Targeted Proteomics and Support Vector Classification Reveal Potential Biomarkers for the Early Detection of High-grade Serous Ovarian Cancer

Tyler T. Cooper[1*], Dylan Z. Dieters-Castator[2*], Jiahui Liu[3], Gabrielle M. Siegers[3], Desmond Pink[3], John D. Lewis[3], Yangxin Fu[3], Helen Steed[4], Gilles A. Lajoie[2†], and Lynne-Marie Postovit[3,4,5†].

1Department of Biochemistry, Western University, London ON, Canada;
2Department of Anatomy and Cell Biology, Western University, London, ON, Canada;
3Department of Oncology, University of Alberta, Edmonton AB, Canada;
4Department of Obstetrics and Gynecology, University of Alberta, Edmonton AB, Canada;
5Department of Biomedical and Molecular Sciences, Queen's University, Kingston ON, Canada
*Authors contributed equally.
†To whom correspondence should be addressed.
E-mail: postovit.lynne@uqueens.ca

Date: March 31st, 2022

Abstract

The 5-year prognosis of late-stage epithelial ovarian cancer (EOC) remains poor, thus the discovery of early-stage EOC biomarkers is of paramount importance. Extracellular vesicles (EVs) circulating in blood are thought to contain proteomic cargo originating from an EOC microenvironment and are thus amenable for clinical biomarker discovery. We profiled the proteome of EVs purified from patient blood plasma, ascites and cell lines using strong cation exchange peptide fractionation and Orbitrap-based tandem mass spectrometry. To further increase sensitivity and specificity of the method, CD9-affinity purification and ultracentrifugation were used to purify EVs. Using parallel reaction monitoring we identified a compendum of 240 proteins that were differentially enrirched in EVs derived from EOC (n=10) patients versus women with non-cancerous gynecological conditions (n=9). Support vector machines were optimized using leave-one-out cross-validation and this methodology was implemented on a test set of malignant (n=4) and control (n=3) donors. Using the relative levels of >450 EV-associated peptides in a cohort of plasma-derived EVs, we identified several combinatorial peptides capable of discriminating high-grade serous EOC with up to 100% accuracy in Stage I, II, and III donors. This study demonstrates an adaptable biomarker discovery pipeline and provides pinoeering evidence of EV-associated biomarkers for the detection of early-stage EOC.

Keywords: High-grade Serous Carcinoma, Extracellular Vesicles, Proteomics, Biomarkers, Ovarian Cancer

1. Introduction

Despite an increasing understanding of epithelial ovarian cancer (EOC) etiology and biology, EOC remains the most lethal gynecological cancer in developed countries[1]. It is estimated that >200,000 women per year will be diagnosed worldwide, and 5-year survival rates below 50% will lead to >100,000 deaths[2]. Early detection of EOC is crucial to improving survival, with 92% and 29% of patients surviving following early versus late-stage detection, respectively[3]. Unfortunately, 75% of women remain asymptomatic until diagnosis in late stages and experience non-specific symptoms (e.g. abdominal discomfort) that may lead to the identification of pelvic masses by transvaginal ultrasound (TVUS) imaging. If abnormal masses are identified, invasive surgical procedures, tissue debulking, and pathohistological analyses are then required to discriminate between benign and malignant disease[1]. High-grade serous carcinoma (HGSC) is the most lethal and aggresive form of EOC, accounting for >75% of EOC cases. The extracellular epitope of MUC16 (CA-125) can be used to monitor the progression of EOC and response to chemotherapeutics in combination with TVUS[4-6]. Unfortunately, tests for CA-125 are not sensitive nor specific enough for early diagnosis of malignant EOC[7]. For example, although ~20% of patients

with late-stage EOC exhibited elevated CA-125 levels (>35 U/mL), increased CA-125 was also observed in women with alternative gynecological conditions[7]. Thus, there remains a dire need to discover alternative biomarkers to aid in the early detection of EOC.

Algorithms, such as the Risk of Malignancy Index (RMI), aim to incorporate menopausal status, CA-125 levels and TVUS imaging[8, 9]. Alternatively, the Risk of Ovarian Cancer Algorithm (ROCA) monitors CA-125 levels over time to assess the risk of developing ovarian cancer. Unfortunately, large randomized control trials (US Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial and UK Collaborative Trial of Ovarian Cancer Screening) involving thousands of females found no significant survival benefit for multimodal screening strategies over standard of care[4-6]. Alternative biomarkers to CA-125 have been proposed for estimating EOC risk. For example, the risk of ovarian malignancy algorithm (ROMA) monitors human epididymis protein 4 (HE4 or WFDC2) in addition to CA-125 [10]. The FDA-approved OVA1 *in vitro* diagnostic multivariate index assay measures five biomarkers (CA-125-II, transferrin [TF], transthyretin (prealbumin), apolipoprotein A1 [APOA1], and beta-2 microglobulin [B2M]) and demonstrates improved prediction accuracy of malignancy risk compared to a physician's pre-operative assessment or CA-125 alone[11]. Moreover, Yip *et al.* screened 259 serum biomarkers from EOC patients and identified nine combinatorial biomarkers with greater specificity than OVA1 (88.9 versus 63.4%)[12]. Høgdall *et al.* screened serum from 150 EOC patients and found B2M, TF, and inter-alpha-trypsin inhibitor heavy chain H4 (ITIH4) robustly predicted overall survival and progression-free survival[13]. These approaches improve cancer classification and monitoring strategies; however, viable biomarkers that are capable of detecting early-stage HGSC are still unavailable.

Blood plasma remains an ideal source for biomarker discovery due to the easy acquisition of patient samples for high-throughput immunoassays. Mass spectrometry (MS)-based proteomics is a medium-throughput technique for biomarker discovery; however, the detection of low abundance proteins in plasma is technically complicated by the presence of high abundance proteins (HAPs) [14-16]. Keshishian *et al*. detected ~5300 plasma proteins by depleting the 14 most abundant plasma proteins as well as ~50 moderately abundant proteins in tandem with high-pH reversed phase fractionation[17]. Alternatively, N-glycopeptide enrichment was recently shown to identify plasma proteins for detecting early ovarian cancer and relapse[18]. It remains to be determined what the optimal strategy is for segregating biomarkers from HAP in primary tissue samples. Extracellular vesicles (EVs), 40-1000nm in diameter, carry bioactive lipid, nucleic acid and proteomic cargo in a lipid membrane that allows for transport through systemic circulation to distant tissues[19]. EVs carry bioactive cargo from or towards a metastatic cancer microenvironment[20], thus the enrichment of EVs may segregate potential biomarkers from HAPs or other liable plasma proteins[21]. A limited number of investigations have attempted to characterize EOC-EV proteomes using biofluids[22].

Herein, we provide evidence of potential biomarkers identified from plasma EVs from donors with malignant EOC (HGSC) using targeted label-free proteomics and support vector machine (SVM) optimization for the classification of HSGC vs non-cancerous donors with clinical presenation of gyneological ailments related to HGSC (e.g. abdominal pain). EV proteomes obtained from cell lines, plasma and ascites fluid samples identified >200 perspective biomarkers associated with plasma EVs in a proof-of-principle study. Label-free parallel reaction monitoring (PRM) proteomics, leave-one-out cross validation (LOOCV) and SVM identified nine peptide combinations that classified malignant EOC vs non-cancerous gynecological conditions with 100% accuracy on a test-set (n=4) containing FIGO stage I, II, III EOC donors. Collectively, this data confirms EV habour prospective biomarkers for early-stage malignant EOC detection.

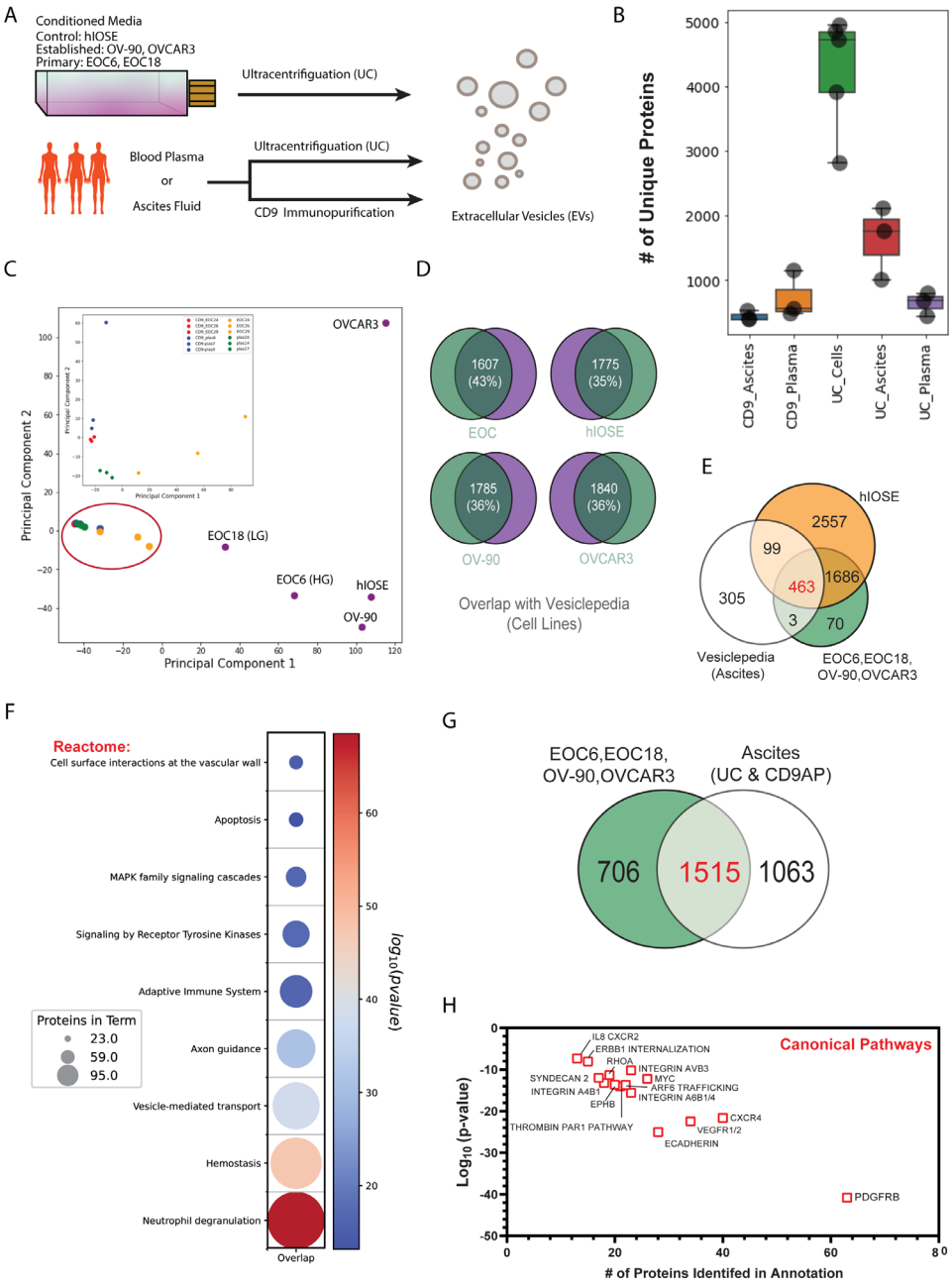

Figure 1. EVs released from established and primary ovarian cancer cell lines harbour disease relevant cargo. (A) EVs were enriched from conditioned media of established or primary cell lines. Alternatively, EVs were enriched from blood plasma or ascites fluid using UC or CD9AP. (B) The number of unique proteins identified was elevated in EVs derived from cell lines compared to EVs enriched from primary sources. Data represented as box plot quartiles (C) Principal component analysis illustrating distinct proteomic 'fingerprints' within cell line-derived EVs and comparisons between EV source or enrichment technique. Biofluid EVs (highlighted by red oval) contain distinct proteomes based on source and method of purification demonstrated by embedded PCA plot. (D) Overlap of EV-proteomes compared to Vesiclepedia database filtered by ovarian cancer cell lines. (E) Overlap of EV proteome of hIOSE cells, cancer cell lines, and Vesiclepedia database filtered for ascites from donors with malignant cancer. (F) Overlapping proteins were enriched with components of neutrophil degranulation according to Reactome database. Size of circles indicate number of proteins identified in each term. (G) Cancer cell lines were compared to primary ascites EVs enriched by UC or CD9AP and overlapping proteins were cross-referenced to (H) Canonical Pathway database. Specifically, cell line and ascites EVs contained more than 20 proteins were identified for Integrin ,E-Cadherin, VEGFR1/2, CXCR4, and PDGFRB signaling axes.

2. Results

2.1 Integrative Proteomic Analysis of Ovarian Cancer Extracellular Vesicles. We first undertook an MS-based approach to characterize EV proteomes from cancer cell lines, healthy donor plasma, and donor ascites fluid for biomarker discovery. Primary (EOC18, EOC6) and established (OVCAR3, OV-90) cell lines were used to model EOC, and a non-malignant ovarian surface epithelial cell line (hIOSE) was also analysed. Plasma from 6 donors (plas 6, 7, 9, 10, 14, 17) and ascites from 3 donors (EOC24, EOC26, EOC29) were also used. EVs were primarily obtained by UC; however, CD9 affinity purification (CD9AP) was also performed on donor plasma and ascites to enrich for smaller EVs **(Fig 1A)**. Notably, EVs derived with CD9 versus UC were matched for the ascites samples. While one may expect that EOC cell lines can not entirely recapitulate the milieu of a tumour microenvironment, EVs derived from cell lines have not yet been compared to patient-derived EVs[23, 24]. SCX fractionation was employed to increase proteomic depth prior LC-MS/MS in all samples. Using this approach, the number of proteins identified in cell line EVs was increased

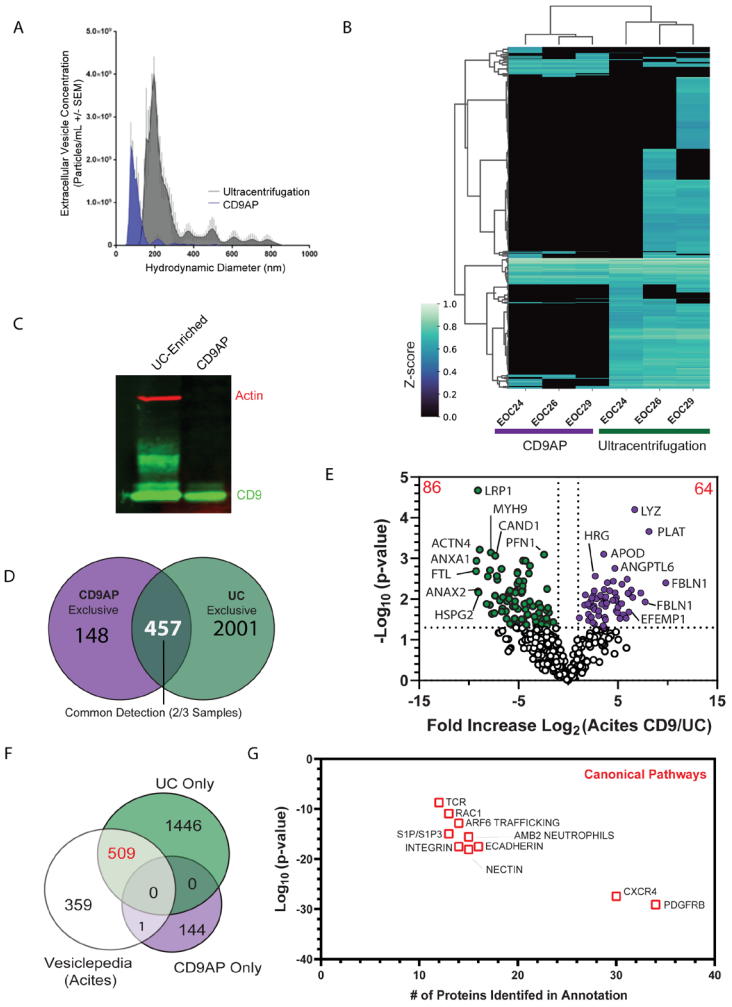

Figure 2. CD9-affinity purification enriches for small subsets of EVs from ascites fluid. (A) Dynamic light scattering of ascites EVs purified by UC or CD9AP demonstrates a subset of EVs is enriched with CD9AP. The distribution of CD9AP-EVs were primarily distributed around <150nm in diameter, whereas UC samples were comprised of a heterogenous mixture of EVs that were primarily distributed around ~200nm, albeit subpopulations of EVs were detectable up to 900nm (see Appendix Table 3). **(B)** Heatmap of identified proteins and dendrogram demonstrate increased proteomic depth obtained by UC, albeit variability of UC-enrichment of ascites EVs was also observable. **(C)** Ectosome and large EV components, such as actin, were depleted using CD9AP, further validating the specified enrichment of small EVs. **(D)** 148 and 2001 proteins were exclusive to EVs enriched by CD9AP or UC (>1 replicate), whereas 457 proteins were common to both CD9AP and UC-enriched EVs (>2 replicates in each condition). **(E)** Volcano plot of common proteins to CD9AP and UC identified 64 and 84 proteins significantly enriched in either CD9AP- or UC-enriched ascites EVs, respectively. **(F)** Overlap of exclusively detected proteins with Vesiclepedia database filtered for ascites EVs. **(G)** The 509 proteins detected in UC-enriched EVs and overlapped with Vesiclepedia were enriched with members of canonical PDGFB, CXCR4, E-Cadherin and Integrin signalling pathways.

compared to primary sources **(Fig 1B)**. Furthermore, PCA analysis confirmed that the proteomes of EVs isolated from biofluids taken from patients was distinct from those derived from cells in culture **(Fig 1C)**. Interestingly, for plasma samples CD9-derived EVs clustered closely to CD9-derived ascites EVs than plasma EVs isolated by UC. Furthermore, the ascites EV samples acquired with UC clustered more closely with ovarian cancer cells grown in culture. Primary cell lines were derived from ascites fluid of patients with low-grade serous (LG/EOC18) and high-grade (HG/EOC6) ovarian cancer. The proteomes of these cell lines reflected an intersect of the ascites microenvironment and EV proteome generated by established cell lines. Similar to proteomic analyses of ovarian cancer cell lysates[25] and in support of our recent characterization[26], cell line EVs harboured unique proteomic cargo compared to each other but primary cell lines cluster along principal components **(Appendix Fig S1A)**. Of note, when we filtered the Vesiclepedia database for EV signatures from EOC cell lines, a 35.1 – 43% overlap was observed with our data **(Fig 1D)**. Overlapping proteins were significantly associated with GO Cellular Component (GOCC) annotations indicative of EV-enrichment **(Appendix Fig S1B)**. hIOSE was separated from cell lines EOC6/18, OV-90 and OVCAR3 and compared to the Vesiclepedia filtered for ascites EVs **(Fig 1E)**. Common proteins (red) were associated with neutrophil degranulation and adaptive immunity **(Fig 1F)**. Only three proteins from EOC cell line EVs exclusively overlapped with Vesiclepedia-ascites **(Appendix Fig S1C)**. These included SLC34A2, a solute transporter upregulated within ovarian cancer tumours [27]. Moreover, 1515 proteins overlapped with primary ascites EVs in our hands and were associated with adaptive immunity and members of the PDGFB, CXCR and VEGF signalling pathways **(Fig 1G and H, Appendix Fig SID).** These results clearly demonstrate that the proteomic 'fingerprint' of EOC cell line-derived EVs is distinct from those contained within biofluid EVs, but that EVs from all sources reflect biological hallmarks of cancer. Moreover, EVs derived from UC are better able to identify a cancer-specific proteome.

2.2 CD9AP increases EV specificity at the expense of proteomic depth. Over the last decade, efforts have been made to compare strategies that enrich EVs from conditioned

media or biological fluids[28-30]. Optimizing EV purity is ideal to identify true EV cargo and elucidate the biological mechanisms dependent on EV biogenesis or uptake. EVs represent a large range of biological vesicles that may reflect anything from 'cellular debris' during apoptotic processes to systematically packaged messages that are able to prime distant microenvironments for cancer metastasis[20, 31]. With these properties in mind, we hypothesized that obtaining high EV purity would uncover additional biomarkers undetected within UC-enriched EV preparations due to EV heterogeneity or residual HAPs. We selected CD9AP to segregate small EVs from large EVs and residual cellular debris in ascites derived from donors with HGSC. Indeed, small EV purity was increased with CD9AP compared to UC when measured by dynamic light scattering **(Fig 2A-C, Appendix Table S3);** however, this occurred at the expense of proteomic depth or number of proteins identified **(Fig 2C,D)**. 148 proteins were exclusively detected in CD9AP-EVs and were enriched with effectors of blood vessel and cancer development, such as TGFB1, BMP2, VEGFC and WNT11 **(Appendix Fig S2A).** On the other hand, >2000 additional proteins were exclusively detected using UC, albeit protein identification across UC-EVs was variable **(Fig 2C,D)**. Notable mediators of cancer biology exclusively detected in UC-EVs included Aldehyde Dehydrogenase 1A1 (ALDH1A1) and epidermal growth factor receptor (EGFR) amongst additional factors associated with wounding or cellular activation during immune response **(Appendix Fig S2B)**. 457 proteins were comparatively detected in UC- and CD9AP-EVs, although differential enrichment was observed between the two isolation methods **(Fig 2E)**. For example, CD9AP-EVs were enriched for 64 proteins such as tissue plasminogen activator (PLAT) and angiopoietin-like 6 (ANGPTL6). Alternatively, UC-EVs were enriched with 84 proteins, such as Annexin1/2 (ANXA1/2) and myosin heavy chain-9 (MYH9). Although both UC- and CD9AP-EVs contain proteins associated with EV biology, several 'classical' EV markers (i.e. CD63) were exclusive to UC-EVs **(Appendix Fig S3)**. These results were not surprising considering CD9 and C63 may represent EVs of distinct biogenic processes[32]. CD9 and integrins are often incorporated into the membrane of EVs and facilitate uptake

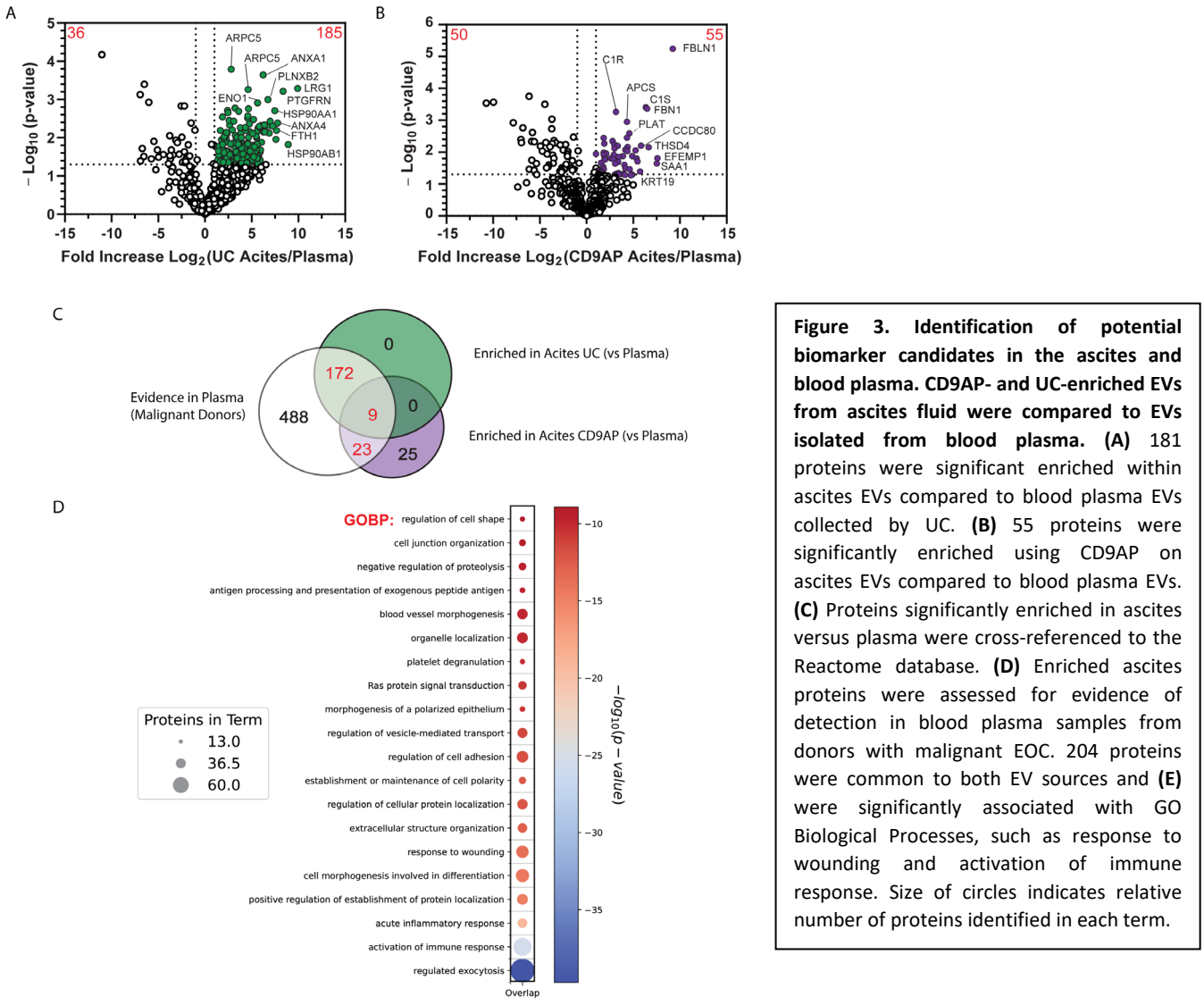

Figure 3. Identification of potential biomarker candidates in the ascites and blood plasma. CD9AP- and UC-enriched EVs from ascites fluid were compared to EVs isolated from blood plasma. (A) 181 proteins were significant enriched within ascites EVs compared to blood plasma EVs collected by UC. **(B)** 55 proteins were significantly enriched using CD9AP on ascites EVs compared to blood plasma EVs. **(C)** Proteins significantly enriched in ascites versus plasma were cross-referenced to the Reactome database. **(D)** Enriched ascites proteins were assessed for evidence of detection in blood plasma samples from donors with malignant EOC. 204 proteins were common to both EV sources and **(E)** were significantly associated with GO Biological Processes, such as response to wounding and activation of immune response. Size of circles indicates relative number of proteins identified in each term.

into recipient cells[33]. Several integrin isoforms were exclusively detected in UC-EVs, supporting the enrichment of EVs subsets likely derived from the plasma membrane using UC. Proteins exclusive to UC or CDAP-EVs accounted for >50% of proteins contained with Vesiclepedia-ascites **(Fig 2F)**. Interestingly, only PARP1 overlapped between CD9AP-EVs and Vesiclepedia-ascites. On the other hand, 509 UC-EV proteins that overlapped with Vesiclepedia-ascites were enriched for growth factors and cytokines in wound response and neutrophil degranulation **(Fig 2G)**. Collectively, our results support previous reports which demonstrate that increased EV purity with CD9AP is likely to identify additional candidates for biomarker analyses[34], albeit putative biomarkers may be lost during this process.

2.3 Quantitative proteomics unveils a large reservoir of putative biomarkers. In an effort to increase the likelihood that selected EOC biomarkers could be used to detect early disease, we next focused our analyses towards proteins that were enriched in ascites EVs or absent in EVs derived from the plasma of healthy donors. We speculated that ascites EVs derived from the tumour microenvironment would harbour proteomic cargo which may be useful for the detection of EOC and that this proteomic cargo may be released into systemic circulation. Moreover, if absent in healthy controls, these EVs could be specifically detected in EOC patients even when disease burden is low. Like ascites, we employed parallel purification strategies, UC and CD9AP, to increase proteomic depth for biomarker discovery in blood plasma. In the UC group, 185 proteins were significantly elevated (2-fold,$p<0.05$) in ascites compared to healthy plasma **(Fig 3A)**. These included proteins associated with cancer cell biology and/or metastasis, such as LRP1, HSP90AA1/AB1, FTH1, CRP, and MUC1. On the other hand, 55 differentially expressed proteins (2-fold, $p<0.05$) were detected between healthy plasma and ascites using CD9AP **(Fig 3B)**. These included cancer-relevant proteins such as FBLN1, MMP14, ANGPTL2, IGBP2, CD14, PLAT. Hence while CD9 may enable the discovery of specific analytes, it is likely that CD9-negative vesicles harbour biomarkers lost during selective enrichment strategies. However, some factors were retained. For example, ascites EV proteins enriched by UC and CD9AP were significantly associated with effectors of neutrophil degranulation **(Appendix Fig S4)**. Next, we sought to determine whether ascites-specific EV proteins could also be detected in the plasma of EOC patients. Several proteins were exclusively detected in EVs from ascites compared to plasma samples. These were considered as potential tumor-specific biomarkers during PRM method development. Over 200 proteins that were enriched within ascites were also detected in plasma samples from donors with EOC and included mediators of immune response and regulated exocytosis **(Fig 3C,D)**. HE4 was not detected in any EV proteomes, which suggested potential EV-independence, similar to that reported in Zhao *et al*[35]. Collectively, these results support the parallel application of UC and CD9AP to 'mine' prospective biomarkers; moreover, they suggest that ascites may be an excellent bio-fluid with which to discover biomarkers and that incorporation into EVs likely enables the presence of these factors in the blood of EOC patients.

2.4 Targeted proteomics of EV-enriched plasma and machine learning classification optimization identify biomarker combinations for the early detection of EOC. Given the large number of proteins significantly enriched in ascites EVs, we next asked whether the abundance of these proteins would be elevated in an independent cohort of plasma EV samples from patients diagnosed with malignant EOC (n=10) versus controls with non-cancerous gynaecological conditions (n=9). We chose this cohort to serve as our control in order to account for markers that may be associated with pathologies or inflammation in general as opposed to ovarian-cancer-specific analytes which may be released from tumour cells or upregulated within the microenvironment of EOC. To enable more accurate, relative label-free quantitative comparisons, a manually curated list of 471 peptides (240 proteins selected from the previous analysis as present in ascites and patient serum) was subsequently targeted in the entire cohort of plasma EVs using a PRM method built in PEAKS[36] and Skyline[37] **(Fig 4)**. Peak areas were normalized to the TIC to correct for technical variability, and additionally normalized to the CD9 peptide EVQEFYK (extracellular region, AAs 120-126) to control for EV purity. Data scaling, support vector machine (SVM) optimization and validation were performed in a Python language environment. A total of 21 peptides were significantly enriched in malignant and non-malignant samples, respectively (Wilcoxon rank-sum test, $p<0.05$) **(Fig 5A; Table 1.)**. Of note, one peptide from CA-125 (MUC16) was included in our PRM method (ELGPYTLDR). Using the Wilcoxon rank-sum test, this peptide achieved $p=0.060$ for an AUC of 0.76 and log2 fold-change 2.12. Despite the selection of these 22 peptides, malignant and non-malignant samples could not be completely segregated using PCA and unsupervised k-means classification **(Fig 5B)**. Machine learning classification models, such as SVMs, have demonstrated immense utility for identifying novel biomarkers for an array of diseases[38]. This is due to their ability to provide high-accuracy classification using high-dimensionality data when sample numbers are limited. Indeed, this is an extremely beneficial and attractive feature of SVMs for biomarker discovery studies where the

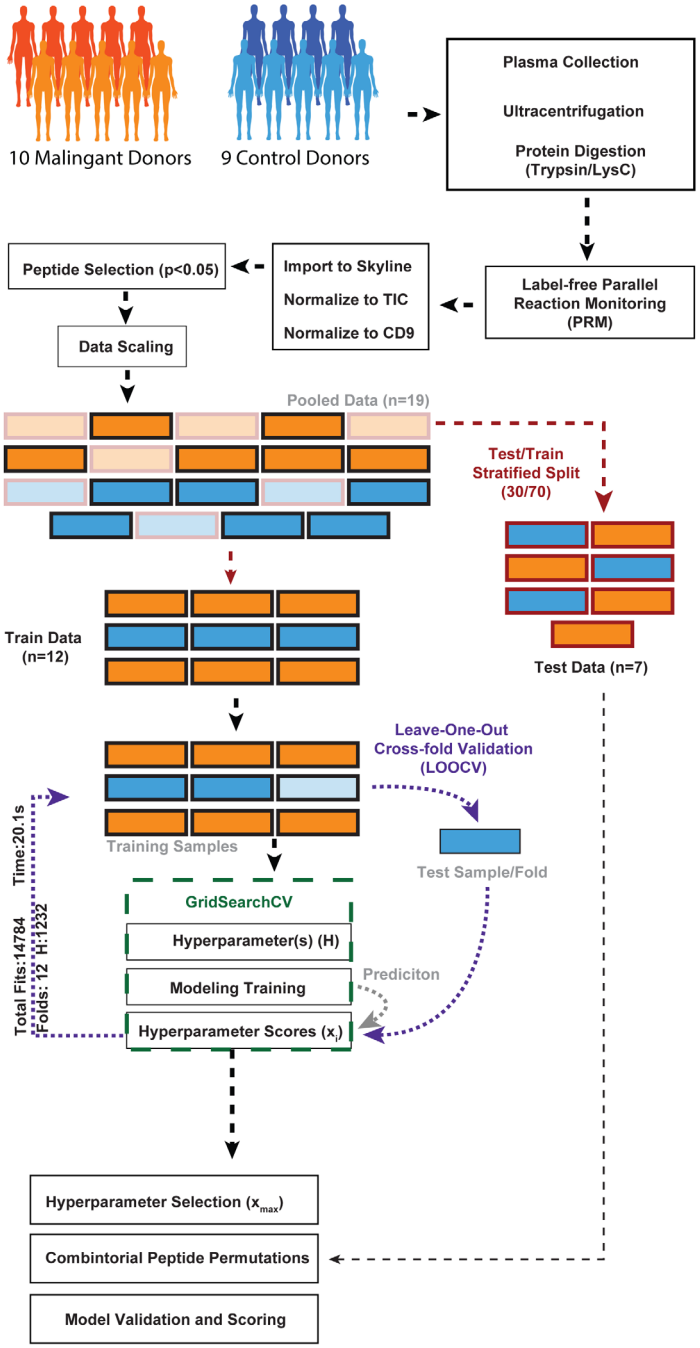

Figure 4. Overview of Leave-One-Out Cross Validation (LOOCV) for the Optimization of Support Vector Classification. Overview of biomarker discovery pipeline utilized within this study. Plasma EVs were isolated from malignant EOC (n=10) or control donors (n=9) using ultracentrifugation and subjected to proteolysis to generate tryptic peptides for label-free PRM. Peptides were selected from proteins either exclusively detected or significantly enriched within ascites compared to blood plasma, see Figure 3. Label-free PRM was quantified in Skyline by normalizing both total ion current (TIC) and CD9 intensity levels. Peptide features were narrowed to those that demonstrated a >2-fold change in HGSC EVs versus control EVs (i.e. 21 (+1) peptides were selected within our study). Normalized intensities of peptide features were scaled by z-scores and a stratified split of donors was performed to generated a training set (~70% of donors) and a test set (~30% of donors). SVMs were optimized using LOOCV to score model hyperparameters and principal components (PC) by removing a single donor from the training pool for to obtain a mean accuracy of model predictions. The 'optimized' hyperparameters are further refined by assess all permutations of features using the reserve test set for model validation and scoring, see Appendix Fig S5. Scoring was performed using a combination of ROC-AUC and Matthew's Correlation Coefficient

acquisition of large donor number is extremely difficult or impossible to obtain. Data features were scaled using z-scores, and randomly split into 10 independent training (70%) and test (30%) sets in a stratified fashion to ensure a comparable number of control and malignant samples were reserved. Donor status, such as FIGO stage, remained blinded until final validations were performed using the reserved test set. As proof-of-principle and for the figures within this manuscript, we retrospectively chose random_state=6 which contained all FIGO stages in both training and test data sets, thus allowing us to speculate on the ability of prospective biomarkers to identify early-stage HGSC. SVM optimization was executed with the GridSearch library that allowed for permutations of feature selection, SVM kernels (linear, poly, rbf) and hyperparameters (i.e. cost/C) to be scored. The optimal kernel and hyperparameter(s) were determined by LOOCV to dampen 'noise' often obtained with low complexity data sets by reserving a single sample for validation[39] **(Fig 4)**. 14,784 total fits or permutations of kernel, principal components, cost or gamma were used to calculate a mean accuracy score.

From these analyses, we identified eight linear SVMs (C=0.025-2) that provided a mean accuracy score >90%

(Fig 5C). Next, we optimized feature selection based on classification accuracy using the reserved test set. The SVM (PC=2, C=0.025) was tested 231 times with paired permutations of all 22 peptides **(Appendix Fig 5)**. Interestingly, nine combinations of peptides were able to classify malignant (n=4) versus non-malignant (n=3) samples with an accuracy score = 1.0 **(Fig 5D, E)**. For example, the combination of CFHR4 and MUC1 was able to accurately classify Stage I, II, and III donors **(Fig 5F)**. Additional peptide combinations provided accuracy scores = 1.0, however GPX3, MUC1, and CFHR4 were represented in the majority of models **(Table 1)** and considered strong drivers of the EOC classification, according to SHapley Additive exPlanations (SHAP) analysis[40] **(Fig 6A)**. For example, CFHR4 and GPX3 were not detected in cell line EVs and were strong drivers of Stage 1 EOC classification **(Fig 6B, Appendix Fig S6)**. Alternatively, MUC1 was not detected within CD9AP-EVs and was a strong driver of Stage III EOC **(Fig 6C)** and control donor classification **(Fig 6D)**.

It should be considered that the selected C hyperparameter (0.025-2) would provide "soft" margins for SVM and high accuracy on the peptide combinations selected during optimization **(Appendix Fig S7A-C)**. SVMs with more conservative margins (C=10) also generated several distinct peptide combinations with high accuracy when optimized using the test set **(Appendix Fig S7 D, E)**. Nonetheless, we demonstrate the robustness of our approach and discovered additional biomarker combinations in EV-depleted plasma **(Appendix Fig S8, Appendix Table S4)**. Interestingly, CFHR4 was considered a strong driver of SVM accuracy in EV-depleted plasma (**Appendix Fig S9)** and was speculated to be constituent of the EV corona[41]. Using a limited number of donor samples, we highlight the use of label-free PRM, SVM optimization using LOOCV and parallel enrichment of EVs to identify combinatorial biomarkers that may be used to detect all stages of EOC.

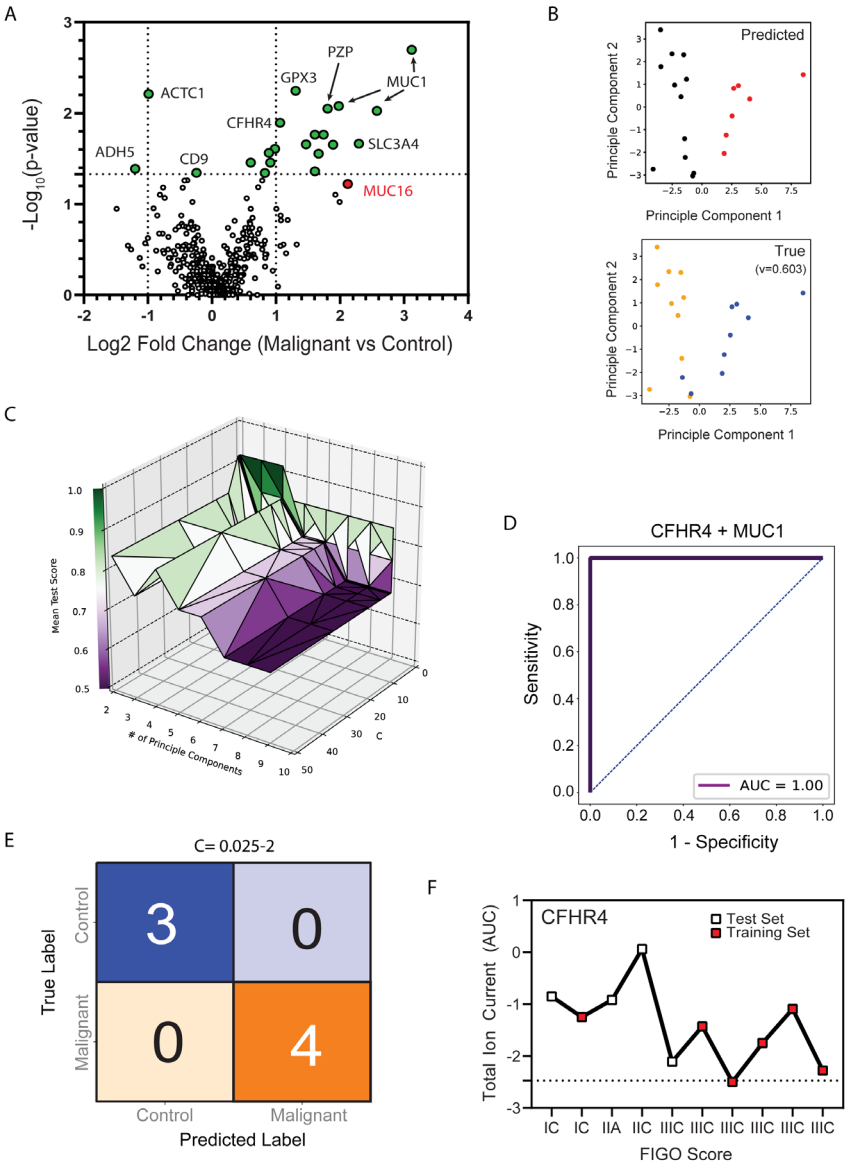

Figure 5. Target Proteomics and Support Vector Classification Identifies Novel Biomarkers for Malignant Ovarian Cancer. 471 peptides corresponding to 240 proteins were analysed in EV-enriched blood plasma from malignant (n=10) versus control (n=9) donors using label-free LC-MS/MS and parallel reaction monitoring. **(A)** Volcano plot highlights peptides that were significantly different between malignant and control donor samples. 21 peptides (p-value <0.05) and EOC antigen MUC16 (red) were selected for further analyses. **(B)** Heatmap of principal component variance ratios of potential biomarkers. **(C-D)** Unsupervised PCA and k-means clustering of pooled samples. Predicted labels (red and black) partially overlapped with true labels (blue = control and orange = malignant). **(E)** Hyperparameter tuning of the linear SVM was performed by LOOCV, leading to hyperparameters C=0.025 and two principal components selected as the 'optimized' SVM based on mean accuracy score (>0.90). Each point of triangulation indicates an SVM combination/fit that was scored using the training set. **(F)** Feature selection was performed using 231 combinations of peptides and test data. From this analysis, nine combinations of peptides provided an accuracy score of 1.0 on the **(G)** test data set. **(H)** For example, the combination of **(I)** CFHR4 and **(J)** MUC1 provided a Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) score of 1.0. Training (red) and test samples (white) were represented by women with Stage I, II, and III EOC.

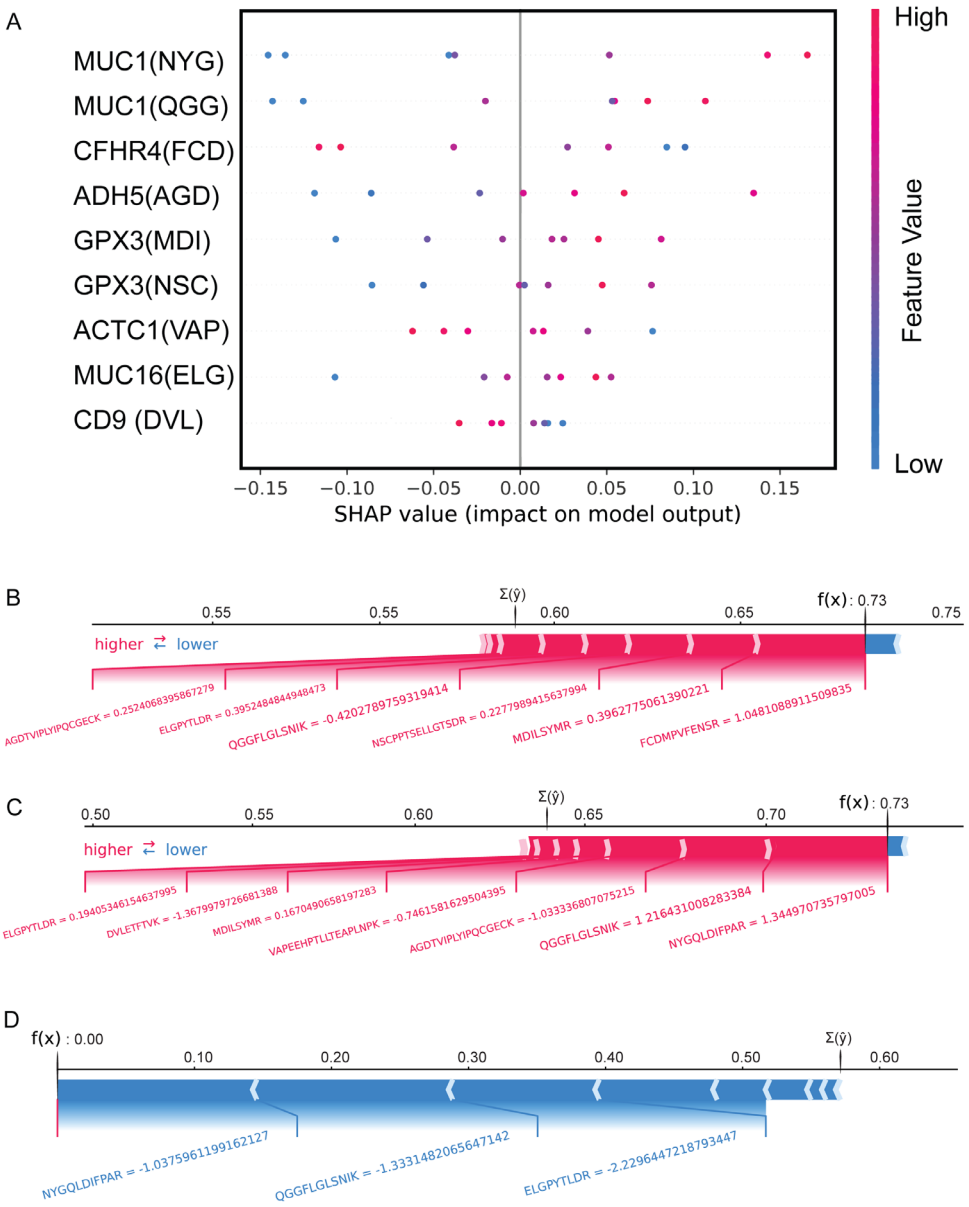

Figure 6. SHAP analysis of EV-enriched biomarkers used for SVM classification. (A) SHAP post-model analyses determined that MUC1 (NYG/QGG), CFHR4 (FCD), ADH5 (AGD) and GPX3 (MDI/NSC) are the most dominant drivers and high value features of EOC classification using the selected SVM (linear, PC=2, C=0.025). Letters inside brackets represent the first 3 amino acids for the associated protein. **(B)** For example, CFHR4 was a strong driver of SVM for early-stage EOC. Alternatively, MUC1 and ADH5 were strong drivers of SVM for **(C)** late-stage EOC or **(D)** non-malignant samples. f(x) = classification score, ∑(ŷ) = average predicted score. Number associated with peptides are SHAP values and indicate the strength of the peptide in determing the classification score f(x) against the average classification score ∑(ŷ).

3. Discussion

In this study, we characterized EV proteomes derived from primary and immortalized cell lines, ascites and plasma using two distinct enrichment strategies (UC and CD9AP) in order to maximize proteomic depth and increase the number of biomarker candidates. Our findings expand upon previous work by several other groups that also utilized mass spectrometry to characterize EVs derived from ascites or cell lines. Significantly and in stark contrast to the previous studies, we were able to build SVM models capable of accurately identifying Stage I, II and III EOC from plasma EVs.

Our comparisons of EV proteomes from ovarian (cancer) cell lines supports previous reports of intercellular heterogeneity, which may reflect differences in tissue of origin or stages of ovarian cancer progression[42]. For example, three distinct proteomic expression profiles were identified during a recent large-scale proteomic analysis of cell lines and primary tumors[43]. We found the EV proteomes of cell lines may reflect the pathophysiology of early-stage EOC, such as inflammation[44], ECM remodeling[45] and angiogenesis[46]. However, many similarities were noted between cancer cells and the non-malignant hIOSE, pointing to potential confounders associated with propagation in tissue culture. Building off the proteome of EOC cell line EVs, we expanded our focus to the proteomic profiling of EVs from primary sources. We executed an in-depth characterization of ascites-EV proteomes using parallel purification strategies, the 'match-between-runs' feature in MaxQuant[47], SCX StageTip fractionation technology[48], and Orbitrap-based instrumentation[49]. Over the last decade, a wave of efforts have attempted to deplete HAPs from biofluids to improve the detection low-abundancy biomarkers[50, 51]. To better delineate proteins specific to EOC, Shender *et al*. compared ascites from patients with ovarian cancer to those with alcohol-induced cirrhosis and identified 424 proteins associated with malignant ascites[52]. More recently, Sinha *et al*. have developed an EOC xenograft model in combination with N-glycopeptide enrichment and PRM to identity potential biomarker candidates in primary patient samples[53]. Considering the proteomic complexity of biofluids, it is unlikely that a single proteomic approach will be able to identify all biomarkers for detecting metastatic EOC.

Within this study, we developed and validated a unique pipeline incorporating EV purification, PRM proteomics, LOOCV and SVM that is tailored for the identification of novel biomarker combinations for early EOC detection. While MUC16 was higher in malignant samples, it was not considered an impactful biomarker using a soft-margin SVM. Combinations of MUC16 and additional peptides were able to provide high accuracy using more conservative SVM margins; however, subsequent investigations with larger cohorts will be necessary to understand the impact of hyperparameter tuning for EOC detection. SHAP analysis can provide additional insight into which peptides drive prediction outcomes within a SVM[40]. Using these analyses, CFHR4 provided high SHAP values in both EV-enriched and EV-depleted models and was exclusively detected in CD9AP-enriched ascites EVs. Two isoforms of CFHR4 have been identified to enhance C-reactive protein (CRP) binding to necrotic cells and tumour tissue, leading to complement activation and opsonisation[54]. The functional role of CFHR4 in EOC progression has not undergone thorough investigation; however, Pedersen et al. demonstrated elevated CFHR4 in small-cell lung cancer using quantitative proteomics[55]. Interestingly, CHFR4 was exclusive to small EVs and was not detected on "microvesicles" in their study, aligned with our results that CHFR4 is likely specific to a subset of EVs. Nonetheless, MUC1 and GPX3 were also relevant to EOC classification and have established roles in cancer progression and metastasis[56-58]. Ultimately, we provide evidence of combinatorial biomarkers that are capable of detecting early stages of EOC. These findings will lead to the development of improved clinical diagnostics for early-stage EOC, in hopes of providing earlier treatment interventions.

3.1 Ideas and Speculation. Despite our efforts, several limitations of this study will need to be addressed by future analyses Our future studies will aim to decomplexify biofluids and isolate extracellular vesicles by integration of size-exclusion chromatography (SEC) into our biomarker discovery pipeline. The integration of SEC will allow us to 1) achieve greater purity of EVs without immunopurification and 2) prospectively identify EV-independent proteins which may be useful for EOC classification. Nonetheless, our future studies will explore the incorporation of heavy-isotope standards during PRM to allow for absolute quantification of biomarkers in plasma. These refined methods should be used to test the diagnostic power of EV biomarkers using an expanded number of control and patient samples, leading eventually to prospective trials.

Within this study, we determined that complement cascade component CFHR4 provided value as a feature for SVM model classification using both EV-enriched and EV-depleted samples. Tóth *et al.* identified complement cascade factors are common components of the EV protein "corona" which may be a result of secondary interactions of EVs and plasma components[41]. Alternatively, the work conducted by Papp *et al.* suggests that complement components may be directly released from the plasma membrane of B-cells and macrophages[59]. This supports the idea that elevated CHFR4 detected in EOC donors may be a reflection of the malignant EOC microenvironment produced by immune and/or tumor

cells[60]. Indeed, Bonavita *et al.* observed complement-dampened mice were protected against epithelial carcinoma[61]. Speculating based off our data and others, enhanced complement activation via CFHR4 may be a distinguishable hallmark of malignant EOC/HGSC.

4. Methodology and Data Analysis

4.1 Cell Culture. OV-90 (ATCC® CRL-11732) and NIH:OVCAR3 (ATCC® HTB-161) were obtained from the ATCC. Human immortalized surface epithelial cells hIOSE (OSE364) were obtained from the Canadian Ovarian Tissue Bank at the BC Cancer Agency and kindly provided by Dr. Ronny Drapkin (Department of Obstetrics and Gynecology, University of Pennsylvania). Primary cell lines EOC6 and EOC18 were isolated from the ascites of patients with high-grade and low-grade serous ovarian cancer, respectively. All cell lines, except OVCAR3, were maintained in M199+MCDB105 supplemented with 5-15% FBS. NIH:OVCAR3 cells were cultured in RPMI-1640 supplemented with 20% FBS and 5µg/mL insulin. Media was exchanged with serum free media for 20-30 hours to generate conditioned media (CM) for EV purification. All work involving the use of patient samples (cell lines, plasma and ascites) was approved by the Health Research Ethics Board of Alberta-Cancer Committee.

4.2 Ultracentrifugation (UC). CM, plasma and ascites samples were first centrifuged at 200-300 x g at 4°C to pellet cells. Supernatants were diluted 1:10 in PBS (except CM) and centrifuged at 3,000 x g for 20 minutes at 4°C to remove cell debris. To remove large membrane fragments, supernatants were spun at 10,000 x g for an additional 20 minutes at 4°C. Lastly, supernatants were ultracentrifuged at 120,000 to 140,000 x g (SW-28 rotor) for 2 hours at 4°C to pellet EVs on an OptimaTM L-100 XP ultracentrifuge (Beckman Coulter). The supernatant was removed and EVs were resuspended in 100-300µL of PBS and stored at -80°C until further use.

4.3 CD9-affinity Purification (CD9AP). Hydrophilic streptavidin magnetic beads (120mg) were washed three times with PBS then resuspended in 5mL PBS (New England Biosystems, S1421S, 20mg/5ml). Beads were mixed with 650µg biotin conjugated anti-CD9 antibody (Abcam, ab28094) at room temperature for 30 minutes and then washed twice with PBS to remove unbound antibody. Beads were resuspended in 6mL PBS and 1mL (~20mg) was added to 10mL plasma or ascites (diluted 1:1 in PBS). Samples were placed on a rotary mixer overnight at 4°C and then rinsed three times with PBS. EVs were eluted from beads with three-500 µl glycine-HCl (0.1M, pH 2.39) washes. A small volume (75µL) of Tris-HCl (1.8M, pH 8.54) was used to neutralize each eluent.

4.4 Western Blotting. EVs were lysed in RIPA buffer. 10 µg protein was loaded onto a 10% SDS-PAGE gel under reducing conditions. Proteins were transferred to PVDF and the membranes were blocked with LI-COR Intercept Blocking solution. Membranes were incubated with anti-CD9 rabbit antibody [CD9 (D8O1A) Rabbit mAb, Cell signaling Tech; #13174S, dilution 1:2000] and an anti-actin mouse antibody [Anti-β-Actin Antibody (C4), Santa Cruz Biotech, sc-47778, dilution 1:1000] overnight at 4°C. Membranes were washed then incubated with IRDye 800CW donkey anti-rabbit (LI-COR# 926-32213, dilution 1:20000) and IRDye-680RD donkey anti-mouse (LI-COR# 926-68072, dilution 1:20000) for 1 hour at room temperature. Membranes were then scanned with the Odyssey Infrared Imager (LI-COR).

4.5 Nanoparticle Tracking Analysis. Samples were diluted 25-fold using filtered 0.2x phosphate buffered saline and then were analyzed using the Nanosight LM10 (405nm laser, 60mW, software version 3.00064). Samples were analyzed for 60 seconds (count range of 20-100 particles per frame). All measurements were done in triplicate.

4.6 EV Protein Extraction and Digestion. To prepare EVs for LC-MS/MS, ~25µg protein quantified by BCA were lyophilized to dryness and reconstituted in 8M Urea, 50mM ammonium bicarbonate (ABC), 10mM dithiothreitol (DTT), 2% SDS lysis buffer. EV proteins were sonicated with a probe sonicator (3 X 0.5s pulses; Level 1) (Fisher Scientific, Waltham, MA), reduced in 10mM DTT for 30 minutes at room temperature (RT), alkylated in 100mM iodoacetamide for 30 minutes at RT in the dark, and precipitated in chloroform/methanol[62]. On-pellet in-solution protein digestion was performed in 100µL 50mM ABC (pH 8) by adding Trypsin/LysC (Promega, 1:50 ratio) to precipitated EV proteins. EV proteins were incubated at 37°C overnight (~18h) in a ThermoMixer C (Eppendorf) at 300 rpm. An additional volume of trypsin (Promega, 1:100 ratio) was added for ~4 hours before acidifying to pH 3-4 with 10% FA.

4.7 SCX Peptide Fractionation and LC-MS/MS. Tryptic peptides were fractionated using strong cation exchange (SCX) StageTips similarly to Kulak *et al* [63].. Briefly, peptides were acidified with 1% TFA and loaded onto a pre-rinsed 12-plug SCX StageTips (Empore™ Supelco, Bellefonte, PA, USA). In total, 6 SCX fractions were collected by eluting in 75, 125, 200, 250, 300 mM ammonium acetate/20% ACN followed by a final elution in 5% ammonium hydroxide/80% ACN. SCX fractions were dried in a SpeedVac (ThermoFisher), re-suspended in ddH2O, and dried again to evaporate residual ammonium acetate. All samples were re-suspended in 0.1% FA prior to LC-MS analysis.

SCX fractions were analyzed using a nanoAquity UHPLC M-class system (Waters) connected to a Q Exactive mass spectrometer (Thermo Scientific) using a nonlinear gradient. Buffer A consisted of water/0.1% FA and Buffer B consisted

of ACN/0.1%FA. Peptides (~1µg estimated by BCA) were initially loaded onto an ACQUITY UPLC M-Class Symmetry C18 Trap Column, 5 µm, 180 µm x 20 mm and trapped for 4 minutes at a flow rate of 5 µl/min at 99% A/1% B. Peptides were separated on an ACQUITY UPLC M-Class Peptide BEH C18 Column (130Å, 1.7µm, 75µm X 250mm) operating at a flow rate of 300 nL/min at 35°C using a non-linear gradient consisting of 1-7% B over 3.5 minutes, 7-19% B over 86.5 minutes and 19-30% B over 30 minutes before increasing to 95% B and washing. Settings for data acquisition on the Q Exactive and Q Exactive Plus are outlined in Supplemental Table 1.

4.8 LC-MS/MS Data Analysis. MS raw files were searched in MaxQuant (1.5.2.8) using the Human Uniprot database (reviewed only, updated May 2014 with 40,550 entries). Missed cleavages were set to 3 and I=L. Cysteine carbamidomethylation was set as a fixed modification. Oxidation (M), N-terminal acetylation (protein), and deamidation (NQ) were set as variable modifications (max. number of modifications per peptide = 5) and all other setting were left as default. Precursor mass deviation was left at 20 ppm and 4.5 ppm for first and main search, respectively. Fragment mass deviation was left at 20 ppm. Protein and peptide FDR was set to 0.01 (1%) and the decoy database was set to revert. The match-between-runs feature was utilized across all sample types to maximize proteome coverage and quantitation. Datasets were loaded into Perseus (1.6.14) and proteins identified by site; reverse and potential contaminants were removed[47]. Protein identifications with quantitative values in >50% samples in each group (cells, plasma or ascites) were retained for downstream analysis unless specified elsewhere. Missing values were imputed using a width of 0.3 and down shift of 1.8 to enable statistical comparisons.

4.9 Label-free Parallel Reaction Monitoring (PRM). 25µg plasma EVs and 50µg EV-depleted plasma from malignant (Supplemental Table 2) and age-matched control donors were digested overnight with Trypsin/LysC (1:50 ratio) and LysC (Wako; 1:100 ratio). To remove large species, digests were filtered through pre-rinsed (100µL 25mM ABC/50% ACN) 10 kDa MWCO YM-10 centrifugal filter units (Millipore) at 14,000xg for 20 min. Centrifugal filter units were washed with an additional 50µL 25 mM ABC/50% ACN for 15 min at 14,000 x g to help recover additional peptides. Filtered samples were dried in a SpeedVac, reconstituted in 0.1% FA and quantified by BCA. To generate spectral data for biomarker candidate (peptides), several unfractionated plasma EV digests (~1µg/sample) were initially analyzed on a Q Exactive Plus using a non-linear 2.5h gradient consisting of 1-7% B over 1 minute, 7-23% B over 134 minutes and 23-35% B over 45 minutes before increasing to 95% B and washing. Raw files were searched against the human Uniprot databased (20, 274 entries) using the de novo search engine PEAKS® (version 8)[36]. Parent and fragment mass error tolerances were set to 20 ppm and 0.05 Da, respectively. Maximum missed cleavages were set to 3 and 1 non-specific cleavage was allowed. Carbamidomethylation was set as a fixed modification, and deamidation, oxidation and acetylation (protein N-term) were included as variable modifications with a maximum of 3 PTMs per peptide allowed. pepXML peptide information and mzXML spectral data were next exported from PEAKS® generate a PRM method in Skyline[37]. Peptides with missed cleavages or containing tryptophan were removed and up to 3 peptides/protein, 7-18 amino acids in length, were chosen for monitoring. In Skyline, the top 5 most intense transitions (b and y ions) were used for quantification and an 8-minute window was chosen to account for deviations in chromatography and minimize the chance of truncation while maximizing the number of MS/MS scans. EV and EV-depleted samples were subsequently analysed using the same gradient but with a targeted PRM method in a randomized fashion. A minimum of 3 transitions were required to measure peak areas, and targets with dotp scores <0.8 or ppm exceeding 20 were assumed to contain interference and initially assigned a peak area of 0. To correct for sample loading and technical variability, peak areas for each peptide were normalized to the total ion current (TIC). Peak areas were additionally normalized to the CD9 peptide EVQEFYK (extracellular region, AAs 120-126) to correct for EV recovery. Normalized peak areas of 0 were assumed to be missing not at random and imputed with the lowest ratio detected for the given peptide.

4.10 Proteomic Data Availability. Proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD023723.

4.11 Data Handling and Statistical Analysis. Differential protein abundance between conditions were determined using a two-tailed Welch's t-test ($p<0.05$) in Perseus (version 1.6.14). Graphing was performed using Python or Prism version 6.01 (GraphPad Software, San Diego, CA). Mann-Whitney rank sum statistical tests were calculated in R (version 3.60) or in RStudio (version 1.2.1335). Data handling and machine learning optimization pipelines were built in Python. Pathway and annotation enrichment analyses were performed using Metascape (metascape.org) using the default settings.

Table 1. EV-enriched Blood Plasma Peptides Selected for Targeted Proteomics and SVM model optimization.

Peptide	Gene	Log_2 Fold Change (M vs. C)	$-Log_{10}$ (p-value)	SVM Model Relevance (% of SVM Model)
MDILSYMR	GPX3	1.31	2.25	33%
QGGFLGLSNIK	MUC1	2.57	2.03	33%
FCDMPVFENSR	CHFR4	1.06	1.89	33%
YVPPSSTDR	MUC1	3.11	2.70	22%
AGDTVIPLYIPQCGECK	ADH5	-1.20	1.40	22%
DVLETFTVK	CD9	-0.24	1.34	22%
NSCPPTSELLGTSDR	GPX3	0.98	1.61	11%
ELGPYTLDR	MUC16	2.12	1.22	11%
VAPEEHPTLLTEAPLNPK	ACTC1	-0.98	2.21	0%
NYGQLDIFPAR	MUC1	1.98	2.08	0%
IISIMDEK	PZP	1.80	2.04	0%
AYAAGFGDR	TNC	1.74	1.76	0%
LDAPSQIEVK	TNC	1.61	1.76	0%
VISQIAMNDEK	SLC34A2	2.29	1.66	0%
FEIENCLANK	PZP	1.47	1.66	0%
ASSFLGEK	C4B	1.89	1.65	0%
AYSLFSYNTQGR	APCS	0.89	1.56	0%
EDSPFALK	PZP	1.66	1.55	0%
ATAQMLEVMFK	LBP	0.91	1.46	0%
SIPQVSPVR	CPN1	0.60	1.45	0%
VATYLPAPEGLK	TNC	1.61	1.36	0%
DNELLVYK	APCS	0.83	1.34	0%

M = Malignant, C=Control, SVM= Support Vector Machine, *Linear SVM (C=0.025)

Acknowledgements

We thank Paula Pittock for technical support. This work was supported by the Sawin-Bladwin Chair in Ovarian Cancer Research and the Dr. Anthony Noujaim Oncology Chair awarded to LMP by the Women and Children Health Research Institute and the Alberta Cancer Foundation, respectively.

Author Contributions

TTC, DDC, and LMP designed the research and wrote the manuscript. TTC, DDC, JL, GMS, and DP conducted experiments. TTC, DDC, JL, GMS, and DP analyzed data and interpreted the results. JDL, GAL, and LMP provided logistic and financial support for experimental work. GAL and LMP supervised the study.

Conflict of Interest

There are no conflicts of interest to report.

References

1. Lheureux S, Gourley C, Vergote I and Oza AM. Epithelial ovarian cancer. *The Lancet*. 2019;393:1240-1253.
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2021;71:209-249.
3. Reid BM, Permuth JB, Sellers TAJCb and medicine. Epidemiology of ovarian cancer: a review. 2017;14:9.
4. Jacobs IJ, Menon U, Ryan A, Gentry-Maharaj A, Burnell M, Kalsi JK, Amso NN, Apostolidou S, Benjamin E and Cruickshank DJTL. Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. 2016;387:945-956.
5. Menon U, Gentry-Maharaj A, Burnell M, Singh N, Ryan A, Karpinskyj C, Carlino G, Taylor J, Massingham SK and Raikou MJTL. 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. 2021;397:2182-2193.
6. Menon U, Gentry-Maharaj A, Hallett R, Ryan A, Burnell M, Sharma A, Lewis S, Davies S, Philpott S and Lopes AJTlo. 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). 2009;10:327-340.
7. Felder M, Kapur A, Gonzalez-Bosquet J, Horibata S, Heintz J, Albrecht R, Fass L, Kaur J, Hu K and Shojaei HJMc. MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress. 2014;13:1-15.
8. Nolen BM and Lokshin AEJFO. Protein biomarkers of ovarian cancer: the forest and the trees. 2012;8:55-71.
9. Nolen BM, Lokshin AEJMd and therapy. Biomarker testing for ovarian cancer: clinical utility of multiplex assays. 2013;17:139-146.
10. Karlsen MA, Sandhu N, Høgdall C, Christensen IJ, Nedergaard L, Lundvall L, Engelholm SA, Pedersen AT, Hartwell D and Lydolph MJGo. Evaluation of HE4, CA125, risk of ovarian malignancy algorithm (ROMA) and risk of malignancy index (RMI) as diagnostic tools of epithelial ovarian cancer in patients with a pelvic mass. 2012;127:379-383.
11. Zhang Z, Chan DWJCE and Biomarkers P. The road from discovery to clinical diagnostics: lessons learned from the first FDA-cleared in vitro diagnostic multivariate index assay of proteomic biomarkers. 2010;19:2995-2999.
12. Yip P, Chen T-H, Seshaiah P, Stephen LL, Michael-Ballard KL, Mapes JP, Mansfield BC and Bertenshaw GPJPO. Comprehensive serum profiling for the discovery of epithelial ovarian cancer biomarkers. 2011;6:e29533.
13. Høgdall C, Fung ET, Christensen IJ, Nedergaard L, Engelholm SA, Petri AL, Risum S, Lundvall L, Yip C and Pedersen ATJGo. A novel proteomic biomarker panel as a diagnostic tool for patients with ovarian cancer. 2011;123:308-313.
14. Hanash SM, Pitteri SJ and Faca VM. Mining the plasma proteome for cancer biomarkers. *Nature*. 2008;452:571-579.
15. Jacobs JM, Adkins JN, Qian W-J, Liu T, Shen Y, Camp DG and Smith RD. Utilizing human blood plasma for proteomic biomarker discovery. *Journal of proteome research*. 2005;4:1073-1085.
16. Nanjappa V, Thomas JK, Marimuthu A, Muthusamy B, Radhakrishnan A, Sharma R, Ahmad Khan A, Balakrishnan L, Sahasrabuddhe NA and Kumar SJNar. Plasma Proteome Database as a resource for proteomics research: 2014 update. 2014;42:D959-D965.
17. Keshishian H, Addona T, Burgess M, Kuhn E, Carr SAJM and Proteomics C. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. 2007;6:2212-2229.
18. Hüttenhain R, Choi M, de la Fuente LM, Oehl K, Chang C-Y, Zimmermann A-K, Malander S, Olsson H, Surinova S and Clough T. A targeted mass spectrometry strategy for developing proteomic biomarkers: A case study of epithelial ovarian cancer*[S]. *Molecular & Cellular Proteomics*. 2019;18:1836-1850.
19. Raposo G and Stahl PDJNRMCB. Extracellular vesicles: a new communication paradigm? 2019;20:509-510.
20. Becker A, Thakur BK, Weiss JM, Kim HS, Peinado H and Lyden DJCc. Extracellular vesicles in cancer: cell-to-cell mediators of metastasis. 2016;30:836-848.
21. Lane R, Korbie D, Hill M, Trau MJC and medicine t. Extracellular vesicles as circulating cancer biomarkers: opportunities and challenges. 2018;7:1-11.
22. Barnabas GD, Bahar-Shany K, Sapoznik S, Helpman L, Kadan Y, Beiner M, Weitzner O, Arbib N, Korach J, Perri TJM and Proteomics C. Microvesicle proteomic profiling of uterine liquid biopsy for ovarian cancer early detection. 2019;18:865-875.
23. Coscia F, Watters K, Curtis M, Eckert M, Chiang C, Tyanova S, Montag A, Lastra R, Lengyel E and Mann M. Integrative proteomic profiling of ovarian cancer cell lines reveals precursor cell associated proteins and functional status. *Nature communications*. 2016;7:1-14.
24. Sinha A, Ignatchenko V, Ignatchenko A, Mejia-Guerrero S, Kislinger TJB and communications br. In-depth proteomic analyses of ovarian cancer cell line exosomes reveals differential enrichment of functional categories compared to the NCI 60 proteome. 2014;445:694-701.
25. Thomas SN, Friedrich B, Schnaubelt M, Chan DW, Zhang H and Aebersold RJI. Orthogonal proteomic platforms and their implications for the stable classification of high-grade serous ovarian cancer subtypes. 2020;23:101079.
26. Ćulum NM, Cooper TT, Lajoie GA, Dayarathna T, Pasternak SH, Liu J, Fu Y, Postovit L-M and Lagugné-Labarthet F. Characterization of ovarian cancer-derived extracellular vesicles by surface-enhanced Raman spectroscopy. *Analyst*. 2021;146:7194-7206.

27. Lin K, Rubinfeld B, Zhang C, Firestein R, Harstad E, Roth L, Tsai SP, Schutten M, Xu K and Hristopoulos MJCCR. Preclinical Development of an Anti-NaPi2b (SLC34A2) Antibody–Drug Conjugate as a Therapeutic for Non–Small Cell Lung and Ovarian Cancers. 2015;21:5139-5150.
28. Xu R, Greening DW, Zhu H-J, Takahashi N and Simpson RJJTJoci. Extracellular vesicle isolation and characterization: toward clinical application. 2016;126:1152-1162.
29. Royo F, Théry C, Falcón-Pérez JM, Nieuwland R and Witwer KWJC. Methods for separation and characterization of extracellular vesicles: results of a worldwide survey performed by the ISEV rigor and standardization subcommittee. 2020;9:1955.
30. Witwer KW, Buzás EI, Bemis LT, Bora A, Lässer C, Lötvall J, Nolte-'t Hoen EN, Piper MG, Sivaraman S and Skog JJJoev. Standardization of sample collection, isolation and analysis methods in extracellular vesicle research. 2013;2:20360.
31. Raposo G and Stoorvogel WJJoCB. Extracellular vesicles: exosomes, microvesicles, and friends. 2013;200:373-383.
32. Mathieu M, Névo N, Jouve M, Valenzuela JI, Maurin M, Verweij FJ, Palmulli R, Lankar D, Dingli F and Loew D. Specificities of exosome versus small ectosome secretion revealed by live intracellular tracking of CD63 and CD9. *Nature communications*. 2021;12:1-18.
33. Mulcahy LA, Pink RC and Carter DRFJJoev. Routes and mechanisms of extracellular vesicle uptake. 2014;3:24641.
34. Kowal J, Arras G, Colombo M, Jouve M, Morath JP, Primdal-Bengtson B, Dingli F, Loew D, Tkach M and Théry CJPotNAoS. Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. 2016;113:E968-E977.
35. Zhao Z, Yang Y, Zeng Y and He MJLoaC. A microfluidic ExoSearch chip for multiplexed exosome detection towards blood-based ovarian cancer diagnosis. 2016;16:489-496.
36. Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A and Lajoie G. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid communications in mass spectrometry*. 2003;17:2337-2342.
37. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC and MacCoss MJ. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*. 2010;26:966-968.
38. Murugan A, Nair SAH and Kumar KJJoms. Detection of skin cancer using SVM, random forest and kNN classifiers. 2019;43:1-9.
39. Cawley GC. Leave-one-out cross-validation based model selection criteria for weighted LS-SVMs. *The 2006 ieee international joint conference on neural network proceedings*. 2006:1661-1668.
40. Lundberg SM and Lee S-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems*. 2017:4765-4774.
41. Tóth EÁ, Turiák L, Visnovitz T, Cserép C, Mázló A, Sódar BW, Försönits AI, Petővári G, Sebestyén A and Komlósi Z. Formation of a protein corona on the surface of extracellular vesicles in blood plasma. *Journal of extracellular vesicles*. 2021;10:e12140.
42. !!! INVALID CITATION !!! {}.
43. Coscia F, Watters K, Curtis M, Eckert M, Chiang C, Tyanova S, Montag A, Lastra R, Lengyel E and Mann MJNc. Integrative proteomic profiling of ovarian cancer cell lines reveals precursor cell associated proteins and functional status. 2016;7:1-14.
44. Matassa D, Amoroso M, Lu H, Avolio R, Arzeni D, Procaccini C, Faicchia D, Maddalena F, Simeon V and Agliarulo I. Oxidative metabolism drives inflammation-induced platinum resistance in human ovarian cancer. *Cell Death & Differentiation*. 2016;23:1542-1554.
45. Sherman-Baust CA, Weeraratna AT, Rangel LB, Pizer ES, Cho KR, Schwartz DR, Shock T and Morin PJJCc. Remodeling of the extracellular matrix through overexpression of collagen VI contributes to cisplatin resistance in ovarian cancer cells. 2003;3:377-386.
46. Bamberger E and Perrett CJMP. Angiogenesis in epithelian ovarian cancer. 2002;55:348.
47. Tyanova S, Temu T and Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature protocols*. 2016;11:2301.
48. Kuljanin M, Dieters-Castator DZ, Hess DA, Postovit LM and Lajoie GA. Comparison of sample preparation techniques for large-scale proteomics. *Proteomics*. 2017;17:1600337.
49. Kelstrup CD, Jersie-Christensen RR, Batth TS, Arrey TN, Kuehn A, Kellmann M and Olsen JVJJopr. Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field Orbitrap mass spectrometer. 2014;13:6187-6195.
50. Kuk C, Kulasingam V, Gunawardana CG, Smith CR, Batruch I, Diamandis EPJM and Proteomics C. Mining the Ovarian Cancer Ascites Proteome for Potential Ovarian Cancer Biomarkers* S. 2009;8:661-669.
51. Drabovich AP and Diamandis EPJJopr. Combinatorial peptide libraries facilitate development of multiple reaction monitoring assays for low-abundance proteins. 2010;9:1236-1245.
52. Shender VO, Pavlyukov MS, Ziganshin RH, Arapidi GP, Kovalchuk SI, Anikanov NA, Altukhov IA, Alexeev DG, Butenko IO and Shavarda AL. Proteome–metabolome profiling of ovarian cancer ascites reveals novel components involved in intercellular communication. *Molecular & Cellular Proteomics*. 2014;13:3558-3571.
53. Sinha A, Hussain A, Ignatchenko V, Ignatchenko A, Tang KH, Ho VW, Neel BG, Clarke B, Bernardini MQ and Ailles LJCS. N-Glycoproteomics of patient-derived xenografts: a strategy to discover tumor-associated proteins in high-grade serous ovarian cancer. 2019;8:345-351. e4.

54. Hebecker M, Okemefuna AI, Perkins SJ, Mihlan M, Huber-Lang M and Józsi M. Molecular basis of C-reactive protein binding and modulation of complement activation by factor H-related protein 4. *Molecular immunology*. 2010;47:1347-1355.
55. Pedersen S, Jensen KP, Honoré B, Kristensen SR, Pedersen CH, Szejniuk WM, Maltesen RG and Falkmer U. Circulating microvesicles and exosomes in small cell lung cancer by quantitative proteomics. *Clinical proteomics*. 2022;19:1-14.
56. Jing X, Liang H, Hao C, Yang X and Cui X. Overexpression of MUC1 predicts poor prognosis in patients with breast cancer. *Oncology reports*. 2019;41:801-810.
57. Beckwith DM and Cudic M. Tumor-associated O-glycans of MUC1: Carriers of the glyco-code and targets for cancer vaccine design. *Seminars in immunology*. 2020;47:101389.
58. Chang C, Worley BL, Phaëton R and Hempel N. Extracellular glutathione peroxidase GPx3 and its role in cancer. *Cancers*. 2020;12:2197.
59. Papp K, Végh P, Prechl J, Kerekes K, Kovács J, Csikós G, Bajtay Z and Erdei A. B lymphocytes and macrophages release cell membrane deposited C3-fragments on exosomes with T cell response-enhancing capacity. *Molecular immunology*. 2008;45:2343-2351.
60. Afshar-Kharghan V. The role of the complement system in cancer. *The Journal of clinical investigation*. 2017;127:780-789.
61. Bonavita E, Gentile S, Rubino M, Maina V, Papait R, Kunderfranco P, Greco C, Feruglio F, Molgora M and Laface I. PTX3 is an extrinsic oncosuppressor regulating complement-dependent inflammation in cancer. *Cell*. 2015;160:700-714.
62. Wessel D and Flügge U. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical biochemistry*. 1984;138:141-143.
63. Kulak NA, Pichler G, Paron I, Nagaraj N and Mann M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature methods*. 2014;11:319-324.