



Proteomic landscape of epithelial ovarian cancer

Received: 7 September 2023

Accepted: 19 July 2024

Published online: 31 July 2024

Check for updates

Lijia Qian^{1,2,3,4,9}, Jianqing Zhu^{1,2,3,4,9}, Zhangzhi Xue^{1,2,3,4,9}, Yan Zhou^{1,2,3,4,9}, Nan Xiang^{1,2,3,4,9}, Hong Xu^{7,9}, Rui Sun^{1,2,3,4}, Wangang Gong^{5,6}, Xue Cai^{1,2,3,4}, Lu Sun^{5,6}, Weigang Ge⁸, Yufeng Liu⁷, Ying Su^{5,6}, Wangmin Lin⁸, Yuecheng Zhan⁸, Junjian Wang^{5,6}, Shuang Song^{1,2,3,4}, Xiao Yi^{1,2,3,4}, Maowei Ni^{5,6}, Yi Zhu^{1,2,3,4,10}✉, Yuejin Hua^{1,2,3,4,10}✉, Zhiguo Zheng^{1,2,3,4,10}✉ & Tiannan Guo^{1,2,3,4,10}✉

Epithelial ovarian cancer (EOC) is a deadly disease with limited diagnostic biomarkers and therapeutic targets. Here we conduct a comprehensive proteomic profiling of ovarian tissue and plasma samples from 813 patients with different histotypes and therapeutic regimens, covering the expression of 10,715 proteins. We identify eight proteins associated with tumor malignancy in the tissue specimens, which are further validated as potential circulating biomarkers in plasma. Targeted proteomics assays are developed for 12 tissue proteins and 7 blood proteins, and machine learning models are constructed to predict one-year recurrence, which are validated in an independent cohort. These findings contribute to the understanding of EOC pathogenesis and provide potential biomarkers for early detection and monitoring of the disease. Additionally, by integrating mutation analysis with proteomic data, we identify multiple proteins related to DNA damage in recurrent resistant tumors, shedding light on the molecular mechanisms underlying treatment resistance. This study provides a multi-histotype proteomic landscape of EOC, advancing our knowledge for improved diagnosis and treatment strategies.

Ovarian carcinoma (OC) is among the deadliest cancers affecting the female reproductive system worldwide¹. Its high mortality is primarily due to difficulties in detecting it at an early stage, both inter- and intra-histological heterogeneity, and a high recurrence rate with limited effective treatment options². Epithelial ovarian carcinoma (EOC) is the most common histological type and is generally diagnosed at an advanced stage due to the intra-abdominal localization of the tumor and the absence of specific early stage symptoms. This late-stage diagnosis contributes to lower five-year relative survival rates compared to early

stage patients³. Recent findings from the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) underscore the value of early stage diagnosis for ovarian cancer. Longitudinal screening utilizing CA125 and ultrasound has exhibited promise in the early detection of ovarian cancer, leading to improved short-term treatment outcomes but limited benefit in mortality reduction. These results emphasize the pressing need for additional diagnostic biomarkers^{4,5}.

EOC can be subdivided into five histological subtypes, namely high-grade serous OC (HGSOC), low-grade serous OC (LGSOC),

¹School of Medicine, Westlake University, Hangzhou, Zhejiang Province, China. ²Westlake Center for Intelligent Proteomics, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang Province, China. ³Research Center for Industries of the Future, School of Life Sciences, Westlake University, Hangzhou, Zhejiang, China. ⁴Affiliated Hangzhou First People's Hospital, School of Medicine, Westlake University, Hangzhou, Zhejiang, China. ⁵Zhejiang Cancer Hospital, Hangzhou, Zhejiang, China. ⁶Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, Zhejiang, China. ⁷MOE Key Laboratory of Biosystems Homeostasis and Protection, Institute of Biophysics, College of Life Science, Zhejiang University, Hangzhou, China. ⁸Westlake Omics (Hangzhou) Biotechnology Co., Ltd., Hangzhou, Zhejiang Province, China. ⁹These authors contributed equally: Lijia Qian, Jianqing Zhu, Zhangzhi Xue, Yan Zhou, Nan Xiang, Hong Xu. ¹⁰These authors jointly supervised this work: Yi Zhu, Yuejin Hua, Zhiguo Zheng, Tiannan Guo.

✉ e-mail: zhuyi@westlake.edu.cn; yjhua@zju.edu.cn; zhengzg@zjcc.org.cn; guotannan@westlake.edu.cn

clear-cell OC (CCOC), endometrioid OC (EMOC), and mucinous OC (MCOC). Each subtype exhibits distinct clinical features, histotype-specific progenitor, and molecular alterations⁶. Notably, even within the same histotype, such as HGSOC, patients at different stages demonstrated critical variations in copy number signature exposures⁷. Furthermore, these tumors may exhibit diverse molecular subtypes, as indicated by transcriptomic data, which are associated with varying prognostic outcomes^{8–10}. These findings highlight the complex landscape of EOC, underscoring the importance of considering histological subtypes, disease stage, and molecular heterogeneity in the diagnosis, treatment, and prognosis of patients with OC.

For operable candidates, primary debulking surgery (PDS) and platinum-based chemotherapy have been the standard treatment regimen for primary EOC patients for over 20 years¹¹. For poor surgical candidates with advanced EOC, neoadjuvant chemotherapy followed by interval debulking surgery has proven to be beneficial in reducing residual disease¹². Unfortunately, nearly 75% of advanced-stage patients develop incurable recurrence¹³, and there are only a few effective treatment options available for such relapsed and chemoresistant cases¹⁴. As a result, there is an urgent need to understand the molecular underpinnings of EOC to facilitate the development of valuable diagnosis biomarkers and treatment strategies.

Genomic and transcriptomic technologies have been extensively utilized in studying EOCs, revealing genomic features among histotypes¹⁵, prognosis-related molecular subtypes within HGSOC^{8,10,15–17} and chemoresistance-associated molecular events^{18–20}. The advent of single-cell transcriptomics has further refined our grasp of HGSOC, elucidating both cell-intrinsic and microenvironmental phenotypes that contribute to its heterogeneity and therapy response^{21,22}. Despite these advancements, the clinical application of these genomic insights remains limited; the only current stratification of EOC patients for treatment, specifically with poly(ADP-ribose) polymerase inhibitors (PARPi), is based on the *BRCA* status^{23,24}.

As direct executors of biological processes, proteins play a pivotal role in comprehensive representations of molecular mechanisms underlying pathological changes. Recent proteome-based studies of EOCs have identified potential biomarkers for malignancy risk^{25–31}, the differential diagnosis between two or among three histotypes^{32–34}, patient stratification^{35–40}, prognosis prediction^{39,41,42}, and potential therapeutic targets^{34,43–45}. However, no systematic study has yet characterized protein alterations in EOC across all five histopathological features or after varied therapeutic regimens to provide additional clues for differential diagnosis, prognosis prediction and precise treatment.

Tumor biomarkers for detecting early stage EOC, monitoring treatment response, and predicting prognosis are crucial for reducing mortality. Currently, available CA125 or HE4 alone is approved to monitor disease progression and evaluate pelvic masses but not for early diagnostic biomarkers for EOC due to their limited specificity and sensitivity^{46,47}. Three multi-marker panels, namely ROMA, OVA1, and OVA2, were approved to assess the malignancy likelihood of ovarian adnexal masses^{47,48}. Remarkably, some of the markers in OVA1 and OVA2 were discovered by mass spectrometry (MS)-based proteomics²⁵. However, these protein panels are still not sensitive and specific enough for independent diagnosis⁴⁶.

Recent studies utilizing MS-based proteomics have discovered multiple potential biomarkers for EOC diagnosis^{26–31}, which require further validation in larger and independent cohorts of early stage EOCs. Various study designs have been employed to discover protein biomarkers for EOC, including blood samples directly from EOC patients versus control ones^{26,29}, integration of varied liquid specimens³¹, and discovery in plasma samples from PDX mouse models to improve identification of low-abundance proteins by concentrating human-derived proteins in the mouse serum^{28,30}. To improve diagnostic specificity, we hypothesize that biomarkers

secreted directly from EOC tissue samples might be more effective. Therefore, we adopted strategies to discover secreted proteins upregulated along with the increased malignancy in ovarian tissues of EOC cohorts and then validate them in plasma samples for potential diagnosis.

In this study, we reported a comprehensive proteomic landscape of 802 EOC tissues, 113 non-carcinoma tissues for control, as well as 180 plasma samples from 813 patients. We characterized the expression of 10,527 tissue proteins using pressure cycling technology (PCT)-Pulse data-independent acquisition (PulseDIA)⁴⁹, as well as 1660 plasma proteins using tandem mass tag (TMT)-based proteomics. We have verified eight biomarker candidates of EOCs using plasma samples from 2551 ascending proteins of tissue samples along tumor progression. Our findings provide insights into the molecular alterations underlying the five distinct histological subtypes of EOCs. Moreover, we have developed protein-based machine learning models capable of predicting one-year recurrence using pre-surgical plasma samples and surgically resected tissue specimens. Lastly, we performed integrative analysis of proteomic data and targeted region sequencing to further elucidate the molecular landscape of both primary and relapsed EOCs. These results provide valuable clues for the development of diagnostic and prognostic tools.

Results

A proteomic landscape of Chinese EOCs

We collected 753 surgically resected EOC tissues from three Chinese cohorts, namely a primary cohort with primary debulking surgery (PDS-EOC, $N=555$), a primary cohort with neoadjuvant chemotherapy (NACT-EOC, $N=74$), and a relapsed cohort (RLP-EOC, $N=124$) (Fig. 1A, Table 1, Supplementary Data 1). The PDS-EOC cohort comprised five histological subtypes of EOC, namely high-grade serous (HGSOC, 80.2%), low-grade serous (LGSOC, 3.4%), clear-cell (CCOC, 6.8%), endometrioid (EMOC, 7.0%) and mucinous (MCOC, 2.5%), while almost all samples in the other two cohorts were HGSOC (Fig. S1A, Table 1, Supplementary Data 1). In the PDS-EOC cohort, most cases (80.9%) were diagnosed as being in the advanced stage, with 30.3% of them becoming increasingly chemotherapy-resistant, resulting in recurrence within one year (Fig. S1A, Table 1, Supplementary Data 1). In addition, we enrolled 108 cases of patients with normal, benign and borderline ovarian tissues for comparison (Fig. 1A, Table 1, Supplementary Data 1).

Then, we performed proteomic profiling of 1114 ovarian tissue samples from 861 patients using PulseDIA⁴⁹, of which 1041 samples passed quality control (details in Method). In total, we quantified 10,527 proteins at high confidence (Fig. 1A, Supplementary Data 2) with negligible batch effects (Fig. S1B–E) and high reproducibility between replicates (Fig. S1F–G). A significantly greater number of proteins were identified in the carcinoma tissues compared with non-carcinoma tissues (Fig. 1B, Supplementary Data 3). Remarkably, unsupervised clustering of global proteome well separated normal and carcinoma tissues (Fig. 1C).

Malignancy-associated proteins in EOC tissue and plasma samples

We hypothesize that a subset of proteins, whose expression levels correlate with the tumor malignancy in tissue biopsies, can also be detected in blood samples. These proteins may serve as viable biomarkers for differentiating between benign and malignant states in a non-invasive manner, thereby augmenting the diagnostic capabilities for cancer management. First, we sought to identify proteins implicated in increasing malignancy degrees in ovarian tissues by comparing the proteome from normal, benign, borderline, early stage, to late-stage primary carcinoma samples. We utilized the Mfuzz method for this analysis (details in the Methods section)⁵⁰. We identified eight clusters of proteins that were consistently upregulated along with the increased malignant degree (Figs. 1D, S2, Supplementary Data 3). Furthermore,

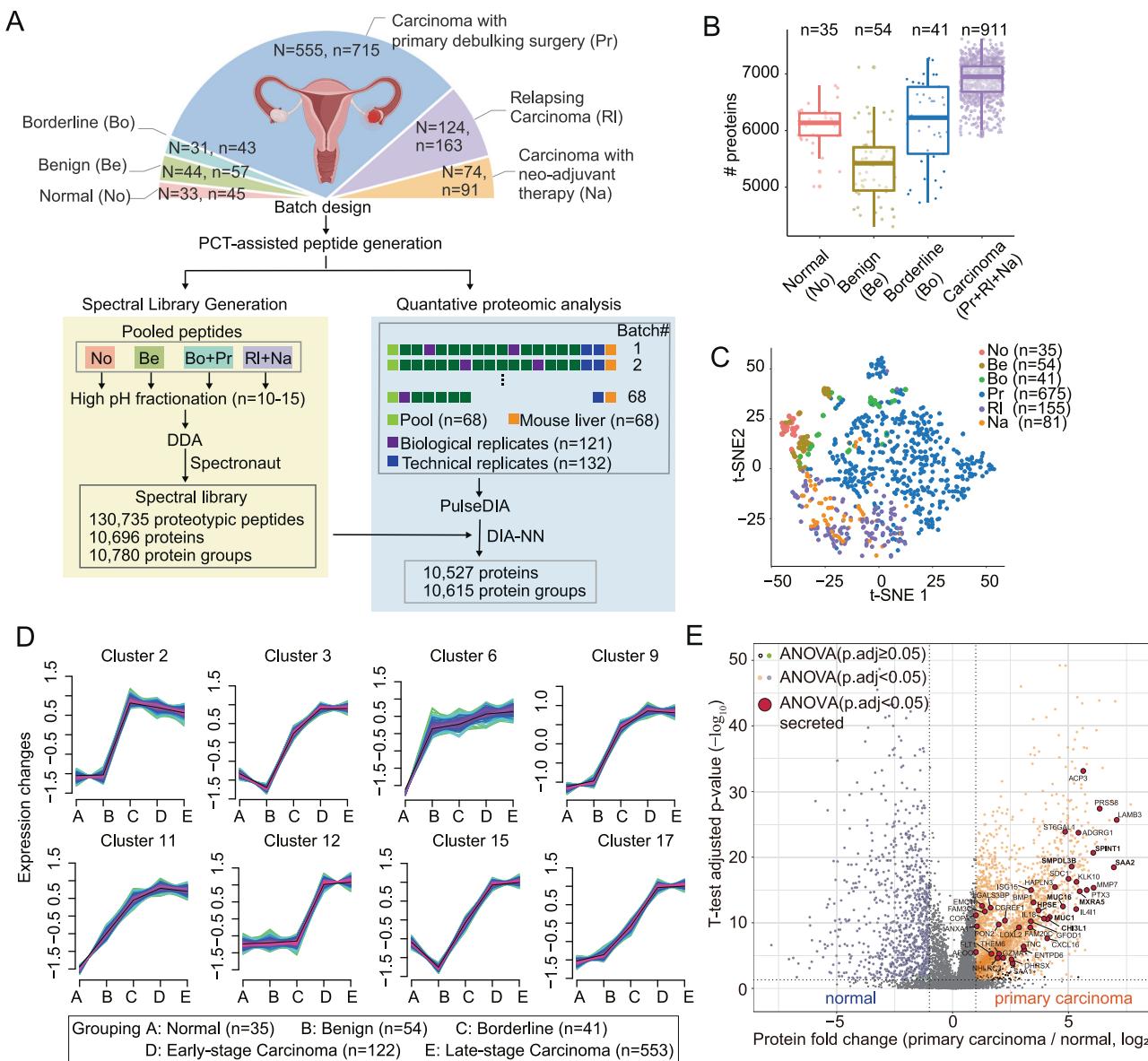


Fig. 1 | The proteomic landscape of Chinese EOCs. A Workflow for the generation of the proteomic landscape of Chinese EOCs. N represents patient number, n represents sample number, DDA represents data-dependent acquisition, DIA represents data-independent acquisition. Created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license (<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>). B The quantified protein numbers of four groups, namely normal, benign, borderline and carcinoma groups. The box in each group is constructed from the first to the third quartile, with a horizontal line indicating the median value. The whiskers extend from the box to represent the range of data within 1.5 times the interquartile range. C The unsupervised clustering of tissue samples using global proteomic data and t-distributed stochastic neighbor embedding (t-SNE). No represents normal group, Be represents benign group, Bo represents borderline group, Pr represents PDS-EOC cohort, RI represents RLP-EOC cohort, Na represents NACT-EOC cohort. D The

eight clusters selected by their ascending trend along with the increased malignancy. Proteins were considered significantly dysregulated across five groups if they presented a Benjamini–Hochberg (B–H) adjusted p -value of less than 0.05, as determined by one-way ANOVA. Following this initial selection, mFuzz clustering was employed to further categorize these proteins. The proteins with membership values by mFuzz less than 0.4 were excluded. E The volcano plot represents differentially expressed proteins (DEPs) between carcinoma samples from PDS-EOC cohort and normal tissues using two-sided unpaired Welch's t test. The colored dots represent those with B–H adjusted p -value less than 0.05 and fold change larger than 2. Those with B–H adjusted p -value greater than 0.05 among five groups, namely normal, benign, borderline, early stage carcinoma, and late-stage carcinoma groups, by One-Way ANOVA were labeled with orange and green. The solid red dots were amplified to highlight the selected DEPs. P_{adj} represents B–H adjusted p -value. Source data are provided as a Source Data file.

2551 of these ascending proteins were significantly upregulated in the primary carcinoma compared to normal tissues (Benjamini–Hochberg [B–H] adjusted p -value < 0.05 and $|\log_2(\text{fold change})| > 1$) (Fig. 1E, Supplementary Data 3). These malignancy-associated proteins were mostly enriched in energy metabolism and tumor necrosis factor (TNF) involved pathways (Supplementary Data 3). TNF is an important component of the malignant cell-autonomous network of inflammatory cytokines⁵¹ and has been reported to induce angiogenesis, infiltration of

myeloid cells, and extracellular matrix remodeling in the ovarian cancer microenvironment^{52,53}. Our data indicate that inflammatory-associated pathways, such as the neutrophil degranulation, tumocidal function of hepatic natural killer cells, Fcγ receptor-mediated phagocytosis, leukocyte extravasation signaling, and metastasis-associated pathways, such as RHO GTPase cycle, were all significantly activated along with increasing degree of tumor malignancy by Ingenuity pathway analysis (IPA) (Supplementary Data 3).

Table 1 | Patient information of EOC cohorts

Variables	Non-carcinoma			Carcinoma		
	Normal (N = 33)	Benign (N = 44)	Borderline (N = 31)	PDS-EOC (N = 555)	RLP-EOC (N = 124)	NACT-EOC (N = 74)
<i>Age—year</i>						
Mean \pm SD.	52.0 \pm 9.4	45.6 \pm 16.4	45.4 \pm 14.4	53.5 \pm 9.3	53.6 \pm 8.6	54.2 \pm 9.0
Median (IQR)	52.0 (49.0–55.0)	45.0 (35.0–59.2)	43.0 (35.0–58.0)	53.0 (47.0–60.0)	53.0 (48.0–59.2)	56.5 (46.0–59.8)
Range	18–68	16–87	22–77	23–80	32–76	36–77
<i>Histology—no. (%)</i>						
HGSOC				445 (80.2)	121 (97.6)	74 (100.0)
LGSOC				19 (3.4)	1 (0.8)	0 (0.0)
CCOC				38 (6.8)	0 (0.0)	0 (0.0)
EMOC				39 (7.0)	2 (1.6)	0 (0.0)
MCOC				14 (2.5)	0 (0.0)	0 (0.0)
<i>FIGO stage—no. (%)</i>						
I				44 (7.9)		0 (0.0)
II				62 (11.2)		2 (2.7)
III				373 (67.2)		43 (58.1)
IV				76 (13.7)		29 (39.2)
<i>Residual tumor—no. (%)</i>						
No residual				310 (55.8)	97 (78.2)	38 (51.4)
Less than 1 cm				182 (32.8)	19 (15.3)	29 (39.2)
Greater than 1 cm				62 (11.2)	5 (4.0)	7 (9.4)
NA				1 (0.2)	3 (2.4)	0 (0.0)
<i>Lymphatic metastasis—no. (%)</i>						
Yes				297 (53.5)	67 (54.0)	42 (56.8)
No				209 (37.6)	21 (16.9)	23 (31.1)
NA				49 (8.8)	36 (29.0)	9 (12.2)
<i>Frequency of chemotherapy—no. (%)</i>						
0–2				32 (5.8)	4 (3.2)	0 (0.0)
3–5				40 (7.2)	26 (21.0)	4 (5.4)
6–8				478 (86.1)	85 (68.5)	62 (83.8)
>8				5 (0.9)	3 (2.4)	8 (10.8)
NA				0 (0.0)	6 (4.8)	0 (0.0)
<i>Relapse-free survival—month</i>						
Mean \pm SD				25.1 \pm 23.7	14.4 \pm 12.2	12.8 \pm 14.4
Median (IQR)				18.0 (9.0–34.0)	10.0 (6.0–19.0)	8.0 (5.0–13.0)
Range				1–154	1–80	1–89

no. (%) number, SD standard deviation, IQR interquartile range.

Next, we focused on tissue-leakage proteins among these malignancy-associated proteins. Among 2551 ascending proteins, we identified 43 annotated as secreted proteins by Human Proteome Atlas⁵⁴. As predicted, we found that carbohydrate antigen 125 (CA125), the well-established EOC biomarker for monitoring the response to chemotherapy and detecting recurrence, was also singled out by this strategy (Fig. 1E, Supplementary Data 3). Moreover, 28 out of these were significantly upregulated between non-carcinoma groups (including benign and borderline samples) and at least one histotypes (B-H adjusted *p*-value <0.05 and FC >1.5) (Supplementary Data 3), indicating their potential to distinguish benign or borderline from carcinoma samples.

To further verify these 43 proteins in blood, we collected 168 plasma samples collected from the PDS-EOC cohort immediately before PDS and conducted proteomic profiling on them (Fig. 2A, Supplementary Data 1). We quantified 1660 proteins at 70% of the NA threshold (Fig. 2A, Supplementary Data 2) with negligible batch effects (Fig. S3A, S3B) and high reproducibility between technical replicates (Fig. S3C). Similar to the results of tissue proteome, unsupervised clustering of the global proteome showed that normal and benign

samples clustered tightly apart from carcinoma samples (Fig. S3D). We confirmed eight of these secreted candidates to be upregulated in the plasma of HGSOC patients compared to non-carcinoma individuals (Fig. 2A–C). These candidates included CA125 (also known as MUC16), mucin-1 (MUC1)⁵⁵, kunitz-type protease inhibitor 1 (SPINT1)⁵⁶, serum amyloid A-2 protein (SAA2)⁵⁷, chitinase-3-like protein 1 (CHI3L1)⁵⁸, matrix-remodeling-associated protein 5 (MXRAS)⁵⁹, and heparanase (HPSE)⁶⁰. All these proteins have been previously reported as overexpressed in clinical specimens (tissues, serum, uterine aspirates, ascites, etc.) from ovarian cancer patients relative to non-carcinoma controls. Notably, acid sphingomyelinase-like phosphodiesterase 3b (SMPDL3B) was also significantly upregulated in HGSOC group (Fig. 2B, C). SMPDL3B, an enzyme involved in lipid modulation of the cell membrane and a negative regulator of Toll-like receptor signaling on macrophages⁶¹, has been associated with impaired migration and a favorable prognosis in localized prostate cancer due to its high expression⁶². Our findings suggest that SMPDL3B may also serve as a potential biomarker for EOC.

To differentiate ovarian carcinoma from non-carcinoma, we constructed 15 machine-learning models. These models were

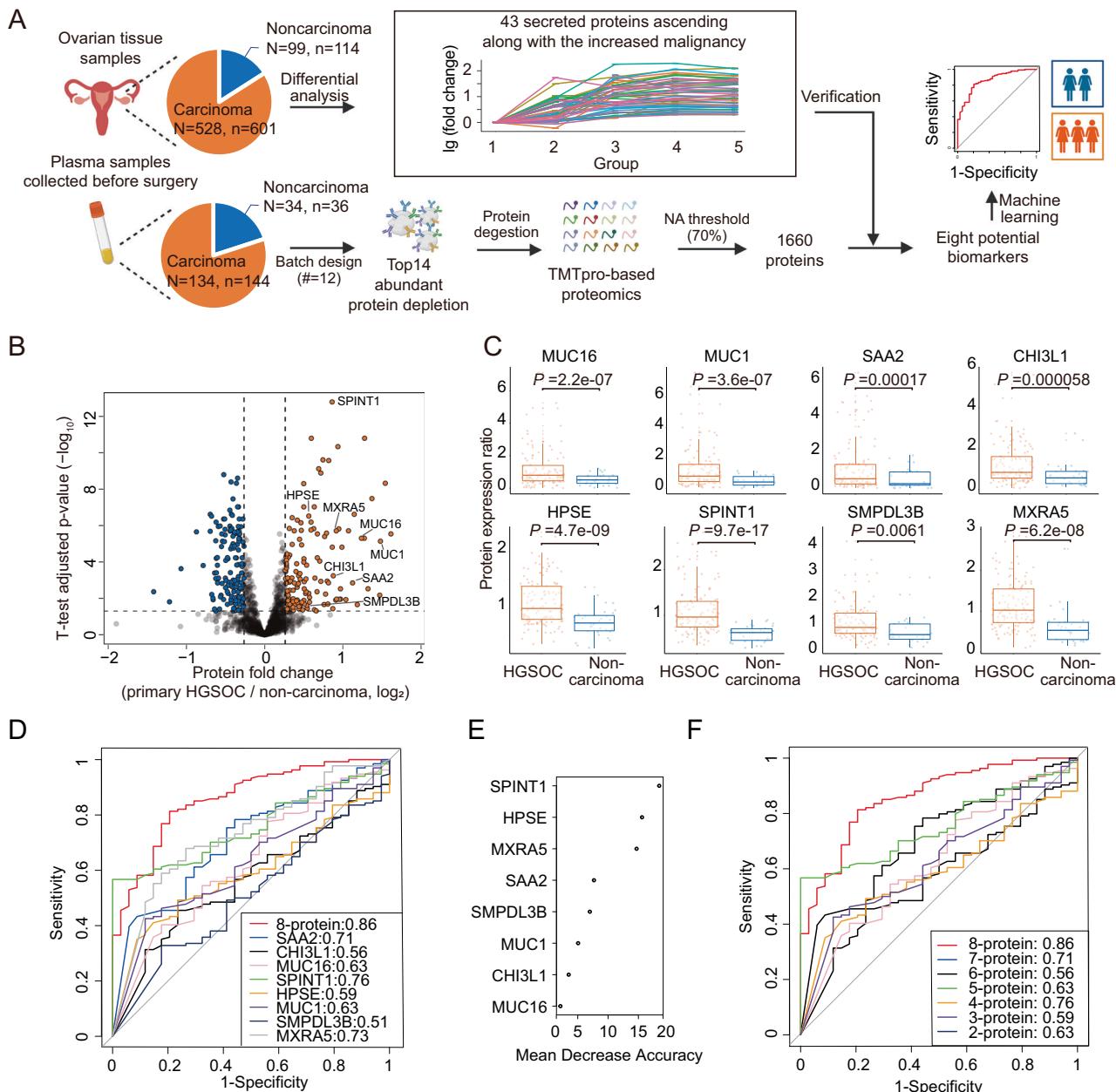


Fig. 2 | The verification of potential protein biomarkers for EOCs in plasma samples. A Proteomic profiling for plasma samples and the verification of potential protein biomarkers upregulated along with the increased malignancy of ovarian tissues in plasma samples. *N* represents patient number, *n* represents sample number, # represents batch number, NA represents missing values. Created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license (<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>). B The DEPs of plasma samples between primary HGSOC and non-carcinoma groups by two-sided unpaired Welch's *t* test. *P*-values were adjusted for multiple comparisons using the B-H correction method. C The expression of the eight validated proteins in plasma samples between primary HGSOC (*n* = 134) and

non-carcinoma (*n* = 34) groups. *P*-values were determined using a two-sided unpaired Welch's *t* test and subsequently adjusted using the B-H correction method. The box in each group is constructed from the first to the third quartile, with a horizontal line indicating the median value. The whiskers extend from the box to represent the range of data within 1.5 times the interquartile range. D The total AUCs by five-fold cross-validation using each feature and eight-feature model. E The average value of Mean Decrease Accuracy among five-fold cross-validation when building the eight-feature model. F The total AUCs by five-fold cross-validation using two to eight features selected by importance ranking. Source data are provided as a Source Data file.

developed using the eight validated proteins aforementioned and their varying combinations, applying random forest algorithms. More details are provided in the Methods. Our analysis prioritized an eight-feature classifier (Fig. 2D–F, Supplementary Data 3). Remarkably, the importance of MUC16/CA125 as detected in our experiment was inferior to the other seven proteins (Fig. 2E).

Our data nominate the eight-feature classifier as a promising approach for distinguishing ovarian carcinoma from non-carcinoma.

However, further validation in independent prospective cohorts is necessary in the future.

Proteomic features among five histological subtypes of Chinese EOC

The five histological subtypes of EOC present varying risk factors, cells of origin, genetic characteristics, clinical features, and treatments⁶. Our analysis of the PDS-EOC cohort revealed a significantly higher

incidence of lymphatic metastasis, advanced stage, and chemoresistance in HGSOC compared to other histological subtypes, as determined by the Fisher's exact test (Fig. 3A and S4A Supplementary Data 1). Yet, the majority of molecular studies have predominantly focused on HGSOC, leaving the systematic molecular characterization of the five histotypes underexplored.

Given that the RLP and NACT cohorts predominantly comprised HGSOC cases, the comparative molecular signatures among histotypes have been constrained to the PDS cohort. Our initial objective was to assess the reproducibility of potential biomarkers distinguished between histological subtypes. We began by comparing the proteomic data of HGSOC with each of the four other histotypes in the PDS-EOC cohort (Supplementary Data 4), as existing published datasets typically only compare two of them. As anticipated, several reported subtype-enriched proteins, such as MC-enriched PIGR, HGD, and CAPN6³³, as well as CC-enriched NAPSA, CRYAB, and NID2^{32,63}, were identified in our analysis (Fig. S4B). This underscores the high reproducibility of these histotype biomarkers.

To systematically investigated the proteomic feature among five histotypes from the PDS cohort, we identified a total of 4313 differentially expressed proteins (DEPs) between each histological subtype and normal ovarian tissue samples (Fig. S4C, Supplementary Data 4), with 2709 of them showing significant dysregulation among the five histotypes (Supplementary Data 4). Among the 2709 DEPs, 727 were commonly dysregulated in all subtypes, while 606 were subtype-specific dysregulated proteins (Fig. 3A, S4D, Supplementary Data 4). Through unsupervised clustering, we found that CCOC-specific upregulated proteins in cluster 4 were enriched in neutrophil degranulation, complement and coagulation cascades and multiple metabolic processes (Fig. 3A and S4E, Supplementary Data 4), which aligns with the observed higher incidence of thromboembolic complications⁶⁴ as well as the morphological enrichment of glycogen within the CCOC group. Due to the genomic similarities between clear cell histological subtypes of ovarian cancer (CCOC) and renal cancer (CCRC)⁶⁵, anti-angiogenic drugs, which are licensed treatments for CCRC, have been tested in multiple clinical trials for CCOC but have shown limited benefit (NCT00979992, NCT02315430 and NCT01914510)⁶⁶⁻⁶⁸. Intriguingly, our CCOC-specific protein patterns revealed upregulated protease inhibitors, such as ITIH1, ITIH3, and SERPINA1 in cluster 4 (Fig. 3A, B, Supplementary Data 4), as well as downregulated VEGF-associated proangiogenic proteins, such as SMAD1⁶⁹ and ARHGEF40⁷⁰ in cluster 2 (Fig. 3A, B, Supplementary Data 4). These findings suggest that angiogenesis is downregulated in CCOC, which may help explain the failure of VEGF inhibitors in clinical trials.

Our study reveals that a systematic analysis of five histotypes from the PDS cohort offers a more comprehensive perspective of differential biomarkers compared to previous studies that have only investigated two histotypes. For instance, earlier research reported polymeric immunoglobulin receptor (PIGR) as a diagnostic marker for EMOC, and its inclusion in the standard clinical marker panel (WT1 and TP53) improved diagnostic potential between HGSOC and EMOC³³. However, in our study, we observed that PIGR was significantly upregulated not only in EMOC but also in HGSOC, CCOC and MCOC groups when compared with normal groups. Although the fold change (FC) in HGSOC (FC = 16.65) was much smaller than that in EMOC (FC = 60.91), the values in CCOC (FC = 163.46) and MCOC (FC = 100.21) were larger than that in EMOC (Fig. 3C). Consequently, PIGR was not included in our histotype-specific DEP list due to its lack of specificity. Similar biases were also found in annexin A4 (ANXA4)⁷¹ and cellular retinoic acid-binding protein 2 (CRABP2)³² (Fig. 3C), indicating that these biomarker candidates might not be suitable for directly distinguishing the five histotypes of PDS-EOC.

Proteins potentially correlated with prognosis of advanced HGSOC

Recurrence was observed in over 80% of women with advanced ovarian cancer, unfortunately, the majority of recurrent cases are

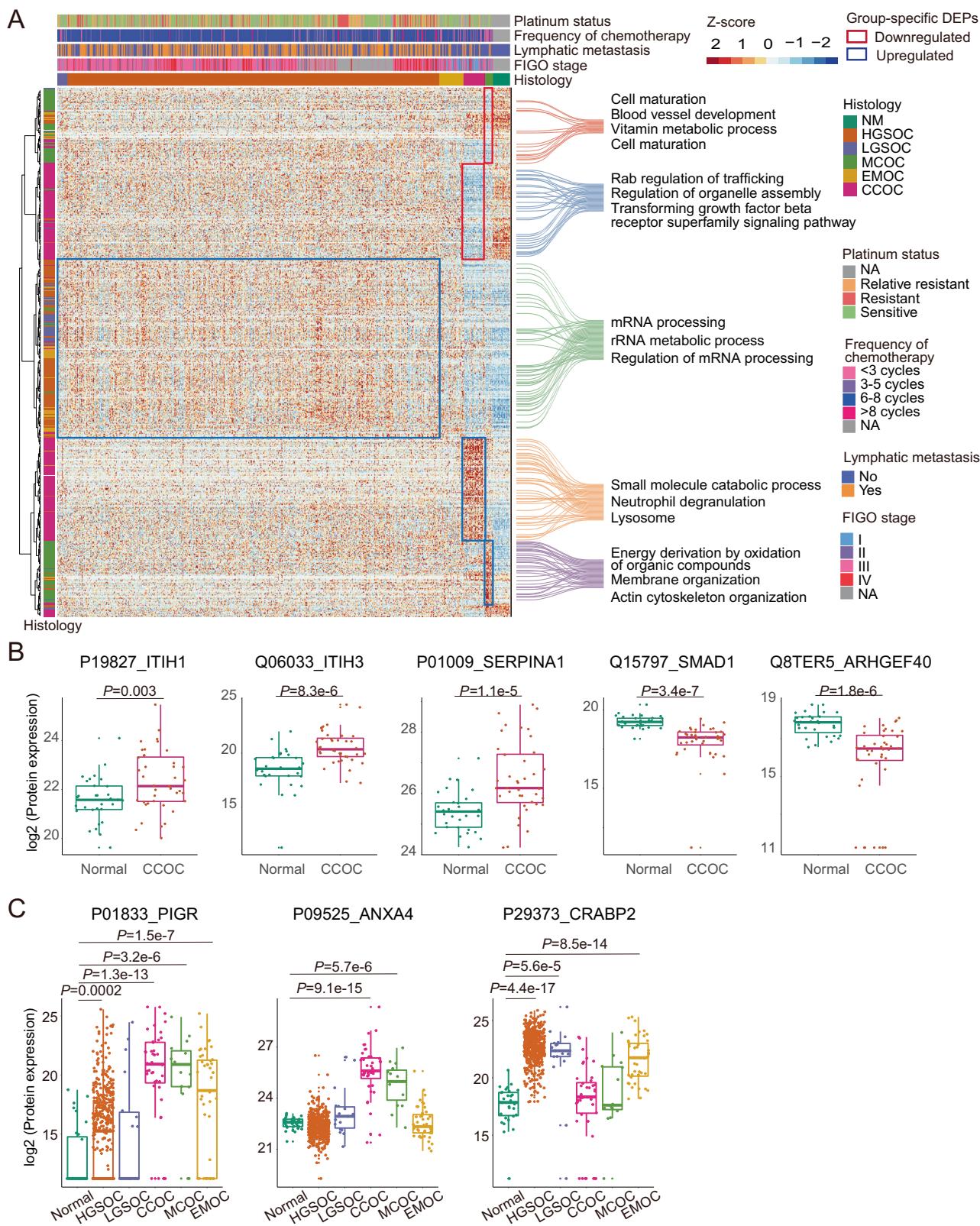
incurable⁶. We identified 572, 346, and 486 potentially prognostic proteins associated with RFS in advanced HGSOC tissues from PDS-, RLP- and NACT-EOC cohorts, respectively (Univariable Cox regression analysis, likelihood-ratio test, *p*-value < 0.05, Fig. 4A, Supplementary Data 5). Notably, most of these potentially prognostic proteins were cohort-specific (Fig. 4A, Supplementary Data 5), highlighting the diverse host responses following resection of primary cancer, relapsed cancer, and cancer with NACT. To further substantiate the prognostic relevance of the identified proteins, we compared our data with the proteomic findings of Zhang et al.³⁹ and Chowdury et al.⁴². We found a significant overlap of potential prognostic proteins between our PDS-HGSOC cohort and those identified in these studies (linear regression model, *p*-value < 0.05) (Supplementary Data 5). In the PDS-EOC cohort, inhibition of angiogenesis by thrombospondin-1 (TSP1) was specifically activated in patients with favorable prognoses (Fig. 4B, Supplementary Data 5). This finding is consistent with previous studies that have shown that the antiangiogenic domain of TSP1 could be utilized to design peptide mimetics for advanced EOC treatment, resulting in survival benefits in mouse models⁷². Although TSP1 expression in tissues is not significantly associated with prognosis, our data showed that its expression in plasma from patients in the PDS-HGSOC cohort correlated with an unfavorable prognostic (Supplementary Data 5).

In the NACT-HS cohort, we observed that spliceosomal cycle was specifically activated in patients with unfavorable prognosis (Fig. 4B, C, Supplementary Data 5). The PD-1/PD-L1 cancer immunotherapy pathway was inhibited in both primary and NACT cohorts with favorable prognosis (Fig. 4B, C, Supplementary Data 5), indicating the survival benefit of PD-L1/PD-1 blockade for these patients.

In the RLP-HGSOC cohort, we found that growth hormone signaling, erythropoietin signaling pathway, ERBB4 signaling, and autophagy were specifically inhibited in patients with unfavorable prognosis (Fig. 4B, Supplementary Data 5). Furthermore, the superpathway of inositol phosphate compounds and D-myo-inositol-5-phosphate metabolism were inhibited in patients with unfavorable prognosis of RLP-HGSOC cohort but activated in those of PDS-HGSOC cohort (Fig. 4B, D, Supplementary Data 5), indicating contrasting anticancer activities of inositol compounds between primary and relapsed cohorts. For instance, phosphoinositide phospholipase C (PLC) is a family of metabolic enzymes in the Phosphoinositide (PI) cycle, catalyzing the generation of two intracellular second messengers, DAG and InsP3. These messengers have been reported to influence major oncogenic signaling pathways, including PI3K/Akt/mTOR and protein kinase C signaling, which in turn regulate cancer cell motility, division, and death^{73,74}. In our data, elevated PIK3R3, PLCB1 and PLCB4 were associated with unfavorable prognoses in the primary HS, while upregulated PIK3R4 and PLCG2 in relapsed HGSOC were associated with favorable prognoses (Fig. 4C, E, Supplementary Data 5), highlighting the complex biology. In line with this, the therapeutic efficacy of multiple PI3K/Akt/mTOR inhibitors in recent clinical trials has demonstrated limited effectiveness in recurrent ovarian cancer patients (NCT01031381, NCT01283035, NCT01833169)⁷⁵⁻⁷⁷, emphasizing the need for more combination treatment strategies.

Characterizing prognostic plasma proteins of primary HGSOC patients

To investigate the prognostic proteins in the plasma of the PDS-HGSOC cohort, we performed a Univariable Cox regression analysis for the plasma proteome and found 124 unfavorable prognostic proteins and 73 favorable prognostic proteins (likelihood-ratio test, *p*-value < 0.05, Supplementary Data 5). These unfavorable prognostic proteins are mainly involved in inflammatory responses and extracellular matrix organization, whereas the favorable prognostic proteins are enriched in the adaptive immune system (Fig. 4F, Supplementary Data 5). Alpha-1-antitrypsin (SERPINA1), platelet-derived growth factor subunit A (PDGFA) and gelsolin (GSN) were



identified as the most significant risk factors, independent of clinical factors (Fig. 4G, Supplementary Data 5). SERPINA1, a protease inhibitor with anti-inflammatory and tissue-protective properties⁷⁸, has been reported to be elevated in both relapsed and nonresponsive patients with ovarian cancer⁷⁹.

Additionally, we identified 17 proteins as potentially prognostic in both tissue and plasma samples of the PDS-HGSOC cohort (likelihood-

ratio test, p -value < 0.05) (Fig. 4H, Supplementary Data 5). Among them, 14 were independent prognostic proteins, predominantly involved in immune responses and extracellular matrix organization (Fig. 4H). Proteasome subunit beta type-9 (PSMB9), a catalytic subunit of the immunoproteasome, outstood as an independent risk factor with favorable prognosis in both plasma and tissue samples (Fig. 4H, Supplementary Data 5). Its favorable prognosis at the mRNA level has

Fig. 3 | Histotype-specific proteins and their functions in the PDS-EOC cohort. A The heatmap visualizes the expression of 606 histotype-specific proteins among normal and five histotypes of primary EOCs. Five clusters of histotype-specific DEPs were highlighted in blue or red frame to label its dysregulation. Top two enriched pathways by Metascape were labeled by connecting lines with corresponding proteins. Outliers of z scores of protein expression were defined according to Tukey's fences where k equals 2. NA represents not available. B Differential expression of proteins involved in angiogenesis between normal ($n = 31$) and CCOC ($n = 40$) group by two-sided unpaired Welch's *t* test. The box in each group is constructed from the first to the third quartile, with a horizontal line indicating the

median value. The whiskers extend from the box to represent the range of data within 1.5 times the interquartile range. C Protein expression among normal group ($n = 31$) and five histotypes from PDS-EOC cohort (485 samples in HGSOC group, 18 samples in LGSOC group, 40 samples in CCOC group, 15 samples in MCOC group and 43 samples in EMOC group) by two-sided unpaired Welch's *t* test. The box in each group is constructed from the first to the third quartile, with a horizontal line indicating the median value. The whiskers extend from the box to represent the range of data within 1.5 times the interquartile range. Source data are provided as a Source Data file.

been reported in multiple cancer types, including ovarian cancer⁸⁰, which might be associated with enhanced tumor-infiltrating lymphocytes⁸¹. Leucine-rich alpha-2-glycoprotein (LRG1) was an independent and unfavorable prognostic protein in both plasma and tissue samples (Fig. 4H, Supplementary Data 5). LRG1, a regulator of pathogenic angiogenesis⁸², has been reported to induce epithelial-mesenchymal transition (EMT), leading to the migration of cancer cells⁸³. Therefore, the higher expression of LRG1 might promote angiogenesis and EMT, resulting in enhanced metastasis and earlier recurrence. The overlap between the prognostic proteins identified in tissue and plasma samples suggests that plasma proteomics might be a useful tool for non-invasive monitoring of ovarian cancer progression. Nonetheless, further validation studies in larger cohorts are needed to confirm the prognostic utility of these biomarkers.

Protein classifiers for prognosis prediction

To validate the prognostic proteins in tissue and plasma samples, we employed targeted proteomic methods and confirmed that 38 tissue proteins and 34 plasma proteins were associated with prognosis (Fig. 5A and S5A, Supplementary Data 6). Using these verified protein features and seven clinical factors, we developed three machine-learning models to predict one-year recurrence following the last cycle of chemotherapy (Fig. 5A and S5A). Model A, based on five clinical factors, failed to distinguish between two groups with significant prognostic differences (Fig. S5B-D). However, when tissue or plasma protein features were incorporated, both tissue model B and plasma model C effectively predict one-year recurrence in advanced primary HGSOC patients from an external validation cohort which comprises clinical samples collected at a different period from discovery and test cohort (Log-rank test, $p = 0.0094$ and 0.012, respectively, Figs. 5B-E and S5E-G). The CPTAC cohort also served as an external validation set to verify the tissue model B's applicability beyond the Chinese patient cohort. Due to the unavailability of two clinical factors and the AGRES protein expression data in the CPTAC cohort, we adjusted our model by retaining the original parameters of the remaining eleven protein features for predicting one-year recurrence. The revised model demonstrated significant predictive power in the CPTAC cohort (Log-rank test, $p = 0.047$, Fig. 5F).

In the tissue model B, multiple proteins have been identified as diagnosis biomarkers for ovarian cancer, such as alpha-2-macroglobulin (A2MG)⁸⁴, serum amyloid A-1 protein (SAA1)^{85,86} and transthyretin (TTHY)^{25,87}. Other proteins have been reported to be associated with chemoresistance and metastasis, including adhesion G protein-coupled receptor E5 (AGRES)^{88,89}, 14-3-3 protein sigma (1433S)⁹⁰, surfactant locus protein 4 (SURF4)⁹¹ and mitochondrial proton/calcium exchanger protein (LETM1)⁹² (Fig. 5B). It is noteworthy that proteasome subunit beta type-6 (PSMB6) in the tissue model B was also included in the prognostic model using copy-number alteration trans-affected proteins, and isochorismatase domain-containing protein 2 (ISOC2) was identified as prognostic proteins in CPTAC dataset³⁹. PSMB6, exhibiting hydrolytic activity for the cleavage of peptide bonds⁹³, is a component of ubiquitin-proteasome system (UPS), which has been demonstrated to be exploited by cancer cells to support their aberrant proliferation⁹⁴. Furthermore, multiple UPS inhibitors have

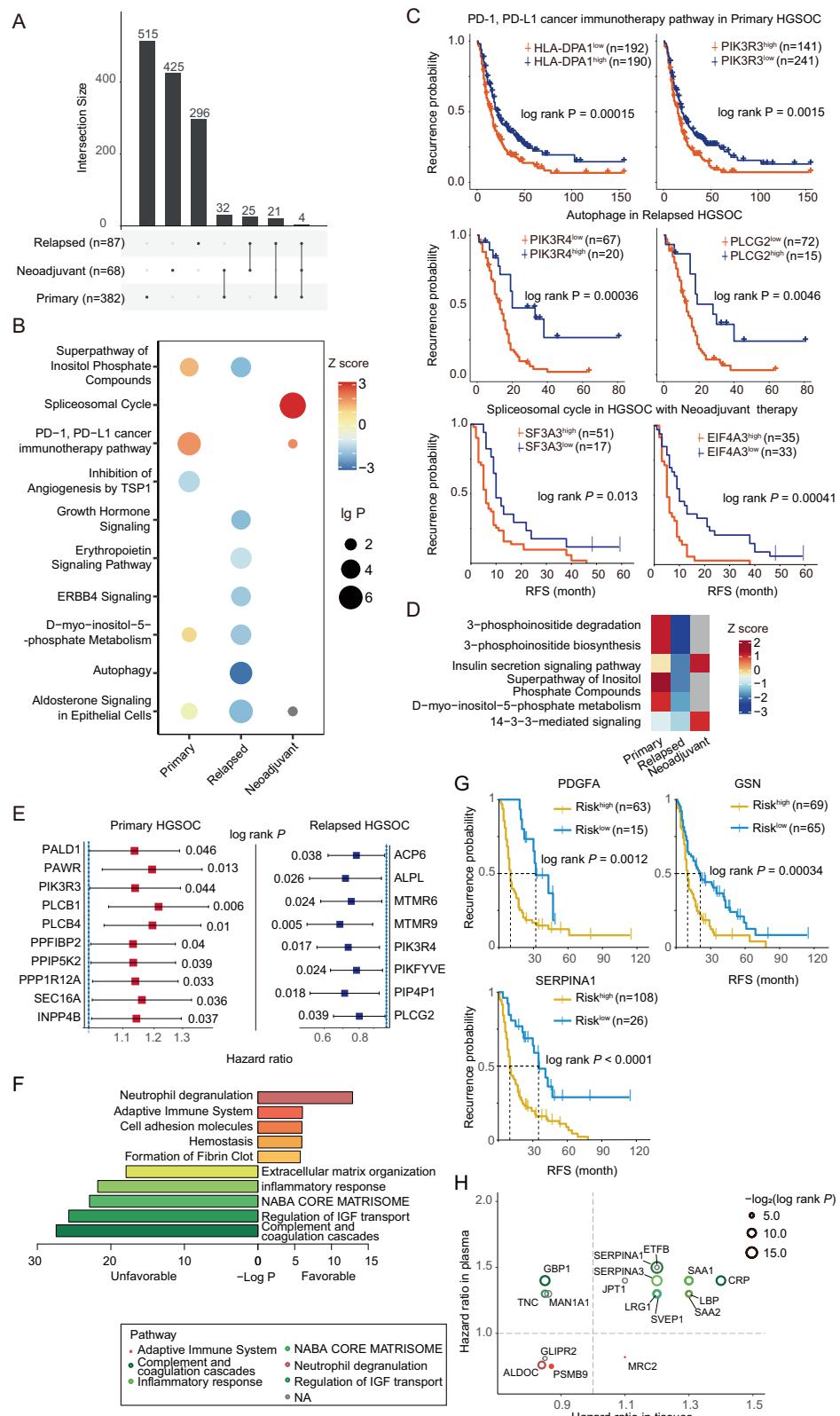
been approved for the treatment of patients with hematological malignancies^{94,95}, and some in vitro and in vivo studies have confirmed their antitumor activity in cell line and mouse models of EOCs^{96,97}. Our study has discovered and verified the prognostic significance of PSMB6 in HGSOC patients, warranting evaluation in clinical trials. In the plasma Model C, three proteins, namely properdin (PROP), C4b-binding protein alpha chain (C4BPA) and complement component C8 (CO8G), belong to the complement system (Fig. 5D), which has been implicated in proliferation and drug resistance in ovarian cancer cells, as well as being associated with unfavorable prognosis in patients⁹⁸. Additionally, coagulation factor XIII A chain (F13A) and alpha-1-antitrypsin (A1AT) are involved in coagulation process (Fig. 5D), which has been reported to have diagnosis⁹⁹ and prognosis¹⁰⁰⁻¹⁰² potential.

Divergent DNA Damage Response in HGSOC: Primary vs. Relapsed

To compare the potential drivers of chemoresistance between PDS and RLPS HGSOC patients, we conducted targeted genomic sequencing of a 295-gene panel for four groups of patients, namely primary sensitive, primary resistant, relapsed sensitive and relapsed resistant groups (Supplementary Data 7, Details in Methods). Consistent with previous reports^{8,103}, TP53 was the most common mutation, occurring in approximately 85% of patients in both PDS and RLPS cohorts (Fig. 6A, Supplementary Data 7). We observed a TP53 mutation prevalence comparable to that reported in Asian populations, approximately 80%^{104,105}. We then focused on the homologous recombination repair (HRR) pathway, as deficiency in HRR has been reported as a target for PARP inhibitors in HGSOC¹⁰⁶. We identified germline or somatic mutations of 14 HRR genes¹⁰⁷ in 56.20% of primary patients and 53.39% of relapsed patients (Fig. 6A, Supplementary Data 7), which mirrors the mutation prevalence reported in the TCGA project⁸. Mutations in HRR genes were significantly more prevalent in the primary sensitive group than the primary resistance group by Fisher's exact test. However, no difference was observed in relapsed patients (Fig. 6A, Supplementary Data 7).

To characterize the effect of HRR pathway mutations on chemosensitivity in primary HGSOC patients, we compared the proteome between chemosensitive HGSOC patients with HRR mutations and chemoresistant ones without any HRR mutations. We identified 53 upregulated and 206 downregulated proteins in chemosensitive HGSOC patients with HRR mutations (Fig. 6B, Supplementary Data 7). Among them, 33 proteins have been annotated to have direct interactions with 14 genes in the HRR pathway, and 32 out of these 33 proteins were downregulated in chemosensitive patients with HRR mutations (Fig. S6A). These 33 proteins were mainly involved in cellular response to DNA damage stimulus, mitotic cell cycle, chromosome organization and mismatch repair (Fig. S6B), suggesting that these processes might drive the chemosensitivity mediated by HRR mutations.

We also performed Fisher's exact tests for all mutated genes identified in our study between the sensitive and resistant groups of relapsed cohorts and found that none of them showed a difference between the two groups (Fig. 6A, Supplementary Data 7). This led us to



hypothesize that proteomic profiling could provide additional insights into the molecular mechanisms underlying chemoresistance in recurrent HGSOC patients, beyond gene mutations.

We then identified DEPs stratified by chemotherapy responses in the relapsed cohort. Only 11 DEPs were found to overlap between the primary and relapsed cohorts (Fig. 6C). Notably, we observed that DNA repair-involved proteins were upregulated in relapsed chemosensitive

patients (Fig. 6D, E). This is in contrast to primary chemosensitive patients with HRR mutations, where these proteins were downregulated (Fig. 6E). These findings suggest that the chemosensitivity of the relapsed cohort may not be primarily driven by HRR-associated processes. Besides, a total of 163 upregulated proteins in relapsed chemosensitive patients were also enriched in histone modification and adaptive immune, while 193 downregulated ones were involved in

Fig. 4 | Prognostic analysis of tissue and plasma samples from three HGSOC cohorts. **A** UpSet plot shows the intersections of prognostic proteins of tissues among advanced carcinoma samples from PDS-HGSOC, RLP-HGSOC and NACT-HGSOC cohorts. n represents samples numbers. **B** Pathway enrichment for prognostic proteins of tissues across three cohorts by IPA. P-values are derived from one-sided Fisher's Exact Test for pathway enrichment. Pathways with $-\lg(p\text{-value})$ larger than 4 in each cohort were selected and those without Z scores were excluded. The radius of the circle represents $-\lg(p\text{-value})$ for each pathway while the color represents the Z score of each pathway. **C** Prognostic differences distinguished by the expression of prognostic proteins from the representative pathway in each cohort. The optimal cutpoint for each protein was determined by

surv_cutpoint function in R package survminer. n represents samples numbers. **D** The heatmap represents the pathways with opposing roles in prognosis among different cohorts. **E** The hazard ratios of the key proteins in metabolism of inositol compounds between PDS-HGSOC ($n=382$) and RLP-HGSOC ($n=87$) cohorts. Data are presented as mean values accompanied by a 95% confidence interval. **F** The enriched pathways for prognostic proteins of plasma using Metascape. P-values are derived from one-sided Fisher's Exact Test for pathway enrichment. **G** The Kaplan–Meier plots represent the prognostic significance of three clinical factor-independent proteins from plasma samples. **H** Prognostic proteins in both tissue and plasma samples. Source data are provided as a Source Data file.

innate immune, wound healing and VEGF-associated signals (Fig. S6C, S6D).

Discussion

In our study, we have successfully identified eight potential tumor-leakage biomarkers in blood through an integrative proteomic analysis of tissue and plasma samples. We also pinpointed 606 differentially expressed proteins specific to histotypes, thus revealing pathological alterations unique to five histotypes and the potential molecular mechanisms influencing clinical outcomes. Notably, when prognostic proteins were compared across three cohorts of HGSOC with differing therapeutic regimens, we observed a clear contrast in the impact of inositol compound metabolism on recurrence between primary and relapsing EOC patients. To predict one-year recurrence, we first validated prognostic proteins using multiple reaction monitoring assays and then constructed two machine learning models, each based on tissue and plasma proteomic data. Both models exhibited significant differences in recurrence-free survival when applied to two groups in the external validation cohort (Log-rank test, p -values = 0.0094 and 0.012, respectively), outshining the model based solely on five clinical factors (p -value = 0.079). Our integrative analysis of sequencing data from 295 genes and proteome data revealed that mutations in genes associated with HRR and resultant proteomic alterations in cellular responses to DNA damage stimuli could potentially drive chemoresistance in primary EOC patients. Interestingly, in relapsing patients, mutations in the HRR pathway seemed to have limited associations with chemoresistance. Simultaneously, proteomic regulation in DNA damage and repair appeared to exert an inverse effect on chemoresistance compared to primary cases.

EOC is a highly fatal gynecologic cancer with stagnant mortality rates, underscoring the need for effective biomarkers for early diagnosis, treatment monitoring, and prognosis prediction. Although CA125 and HE4 are approved for monitoring disease progression, their diagnostic specificity and sensitivity remain limited^{46,47}. Recent trials have shown some promise in early detection through longitudinal screening, yet the impact on mortality reduction remains minimal, highlighting the urgent need for additional diagnostic biomarkers^{4,5}. To enhance diagnostic specificity, we focused on secreted proteins upregulated with increased malignancy in EOC tissues, validated in plasma samples for potential diagnosis.

Previous research has underscored the utility of multiple biomarkers in detecting ovarian cancer, with multi-marker panels like ROMA, OVA1, and OVA2 being pivotal for assessing the malignancy of ovarian adnexal masses^{25,47,48}. In our study, we constructed machine learning models based on individual proteins and a combination of multiple proteins. The eight-protein model notably outperformed the single protein model in AUC, emphasizing the advantages of multiple biomarkers. Among these proteins, SPINT1, HPSE, and MXRA5 were the top performers in terms of Mean Decrease Accuracy, reflecting their crucial role in differentiating between carcinoma and non-carcinoma conditions. Interestingly, the MXRA5-based model demonstrated the highest AUC (0.82) in single protein modeling, surpassing the AUC of the CA125-based model and indicating the

potential discriminatory power of MXRA5. Although MXRA5 is known to be overexpressed in the ascites of ovarian cancer patients⁵⁹, ascitic fluid is not ideal for routine risk assessment due to its prevalence in advanced-stage patients. The potential diagnostic significance of plasma MXRA5 level has not been reported in the literature. However, additional validation is needed in multicenter cohorts, including early stage EOC patients and samples from other pelvic conditions.

HGSOC is the most prevalent and aggressive histological subtype of ovarian cancer, accounting for approximately 70% of all ovarian cancer-related fatalities. While the initial response to chemotherapy in most HGSOC patients is often positive, a significant number of patients unfortunately experience a relapse within a short time frame, with some even developing resistance to treatment¹⁴. This has led us to focus the following analysis on HGSOC, aiming to identify potential prognostic biomarkers and reveal the underlying molecular mechanisms responsible for chemoresistance. The capability to foresee chemoresistance before commencing chemotherapy could be crucial in tailoring treatment strategies, thereby improving patient outcomes.

To address this challenge, we have developed two predictive models based on the multi-protein expression of surgical tissue specimens and plasma samples collected before surgery. The models imply that if a patient's score exceeds 0.5, they may be more responsive to chemotherapy. These biomarkers could potentially segregate patients who are likely to be more sensitive to chemotherapy. The success of these predictive models underscores the promising potential of proteomic profiling of peripheral blood in guiding patients towards specific therapies, including immunotherapies, thereby contributing significantly to the advancement of precision medicine in the context of heterogeneous EOC.

A recently published proteogenomic study aimed to distinguish between refractory EOC cases (those with disease progression or stability during chemotherapy) and sensitive cases (those without disease progression within 2 years) in a dataset of 242 high-grade serous ovarian cancer (HGSOC) tissue samples⁴². The study unveiled a 64-protein model with potential predictive power for refractory diseases. Our study, in contrast, aspired to construct chemoresistance predictive models with more substantial clinical translational potential within a Chinese cohort. Firstly, we sought to predict one-year recurrence in an expanded cohort of 380 HGSOC patients without applying any selection criteria based on clinical factors. Secondly, we examined the feasibility of predicting chemoresistance using both surgical tissue samples and blood samples collected immediately before surgery, thereby exploring the potential of liquid biopsy-based predictions as a less invasive and more accessible approach. Thirdly, to ensure the practical applicability of our predictive models, we developed high-throughput and clinical-grade assays for both tissue and plasma protein features. These assays employ a 15-minute gradient using multiple reaction monitoring (MRM), implemented on a clinically approved medical-device mass spectrometer, facilitating efficient and rapid analysis of protein biomarkers. Furthermore, the MRM assays were rigorously validated in independent cohorts, guaranteeing the reliability and robustness of our predictive models.

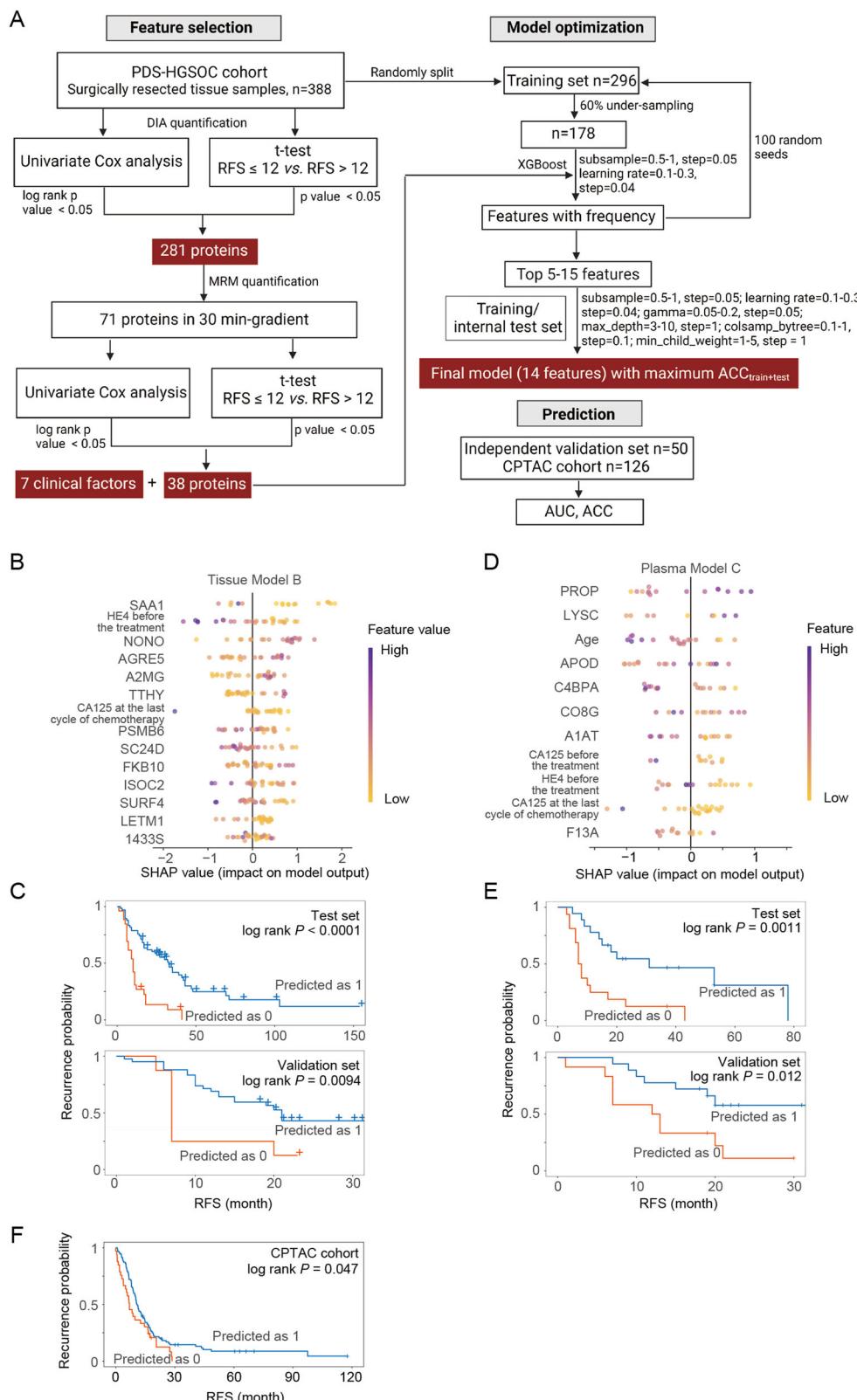
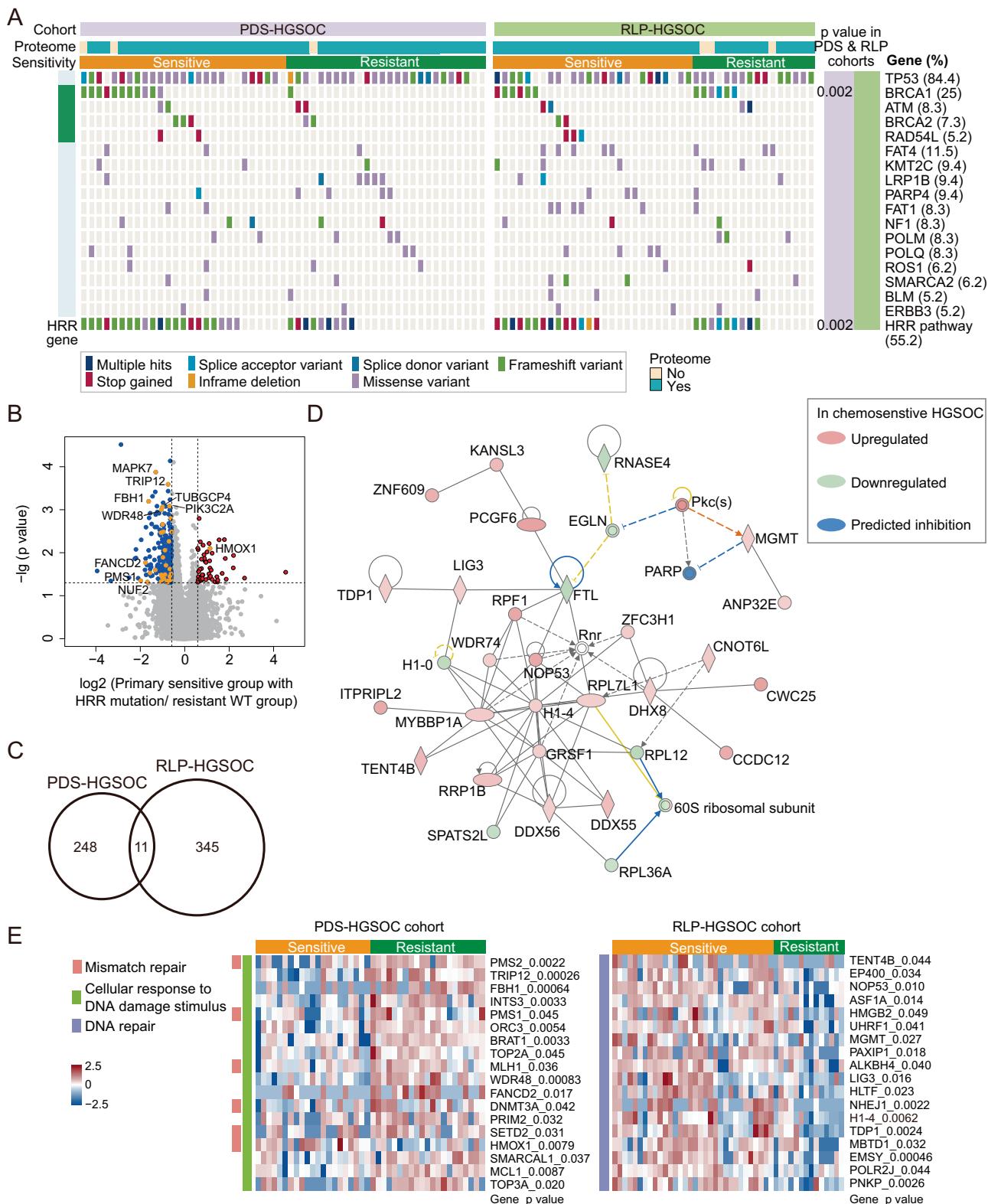


Fig. 5 | Prognostic models to predict one-year recurrence in primary HGSOC patients. **A** The workflow of machine learning model based on tissue proteomic data. *n* represents sample numbers. **B** The impact of 12 tissue proteins and two clinical factors on tissue Model B output. **C** The Kaplan-Meier plot of the predicted groups by the tissue Model B in the internal test and external validation sets. **D** The impact of 7 plasma proteins and four clinical factors on plasma Model C output.

E The Kaplan-Meier plot of the predicted groups by the plasma Model C in the internal test and external validation sets. **F** The Kaplan-Meier plot for groups predicted by the classifier, which incorporates the remaining 11 features from tissue Model B, as applied to the CPTAC dataset. Source data are provided as a Source Data file.



Our validation efforts extended to cohorts beyond China, including the CPTAC³⁹ and Chowdhury cohorts⁴². Despite only being able to utilize 11 of the 14 features from our model, the CPTAC cohort was still instrumental in stratifying patients into two groups with significantly different Recurrence-Free Survival (RFS), albeit with a *p*-value in close proximity to 0.05³⁹. The two factors that were excluded from the model were both clinical factors, namely CA125 and HE4 levels, both of which are routinely measured indicators in clinical

practice. Additionally, the absent protein, AGRE5, can be quantified using MRM, underscoring its feasibility of detection in a clinical environment. This reinforces the robustness and validity of our model. However, our model did not demonstrate the same efficacy in distinguishing between chemotherapeutic refractory and sensitive patients within the Chowdhury cohort, even though there was some overlap in prognostic proteins between the two cohorts⁴². This discrepancy could be attributed to several factors: (1) variations in the methods employed

Fig. 6 | Integrative analysis of target sequencing and proteome. A Genetic variants of 295-gene panel in 53 primary HGSOC (PDS-HGSOC) patients and 43 relapsed HGSOC (RLP-HGSOC) patients. The associations between gene mutations and chemoresistance were performed by Fisher's exact test: **, $0.001 \leq p\text{-value} < 0.01$. B The volcano plot showed comparison of proteome between chemosensitive HGSOC patients with HRR mutations and chemoresistant ones without any HRR mutations by two-sided unpaired Welch's t test. The colored dots represent DEPs. The orange dots represent 33 DEPs which were annotated to have direct interactions with 14 genes in the HRR pathway, and those with significant p -value or fold change were labeled with gene names. HRR, homologous recombination repair; WT, patients without gene mutations in HRR pathway. C The overlapping relationship of chemoresistance associated DEPs between primary and relapsed

HGSOC cohorts. PDS-HGSOC, primary HGSOC patients; RLP-HGSOC, relapsed HGSOC patients. D The most enriched network using DEPs between chemoresistant and chemosensitive groups of relapsed HGSOC cohort. P -values are derived from one-sided Fisher's Exact Test for pathway enrichment. The green and red patterns represent downregulated and upregulated proteins in chemosensitive groups, while the blue ones were predicted to be inhibited by IPA. The shape of each molecule is according to legends of IPA. E The heatmap showed the expression of DNA damage repair associated DEPs between chemosensitive and chemoresistant groups of primary and relapsed patients by two-sided unpaired Welch's t test. The outliers of protein expression were defined as those with absolute values of Z-score larger than 2.5, and these outliers were imputed as 2.5 in the heatmap. Source data are provided as a Source Data file.

for mass spectrometry quantification; (2) the inherent high heterogeneity observed in high-grade serous ovarian cancers; and (3) the complex interplay of mechanisms that underpin responses to platinum-based treatments. Despite these promising results, further validation in prospective cohorts, as well as absolute quantification employing stable isotope-labeled peptides, remains necessary before these models can be integrated into clinical practice.

Our study, while offering insights, is not without its constraints, which are necessary to acknowledge for a holistic understanding of our findings. Firstly, ovarian carcinoma can originate either from the Ovarian Surface Epithelium (OSE) or the Fallopian Tube Epithelial (FTE)¹⁰⁸. Given the inherent challenges in distinguishing the precise origin of each carcinoma and obtaining FTE samples from the same patients in a clinical setting, we have attempted to reduce potential biases. This was achieved by excluding DEPs that could be attributed to differences in tissue type³⁶ when comparing ovarian carcinoma and normal ovarian tissue samples (details in Method). Secondly, this study provides a comprehensive collection of clinical specimens from EOC samples. We have endeavored to interpret the biological significance of differentially expressed proteins from each comparison as thoroughly as possible. However, due to the vast amount of data and the exploratory nature of this study, we have not provided an exhaustive analysis of the biological significance of all identified differentially expressed proteins. Thirdly, while numerous histotype-specific biomarkers have been identified in our cohort, the predominance of HGSOC tumors over the four other less common histological subtypes of EOC may introduce potential biases in the analysis of DEPs. Fourthly, while our study reveals a higher frequency of chemoresistance in HGSOC compared to other subtypes, we acknowledge the established chemoresistant nature of most mucinous and LGSOC tumors as reported by Lheureux et al.¹³. The lower frequency of chemoresistance in these subtypes observed in our study may be attributed to their relative rarity, resulting in a smaller sample size. Lastly, clinical specimens for the NACT cohort were collected at the time of interval debulking surgery. Consequently, neoadjuvant chemotherapies could potentially influence the identification of prognostic proteins across the three cohorts. To further elucidate these changes, the collection of serial samples would be beneficial for future studies.

Methods

Ethics statement

This study was approved by the Medical Ethics Committee of Zhejiang Cancer Hospital (IRB-2020-155) and Medical Ethical Committee of Westlake University (20190401GTN0009). The methodologies employed in this study adhered to the ethical standards outlined in the Declaration of Helsinki. Informed consent was signed before sample collection.

For the purposes of quality control within our mass spectrometry analysis, liver samples were sourced from a single eight-week-old male C57BL/6 mouse. Upon collection, these samples were promptly stored at -80°C to ensure their preservation until required for protein

extraction. It is important to note that these mouse liver peptides were solely employed as a technical reference to validate the accuracy and reproducibility of our mass spectrometry procedures and were not utilized for any comparative proteomic analyses. As such, their use does not bear on the scientific findings related to the human clinical samples that are central to our study. The use of a single male mouse was deemed sufficient for the technical purpose it served, which is why additional consideration of sex in the study's design and analysis was not applicable.

All experimental procedures involving animals were conducted in strict accordance with ethical standards and were approved by the Laboratory Animal Resources Center of Westlake University, under the approval number 19-027-GTN.

Sample collection and histological analysis

In our study on ovarian cancer, we focused solely on female patients because this type of cancer occurs only in women. Therefore, our research did not include a comparison of sexes or genders. We did not collect gender identity through self-report since our study was based on the biological characteristics of the disease, which are specific to female reproductive organs.

Surgically resected EOC tissue samples were collected from 753 patients, comprising 555 primary EOC patients who underwent primary debulking surgery (PDS cohort), 74 primary EOC treated with neoadjuvant chemotherapy (NACT cohort), and 124 relapsed EOC patients (RLP cohort; Supplementary Data 1). Additionally, 108 non-carcinoma patients were included for comparison, consisting of 33 normal cases with uterine myoma or cervical cancer but without histologically documented ovarian involvement, 48 benign cases and 31 borderline cases. All patients were diagnosed between 2006 and 2018 following the WHO classification of Tumors: Female genital Tumors (fifth edition). Tissue specimens were collected without any clinical factor-based selection criteria other than availability and were stored at -80°C . Samples of NACT cohort were collected at the interval debulking surgery. Prior to pathological examination, the tissue specimens were embedded in optimal cutting temperature (OCT) compound, and subsequently subjected to freezing microtome sectioning. Two senior pathologists independently confirmed the histologic subtypes and proportions of tumor nuclei ($>60\%$) using the hematoxylin and eosin-stained histological slices. Surgical staging was determined according to the 1988 International Federation of Gynecology and Obstetrics (FIGO) staging system. A total of 180 plasma samples were collected immediately prior to surgery from 34 cases of patients with normal or benign ovarian tissues and 134 cases of primary patients with high-grade serous carcinoma of PDS-EOC cohort.

Supplementary Data 1 contains comprehensive patient data, including age at diagnosis, residual tumor size, histotype, tumor grade, International Federation of Gynecology and Obstetrics (FIGO) stage, lymph metastasis status, chemotherapy frequency, recurrence status, recurrence-free survival time, pre-treatment levels of CA125 and HE4, CA125 levels after the last chemotherapy cycle, and the administration

of Bevacizumab or PARP inhibitor therapy. All patients included in this study received platinum and taxane therapy. Any additional treatments with Bevacizumab or PARP inhibitors are indicated in Supplementary Data 1. Patients were divided into two groups based on their response to adjuvant therapy. Those who relapsed within 6 months following the last cycle were classified as the resistant group, while those who relapsed more than 6 months after the last cycle were identified as the sensitive group.

For external validation, 57 tissue samples from 49 primary HGSOC patients with PDS and 30 paired plasma samples were collected between 2018 and 2019. All patients were diagnosed as FIGO stage III and IV. These patients received a minimum of six cycles of platinum-based chemotherapy following PDS. Detailed clinical information is listed in Supplementary Data 1.

Batch design

In the discovery cohort, tissue samples were randomly distributed to 68 batches to minimize the batch effect. Multiple replicate samples were designed to monitor the quality during Pressure Cycling Technology (PCT)-assisted sample preparation and PulseDIA on the same Q Exactive HF hybrid Quadrupole-Orbitrap (QE-HF) (Thermo Fisher Scientific). During sample preparation, each batch contained one mouse liver sample and 121 cases of biological replicates, which represent different ovarian tissue samples dissected from the same patient. During MS acquisition, each batch included one pooled peptide sample combined from all samples and 132 cases of technical replicates, which represent the same peptide sample were run twice (Supplementary Data 1). Plasma samples were randomly distributed into 12 batches. Twelve biological replicates were designed during sample preparation, while during MS acquisition, each batch contained one pooled peptide sample labeled with the TMT126 channel.

Proteomics data acquisition

Approximately 1 mg of fresh frozen specimens was weighed and washed sequentially with ethanol solutions to remove OCT compound¹⁰⁹. The process included an initial wash with 70% ethanol, followed by a rinse with water, and subsequent washes with increasing ethanol concentrations (70%, 85%, and 100%) for efficient OCT removal, each involving vortexing and supernatant discarding steps. A four-step PCT-assisted lysis and digestion were then performed, generating peptide samples for each tissue specimen^{49,109,110}. The procedure involved lysing tissues in urea-thiourea buffer within PCT-MicroTubes under pressure cycling, followed by reduction and alkylation with TCEP and IAA. Lys-C (enzyme-to-substrate ratio = 1:40) and trypsin (enzyme-to-substrate ratio = 1:50) were then sequentially added for proteolytic digestion under pressure cycling. The reaction was quenched with TFA, and peptides were obtained by centrifugation, yielding the peptide samples for analysis. The peptide samples were desalted, dried, redissolved in buffer A (2% ACN, 0.1% formic acid), and their concentrations were measured prior to MS analysis¹⁰⁹. Generated peptides were injected and separated over a 30-minute LC gradient on a nanoflow DIONEX UltiMate 3000 RSLC nano System connected to a Q Exactive HF-X hybrid Quadrupole-Orbitrap (Thermo Fisher ScientificTM, San Jose, USA). PulseDIA parameters were set with two schemes of complementary and discontinuous isolation windows across two injections, each with 1 *m/z* overlap between adjacent windows⁴⁹. MS1 scans covered a range of 390–1210 *m/z* at 60,000 resolution, with an AGC target of 3e6 and a maximum ion injection time of 80 ms. MS2 scans were performed at 30,000 resolution with an AGC target of 1e6 and a maximum ion injection time of 50 ms. The two parts of PulseDIA raw files were analyzed using DIA-NN (1.7.12) against the spectral library, respectively. The spectral library for ovarian tissue specimens contains 130,735 proteotypic peptides and 10,780 protein groups as previously released¹¹⁰, while spectral library for mouse liver samples for quality control contains 134,856 proteotypic peptides and

8764 protein groups. In the DIA-NN setting, RT profiling was performed, and other parameters were set to default. Quantitative data for precursor and proteins were both controlled below 1%, and those for precursors in two injections were merged by their average values using the R program named Pulsedia_DIANN_OpenSWATH_SpectronautResult_combine (<https://github.com/guomics-lab/PulseDIA>). The combined peptide matrix was converted into a protein matrix using the mean of the top 3 precursor intensities in ProteomeExpert¹¹¹.

Peptides were generated from plasma samples after depleting 14 high-abundance plasma proteins. Plasma was mixed with the High Select™ Top14 Abundant Protein Depletion Resin (Thermo Fisher Scientific, San Jose, USA) and incubated to deplete high-abundance proteins¹¹². Post-incubation, the proteins were digested using TCEP and IAA for reduction and alkylation, followed by a two-step trypsin digestion at a 1:100 enzyme-to-substrate ratio, and the reaction was halted by adding TFA. The peptides were then desalted using SOLAμ™ HRP columns (Thermo Fisher Scientific, San Jose, USA), dried in a vacuum concentrator, and resuspended in MS buffer A for concentration measurement. A 16-plex labeling using TMTpro reagents (Thermo Fisher Scientific, San Jose, USA) was performed for 5 μg of peptides¹¹², and 16 samples from each batch were pooled together for high-pH fractionation using basic pH reversed-phase liquid chromatography¹¹². The 30 concatenated fractions per batch were initially separated over a 60-min gradient from 7% to 30% buffer B (buffer A: 2% ACN, 0.1% formic acid; buffer B: 98% ACN, 0.1% formic acid) and then analyzed using data-dependent acquisition (DDA) mode on a nanoflow DIONEX UltiMate 3000 RSLC nano System (Thermo Fisher ScientificTM, San Jose, USA) connected to an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher ScientificTM, San Jose, USA). The mass spectrometer was operated in positive mode, equipped with a FAIMS Pro interface. Optimal compensation voltages were set at -48V and -68V with a cycle time of 1 s per FAIMS experiment. MS1 scans were performed at a resolution of 60,000 with a normalized AGC target of 300% over a mass range of 375–1800 *m/z*. Dynamic exclusion was customized with an exclusion duration of 40 s. MS2 scans were carried out at a resolution of 30,000 with a normalized AGC target of 200%, using an isolation window of 0.7 *m/z* and setting the first mass at 100 *m/z*. Normalized HCD collision energy was set to 36%, Turbo-TMT was enabled, and MS/MS data were recorded in centroid mode. The mass raw data were analyzed by Proteome Discoverer (Version 2.5.0.400, Thermo Fisher Scientific) using a FASTA file (downloaded on 2018-02-09) containing 20,259 reviewed Homo sapiens protein sequences. The Proteome Discoverer settings were configured with trypsin as the protease, allowing up to two missed cleavages. Static modifications included carbamidomethylation (+57.021464) on cysteine, TMTpro (+304.207145) on lysine residues, and acetylation (+42.010565) on peptides' N-termini. Variable modifications were oxidation (+15.994915) on methionine and acetylation (+42.010565) on peptides' N-termini. Precursor and product ion mass tolerances were set to 10 ppm and 0.02 Da, respectively, with peptide-spectrum match validation at 1% target FDR (strict) and 5% target FDR (relaxed). Normalization was conducted against the total peptide amount, and all other parameters were maintained at default settings. Protein expression levels were calculated as grouped abundance ratios using the pooled sample labeled by the TMT126 channel for batch alignment.

Preprocessing of the protein matrix and quality control

Protein counts for each sample were summarized into four groups, namely normal, benign, borderline, and carcinoma groups. Outliers with fewer proteins in each group were identified using Tukey's fences, where *k* equals 1.5, resulting in the exclusion of 50 samples from 34 patients. Then, the protein matrix of solid specimens was then standardized by quantile normalization, and the missing values were imputed as 0.8 times the minimum value. Unsupervised clustering of

six groups, namely normal, benign, borderline, PDS-EOC, RLP-EOC, and NACT-EOC groups, was performed using global proteome, and 23 samples with incorrect grouping were excluded.

To evaluate the reproducibility during sample preparation and MS acquisition, Pearson correlation coefficients were calculated (a) among mouse liver samples, (b) among pooled samples, (c) between technical replicates, and (d) between biological replicates using \log_2 (intensity). Potential batch effects derived from designed batches, different columns, and injected peptide amounts were assessed through unsupervised clustering of pooled samples and ovarian tissue specimens.

In the protein matrix of plasma samples, proteins with a missing value rate higher than 70% were excluded. Batch correction based on the designed batch was then performed using Combat in BatchServer for the remaining protein matrix of 1660 proteins¹³. After batch correction, reproducibility during sample preparation and MS acquisition was evaluated by calculating the median coefficient of variation (CV) between biological replicates (using the ratio) and among pooled samples (using \log_2 (abundance)), respectively. The batch effect of the designed batch and MS machines was assessed through unsupervised clustering of Principal component analysis (PCA) for the proteomics data.

The selection of upregulated proteins along with the increased malignancy and their validation in plasma

One-Way Analysis of variance (ANOVA) was performed among five ovarian tissue groups: normal, benign, borderline, early stage carcinoma (FIGO stage I and II) of the PDS cohort and late-stage carcinoma (FIGO stage III and IV) of the PDS cohort. Proteins with Benjamini–Hochberg [B–H] adjusted p -value < 0.05 were selected for Mfuzz clustering (Supplementary Data 3). As a result, 8741 proteins were classified into 20 clusters among five groups. Two-sided unpaired Welch's t test was also performed to identify dysregulated proteins (a) between normal and carcinoma groups of PDS cohort using \log_2 (abundance); (b) between non-carcinoma and carcinoma groups of primary HGSOOC plasma samples using the ratio. Considering that ovarian carcinoma can originate not only from ovarian cells but also from FTE cells, we implemented a control measure to reduce potential biases. Five proteins (LYPLA2, MED17, RAB27B, and VMP1), which have been previously reported to exhibit significant upregulation in FTE compared to OSE (with a p -value < 0.05)³⁶, were excluded from our list of identified dysregulated proteins.

The criteria for the potential biomarkers of ovarian cancer were as follows: (a) Seven clusters exhibited upregulation along with increased malignancy with membership values > 0.4 ; (b) B–H adjusted p -value < 0.05 by two-sided unpaired Welch's t test and fold change > 2 between normal and carcinoma groups of ovarian tissue samples (Supplementary Data 3); (c) The human secretome and membrane proteome annotated by The Human Protein Atlas (Supplementary Data 3); (d) B–H adjusted p -value < 0.05 by two-sided unpaired Welch's t test and fold change > 1.2 between non-carcinoma and carcinoma groups of ovarian plasma samples (Supplementary Data 3).

Plasma protein classifiers to distinguish ovarian carcinoma and non-carcinoma patients

Firstly, we identified secreting proteins associated with malignancy and validated these proteins in plasma samples. Subsequently, utilizing the Random Forest package, models were built using either single or combinations of two to eight proteins to distinguish carcinoma patients from non-carcinoma patients.

For the protein matrix of the eight selected potential biomarkers, missing values were imputed as 0. We employed the R package randomForest (version 4.6.14) to build a thousand trees with five-fold cross-validation. Initially, we constructed nine models: one using all eight features and eight additional models, each employing one of these features individually. For the model encompassing all eight

features, we calculated the average value of the mean decrease accuracy for each feature across the five-fold cross-validation as an importance value. Subsequently, we excluded the least important protein in sequence to construct models using seven to two features. The total area under the curve (AUC) was calculated for 168 plasma samples when each was grouped into test set. Statistical differences between receiver operating characteristic (ROC) curves of different models were evaluated using bootstrap test with the pROC package¹¹⁴.

Histotype-specific differentially expressed proteins (DEPs) and pathways

First, two-sided unpaired Welch's t test was performed to identify dysregulated proteins (B–H adjusted p -value < 0.05 and fold change > 2) between each histological subtype of primary carcinoma by PDS and normal ovarian tissues. Considering that ovarian carcinoma can originate not only from ovarian cells but also from FTE cells, we implemented a control measure to reduce potential biases. Fourteen proteins (CDKN2AIPNL, DDB2, H1-0, H1-10, H1-1, HMGB2, LYPLA2, MED17, PHGDH, PRKAG1, PTMS, RAB27B, TNRC6B, and VMP1), which have been previously reported to exhibit significant dysregulation between FTE and OSE (with a p -value < 0.05)³⁶, were excluded from our list of identified dysregulated proteins. Second, One-Way ANOVA was carried out among the five histological subtypes of the PDS cohort and found that 4534 proteins were differentially expressed among five group (B–H adjusted p -value < 0.05). Additionally, 2709 proteins were identified as dysregulated both by two-sided unpaired Welch's t test and One-Way ANOVA.

The criteria for the histotype-specific DEPs were as follows: (a) DEPs were defined as those with a B–H adjusted p -value < 0.05 by two-sided unpaired Welch's t test and fold change > 2 between normal and each histotype group; (b) B–H adjusted p -value < 0.05 by One-Way ANOVA; (c) DEPs present in only one histotype.

Unsupervised clustering was then performed for these histotype-specific DEPs using Ward's minimum variance method. In each cluster, the major histotype to which the DEPs belong was chosen, and pathway enrichment was performed for DEPs of these major histotype using Metascape.

Univariable and Multivariable Cox regression analysis

Proteins with a missing value ratio of less than 70% were included for univariable Cox regression analysis. Residuals of the linear regression models were calculated to remove the potential effect of age at diagnosis on protein expressions. These residuals were then standardized using rank-based inverse normal transformation. After standardization, univariable Cox regression analysis was performed to identify prognostic proteins with a p -value < 0.05 based on the likelihood p -value. Kaplan–Meier plots were drawn for representative proteins to show their significant relationship between protein expression and the optimal cut point for each protein, determined by surv_cutpoint.

Univariate Cox regression analysis was also performed for clinical factors, and missing values in CA125 and HE4 levels were imputed as median values. To determine the prognostic proteins' independence of clinical factors, multivariable Cox regression analysis was performed for each protein to adjust the effects of four prognostic clinical factors.

To validate the prognostic proteins identified in our study, we performed a comparative analysis with the potential prognostic proteins pinpointed in Chowdury et al.'s paper (linear regression model, p -value < 0.05)⁴².

Targeted genomic sequencing

For the targeted genomic sequencing, a 295-gene panel was employed for four balanced patient groups: primary sensitive ($N=27$), primary resistant ($N=26$), relapsed sensitive ($N=26$), and relapsed resistant ($N=17$). No other specific inclusion criteria were applied for these

samples, and we did not include patients who underwent NACT. Somatic DNA was extracted from fresh frozen tumor tissues using the NucleoSpin TriPrep Kit (Macherey-Nagel, Germany), and patient-matched genomic DNA was extracted from peripheral blood lymphocytes using NucleoSpin Blood Kit (Macherey-Nagel, Germany) according to manufacturer's instructions. The quality of isolated genomic DNA was verified through agarose gel electrophoresis and concentration measurement using Qubit® DNA Assay Kit in Qubit® 3.0 Flurometer (Invitrogen, USA).

Extracted DNA was fragmented into 180–280 bp by hydrodynamic shearing system (Covaris, Massachusetts, USA). DNA fragments underwent end repair, 3' ends adenylation and ligation-mediated PCR (LM-PCR). The fragments were then hybridized to probes designed for each targeted gene, and non-hybridized ones were washed out. Real-time PCR was performed to estimate the product magnitude from LM-PCR. After library quality assessment, the clustering of the index-coded samples was generated using Illumina PE Cluster Kit (Illumina, USA) on a cBot Cluster Generation System, and then high-throughput sequencing was conducted on an Illumina platform to generate 150 bp paired-end reads.

Sequence artifacts, including those paired reads in either read containing adapter contamination (>10 nucleotides aligned to the adapter, allowing $\leq 10\%$ mismatches), uncertain bases (more than 10%) or low-quality bases (Phred quality <5, proportion > 50%), were discarded. More detailed quality control statistics are summarized in Supplementary Data 7.

Valid sequencing data were mapped to the reference genome (GRCh37/hg19) using Burrows-Wheeler Aligner (BWA) software (<http://github.com/Ih3/bwa>)¹¹⁵. BAM files were sorted, and duplicate-marking was done using SAMtools¹¹⁶ and Sambamba¹¹⁷. Somatic single-nucleotide variants (SNVs) and insertions/deletions (indels) were retrieved with MuTect (v 3.1.0-g72492bb) (<http://github.com/broadinstitute/mutect>) and Strelka (v 1.0.14) (<http://github.com/Illumina/strelka>), respectively. Germline SNVs and indels were called using Genome Analysis Toolkit (GATK, v 3.1.0-g72492bb). Mutations in coding regions were manually checked using Integrative Genomics Viewer (IGV, version 2.3.34), and filtered variants were annotated using Oncotator (version 1.5.1.0) (<http://github.com/broadinstitute/oncotator>) and Variant Effect Predictor (VEP, v 83) (<http://github.com/Ensembl/ensembl-vep>). Copy number variations were analyzed using Cnvkit v0.9.9¹¹⁸.

Bioinformatic analysis for genomic and proteomic data

Firstly, Fisher's exact test was performed to evaluate the associations between each gene mutation (combining germline and somatic mutations) and chemoresistance in the 295-gene panel. This panel included 14 genes with a direct or indirect role in homologous recombination repair (HRR)¹⁰⁷. The associations between HRR mutations and chemoresistance were also evaluated.

Next, two-sided unpaired Welch's *t* test was performed to identify dysregulated proteins between (i) chemosensitive HGSOC patients with HRR mutations versus chemoresistant ones without any HRR mutations; (ii) chemosensitive versus chemoresistant HGSOC patients in the relapsed cohort. Dysregulated proteins were defined as those with a *p*-value less than 0.05 and a fold change greater than 1.5. Lastly, pathway enrichment for these dysregulated proteins was performed using Metascape and String.

Targeted proteome by MRM

Quantification of prognostic proteins was performed using multiple reaction monitoring (MRM) in tissue and plasma samples. For tissue samples, 71 out of 281 prognostic proteins were quantified by MRM, while for plasma samples, 51 out of 241 prognostic proteins were quantified by MRM (Supplementary Data 6). Common internal retention time (CiRT) standard peptides were used for retention time

prediction, with 13 and 12 peptides selected from OVLib¹¹⁰ and a published blood spectral library¹¹⁹, respectively (Supplementary Data 6). Peptides were separated at a flow rate of 0.2 mL/min over a 15-min LC gradient from 10% to 40% buffer B (buffer A: 0.1% formic acid aqueous solution; buffer B: 0.1% formic acid in acetonitrile solution) in JasperTM HPLC system (SCIEX, CA, USA). The ionized peptides were transferred into TRIPLE QUADTM 4500MD (SCIEX, CA, USA) for analysis.

A total of 388 transitions of 100 peptides from tissue samples and 389 transitions of 101 peptides from plasma samples were selected and analyzed within a ± 1 min time window using time-scheduled acquisition. The target scan time per cycle was set as 2.5 s for tissue samples and 1.7 s for plasma samples.

Machine learning

To predict one-year relapse after the last chemotherapy, we first identified prognostic proteins in the global proteomic data of the discovery cohort, and verified these prognostic proteins using targeted proteomics and optimized models by machine learning. Finally, we evaluated the predictive utility of the final model using an independent validation cohort. The discovery cohort consisted of primary HGSOC patients with at least six cycles of platinum-based chemotherapy from PDS cohort. We excluded patients with an inconclusive outcome of recurrence within one year, resulting in 400 tissue samples from 347 patients and 141 plasma samples from 131 patients (Supplementary Data 1).

Prognostic proteins were identified by univariate Cox analysis and two-sided unpaired Welch's *t* test. For ovarian tissues, 281 prognostic proteins met both criteria (*p*-value < 0.05 by univariate Cox analysis and *p*-value < 0.05 by two-sided unpaired Welch's *t* test between patients relapsing within one year and those after one year). For plasma samples, 241 prognostic proteins met either of the two criteria mentioned above. Then, 71 out of 281 prognostic proteins from ovarian tissues and 51 out of 241 prognostic proteins from plasma were quantified by MRM. Forty tissue proteins and 34 plasma proteins were verified using MRM assay (Supplementary Data 6). Two immunoglobins among verified tissue proteins were excluded. Thus, 38 tissue proteins and 34 plasma proteins were left to build the predictive model by eXtreme Gradient Boosting (XGBoost) algorithm.

Seven clinical factors, including age at diagnosis, residual tumor size, FIGO stage, metastasis of lymph, CA125 and HE4 levels before the treatment, and CA125 at the last cycle of chemotherapy, and verified prognostic proteins quantified by MRM were used to select features to optimize three predictive models (A, B, and C). Model A was based on clinical factors only, while Model B and C were based on clinical and protein features from tissue and plasma samples, respectively.

We randomly split the discovery cohort into a training set and an internal test set at a ratio of 3:1. Then, one hundred iterations of 60% under-sampling of the training set were performed to build models using XGBoost. Two parameters, namely subsample (from 0.5 to 1 at a step of 0.05) and leaning rate (from 0.1 to 0.3 at a step of 0.04), were optimized. The features were ranked by frequency in each model and top 5 to 15 features were selected to build models for the entire training set using XGBoost. The other four parameters, namely gamma (from 0 to 0.2 with a step at 0.05), max_depth (from 3 to 10 with a step at 1), colsample_bytree (from 0.1 to 1 with a step at 0.1), min_child_weight (from 1 to 5 with a step at 1), were optimized. The model with maximal accuracy of discovery cohort was selected finally. The independent validation set was used to evaluate the predictive utility of the final model.

We utilized the CPTAC cohort as an external validation set to verify the generalizability of our model beyond the Chinese population. This cohort had 32 samples with measurements obtained by both Johns Hopkins University (JHU) and Pacific Northwest National Laboratory (PNNL). To avoid redundancy, we removed the 32 duplicate samples assayed by PNNL, resulting in a final set of 126 unique

samples. The protein matrix of these samples underwent Z-score normalization for standardization. As the CPTAC dataset lacked two clinical factors and the expression data for the AGRE5 protein, we adapted tissue model B by retaining the original parameters of the remaining eleven protein features. This revised model was then applied to predict one-year recurrence in the CPTAC cohort.

Statistics and reproducibility

All patient diagnoses were established between 2006 and 2018, adhering to the WHO Classification of Tumors: Female Genital Tumors (5th edition). Tissue specimens were collected based solely on their availability, without any clinical factor-based selection criteria. While a formal sample-size calculation was not performed, we ensured that each analyzed group contained at least 10 samples, a number deemed sufficient for statistical purposes.

To verify the reproducibility of our proteomic data, biological replicates were utilized during experimentation. Any additional replication data not reported in the manuscript, whether successful or unsuccessful, are not available.

Protein counts from individual samples were categorized into four groups: normal, benign, borderline, and carcinoma. We identified and excluded outliers with abnormally low protein counts in each category using Tukey's fences ($k=1.5$), resulting in the removal of 50 samples from 34 patients. Unsupervised clustering was applied to six designated groups—normal, benign, borderline, primary debulking surgery epithelial ovarian cancer (PDS-EOC), recurrent low platinum-sensitive epithelial ovarian cancer (RLP-EOC), and neoadjuvant chemotherapy epithelial ovarian cancer (NACT-EOC)—based on global proteome profiles; 23 samples that clustered incorrectly were excluded from the analysis.

Statistical analyses were conducted using R software (versions 4.0.5 and 4.3.1). Within the stats package (v4.3.1), we performed Analysis of Variance (ANOVA), Welch's *t* test, and Principal Component Analysis (PCA). The Benjamini-Hochberg procedure was utilized to adjust *p*-values for multiple comparisons using the *p.adjust* function. We calculated correlation coefficients with the *corrplot* package (v0.92) and conducted soft clustering using the *Mfuzz* package (v2.60.0). The *randomForest* package (v4.6.14) was employed to develop plasma protein classifiers to differentiate between carcinoma and non-carcinoma cases. Cox proportional hazards regression analysis was carried out using the *survival* package (v3.5-7). Heatmaps were generated with the *pheatmap* package (v1.0.12), employing *ward.D2* linkage for protein clustering. Prognostic predictive models were built using the *xgboost* function in the *xgboost* package (v1.6.0.1), with SHAP values derived from the *SHAPforxgboost* package (v0.1.3).

Language polishing

During the preparation of this work the authors used ChatGPT in order to improve language and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw proteomics data generated during this study have been archived in the iProX database and are accessible via the following identifier: [PXD053625](#), also available as [IPX0004169000](#). The raw targeted genomic data are available at the Genome Sequence Archive (GSA) under accession ID [HRA007126](#); also available under accession number [PRJCA024971](#). The remaining data are available within the Article, Supplementary Information or Source Data file.

Code availability

The specific R codes used in this study is publicly available on Zenodo at <https://doi.org/10.5281/zenodo.1128421>¹²⁰.

References

1. Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
2. Qian, L., Sun, R., Xue, Z. & Guo, T. Mass spectrometry-based proteomics of epithelial ovarian cancers: a clinical perspective. *Mol. Cell Proteom.* **22**, 100578 (2023).
3. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer Statistics, 2021. *CA Cancer J. Clin.* **71**, 7–33 (2021).
4. Menon, U. et al. 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 (2021).
5. Menon, U. et al. Tumour stage, treatment, and survival of women with high-grade serous tubo-ovarian cancer in UKCTOCS: an exploratory analysis of a randomised controlled trial. *Lancet Oncol.* **24**, 1018–1028 (2023).
6. Matulonis, U. A. et al. Ovarian cancer. *Nat. Rev. Dis. Prim.* **2**, 16061 (2016).
7. Cheng, Z. et al. The genomic landscape of early-stage ovarian high-grade serous carcinoma. *Clin. Cancer Res.* **28**, 2911–2922 (2022).
8. Cancer Genome Atlas Research N. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
9. Tothill, R. W. et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin. Cancer Res.* **14**, 5198–5208 (2008).
10. Konecny, G. E. et al. Prognostic and therapeutic relevance of molecular subtypes in high-grade serous ovarian cancer. *J. Natl Cancer Inst.* **106** (2014).
11. Ledermann, J. A. First-line treatment of ovarian cancer: questions and controversies to address. *Ther. Adv. Med Oncol.* **10**, 1758835918768232 (2018).
12. Armstrong, D. K. et al. NCCN Guidelines® Insights: Ovarian Cancer, Version 3.2022. *J. Natl Compr. Canc Netw.* **20**, 972–980 (2022).
13. Lheureux, S., Gourley, C., Vergote, I. & Oza, A. M. Epithelial ovarian cancer. *Lancet* **393**, 1240–1253 (2019).
14. Holmes, D. Ovarian cancer: beyond resistance. *Nature* **527**, S217 (2015).
15. Wang, Y. K. et al. Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes. *Nat. Genet.* **49**, 856–865 (2017).
16. Macintyre, G. et al. Copy number signatures and mutational processes in ovarian carcinoma. *Nat. Genet.* **50**, 1262–1270 (2018).
17. Funnell, T. et al. Single-cell genomic variation induced by mutational processes in cancer. *Nature* **612**, 106–115 (2022).
18. Patch, A. M. et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489–494 (2015).
19. Burdett, N. L. et al. Multiomic analysis of homologous recombination-deficient end-stage high-grade serous ovarian cancer. *Nat. Genet.* **55**, 437–450 (2023).
20. Etemadmoghadam, D. et al. Integrated genome-wide DNA copy number and expression analysis identifies distinct mechanisms of primary chemoresistance in ovarian carcinomas. *Clin. Cancer Res.* **15**, 1417–1427 (2009).
21. Izar, B. et al. A single-cell landscape of high-grade serous ovarian cancer. *Nat. Med.* **26**, 1271–1279 (2020).
22. Vazquez-Garcia, I. et al. Ovarian cancer mutational processes drive site-specific immune evasion. *Nature* **612**, 778–786 (2022).

23. Moore, K. et al. Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. *N. Engl. J. Med.* **379**, 2495–2505 (2018).

24. Banerjee, S. et al. Maintenance olaparib for patients with newly diagnosed advanced ovarian cancer and a BRCA mutation (SOLO1/GOG 3004): 5-year follow-up of a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* **22**, 1721–1731 (2021).

25. Zhang, Z. et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res.* **64**, 5882–5890 (2004).

26. Wang, Q. et al. Selected reaction monitoring approach for validating peptide biomarkers. *Proc. Natl Acad. Sci. USA* **114**, 13519–13524 (2017).

27. Sinha, A. et al. N-glycoproteomics of patient-derived xenografts: a strategy to discover tumor-associated proteins in high-grade serous ovarian cancer. *Cell Syst.* **8**, 345–351.e344 (2019).

28. Huttenthaler, R. et al. A targeted mass spectrometry strategy for developing proteomic biomarkers: a case study of epithelial ovarian cancer. *Mol. Cell Proteom.* **18**, 1836–1850 (2019).

29. Wang, Q. et al. Evaluation of proteomics-identified CCL18 and CXCL1 as circulating tumor markers for differential diagnosis between ovarian carcinomas and benign pelvic masses. *Int J. Biol. Markers* **26**, 262–273 (2011).

30. Tang, H. Y. et al. A xenograft mouse model coupled with in-depth plasma proteome analysis facilitates identification of novel serum biomarkers for human ovarian cancer. *J. Proteome Res.* **11**, 678–691 (2012).

31. Amon, L. M. et al. Integrative proteomic analysis of serum and peritoneal fluids helps identify proteins that are up-regulated in serum of women with ovarian cancer. *PLoS ONE* **5**, e11137 (2010).

32. Hughes, C. S. et al. Quantitative profiling of single formalin fixed tumour sections: proteomics for translational research. *Sci. Rep.* **6**, 34949 (2016).

33. Dieters-Castator, D. Z. et al. Proteomics-derived biomarker panel improves diagnostic precision to classify endometrioid and high-grade serous ovarian carcinoma. *Clin. Cancer Res.* **25**, 4309–4319 (2019).

34. Gong, T. T. et al. Proteomic characterization of epithelial ovarian cancer delineates molecular signatures and therapeutic targets in distinct histological subtypes. *Nat. Commun.* **14**, 7802 (2023).

35. Coscia, F. et al. Integrative proteomic profiling of ovarian cancer cell lines reveals precursor cell associated proteins and functional status. *Nat. Commun.* **7**, 12645 (2016).

36. Francavilla, C. et al. Phosphoproteomics of primary cells reveals druggable kinase signatures in ovarian cancer. *Cell Rep.* **18**, 3242–3256 (2017).

37. Shrestha, R. et al. Multiomics characterization of low-grade serous ovarian carcinoma identifies potential biomarkers of MEK inhibitor sensitivity and therapeutic vulnerability. *Cancer Res.* **81**, 1681–1694 (2021).

38. Tong, M. et al. Molecular subtyping of cancer and nomination of kinase candidates for inhibition with phosphoproteomics: Reanalysis of CPTAC ovarian cancer. *EBioMedicine* **40**, 305–317 (2019).

39. Zhang, H. et al. Integrated proteogenomic characterization of human high-grade serous ovarian cancer. *Cell* **166**, 755–765 (2016).

40. Ji, J. X. et al. The proteome of clear cell ovarian carcinoma. *J. Pathol.* **258**, 325–338 (2022).

41. Yang, J. Y. et al. Predicting time to ovarian carcinoma recurrence using protein markers. *J. Clin. Invest.* **123**, 3740–3750 (2013).

42. Chowdhury, S. et al. Proteogenomic analysis of chemo-refractory high-grade serous ovarian cancer. *Cell* **186**, 3476–3498.e3435 (2023).

43. Coscia, F. et al. Multi-level proteomics identifies CT45 as a chemosensitivity mediator and immunotherapy target in ovarian cancer. *Cell* **175**, 159–170.e116 (2018).

44. Eckert, M. A. et al. Proteomics reveals NNMT as a master metabolic regulator of cancer-associated fibroblasts. *Nature* **569**, 723–728 (2019).

45. Qu, H. et al. Identification and validation of differentially expressed proteins in epithelial ovarian cancers using quantitative proteomics. *Oncotarget* **7**, 83187–83199 (2016).

46. Bast, R. C. Jr et al. New tumor markers: CA125 and beyond. *Int J. Gynecol. Cancer* **15**, 274–281 (2005).

47. Bottoni, P. & Scatena, R. The role of CA 125 as tumor marker: biochemical and clinical aspects. *Adv. Exp. Med. Biol.* **867**, 229–244 (2015).

48. Coleman, R. L. et al. Validation of a second-generation multivariate index assay for malignancy risk of adnexal masses. *Am. J. Obstet. Gynecol.* **215**, 82.e81–82.e11 (2016).

49. Cai, X. et al. PulseDIA: data-independent acquisition mass spectrometry using multi-injection pulsed gas-phase fractionation. *J. Proteome Res.* **20**, 279–288 (2021).

50. Kumar, L. M. EF. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* **2**, 5–7 (2007).

51. Kulbe, H. et al. The inflammatory cytokine tumor necrosis factor-alpha generates an autocrine tumor-promoting network in epithelial ovarian cancer cells. *Cancer Res.* **67**, 585–592 (2007).

52. Kulbe, H. et al. A dynamic inflammatory cytokine network in the human ovarian cancer microenvironment. *Cancer Res.* **72**, 66–75 (2012).

53. Hagemann, T. et al. Macrophages induce invasiveness of epithelial cancer cells via NF-kappa B and JNK. *J. Immunol.* **175**, 1197–1205 (2005).

54. Uhlen, M. et al. The human secretome. *Sci. Signal.* **12** (2019).

55. Wang, L. et al. Expression of MUC1 in primary and metastatic human epithelial ovarian cancer and its therapeutic significance. *Gynecol. Oncol.* **105**, 695–702 (2007).

56. Martinez-Garcia, E. et al. Development of a sequential workflow based on LC-PRM for the verification of endometrial cancer protein biomarkers in uterine aspirate samples. *Oncotarget* **7**, 53102–53115 (2016).

57. Moshkovskii, S. A. et al. Acute phase serum amyloid A in ovarian cancer as an important component of proteome diagnostic profiling. *Proteom. Clin. Appl.* **1**, 107–117 (2007).

58. Hogdall, E. V. et al. YKL-40 tissue expression and plasma levels in patients with ovarian cancer. *BMC Cancer* **9**, 8 (2009).

59. Elschenbroich, S. et al. In-depth proteomics of ovarian cancer ascites: combining shotgun proteomics and selected reaction monitoring mass spectrometry. *J. Proteome Res.* **10**, 2286–2299 (2011).

60. Zhang, W. et al. Overexpression of heparanase in ovarian cancer and its clinical significance. *Oncol. Rep.* **30**, 2279–2287 (2013).

61. Heinz, L. X. et al. The Lipid-Modifying Enzyme SMPDL3B Negatively Regulates Innate Immunity. *Cell Rep.* **11**, 1919–1928 (2015).

62. Waldbillig, F. et al. Phosphodiesterase SMPDL3B gene expression as independent outcome prediction marker in localized prostate cancer. *Int. J. Mol. Sci.* **21** (2020).

63. Kobel, M. et al. A limited panel of immunomarkers can reliably distinguish between clear cell and high-grade serous carcinoma of the ovary. *Am. J. Surg. Pathol.* **33**, 14–21 (2009).

64. Duska, L. R. et al. When 'never-events' occur despite adherence to clinical guidelines: the case of venous thromboembolism in clear cell cancer of the ovary compared with other epithelial histologic subtypes. *Gynecol. Oncol.* **116**, 374–377 (2010).

65. Zorn, K. K. et al. Gene expression profiles of serous, endometrioid, and clear cell subtypes of ovarian and endometrial cancer. *Clin. Cancer Res.* **11**, 6422–6430 (2005).

66. Chan, J. K. et al. A phase II evaluation of sunitinib in the treatment of persistent or recurrent clear cell ovarian carcinoma: An NRG Oncology/Gynecologic Oncology Group Study (GOG-254). *Gynecol. Oncol.* **150**, 247–252 (2018).

67. Konstantinopoulos, P. A. et al. Phase II study of single-agent cabozantinib in patients with recurrent clear cell ovarian, primary peritoneal or fallopian tube cancer (NRG-GY001). *Gynecol. Oncol.* **150**, 9–13 (2018).

68. Lheureux, S. et al. A clinical and molecular phase II trial of oral ENMD-2076 in ovarian clear cell carcinoma (OCCC): a study of the princess margaret phase II consortium. *Clin. Cancer Res.* **24**, 6168–6174 (2018).

69. Moya, I. M. et al. Stalk cell phenotype depends on integration of Notch and Smad1/5 signaling cascades. *Dev. Cell* **22**, 501–514 (2012).

70. Hernandez-Garcia, R., Iruela-Arispe, M. L., Reyes-Cruz, G. & Vazquez-Prado, J. Endothelial RhoGEFs: A systematic analysis of their expression profiles in VEGF-stimulated and tumor endothelial cells. *Vasc. Pharm.* **74**, 60–72 (2015).

71. Zhu, Y. et al. Classifications of ovarian cancer tissues by proteomic patterns. *Proteomics* **6**, 5846–5856 (2006).

72. Russell, S. et al. Combined therapy with thrombospondin-1 type I repeats (3TSR) and chemotherapy induces regression and significantly improves survival in a preclinical model of advanced stage epithelial ovarian cancer. *FASEB J.* **29**, 576–588 (2015).

73. Owusu Obeng, E. et al. Phosphoinositide-dependent signaling in cancer: a focus on phospholipase C isozymes. *Int J. Mol. Sci.* **21**, 2581 (2020).

74. Poli, A., Mongiorgi, S., Cocco, L. & Follo, M. Y. Protein kinase C involvement in cell cycle modulation. *Biochem. Soc. Trans.* **42**, 1471–1476 (2014).

75. Taylor, S. E., Chu, T., Elvin, J. A., Edwards, R. P. & Zorn, K. K. Phase II study of everolimus and bevacizumab in recurrent ovarian, peritoneal, and fallopian tube cancer. *Gynecol. Oncol.* **156**, 32–37 (2020).

76. Lee, E. K. et al. Results of an abbreviated phase II study of AKT inhibitor MK-2206 in the treatment of recurrent platinum-resistant high grade serous ovarian, fallopian tube, or primary peritoneal carcinoma (NCT 01283035). *Gynecol. Oncol. Rep.* **32**, 100546 (2020).

77. Piha-Paul, S. A. et al. Efficacy and safety of buparlisib, a PI3K inhibitor, in patients with malignancies harboring a PI3K pathway activation: a phase 2, open-label, single-arm study. *Oncotarget* **10**, 6526–6535 (2019).

78. de Serres, F. & Blanco, I. Role of alpha-1 antitrypsin in human health and disease. *J. Intern. Med.* **276**, 311–335 (2014).

79. Thompson, S., Guthrie, D. & Turner, G. A. Fucosylated forms of alpha-1-antitrypsin that predict unresponsiveness to chemotherapy in ovarian cancer. *Br. J. Cancer* **58**, 589–593 (1988).

80. Fejzo, M. S. et al. Analysis in epithelial ovarian cancer identifies KANSL1 as a biomarker and target gene for immune response and HDAC inhibition. *Gynecol. Oncol.* **160**, 539–546 (2021).

81. Rouette, A. et al. Expression of immunoproteasome genes is regulated by cell-intrinsic and -extrinsic factors in human cancers. *Sci. Rep.* **6**, 34019 (2016).

82. Wang, X. et al. LRG1 promotes angiogenesis by modulating endothelial TGF-beta signalling. *Nature* **499**, 306–311 (2013).

83. Zhang, J., Zhu, L., Fang, J., Ge, Z. & Li, X. LRG1 modulates epithelial-mesenchymal transition and angiogenesis in colorectal cancer via HIF-1alpha activation. *J. Exp. Clin. Cancer Res.* **35**, 29 (2016).

84. Miyamoto, S. et al. Multiple reaction monitoring for the quantitation of serum protein glycosylation profiles: application to ovarian cancer. *J. Proteome Res.* **17**, 222–233 (2018).

85. Moshkovskii, S. A. et al. Ovarian cancer marker of 11.7 kDa detected by proteomics is a serum amyloid A1. *Proteomics* **5**, 3790–3797 (2005).

86. Li, Z. et al. Serum amyloid a, a potential biomarker both in serum and tissue, correlates with ovarian cancer progression. *J. Ovarian Res.* **13**, 67 (2020).

87. Moore, L. E. et al. Evaluation of apolipoprotein A1 and post-translationally modified forms of transthyretin as biomarkers for ovarian cancer detection in an independent study population. *Cancer Epidemiol. Biomark. Prev.* **15**, 1641–1646 (2006).

88. Park, G. B. & Kim, D. MicroRNA-503-5p inhibits the CD97-mediated JAK2/STAT3 pathway in metastatic or paclitaxel-resistant ovarian cancer cells. *Neoplasia* **21**, 206–215 (2019).

89. Jinawath, N. et al. Oncoproteomic analysis reveals co-upregulation of RELA and STAT5 in carboplatin resistant ovarian carcinoma. *PLoS ONE* **5**, e11198 (2010).

90. Akahira, J. et al. Decreased expression of 14-3-3 sigma is associated with advanced disease in human epithelial ovarian cancer: its correlation with aberrant DNA methylation. *Clin. Cancer Res.* **10**, 2687–2693 (2004).

91. Yue, Y. et al. SURF4 maintains stem-like properties via BIRC3 in ovarian cancer cells. *J. Gynecol. Oncol.* **31**, e46 (2020).

92. Wang, J. et al. Long non-coding RNA RHPN1-AS1 promotes tumorigenesis and metastasis of ovarian cancer by acting as a ceRNA against miR-596 and upregulating LEMT1. *Aging (Albany NY)* **12**, 4558–4572 (2020).

93. Murata, S., Yashiroda, H. & Tanaka, K. Molecular mechanisms of proteasome assembly. *Nat. Rev. Mol. Cell Biol.* **10**, 104–115 (2009).

94. Manasanch, E. E. & Orlowski, R. Z. Proteasome inhibitors in cancer therapy. *Nat. Rev. Clin. Oncol.* **14**, 417–433 (2017).

95. Rajkumar, S. V., Richardson, P. G., Hideshima, T. & Anderson, K. C. Proteasome inhibition as a novel therapeutic target in human cancer. *J. Clin. Oncol.* **23**, 630–639 (2005).

96. Frankel, A., Man, S., Elliott, P., Adams, J. & Kerbel, R. S. Lack of multicellular drug resistance observed in human ovarian and prostate carcinoma treated with the proteasome inhibitor PS-341. *Clin. Cancer Res.* **6**, 3719–3728 (2000).

97. Guo, N., Peng, Z. & Zhang, J. Proteasome inhibitor MG132 enhances sensitivity to cisplatin on ovarian carcinoma cells in vitro and in vivo. *Int J. Gynecol. Cancer* **26**, 839–844 (2016).

98. Bareke, H. & Akbuga, J. Complement system's role in cancer and its therapeutic potential in ovarian cancer. *Scand. J. Immunol.* **88**, e12672 (2018).

99. Zhang, W., Peng, P., Ou, X., Shen, K. & Wu, X. Ovarian cancer circulating extracellular vesicles promote coagulation and have a potential in diagnosis: an iTRAQ based proteomic analysis. *BMC Cancer* **19**, 1095 (2019).

100. Polterauer, S. et al. Plasma fibrinogen levels and prognosis in patients with ovarian cancer: a multicenter study. *Oncologist* **14**, 979–985 (2009).

101. Wang, L.-B. et al. Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer Cell* **39** (2021).

102. Tas, F. et al. Clinical and prognostic significance of coagulation assays in advanced epithelial ovarian cancer. *Int. J. Gynecol. Cancer* **23**, 276–281 (2013).

103. Ahmed, A. A. et al. Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. *J. Pathol.* **221**, 49–56 (2010).

104. Choi, M. C. et al. Clinical impact of somatic variants in homologous recombination repair-related genes in ovarian high-grade serous carcinoma. *Cancer Res. Treat.* **52**, 634–644 (2020).

105. Sugino, K. et al. Germline and somatic mutations of homologous recombination-associated genes in Japanese ovarian cancer patients. *Sci. Rep.* **9**, 17808 (2019).

106. Ledermann, J. A., Drew, Y. & Kristeleit, R. S. Homologous recombination deficiency and ovarian cancer. *Eur. J. Cancer* **60**, 49–58 (2016).

107. de Bono, J. et al. Olaparib for metastatic castration-resistant prostate cancer. *N. Engl. J. Med.* **382**, 2091–2102 (2020).

108. Kurman, R. J. & Shih, Ie. M. The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. *Am. J. Surg. Pathol.* **34**, 433–443 (2010).

109. Cai, X. et al. High-throughput proteomic sample preparation using pressure cycling technology. *Nat. Protoc.* **17**, 2307–2325 (2022).

110. Qian, L. et al. Resistance prediction in high-grade serous ovarian carcinoma with neoadjuvant chemotherapy using data-independent acquisition proteomics and an ovary-specific spectral library. *Mol. Oncol.* **17**, 1567–1580 (2023).

111. Zhu, T. et al. ProteomeExpert: a Docker image-based web server for exploring, modeling, visualizing and mining quantitative proteomic datasets. *Bioinformatics* **37**, 273–275 (2021).

112. Zhou, Y. et al. High-throughput and in-depth proteomic profiling of 5 μL plasma and serum using TMTpro 16-plex. *Methods Mol. Biol.* **2628**, 81–92 (2023).

113. Zhu, T. et al. BatchServer: a web server for batch effect evaluation, visualization, and correction. *J. Proteome Res.* **20**, 1079–1086 (2021).

114. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **12**, 77 (2011).

115. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

116. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

117. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).

118. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* **12**, e1004873 (2016).

119. Zhang, Y. et al. Potential use of serum proteomics for monitoring COVID-19 progression to complement RT-PCR detection. *J. Proteome Res.* **21**, 90–100 (2022).

120. Qian, L. et al. Proteomic landscape of epithelial ovarian cancer. *Zenodo* <https://doi.org/10.5281/zenodo.11284219> (2024).

Acknowledgements

This work was supported by grants from the National Key R&D Program of China (2022YFF0608403 to Y.Zhu. and 2021YFA1301600 to T.G.). We thank Westlake University Supercomputer Center for assistance in data generation and storage, and the Mass Spectrometry & Metabolomics Core Facility at the Center for Biomedical Research Core Facilities of Westlake University for sample analysis. We thank the Research Center for Industries of the Future (RCIF) at Westlake University for partially supporting this work.

Author contributions

T.G., Z.Z., Y.H., Y.Zhu., J.Z., and H.X. designed and supervised the project. J.Z., L.S., and J.W. collected tissue and plasma samples. Z.Z.,

W.Gong, Y.S., and M.N. are responsible for quality assessment of tissue samples. Z.Z. collected patient information. L.Q., N.X., and X.Y. conducted global proteomics analysis. L.Q., X.C., W.L., and Y.Zhan. conducted targeted proteomics analysis. L.Q., Z.X., Y.Zhou., R.S., W.Ge, and T.G. interpreted the proteomic data. L.Q., Z.X., Y.L., and S.S. interpreted the genomic data. L.Q. and T.G. wrote the manuscript with inputs from co-authors. T.G. supported and supervised the study.

Competing interests

T.G. and Y.Zhu. are shareholders of Westlake Omics Inc., W.L., W.G., and Y.Zhan. are employees of Westlake Omics Inc. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-50786-z>.

Correspondence and requests for materials should be addressed to Yi Zhu, Yuejin Hua, Zhiguo Zheng or Tiannan Guo.

Peer review information *Nature Communications* thanks Line Bjørge who co-reviewed with Katrin Kleinmanns and Ignacio Vázquez-García who co-reviewed with Estelle Yao and Thomas Kislinger for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024