 | Mass spectrometry data processing and statistical analysis

The generated protein list for each lectin batch was filtered to remove contaminants, reverse identified protein IDs, proteins with < 2 peptide IDs and score < 5. Proteins which were missing in < 25% of all samples were considered missing at random and imputed using localized least square regression (llsimpute) [22]. Proteins missing in > 25% were imputed with the minimum detected value (values drawn randomly from a normal distribution centred at sample minimum and with SD estimated from non-missing proteins). Quantitative analysis was conducted at the protein level using the summed intensity of all peptides mapped to each protein. Log2 transformed data were analysed using the R limma package [23] and Student's T-test. As candidates

from the discovery phase will be further confirmed by targeted MS, and application of false discovery rate to the dataset yielded very few significant differences, we shortlisted candidates based on non-adjusted p-values. Differentially abundant proteins were visualised by volcano plots, using the criteria p-value < 0.05 and Log_2 fold change > 1, and the nomenclature 'lectin-UniProt entry name'. All graphical output has been generated using R or GraphPad Prism v9 (San Diego, USA) and figures prepared using Illustrator v26.3.1 (Adobe Inc, USA) and Biorender (www.biorender.com).

2.3 | Biomarker validation phase

2.3.1 | Multiple reaction monitoring assay development

A custom multiple reaction monitoring (MRM) assay was developed for the list of protein biomarker candidates discovered from UKCTOCS and UKOPS clinical sets, after manually removing immunoglobulins from the list. An initial transition list was selected by matching to a spectral library generated from *in silico* trypsin digest of the discovery phase raw spectral files ($n = 420$) in Skyline v 21.1.0.278 (http://skyline.maccosslab.org/), using human proteome as background proteome. For each candidate biomarker protein, a minimum of 10 peptides were selected, each consisting of at least six transitions (b and y ions). Additionally, the 'Unique Peptides' parameter in Skyline was applied to check for peptide uniqueness to a single protein.

For MRM method optimisation, digested peptide samples from all the lectin batches and across both discovery sets were pooled into a single sample. To monitor retention time across all runs, three stable isotope standard (SIS) peptides (VTSIQDWVQK, NLAVSQVVHK, LSPIYNLVPVK) were added. The final dynamic MRM method consisted of 60 proteins (59 candidate biomarker proteins + one chicken ovalbumin protein), 176 peptides (170 candidate peptides + three SIS peptides + three chicken ovalbumin peptides) and 860 transitions with a delta retention time of 1 min.

2.3.2 | LeMBA-MRM-MS

Candidate biomarker validation was performed on the independent clinical sample set from AOCS. Based on the discovery phase results, AAL, SNA and STL lectins were selected for the validation phase and the serum samples were subjected to the LeMBA workflow as described in the discovery phase. Prior to MS injection, the above three SIS peptides were spiked-in to the samples for monitoring retention time stability across runs. MRM-MS was performed on an Agilent 6490 triple quadrupole mass spectrometer coupled to an Agilent 1290 Infinity UHPLC system, equipped with an Agilent jet stream + ESI source. The mass spectrometer was controlled by MassHunter software (Agilent Technologies, USA). Digested peptides (10 µL, 20 µg) were injected onto a reverse phase AdvanceBio Peptide Mapping analytical column (150 × 2.1 mm i.d., 2.7 µm, part number 653750–902, Agilent Technologies, USA) connected to a 5 mm long guard column and the sample

injection order was randomised in the worklist. The column compartment was maintained at 50°C. The peptides were eluted using mobile phase A (0.1% v/v FA) over the specified gradient of mobile phase B (100% acetonitrile, 0.1% v/v FA) for 35 min at a flow rate of 0.4 mL/min (3% B at 0 min; 30% B at 20 min; 40% B at 24 min; 95% B at 24.5 min; 95% B at 28.5 min and 3% B at 29 min until end of run). The mass spectrometer operated in positive ion mode and the source parameters were set as 150 V high pressure RF, 60 V low pressure RF, 4000 V capillary voltage, 300 V nozzle voltage, 30 psi nebulizer gas flow, 15 L/min drying gas flow at a temperature of 150°C, 11 L/min sheath gas flow at a temperature of 250°C and 200 V delta EMV. The quadrupole was set at unit resolution [0.7 Da full width at half maximum in the first quadrupole (Q1) and the third quadrupole (Q3)], fragmentor at 380 V and cell accelerator voltage at 4 V.

Due to a mass spectrometer software failure, data for the first batch of AAL-pulldown samples were not saved. The entire plate had to be re-run using remaining sample volume, and 25 samples were noted to have lower remaining volume. In addition, an injection problem was noted for one SNA-pulldown sample.

2.3.3 | Data analysis

The data analysis for each lectin MRM-MS was performed independently with no comparisons performed across the three lectins. MRM-MS raw data for the three lectins were exported to Skyline v 21.1.0.278 (downloaded January 2022, http://skyline.maccosslab.org/) to manually check for correct peak integration and the peak area of each measured transition was exported and further analysed in R (v1.4.1103). For each measured peptide in each sample, the transition peak areas were summed and then Log_2 transformed.

Data quality control was conducted using peak area for the three internal standard chicken ovalbumin peptides (Figure S1) and the three spiked-in SIS peptides (Figure S2). While most chicken ovalbumin peptides were consistent, the 25 AAL and one SNA samples with noted aberrations at the MS step showed larger variability, and were removed as outliers (Figure S1, S2). After outlier removal, the calculated %CV for all peptide standards was less than 10% (except peptide AAL-VASMASEK, 11.3%) (Table S3) and the peak area distribution of all samples remaining in the analysis exhibited a normal data distribution (Figure S3). As the dataset has low %CV, we decided normalisation is not required. Student's *t*-test was performed between the case control groups and false discovery rate using Benjamini–Hochberg method was applied to identify significantly differing proteins at *p*-adjusted-value < 0.05. All graphical output has been generated using R or GraphPad Prism and figures prepared using Adobe Illustrator and Biorender.

2.3.4 | Multi-marker panel development

Generalized regression with binomial distribution and lasso estimation was used to develop multi-marker panels using JMP Pro version 16.2.0 (JMP Pro Inc., Carey, NC, USA). Performance of the multi-marker panels were assessed using leave-one-out cross-validation, where area under the receiver operating curve, specificity, sensitivity was calculated on the left-out observations. The number of times each parameter (peptide) was chosen in the cross-validation models is presented as an indication of the relative importance of the markers for prediction. All models were also run using peptides standardised by subtraction of the mean of the three internal standard chicken ovalbumin peptides, but are not reported as they yielded comparable results. Inclusion of the 25 AAL outliers also was tested and led to similar albeit slightly worse prediction models.

3 | RESULTS

3.1 | Discovery of candidate serum glycoprotein biomarkers

In the UKOPS discovery set, we found 15 and 16 differentially abundant proteins when HGSOC samples were compared to benign and healthy samples, respectively (Figure 2, Table 1, Table S4). Compared to benign samples, UKOPS HGSOC samples exhibited an increase in AAL-A1AG1 (alpha-1-acid-glycoprotein 1), AAL-A1AG2 (alpha-1-acid-glycoprotein-2), AAL-FA5 (coagulation factor V), AAL-HGF (hepatocyte growth factor), STL-FIBA (fibrinogen alpha chain), AAL-FIBB (fibrinogen beta chain), LPHA-ITIH3 (inter-alpha-trypsin-inhibitor heavy chain 3), as well as decreased SNA-CO8G (complement component C8 gamma chain), SNA-FA10 (coagulation factor X), LPHA-PON1 serum paraoxonase/arylesterase 1) and STL-PON1 (Figure 2A). When compared to healthy samples, the UKOPS HGSOC samples exhibited an increase in ECA-CO9 (complement component C9) and ECA-LG3BP (galectin-3-binding protein), and a reduction of ECA-C8B (complement component C8B), ECA-CO8G, ECA-LUM (lumican) and ECA-THBG (thyroxine-binding globulin), STL-ACTG (actin), STL-CBG (corticosteroid-binding globulin, STL-CHLE (cholinesterase) and STL-PON1 (Figure 2B).

In the UKCTOCS discovery set, we found 14 and 12 differentially abundant proteins in HGSOC samples when compared to benign and healthy samples, respectively (Figure 2, Table 1, Table S5). Compared to benign samples, UKCTOCS HGSOC samples showed an increase in WFA-ALDOA (fructose bisphosphates aldolase A), WFA-CO9 and WFA-HV307 (immunoglobulin heavy-variable 3–7), SNA-F13B (coagulation factor XIII B chain) and SNA-ZA2G (zinc-alpha-2-glycoprotein), as well as reduction in AAL-MASP1 (mannan-binding lectin serine protease 1), ECA-APOH (beta-2 glycoprotein-1), ECA-CO7 (complement component C7), ECA-CO8G, ECA-FBLN1 (fibulin-1) and ECA-PON1 (Figure 2C). On the other hand, when compared to healthy samples, HGSOC samples showed increased STL-A1AT (alpha-1-antitrypsin), STL-HEP2 (heparin cofactor 2), STL-ITIH3, STL-IGLL5 (immunoglobulin lambda like growth factor), STL-LG3BP, and reduced AAL-MASP1, ECA-CO8G, ECA-FBLN1 and ECA-PGRP2 (peptidoglycan recognition protein-2), WFA-CO7 and WFA-FA12 (coagulation factor XII) (Figure 2D).
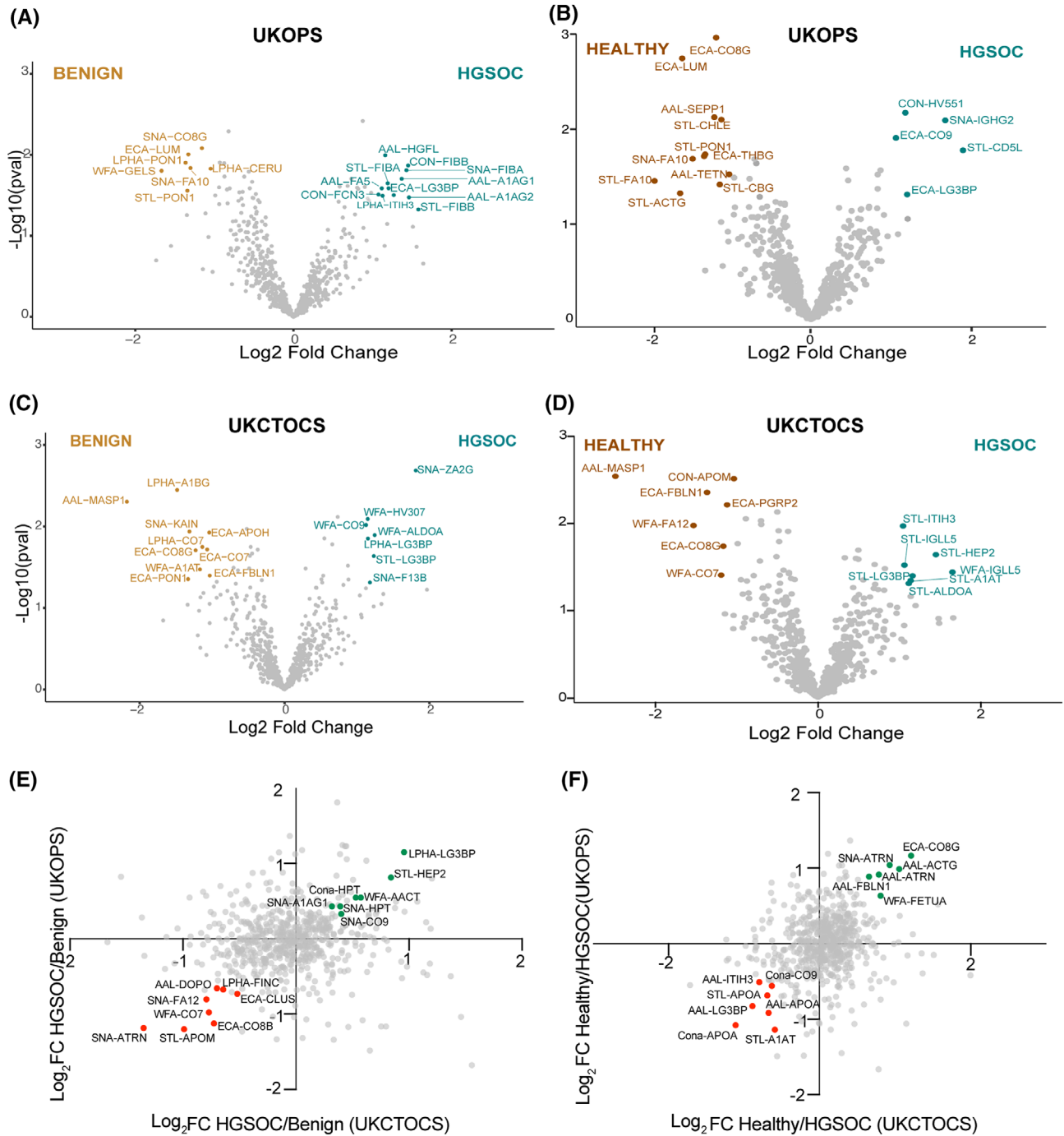
FIGURE 2 Biomarker discovery data. Volcano and two-way scatter plots visualising the differentially abundant proteins and correlated proteins, respectively, between the benign and HGSOC (A, C, E) and healthy and HGSOC (B, D, F) clinical comparisons for UKOPS and UKCTOCS sample sets. The volcano plots highlight all differential glycoproteins according to the criteria $p < 0.05$, $\mathrm{Log_2}$ Fold Change > 1. The scatter plots highlight select glycoproteins ($\mathrm{Log_2}$ Fold Change > 0.5) that are upregulated (green dots) and downregulated (red dots) in both sample sets. All candidates are indicated using the nomenclature 'lectin-Uniprot entry name'.

Additionally, there was an overlap of a subset of candidates that displayed the same expression trend in both the sample sets such as LPHA-LG3BP, STL-HEP2 and SNA-CO9 in the HGSOC versus benign comparison (Figure 2E) and ECA-C08G, SNA-ATRN and STL-A1AT in the healthy versus HGSOC comparison (Figure 2F).

3.2 | Validation of candidate biomarkers

Three lectins (AAL, SNA and STL) with the largest number of candidate proteins discovered in both UKOPS and UKCTOCS sets (Table 1) were selected for validation in the independent clinical AOCS cohort.

**Proteomics
Clinical Applications**

TABLE 1 Discovery phase biomarker candidates.

LECTIN	HGSOC versus Benign		HGSOC versus Healthy		Total number
	UKOPS	UKCTOCS	UKOPS	UKCTOCS	
AAL	A1AG1, A1AG2, FA5, HGFL	MASP1	SEPP1	MASP1	6
CON-A	FIBB, FCN3	–	HV551	APOM	4
ECA	LUM, LG3BP	APOH, CO7, CO8G, FBLN1, PON1	CO9, LG3BP, CO8B, CO8G, LUM, THBG	CO8G, FBLN1, PGRP2	11
L-PHA	PON1, CERU, ITIH3	-	–	–	6
SNA	FA10, CO8G, FIBA	KAIN, F13B, ZA2G	IGHG2, CBPN, FA10	–	9
STL	PON1, FIBA, FIBB	LG3BP	CD5L, CHLE, PON1, FA10, ACTG, CBG	A1AT, ITIH3, HEP2, IGLL5, LG3BP	13
WFA	GELS	A1AT, CO9, ALDOA, HV307	-	IGLL5, CO7, FA12	8
Total number	15	15	14	12	

Note: For each lectin, significant proteins with *p*-value < 0.05 and $\log_2$FC > 1 identified in each clinical cohort have been detailed out below. The total number of candidate proteins across each lectin and clinical cohort accounts for overlaps. The proteins are labelled by their UniProt entry name.

The list of protein candidates discovered from UKOPS and UKCTOCS were combined to generate a list of 44 proteins, which fell short of the target number of ~60 candidate proteins that we previously used for biomarker validation [17, 18]. In order to assess additional candidates which may be just outside of the *p* < 0.05 cut-off, we expanded the selection threshold to *p*-value < 0.1 and removed $\log_2$ fold change filtering. This resulted in an initial list of 102 proteins which was filtered down to a final MRM target list of 59 candidates based on suitability of protein tryptic peptides for MRM. The developed custom MRM assay measured 170 peptides from the 59 candidate proteins, with at least three peptides per protein and 4–5 transitions per peptide. The full MRM-MS data are provided in Table S6.

Univariate analysis for HGSOC versus benign and HGSOC versus healthy samples was conducted at the peptide level on each lectin dataset, with significance cut-offs set at adjusted *p*-value < 0.05 and $\log_2$ fold change > 0.5 (Table S7). Overall, for the HGSOC versus benign comparison, we found 53 (AAL), 80 (SNA) and 49 (STL) peptides with an overlap of 15 common peptides that were mapped to seven proteins (Figure 3A, C, E, G). Likewise, for the HGSOC versus healthy comparison, we found 58 (AAL), 88 (SNA) and 74 (STL) peptides, with an overlap of 38 peptides that were mapped to 18 proteins (Figure 3B, D, F, H). To summarise the peptide differential expression data into HGSOC glycoprotein biomarkers, we next looked for consistency in the peptide differential expression. Each glycoprotein candidate was measured by three or more non-glycosylated peptides, but not all peptides showed significance or consistent direction of change, possibly related to proteoforms or protein cleavage. Filtering for consistent direction of change (up/down) across all measured peptides revealed 21%–31% of proteins had all consistent peptides, with five proteins elevated in HGSOC (A1AT, AACT, CO9, HPT and ITIH3), and four down-regulated proteins (A2MG, ALS, IBP3 and PON1) across the three lectins (Table 2). The validated glycoproteins had diverse lectin-

binding affinities, from binding to all three lectins (CO9, A2MG) to a single lectin (AAL-HPT, STL-PON1) (Table 2).

We used STRING v 11.5 to investigate the interactions between the nine validated biomarker proteins. The developed protein-protein interaction (PPI) network had 16 edges (expected 0), and significantly more interactions than expected (PPI enrichment *p*-value of <1 × 10^{-16}) (Figure S4A). Functional enrichment analyses revealed significant enrichments in Gene Ontology Cellular Component terms Blood microparticle (five out of 115 genes, FDR 1.67 × 10^{-6}), extracellular exosome (eight out of 2099 genes, FDR 8.21 1.67 × 10^{-5}), insulin-like growth factor ternary complex (two out of four genes, FDR 0.0048), amongst others, as well as KEGG pathway, Complement and coagulation cascades (three out of 82 genes, FDR 0.0022) (full analysis in Table S8).

3.3 | Development of multi-marker signature for HGSOC

The receiver operating curve (AUC), specificity and sensitivity of the developed multi-peptide models are detailed in Table 3, Table S9. All four models performed similarly with the AAL signature having the highest AUC (87.5%), sensitivity (70.4%) and specificity (90.7%). To further inspect the stable peptides for each of the models, we filtered peptides chosen in at least 50% of the cross-validation runs (Table 3). This analysis revealed several interesting observations. The IBP3 peptide ALAQCAPPPAVCAELVR was always selected for each lectin signature, indicating strong predictive value for HGSOC. Two peptides were highly stable for SNA, STL and the combined signatures, namely, A2MG_NEDSLVFVQTDK and CHLE_NIAAFGGNPK. For the combined signature, both SNA and STL binding IBP3_ALAQCAPPPAVCAELVR were selected with high stability (100% and 91.6%, respectively).
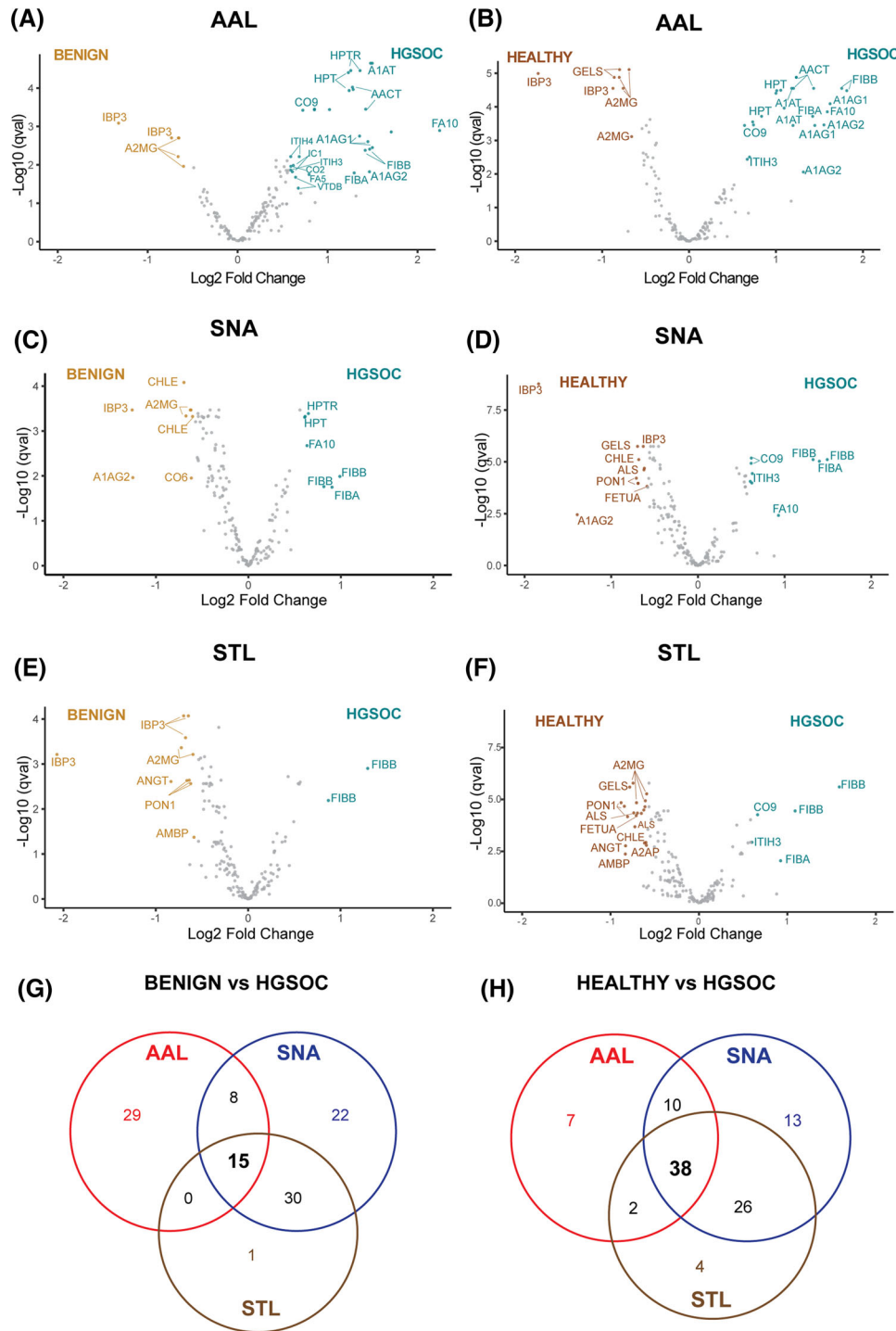
FIGURE 3 Biomarker validation data. LeMBA-MRM data were analysed for differentially abundant peptides between benign and HGSOC (A, C, E, G), and healthy and HGSOC (B, D, F, H) for AAL (A, B), SNA (C, D), and STL (E, F) lectins, respectively. Each dot in the volcano plot indicates a peptide, labelled only by the corresponding UniProt ID for the protein for visualization. The overlap between candidates for each lectin is shown in G, and H.

Strikingly, the stable peptides in the combined signature comprised two SNA and five STL peptides, with no AAL peptides. PPI analysis again revealed significant interactions with enrichment value of 1.29×10^{-7} (Figure S4B), and enrichment of the GO Cellular Component term Blood microparticle (four out of 115 genes, FDR 3.28×10^{-5}), as well as the KEGG pathway Complement and coagulation cascades (four out of 82 genes, FDR 1.74×10^{-6}). Additionally, the GO biological process Blood coagulation, Fibrin clot formation was also highly enriched (four out of 26 genes, FDR 8.62×10^{-7}, Table S8).

TABLE 2 Validated biomarker glycoproteins.

	AAL	SNA	STL
Number of significant peptides	47	41	38
Protein numbers with any significant peptide	23	20	19
Protein numbers with all peptides consistent	5 (21.7%)	6 (30%)	6 (31.6%)
Increased in HGSOC	A1AT, AACT, CO9, HPT	AACT, CO9, ITIH3	CO9, ITIH3
Reduced in HGSOC	A2MG	A2MG, ALS, IBP3	A2MG, ALS, IBP3, PON1

Note: Table shows the number of significant peptides for either HGSOC versus benign or HGSOC versus healthy comparison (q-value < 0.05 and $\log_2$FC > 0.5), the number of proteins with any significant peptide, and the number of proteins with all measured peptides significant and consistent in direction. Protein with all peptides consistent are arranged by alphabetical order of their Uniprot entry name, according to the direction of change in HGSOC.

3.4 | Evaluation of validated biomarker candidates for early HGSOC detections

To determine if any of the nine validated univariate protein biomarkers were altered in the pre-clinical samples, we re-examined the UKCTOCS discovery LeMBA-DDA-MS data set for the nine proteins (Table 2). Three proteins, namely, CO9, ITIH3 and A2MG were significantly altered in the UKCTOCS samples. Figure 4 illustrates the comparative data from discovery UKCTOCS (protein level) and validation (peptide level) phases. AAL-CO9 (Figure 4A) and STL-ITIH3 (Figure 4B) were significantly higher in HGSOC group compared to benign and healthy groups in the UKCTOCS set, while SNA-A2MG was lower in HGSOC group (Figure 4C). STL-CO9 was also elevated in HGSOC versus other groups in the UKCTOCS set, albeit not statistically significant (Figure 4A), while STL-A2MG was not detected in the UKCTOCS or UKOPS datasets likely due to the lower sensitivity of DDA-MS compared to MRM-MS. The three early detection biomarkers CO9, ITIH3 and A2MG interacted in a tight network (enrichment p-value 3.91e−06) (Figure S4C), which was functionally enriched in the KEGG pathway Complement and coagulation cascades (FDR 0.00183) and in the Uniprot annotated keyword of Serine protease inhibitors (FDR 0.0351, Table S8).

4 | DISCUSSION

This study, involving multiple independent clinical and pre-clinical sample sets, provides evidence of glycoproteins as serum biomarkers for HGSOC and lays the foundation for further research on larger patient cohorts. Excitingly, three glycoproteins (CO9, ITIH3 and A2MG) were altered in pre-clinical serum samples collected 11.1 ± 5.1 months prior to HGSOC diagnosis, suggesting promise in early detection. The observed changes in proteins pulled down by the three lectins (AAL, SNA and STL) suggest alterations in α-fucose, sialic acid and N-acetylglucosamine during HGSOC development that require further characterization.

To increase likelihood of successful biomarker development, our glycoprotein-focused biomarker pipeline addressed the issue of tech-nical and biological variations by using LeMBA as a common platform across all phases, and developing multi-marker panels, respectively. The LeMBA-MRM platform also has the advantage of being able to be deployed as a clinical assay [24] reducing the time it takes for the findings to be translated for patient use. Alternatively, lectin-immunoassays can be developed for the discovered biomarkers [25].

We employed a phased biomarker study design to operate within budget [26] where the discovery phase screen uses a relatively small sample size to generate a shortlist of lectins and protein candidates for validation in a larger cohort. In view of the small discovery samples size, our choice of low-stringency statistics on the discovery data, and experimental design to analyse all candidate proteins against the three selected lectins (using the single MRM assay) was ultimately critical for successful biomarker validation. Notably, only one specific lectin-protein discovery phase candidate (STL-PON1, Table 1) was ultimately confirmed in the validation cohort (Table 2). The validated biomarkers were comprised mostly of lectin-protein combinations that were just outside the initial cut-off for the discovery analysis but were analysed in the validation cohort as the custom MRM assay was used on all three selected lectins. This outcome highlights the need for data-specific statistical approaches and considered (non-)use of multiple-testing adjustment in discovery science. Aside from statistics and sample size, biological variation related to evolution of the cancer between the discovery clinical (UKOPS) and validation clinical (AOCS) samples is likely to have contributed to the observed differences.

Strikingly, the validated biomarker proteins physically interact and are functionally enriched in the Complement and coagulation cascades, two components of the innate immune system. This finding is in line with the higher risk of venous thromboembolism (VTE) in cancer patients [27, 28] and the emerging concept of a tumour 'coagulome', a cancer-driven network of molecular effectors favouring thrombosis or bleeding [29, 30]. The analysis of the cancer coagulome found expression of genes encoding six pro-coagulant and fibrinolytic factors (*F3, PLAU, PLAT, PLAUR, SERPINB2*, and *SERPINE1)* in The Cancer Genome Atlas [29], and correlated with VTE incidence in 32 cancer types in a previously reported Dutch study [27]. While a moderate correlation was found between VTE risk and expression of Tissue Factor (F3), a major pro-coagulant factor [29], the authors noted heterogeneity

TABLE 3 Signature peptides from lectin-pulldowns for distinguishing HGSOC from benign/healthy and the stability of each peptide assessed by number of times selected during cross-validations.

Lectin (sample size)	AAL (n = 70)	SNA (n = 95)	STL (n = 95)	Combined (n = 95)
Model performance (range)	AUC = 87.6% (78.4%, 96.8%)	AUC = 84.6% (76.1%, 93.0%)	AUC = 85.0% (76.6%, 93.4%)	AUC = 84.8% (76.4%, 93.3%)
	Sens = 70.4% (51.5%, 84.1%)	Sens = 64.1% (48.4%, 77.3%)	Sens = 64.1% (48.4%, 77.3%)	Sens = 64.1% (48.4%, 77.3%)
	Spec = 90.7% (78.4%, 96.3%)	Spec = 87.5% (76.4%, 93.8%)	Spec = 85.7% (74.3%, 92.6%)	Spec = 89.3% (78.5%, 95.0%)
UniProt Name_Peptide	**Times selected**	**Times selected**	**Times selected**	**Times selected**
IBP3_ALAQCAPPPAVCAELVR	70 (100%)	95 (100%)	95 (100%)	SNA 95 (100%) STL 87 (91.6%)
A1AT_QINDYVEK	70 (100%)			
HPT_VTSIQDWVQK	69 (98.5%)		83 (87.4%)	
IBP3_FLNVLSPR	69 (98.5%)			
GELS_QTQVSVLPEGGETPLFK	68 (97.1%)			
FETUA_HTLNQIDEVK	64 (91.4%)			
FIBB_YQJSVNK	46 (65.7%)			
FA10_ETYDFDIAVLR		94 (98.9%)		
A2MG_NEDSLVFVQTDK		93 (97.9%)	95 (100%)	STL 95 (100%)
CHLE_NIAAFGGNPK		93 (97.9%)	67 (70.5%)	SNA 92 (96.8%)
KNG1_ENFLFLTPDCK		93 (97.9%)		
ITIH2_VQSTITSR		90 (94.7%)		
ITIH3_EHLVQATPENLQEAR		86 (90.5%)	91 (95.8%)	
C1QC_QTHQPPAPNSLIR		85 (89.5%)		
HPT_DYAEVGR		84 (88.4%)		
THRB_SGIECQLWR		79 (83.2%)		
KNG1_EGDCPVQSGK			95 (100%)	STL 95 (100%)
FIBB_QDGSVDFGR			78 (82.1%)	
FA10_IVGGQECK				STL 66 (69.5%)
FIBB_SILENLR				STL 60 (63.2%)

Abbreviations: AUC, area under the receiver operating curve; Sens, sensitivity; Spec, specificity.

(A) COMPLEMENT COMPONENT C9 (CO9)
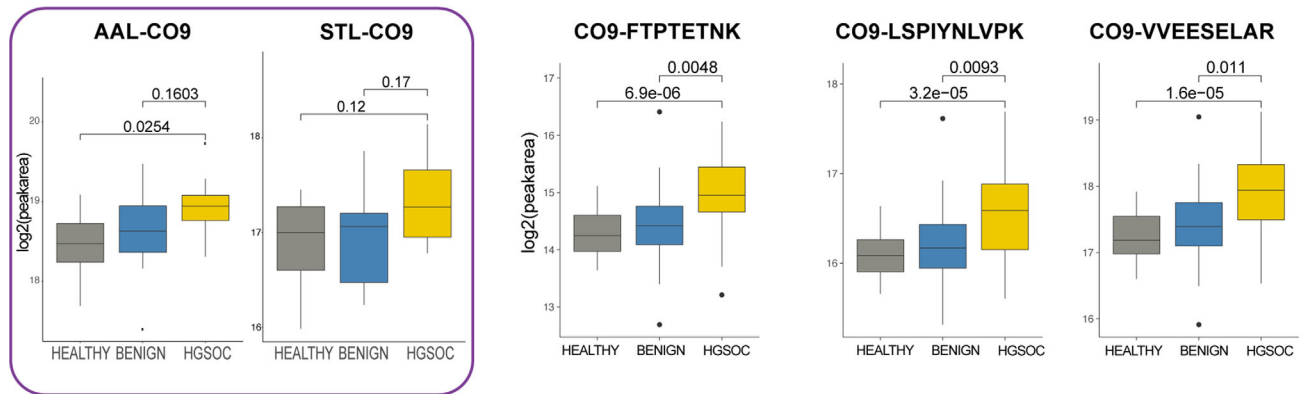
(B) INTER-ALPHA-TRYPSIN-INHIBITOR (ITIH3)
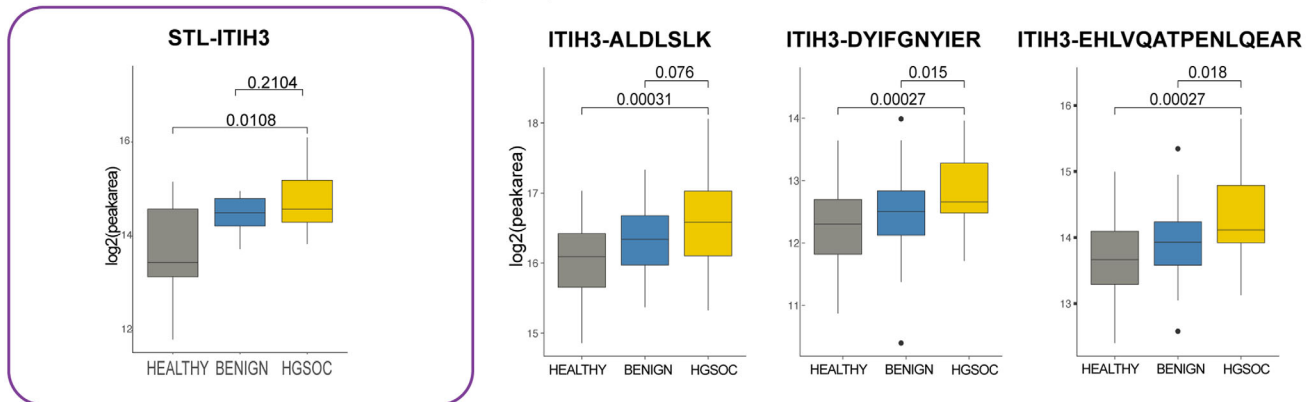
(C) ALPHA-2-MACROGLOBULIN (A2MG)
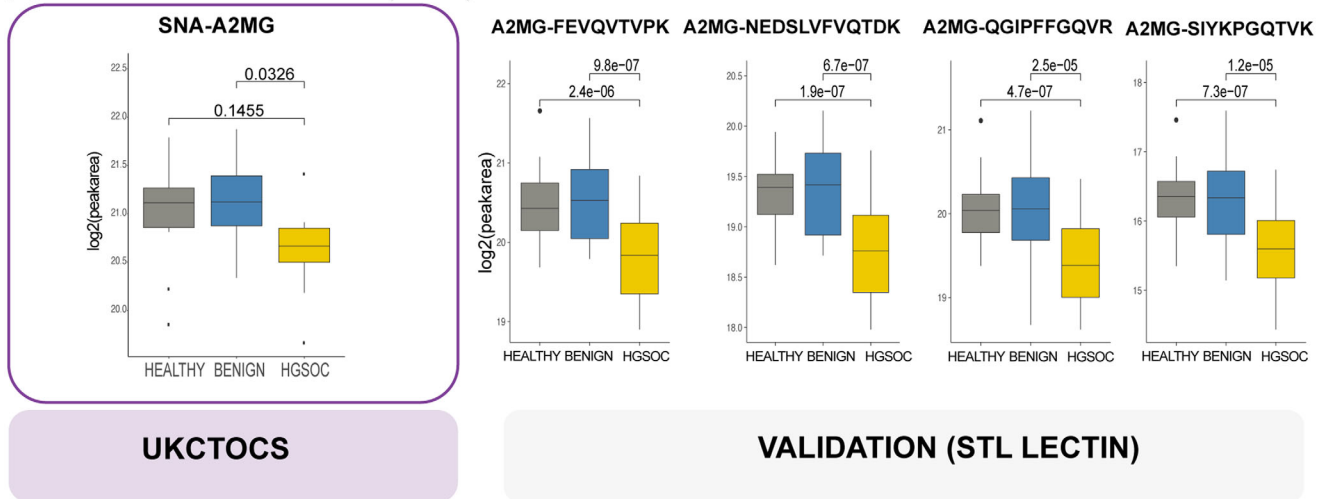
FIGURE 4 Potential HGSOC early detection biomarkers. Comparison of glycoprotein biomarker data between UKCTOCS discovery case control set (left panel, LeMBA-DDA-MS protein level data), and STL-pulldown validation case control set (right panel, LeMBA-MRM-MS, peptide level data) for (A) CO9, (B) ITIH3, and (C) A2MG. For comparison, the unadjusted *p*-values are shown for both data sets.

across tumour types. Intriguingly, while ovarian cancer had one of the highest VTE incidences in this study, the six examined genes showed moderate expression levels in this cancer [29], indicating important roles of other mechanisms and procoagulant factors. Indeed, our findings identified altered glycosylation of two additional proteins not yet considered in the cancer coagulome, namely, thrombin (THRB) and coagulation factor X (FA10, Table 3), which should be evaluated for inclusion in the coagulome.

Proteins of the Complement and coagulation cascades interact at several levels, share common regulator proteins and both systems

act on immune and endothelial cells [31, 32]. Interestingly, both cascades are activated or regulated by extracellular vesicles [30, 32], small membrane-enclosed vesicles released by cells for inter-cellular communication. Our biomarkers were enriched in the Gene Ontology term 'Blood microparticles', defined as a type of extracellular vesicle (EV) devoid of nucleic acids, released from several cell types including platelets, blood cells and endothelial cells. This finding is supported by two previous independent cohort studies reporting elevated serum/plasma EV procoagulant activity in ovarian cancer patients [33, 34]. While a recent meta-analysis reports VTE to be higher in advanced serous ovarian cancer, clear cell histology and ascites at diagnosis [35], our study has found evidence of perturbations in the glycosylation of Complement and coagulation cascade several months prior to ovarian cancer diagnosis.

We were highly encouraged to find three validated biomarkers to be altered several months prior to cancer diagnosis in the UKCTOCS set. Levels of two candidates were elevated (glycoforms of C9 and ITIH3) and one was lower (SNA-A2MG) compared to healthy and benign samples. All three proteins are associated with inflammation, with C9 and A2MG having known roles in the Complement and coagulation pathways. Of the three potential early detection biomarkers, only A2MG has previously been reported to be altered in ovarian cancer. Similar to our findings, Miyamoto et al. reported decreased A2MG protein in ovarian cancer patients compared to healthy controls [36]. A2MG is a broad-spectrum protease inhibitor which inhibits thrombin and the complement pathway [37]. Reduced A2MG levels may indicate elevated Coagulation and complement pathway activity, although the role of the SNA-A2MG glycoform is currently unknown.

C9 is the terminal Component of the complement cascade, which has also been found to be elevated in serum of gastric [38], lung [39], colorectal [40, 41] and esophageal cancers [17, 18] through proteomics or glycoproteomic approaches. In our esophageal adenocarcinoma biomarker study, C9 glycoforms binding to each of the six short-listed lectins (AAL, EPHA, JAC, NPL, PSA, WGA) was also significantly elevated in esophageal adenocarcinoma compared to the precursor benign condition, Barrett's esophagus but with some variability in healthy groups [18]. The current study shortlisted three lectins (AAL, SNA, STL) for the validation phase, and all three C9 glycoforms measured were significantly elevated in HGSOC, suggesting that C9 is subjected to increased aberrant glycosylation that may contribute to disease progression. Interestingly, AAL-C9 glycoform levels have been reported to be elevated in lung and stomach cancers, intermediate in hepatocellular carcinoma and low in breast cancer [39]. Recently, we reported the release of C9$^+$ EVs by esophageal adenocarcinoma cells as a potential mechanism of the elevated serum C9 glycoform in esophageal cancer [42]. However, the specific glycosylation differences in cancer serum and EVs, as well as the molecular mechanisms underpinning the altered C9 glycosylation in different cancer types remains to be determined.

Inter-alpha-trypsin inhibitor family members, including ITIH3, have been implicated in inflammation and carcinogenesis [43]. In addition to its protease inhibitor activity, ITIH3 is thought to stabilize the extracellular matrix through binding to hyaluronic acid [44]. Interestingly,

previous reports suggest blood ITIH3 to be elevated for a similar range of cancers as C9, namely, lung [45], gastric [46], pancreatic [47] and colorectal [48] cancers.

In conclusion, we report the discovery and validation of serum glycoprotein markers for HGSOC using a lectin-assisted proteomics workflow that is directly translatable to blood tests. The validated markers show high specificity when bench-marked against the existing ovarian cancer biomarker, CA125. Their utility in ovarian cancer diagnosis and monitoring will need to be evaluated in additional cohorts, such as at diagnosis and following surgery/chemotherapy. Furthermore, several markers were elevated months prior to cancer diagnosis, and should be further evaluated as for ovarian cancer screening. Functional enrichment of the validated markers highlights blood microparticle (EV)-mediated complement and coagulation activity ahead of clinical diagnosis of HGSOC. Further investigation on the contribution of EV-mediated complement and coagulation in ovarian cancer development may provide mechanisms for prevention.

5 | ASSOCIATED DATA

The MS proteomics data for the discovery phase have been deposited to the ProteomeXchange Consortium via the PRIDE [49] partner repository with the dataset identifier PXD032299. The targeted mass spectrometry data for the validation phase have been deposited to PRIDE via Panorama Public database with the dataset identifier PXD033108.

CONFLICT OF INTEREST STATEMENT

Usha Menon had shares (2011–2020) awarded by UCL in Abcodia which had an interest in early detection of cancer. SA had salary support from Abcodia between 2011 and 2021. Usha Menon, Sophia Apostolidou and Aleksandra Gentry-Maharaj are involved in institutional research collaborations with iLOF, Micronoma, RNA Guardian, Mercy Bioanalytics and Synteny Biotechnology. All other authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The mass spectrometry proteomics data for the discovery phase have been deposited to the ProteomeXchange Consortium via the PRIDE [49] partner repository with the dataset identifier PXD032299. The targeted mass spectrometry data for the validation phase have been deposited to PRIDE via Panorama Public database with the dataset identifier PXD033108.

REFERENCES

1. Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, *70*, 7–30. https://doi.org/10.3322/caac.21590
2. Menon, U., Gentry-Maharaj, A., Burnell, M., Singh, N., Ryan, A., Karpinskyj, C., Carlino, G., Taylor, J., Massingham, S. K., Raikou, M., Kalsi, J. K., Woolas, R., Manchanda, R., Arora, R., Casey, L., Dawnay, A., Dobbs, S., Leeson, S., Mould, T., … Parmar, M. (2021). Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *Lancet*, *397*, 2182–2193. https://doi.org/10.1016/S0140-6736(21)00731-5
3. Pinsky, P. F., Yu, K., Kramer, B. S., Black, A., Buys, S. S., Partridge, E., Gohagan, J., Berg, C. D., & Prorok, P. C. (2016). Extended mortality results for ovarian cancer screening in the PLCO trial with median 15 years follow-up. *Gynecologic Oncology*, *143*, 270–275. https://doi.org/10.1016/j.ygyno.2016.08.334
4. Henderson, J. T., Webber, E. M., & Sawaya, G. F. (2018). Screening for ovarian cancer. *Jama*, *319*, 595–606. https://doi.org/10.1001/jama.2017.21421
5. Nash, Z., & Menon, U. (2020). Ovarian cancer screening: Current status and future directions. *Best Practice & Research. Clinical Obstetrics & Gynaecology*, *65*, 32–45. https://doi.org/10.1016/j.bpobgyn.2020.02.010
6. Pinho, S. S., & Reis, C. A. (2015). Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer*, *15*, 540–555. https://doi.org/10.1038/nrc3982
7. Chen, K., Gentry-Maharaj, A., Burnell, M., Steentoft, C., Marcos-Silva, L., Mandel, U., Jacobs, I., Dawnay, A., Menon, U., & Blixt, O. (2013). Microarray glycoprofiling of CA125 improves differential diagnosis of ovarian cancer. *Journal of Proteome Research*, *12*, 1408–1418. https://doi.org/10.1021/pr3010474
8. Gidwani, K., Huhtinen, K., Kekki, H., Van Vliet, S., Hynninen, J., Koivuviita, N., Perheentupa, A., Poutanen, M., Auranen, A., Grenman, S., Lamminmäki, U., Carpen, O., Van Kooyk, Y., & Pettersson, K. (2016). A Nanoparticle-Lectin Immunoassay Improves Discrimination of Serum CA125 from Malignant and Benign Sources. *Clinical Chemistry*, *62*, 1390–1400. https://doi.org/10.1373/clinchem.2016.257691
9. Bayoumy, S., Hyytiä, H., Leivo, J., Talha, S. M., Huhtinen, K., Poutanen, M., Hynninen, J., Perheentupa, A., Lamminmäki, U., Gidwani, K., & Pettersson, K. (2020). Glycovariant-based lateral flow immunoassay to detect ovarian cancer–associated serum CA125. *Communications Biology*, *3*, 460. https://doi.org/10.1038/s42003-020-01191-x
10. Biskup, K., Braicu, E. I., Sehouli, J., Fotopoulou, C., Tauber, R., Berger, M., & Blanchard, V. (2013). Serum glycome profiling: a biomarker for diagnosis of ovarian cancer. *Journal of Proteome Research*, *12*, 4056–4063. https://doi.org/10.1021/pr400405x
11. Dědová, T., Braicu, E. I., Sehouli, J., & Blanchard, V. (2019). Sialic acid linkage analysis refines the diagnosis of ovarian cancer. *Frontiers in Oncology*, *9*, 261. https://doi.org/10.3389/fonc.2019.00261
12. Hu, Y., Pan, J., Shah, P., Ao, M., Thomas, S. N., Liu, Y., Chen, L., Schnaubelt, M., Clark, D. J., Rodriguez, H., Boja, E. S., Hiltke, T., Kinsinger, C. R., Rodland, K. D., Li, Q. K., Qian, J., Zhang, Z., Chan, D. W., Zhang, H., … Shi, Z. (2020). Integrated proteomic and glycoproteomic characterization of human high-grade serous ovarian carcinoma. *Cell Reports*, *33*, 108276.
13. Shetty, V., Hafner, J., Shah, P., Nickens, Z., & Philip, R. (2012). Investigation of ovarian cancer associated sialylation changes in N-linked glycopeptides by quantitative proteomics. *Clinical Proteomics*, *9*, 10. https://doi.org/10.1186/1559-0275-9-10
14. Sinha, A., Hussain, A., Ignatchenko, V., Ignatchenko, A., Tang, K. H., Ho, V. W. H., Neel, B. G., Clarke, B., Bernardini, M. Q., & Ailles, L., Kislinger, T. (2019). N-glycoproteomics of patient-derived xenografts: A strategy to discover tumor-associated proteins in high-grade serous ovarian cancer. *Cell Systems*, *8*, 345–351.e4 e344.
15. Wu, J., Xie, X., Liu, Y., He, J., Benitez, R., Buckanovich, R. J., & Lubman, D. M. (2012). Identification and confirmation of differentially expressed fucosylated glycoproteins in the serum of ovarian cancer patients using a lectin array and LC–MS/MS. *Journal of Proteome Research*, *11*, 4541–4552. https://doi.org/10.1021/pr300330z
16. Choi, E., Loo, D., Dennis, J. W., O'leary, C. A., & Hill, M. M. (2011). High-throughput lectin magnetic bead array-coupled tandem mass spectrometry for glycoprotein biomarker discovery. *Electrophoresis*, *32*, 3564–3575. https://doi.org/10.1002/elps.201100341
17. Shah, A. K., Hartel, G., Brown, I., Winterford, C., Na, R., Cao, K.-A. L., Spicer, B. A., Dunstone, M. A., Phillips, W. A., Lord, R. V., Barbour, A. P., Watson, D. I., Joshi, V., Whiteman, D. C., & Hill, M. M. (2018). Evaluation of serum glycoprotein biomarker candidates for detection of esophageal adenocarcinoma and surveillance of Barrett's esophagus. *Molecular & Cellular Proteomics*, *17*, 2324–2334. https://doi.org/10.1074/mcp.RA118.000734
18. Shah, A. K., Cao, K.-A. L., Choi, E., Chen, D., Gautier, B., Nancarrow, D., Whiteman, D. C., Saunders, N. A., Barbour, A. P., Joshi, V., & Hill, M. M. (2015). Serum glycoprotein biomarker discovery and qualification pipeline reveals novel diagnostic biomarker candidates for esophageal adenocarcinoma. *Molecular & Cellular Proteomics*, *14*, 3023–3039. https://doi.org/10.1074/mcp.M115.050922
19. Oungsakul, P., Choi, E., Shah, A. K., Mohamed, A., O'leary, C., Duffy, D., Hill, M. M., & Bielefeldt-Ohmann, H. (2021). Candidate glycoprotein biomarkers for canine visceral hemangiosarcoma and validation using semi-quantitative lectin/immunohistochemical assays. *Veterinary Sciences*, *8*, 38. https://doi.org/10.3390/vetsci8030038
20. Menon, U., Gentry-Maharaj, A., Hallett, R., Ryan, A., Burnell, M., Sharma, A., Lewis, S., Davies, S., Philpott, S., Lopes, A., Godfrey, K., Oram, D., Herod, J., Williamson, K., Seif, M. W., Scott, I., Mould, T., Woolas, R., Murdoch, J., … Jacobs, I. (2009). Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage

distribution of detected cancers: Results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *The Lancet Oncology*, 10, 327–340. https://doi.org/10.1016/S1470-2045(09)70026-9

21. Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26, 1367–1372. https://doi.org/10.1038/nbt.1511

22. Kim, H., Golub, G. H., & Park, H. (2005). Missing value estimation for DNA microarray gene expression data: Local least squares imputation. *Bioinformatics*, 21, 187–198. https://doi.org/10.1093/bioinformatics/bth499

23. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43, e47. https://doi.org/10.1093/nar/gkv007

24. Bringans, S., Duong, M., Laming, C., Di Prinzio, P., Hill, M., & Lipscombe, R. (2022). Establishing a mass spectrometry based diagnostic test for oesophageal cancer. *Diseases of the Esophagus*, 35, doac051203. https://doi.org/10.1093/dote/doac051.203

25. Webster, J. A., Wuethrich, A., Shanmugasundaram, K. B., Richards, R. S., Zelek, W. M., Shah, A. K., Gordon, L. G., Kendall, B. J., Hartel, G., Morgan, B. P, Trau, M., & Hill, M. M. (2021). Development of endo-screen chip, a microfluidic pre-endoscopy triage test for esophageal adenocarcinoma. *Cancers (Basel)*, 13, 2865.

26. Rifai, N., Gillette, M. A., & Carr, S. A. (2006). Protein biomarker discovery and validation: The long and uncertain path to clinical utility. *Nature Biotechnology*, 24, 971–983. https://doi.org/10.1038/nbt1235

27. Blom, J. W., Vanderschoot, J. P. M., Oostindiër, M. J., Osanto, S., Van Der Meer, F. J. M., & Rosendaal, F. R. (2006). Incidence of venous thrombosis in a large cohort of 66 329 cancer patients: Results of a record linkage study. *Journal of Thrombosis and Haemostasis*, 4, 529–535. https://doi.org/10.1111/j.1538-7836.2006.01804.x

28. Stein, P. D., Beemath, A., Meyers, F. A., Skaf, E., Sanchez, J., & Olson, R. E. (2006). Incidence of venous thromboembolism in patients hospitalized with cancer. *American Journal of Medicine*, 119, 60–68. https://doi.org/10.1016/j.amjmed.2005.06.058

29. Saidak, Z., Soudet, S., Lottin, M., Salle, V., Sevestre, M.-A., Clatot, F., & Galmiche, A. (2021). A pan-cancer analysis of the human tumor coagulome and its link to the tumor immune microenvironment. *Cancer Immunology, Immunotherapy*, 70, 923–933. https://doi.org/10.1007/s00262-020-02739-w

30. Galmiche, A., Rak, J., Roumenina, L. T., & Saidak, Z. (2022). Coagulome and the tumor microenvironment: an actionable interplay. *Trends in Cancer*, 8, 369–383. https://doi.org/10.1016/j.trecan.2021.12.008

31. Oncul, S., & Afshar-Kharghan, V. (2020). The interaction between the complement system and hemostatic factors. *Current Opinion in Hematology*, 27, 341–352. https://doi.org/10.1097/MOH.0000000000000605

32. Ratajczak, M. Z., & Ratajczak, J. (2021). Innate immunity communicates using the language of extracellular microvesicles. *Stem Cell Reviews and Reports*, 17, 502–510. https://doi.org/10.1007/s12015-021-10138-6

33. Claussen, C., Rausch, A.-V., Lezius, S., Amirkhosravi, A., Davila, M., Francis, J. L., Hisada, Y. M., Mackman, N., Bokemeyer, C., Schmalfeldt, B., Mahner, S., & Langer, F. (2016). Microvesicle-associated tissue factor procoagulant activity for the preoperative diagnosis of ovarian cancer. *Thrombosis Research*, 141, 39–48. https://doi.org/10.1016/j.thromres.2016.03.002

34. Zhang, W., Peng, P., Ou, X., Shen, K., & Wu, X. (2019). Ovarian cancer circulating extracelluar vesicles promote coagulation and have a potential in diagnosis: An iTRAQ based proteomic analysis. *BMC Cancer*, 19, 1095. https://doi.org/10.1186/s12885-019-6176-1

35. Weeks, K. S., Herbach, E., Mcdonald, M., Charlton, M., & Schweizer, M. L. (2020). Ovarian cancer patients by stage, histology, cytoreduction, and ascites at diagnosis. *Obstetrics and Gynecology International*, 2020, 1.

36. Miyamoto, S., Stroble, C. D., Taylor, S., Hong, Q., Lebrilla, C. B., Leiserowitz, G. S., Kim, K., & Ruhaak, L. R (2018). Multiple reaction monitoring for the quantitation of serum protein glycosylation profiles: application to ovarian cancer. *Journal of Proteome Research*, 17, 222–233. https://doi.org/10.1021/acs.jproteome.7b00541

37. Vandooren, J., & Itoh, Y. (2021). Alpha-2-macroglobulin in inflammation, immunity and infections. *Frontiers in Immunology*, 12, 803244.

38. Chong, P.-K., Lee, H., Loh, M. C. S., Choong, L.-Y., Lin, Q., So, J. B. Y., Lim, K. H., Soo, R. A., Yong, W. P., Chan, S. P., Smoot, D. T., Ashktorab, H., Yeoh, K. G., & Lim, Y. P. (2010). Upregulation of plasma C9 protein in gastric cancer patients. *Proteomics*, 10, 3210–3221. https://doi.org/10.1002/pmic.201000127

39. Narayanasamy, A., Ahn, J.-M., Sung, H.-J., Kong, D.-H., Ha, K.-S., Lee, S.-Y., & Cho, J.-Y. (2011). Fucosylated glycoproteomic approach to identify a complement component 9 associated with squamous cell lung cancer (SQLC). *Journal of Proteomics*, 74, 2948–2958. https://doi.org/10.1016/j.jprot.2011.07.019

40. Chantaraamporn, J., Champattanachai, V., Khongmanee, A., Verathamjamras, C., Prasongsook, N., Mingkwan, K., Luevisadpibul, V., Chutipongtanate, S., & Svasti, J. (2020). Glycoproteomic analysis reveals aberrant expression of complement C9 and fibronectin in the plasma of patients with colorectal cancer. *Proteomes*, 8, 26.

41. Murakoshi, Y., Honda, K., Sasazuki, S., Ono, M., Negishi, A., Matsubara, J., Sakuma, T., Kuwabara, H., Nakamori, S., Sata, N., Nagai, H., Ioka, T., Okusaka, T., Kosuge, T., Shimahara, M., Yasunami, Y., Ino, Y., Tsuchida, A., Aoki, T., … Yamada, T. (2011). Plasma biomarker discovery and validation for colorectal cancer by quantitative shotgun mass spectrometry and protein microarray. *Cancer Science*, 102, 630–638. https://doi.org/10.1111/j.1349-7006.2010.01818.x

42. Kolka, C. M., Webster, J., Lepletier, A., Winterford, C., Brown, I., Richards, R. S., Zelek, W. M., Cao, Y., Khamis, R., Shanmugasundaram, K. B., Wuethrich, A., Trau, M., Brosda, S., Barbour, A., Shah, A. K., Eslick, G. D., Clemons, N. J., Morgan, B. P, & Hill, M. M. (2022). C5b-9 membrane attack complex formation and extracellular vesicle shedding in barrett's esophagus and esophageal adenocarcinoma. *Frontiers in Immunology*, 13, 842023.

43. Hamm, A., Veeck, J., Bektas, N., Wild, P. J., Hartmann, A., Heindrichs, U., Kristiansen, G., Werbowetski-Ogilvie, T., Del Maestro, R., Knuechel, R., & Dahl, E. (2008). Frequent expression loss of Inter-alpha-trypsin inhibitor heavy chain (ITIH) genes in multiple human solid tumors: a systematic expression analysis. *BMC Cancer*, 8, 25. https://doi.org/10.1186/1471-2407-8-25

44. Huang, L., Yoneda, M., & Kimata, K. (1993). A serum-derived hyaluronan-associated protein (SHAP) is the heavy chain of the inter alpha-trypsin inhibitor. *Journal of Biological Chemistry*, 268, 26725–26730. https://doi.org/10.1016/S0021-9258(19)74373-7

45. Heo, S.-H., Lee, S.-J., Ryoo, H.-M., Park, J.-Y., & Cho, J.-Y. (2007). Identification of putative serum glycoprotein biomarkers for human lung adenocarcinoma by multilectin affinity chromatography and LC-MS/MS. *Proteomics*, 7, 4292–4302. https://doi.org/10.1002/pmic.200700433

46. Chong, P. K., Lee, H., Zhou, J., Liu, S.-C., Loh, M. C. S., Wang, T. T., Chan, S. P., Smoot, D. T., Ashktorab, H., So, J. B. Y., Lim, K. H., Yeoh, K. G., & Lim, Y. P. (2010). ITIH3 is a potential biomarker for early detection of gastric cancer. *Journal of Proteome Research*, 9, 3671–3679. https://doi.org/10.1021/pr100192h

47. Liu, X., Zheng, W., Wang, W., Shen, H., Liu, L., Lou, W., Wang, X., & Yang, P. (2017). A new panel of pancreatic cancer biomarkers discovered using a mass spectrometry-based pipeline. *British Journal of Cancer*, 117, 1846–1854. https://doi.org/10.1038/bjc.2017.365

48. Kopylov, A. T., Stepanov, A. A., Malsagova, K. A., Soni, D., Kushlinsky, N. E., Enikeev, D. V., Potoldykova, N. V., Lisitsa, A. V., & Kaysheva, A. L. (2020). Revelation of proteomic indicators for colorectal cancer in initial stages of development. *Molecules (Basel, Switzerland)*, 25, 619.

49. Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D. J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., Walzer, M., Wang, S., Brazma, A., & Vizcaíno, J. A. (2022). The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Research*, *50*, D543–D552. https://doi.org/10.1093/nar/gkab1038

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.