Plasma cfDNA methylation markers for the detection and prognosis of ovarian cancer

Leilei Liang,1,† Yu Zhang,1† Chengcheng Li,1† Yuchen Liao,2 Guoqiang Wang,3 Jiayue Xu,2 Yifan Li,a Guangwen Yuan,a Yangchun Sun,4 Rong Zhang,4 Xiaoguang Li,a Weiqi Nian,4 Jing Zhao,1 Yuzi Zhang,1 Xin Zhu,1 Xiaofang Wen,1 Shangli Cai,1 Ning Li,a* and Lingying Wua*

aDepartment of Gynecologic Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China
bDepartment of Gynecology, Xiangya Hospital, Central South University, Changsha, China
cBurning Rock Biotech, Guangdong, China
dChongqing University Cancer Hospital, Chongqing, China

Summary

Background Plasma cell-free DNA (cfDNA) methylation has shown the potential in the detection and prognostic testing in multiple cancers. Herein, we thoroughly investigate the performance of cfDNA methylation in the detection and prognosis of ovarian cancer (OC).

Methods The OC-specific differentially methylated regions (DMRs) were identified by sequencing ovarian tissue samples from OC (n = 61), benign ovarian disease (BOD, n = 49) and healthy controls (HC, n = 37). Based on 1,272 DMRs, a cfDNA OC detection model (OC-D model) was trained and validated in plasma samples from patients of OC (n = 104), BOD (n = 56) and HC (n = 56) and a prognostic testing model (OC-P model) was developed in plasma samples in patients with high-grade serous OC (HG-SOC) in the training cohort and then tested the rationality of this model with International Cancer Genome Consortium (ICGC) tissue methylation data. Mechanisms were investigated in the TCGA-OC cohort.

Findings In the validation cohort, the cfDNA OC-D model consisting of 18 DMRs achieved a sensitivity of 94.7% (95% CI: 85.4%–98.9%) at a specificity of 88.7% (95% CI: 78.7%–94.9%), which outperformed CA 125 (AUC: 0.967 vs 0.905, P = 0.03). Then the cfDNA OC-P model consisting of 15 DMRs was constructed and associated with a better prognosis of HG-SOC in multivariable Cox regression analysis (HR: 0.29, 95% CI, 0.11–0.78, P = 0.01) in the training cohort, which was also observed in the ICGC cohort using tissue methylation (HR: 0.56, 95% CI, 0.32–0.98, P = 0.04). Investigation into mechanisms revealed that the low-risk group had higher homologous recombination deficiency and immune cell infiltration (P < 0.05).

Interpretation Our study demonstrated the potential utility of cfDNA methylation in the detection and prognostic testing in OC. Future studies with a larger population are warranted.

Funding This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sector.

Copyright © 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Abbreviations: AUC, area under the curve; BOD, benign ovarian disease; CI, confidence interval; CAP, College of American Pathologists; CLIA, Clinical Laboratory Improvement Amendments; cfDNA, cell-free DNA; CA125, cancer antigen 125; ctDNA, circulating tumour DNA; DMRs, differentiated methylation regions; FFPE, formalin-fixed and paraffin-embedded; GO, gene ontology; HG-SOC, high-grade serous ovarian cancer; ICGC, International Cancer Genome Consortium; HC, healthy controls; HR, homologous recombination repair; HRD, homologous recombination deficiency; LST, large-scale state transition; LOH, loss of heterozygosity; LASSO, least absolute shrinkage and selection operator; NtAI, telomeric allelic imbalance; QC, quality control; TVS, transvaginal sonography; TCGA-OC, the cancer genome atlas ovarian cancer cohort; WBC, white blood cell

*Corresponding authors at: Department of Gynecologic Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, No. 17 Panjiayuan Nanli, Chaoyang District, Beijing, China.

E-mail addresses: lining1502@cscoc.ac.cn (N. Li), wulingying@cscoc.org.cn (L. Wu).

† These authors contribute equally to the manuscript.
Keywords: Ovarian cancer; Methylation; Circulating cell-free DNA; Ovarian cancer detection; Prognosis; Liquid biopsy

Research in context

Evidence before this study

Approximately 70% of ovarian cancer (OC) patients are diagnosed with advanced stages. Currently, no standard screening approach is recommended for ovarian cancer (OC) and there are no effective methods to evaluate the prognosis of OC. Although the cancer antigen 125 (CA125), a tumour protein biomarker, is the most used marker for OC screening, the application is limited due to its poor sensitivity and specificity, and failure to improve survival in a large randomized controlled trial for OC screening. Moreover, CA125 has also been investigated to evaluate chemotherapeutic efficacy and prognosis but has poor accuracy. A few studies have explored the rationality and accuracy of liquid biopsy such as cfDNA methylation in detecting OC. However, there is still room for the improvement of sensitivity in detecting OC. Moreover, limited studies have focused on the clinical utility of cfDNA methylation in the prognostic stratification of OC.

Added value of this study

We built the cfDNA methylation models for the detection and prognostic stratification of OC patients, respectively. An independent validation set was used to ensure the robustness of the detection model. The OC detection model exhibited excellent performance in detecting OC, and outperformed CA125, which would be meaningful in clinical application considering the remarkably poor outcomes of advanced OC. Beyond detection of OC, we also explored the possible clinical utility of cfDNA methylation in the prognostic stratification of OC. For prognostic markers, mechanisms investigation into homologous recombination repair (HRR) pathway and immune characteristics revealed higher immunogenicity and immune cell infiltrations in the low-risk group, respectively.

Implications of all the available evidence

This study demonstrated the potential utility of cfDNA methylation markers in the detection and prognostic testing in OC. Moreover, elucidating the mechanisms underlying the prognosis of OC at molecular level is significant to facilitate the treatment of ovarian cancer and to improve the survival of patients.

Introduction

Ovarian cancer (OC) remains the most lethal gynaecological malignancy in women, and the number of newly diagnosed cases is increasing worldwide.1 High-grade serous ovarian cancer (HG-SOC) is the most common histologic subtype of OC, accounting for 70%–80% of OC.2 OC is usually insidious onset and rapid intraperitoneal spread. Approximately 70% of OC patients are diagnosed at advanced stages whose 5-year survival rate is lower than those diagnosed at earlier stages.3 Given the malignant properties of OC, detection and prognostic estimation of OC are imperative for early intervention, stratification, treatment determination and management of patients. Screening strategies have been widely studied for OC, such as transvaginal sonography (TVS) and cancer antigen 125 (CA125).4 However, these methods had poor sensitivities (69.0%–93.8%) and specificities (58.0%–94.6%)5 and failed to improve survival in the randomized controlled trial for OC screening.4,7 Therefore, no recommended screening approach is available for OC nowadays.8 Moreover, CA125 has also been investigated to evaluate chemotherapeutic efficacy and prognosis but has poor accuracy, which may be due to inflammation-induced CA125 secretion.9

Circulating cell-free DNA (cfDNA) is extracellular nucleic acid fragments shed into plasma via necrosis, apoptosis, or active release of cells, which has shown the potential to revolutionize the screening, diagnosis, prognosis and treatment of multiple cancers.10,11 The genomic and epigenomic alterations in circulating tumour DNA (ctDNA) can be detected by analysing mutations, copy number variations, chromosomal rearrangements, and methylation alterations in plasma.12 Among them, methylation alteration stands out in the detection and surveillance of cancers due to its early occurrence during tumorigenesis and abundant signals for analysis.13,14 cfDNA methylation has been explored in multiple types of cancers for detection including OC.15–19 A previous study reported that cfDNA methylation patterns could discriminate HG-SOC patients from healthy controls and benign pelvic mass with a sensitivity of 41.4% at a specificity of 90.7%.18 Another study using 7 methylation markers separated epithelial OC from benign pelvic masses with a sensitivity of 90.6% at a specificity of 89.7%; however, the external validation to further prove the clinical utility was lacking. Meanwhile, prognostic testing by cfDNA methylation is seldom studied.

In this study, we aimed to evaluate the performance of cfDNA methylation alteration in the detection and prognostic testing of OC. We first identified differentially methylated regions (DMRs) by comparing methylation profiles of ovarian tissues from patients who were diagnosed with malignant and benign ovarian tumours
and healthy individuals. Subsequently, we employed least absolute shrinkage and selection operator (LASSO) regression analysis to develop cfDNA ovarian cancer detection (OC-D) and prognostic (OC-P) testing models in plasma, separately. The cfDNA OC-D model was further validated in an independent validation cohort with plasma samples, and the rationality of the OC-P model was tested in the International Cancer Genome Consortium (ICGC) cohort. Based on those candidate markers, mechanisms were investigated in the TCGA-OC cohort, providing insight into therapeutic implications of OC in future.

Methods

Study design and participants
This study aimed to build cfDNA detection and prognostic models for OC through the following sections: marker selection, detection model development and validation, prognostic model development and validation, and mechanistic investigation (Figure 1).

OC-specific methylation markers selection. A customized methylation panel of 161,984 CpG sites was used in the present study, spanning ~2.7 Mb of the human genome in 6 common cancer types including OC, lung cancer, colorectal cancer, pancreatic cancer, liver cancer, and oesophageal cancer. The panel was originally developed for the detection and tissue-of-origin of multi-type cancers.21,22

Tissue samples from patients with OC and benign ovarian diseases were collected from Cancer Hospital, Chinese Academy of Medical Sciences Hospital. Benign ovarian diseases were included but were not limited to teratoma, ovarian serous cystadenoma, endometrial cyst and fibroma of the ovary. Normal ovarian tissue samples were obtained from individuals who underwent surgery for other non-gynecologic malignancies such as patients with myoma of the uterus, fibroid who were postmenopausal with hysterectomy with elective bilateral salpingo-oophorectomy. The diagnosis of all formalin-fixed and paraffin-embedded (FFPE) tissues was confirmed by an independent pathologist before DNA extraction. Tumour tissues that had tumor cells less than 30% or failed the DNA quality control (QC) criterion were excluded. The detailed methods for sample collection, DNA extraction and methylation sequencing are provided in Supplementary Methods.

OC-specific DMRs were selected by modified Waldtest with an adjusted P-value < 0.05 and absolute mean difference ≥ 0.2.

Detection model training and validation. In the training set, blood samples of patients with pathologically diagnosed benign, and malignant ovarian tumours were obtained from Xiangya Hospital Central South University from December 2017 to December 2019. Blood samples for the above patients were all collected prior to any treatment including local/regional therapy and surgery. Healthy controls were recruited from Chongqing University Cancer Hospital from November 2019 to December 2019 and defined as participants who were free from a history of malignancy and critical illnesses including hepatitis, liver cirrhosis, chronic obstructive pulmonary disease, and colorectal disease. The detailed inclusion and exclusion criteria are described in Supplementary Methods.

In the validation set, blood samples from patients with benign ovarian diseases, borderline and OC and healthy controls were independently collected from the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences from December 2017 to January 2020. The inclusion and exclusion criteria were the same as the above.

Prognosis model training and validation. In the cfDNA prognostic model development, blood samples from patients with OC were obtained from the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and used as the training cohort. The inclusion and exclusion criteria are described in the Supplementary methods. cfDNA from plasma was collected and sequenced by ELSA-seq from the eligible participants. Totally, 51 patients were pathologically diagnosed with HGS-OC, with methylome data and available survival data were analysed (named National Cancer Center cohort [NCC cohort]). All patients received 6–8 cycles of the platinum-based chemotherapy regimen. Progression-free survival (PFS) was defined as the time interval from diagnosis to progression or death due to any cause. Resistance to platinum-based chemotherapy was defined as progression/relapse within 6 months.

A public cohort of 80 patients with HGS-OC who had clinical survival data and tissue methylation profiles generated by Infinium HumanMethylation 450K array was obtained from the ICGC dataset (ICGC cohort)23 and was analysed to validate the performance of the OC-P model. The TCGA-OC cohort (n = 558) with clinical, methylomic data sequenced by 27K array and transcriptomic data was downloaded from https://xenabrowser.net/datapages?cohort and was used to investigate the mechanism of the candidate methylation markers in the prognostic model.

The patient-level data in our study including sequencing data and diagnosis, of which readers may contact the corresponding authors for the access for non-commercial purposes.

Ethics approval and consent to participate
This study was approved by the Ethics Committees of Cancer Hospital, Chinese Academy of Medical Sciences
Figure 1. Flow diagram of study.

~450,000 CpG sites TCGA & GEO database (cancer = 2,018, non-cancer = 195), GEO white blood cell samples (n=656)

~5.5 million CpG sites Inhouse data (cancer = 249, non-cancer = 288)

Panel design (tissue)

Customized targeted methylation panel covering 161,984 CpG sites

OC tissue (n = 61)
Ovarian benign tissue (n = 49)
Normal controls (n = 37)

Marker selection (tissue)

1,272 DMRs were identified

Detection Model development

Training set (plasma cfDNA targeted methylation panel)
- Cancer (n = 47)
  stage I (n = 10)
  stage II (n = 6)
  stage III (n = 19)
  stage IV (n = 12)
- Healthy control (n = 25)
- Benign (n = 25)

LASSO 5-fold cross validation

Validation set
- (plasma cfDNA targeted methylation panel)
  Cancer (n = 57)
  stage I (n = 6)
  stage II (n = 7)
  stage III (n = 32)
  stage IV (n = 8)
  unknown (n = 4)
  Healthy control (n = 31)
  Benign (n = 31)

Exploratory analysis
- Borderline (n = 8)

Prognostic Model development

Training set (NCC cohort)
- (plasma cfDNA targeted methylation panel)
  HGS-OC (n = 51)
  stage II (n = 5)
  stage III (n = 37)
  stage IV (n = 6)
  unknown (n = 3)

Validation set (ICGC dataset)
- (tissue sample 450K array)
  HGS-OC (n = 80)
  stage III (n = 68)
  stage IV (n = 12)

Mechanisms investigation
- (TCGA-OC dataset)
- (tissue 27K array)
  HGS-OC (n = 558)
Enhanced linear-splinter amplification sequencing (ELSA-seq)
The detailed description of ELSA-seq including sequencing approach, data-processing pipeline, and co-methylation region segregation and the calculation formula of methylation region score was described in Supplementary Methods and in our previous work.25

Construction of the OC-D and OC-P models
For the OC-D model, a five-cross-validation model with binomial deviance minimization criteria was implemented in the training set. The lambda with 1 standard error (SE) was used for feature selection. Based on the selected methylation markers, the predicted risk score of each patient was calculated according to the methylation rate, and the estimated regression coefficient of the selected markers. The performance of the model was further evaluated in the independent validation cohort. Based on the OC-D model and CA125 (cutoff = 35 IU/ml), a combined model was constructed by logistic regression.

For the OC-P model, LASSO-penalized Cox regression analysis with binomial deviance minimization criteria was performed to reduce the number of methylation markers using three-fold cross-validation. The minimum lambda was selected to screen the optimal methylation markers. LASSO-penalized Cox regression analysis was used to select the candidate markers. DMRs located within the 1 kb region upstream or downstream of the transcription start sites of functional annotations genes were selected. The OC-P model was constructed based on a linear combination of the regression coefficients (β) derived from the Cox proportional hazards regression model multiplied by its methylation level. Patients were divided into high- and low-risk groups based on the median score of the prognostic signature. The prognostic markers were further validated in the ICGC and TCGA-OC cohorts.

Statistical analysis
Continuous variables were described with median (interquartile range [IQR]) and were compared by Mann-Whitney U test. Categorical variables were described with the number (percentages) and compared by Chi-square test or Fisher’s exact test. Gene Ontology (GO) enrichment analysis of genes with DMRs was performed using Database for Annotation, Visualization and Integrated Discovery (DAVID, https://david.ncifcrf.gov/).24 Single sample gene set enrichment analysis (ssGSEA) was performed to estimate the score of immune signalling pathways defined by Puleo et al previously.25 The infiltrations of 28 immune cells in the TCGA cohort were quantified using the ssGSEA implementation R package “gsva” analysis.25 Receiver operating characteristic (ROC) analysis was performed using the function “roc” in the R package pROC. The area under the curve (AUC) and 95% confidence interval (CI) were generated to evaluate the model performance. The cut-off value for the detection model was determined by Youden’s index. The 95% CIs for sensitivity and specificity were generated using the Clopper-Pearson method.25,26 The comparisons of AUCs in different groups were performed using the DeLong method.26,27 Survival curves were estimated by Kaplan-Meier (KM) curves and compared by log-rank test with hazard ratio determined by Cox regression. Variables with P < 0.1 in the univariable Cox regression were included in the multivariable Cox regression. A two-sided P value of 0.05 was set as the level of significance. The statistical analyses were performed using R 3.4.2 and MedCalc v19.3.1

Role of funders
Funders had no role in study design, data collection, data analyses, interpretation, or writing of report.

Results
Patient characteristics
The demographics and clinical characteristics of all participants are summarized in Supplementary Table S1. In brief, the methylation profiles of tissue samples were obtained from 66 malignant OC, 49 benign ovarian disease (BOD) and 37 healthy controls (HC) to identify OC-specific methylation markers. The age was relatively balanced between OC and healthy/benign controls (median, 51 [IQR, 37–57] years vs. 49 [IQR, 45–60] years). Patients in stages I/II account for 20.3% of all patients. HG-SOC was the main histological type (87.9%).

The plasma samples of 47 OC patients, 25 BOD and 25 HC were used to train the cfDNA OC-D model. The stage was relatively balanced with 21.3% stage I, 12.8% stage II, 40.4% stage III and 25.5% stage IV. The independent validation cohort included 57 OC plasma samples, 31 BOD and 31 HC, and 13 OC patients were in stages I–II (26.6%). Most patients were HG-SOC (78.9%), and the others were endometrioid cancer (5.3%), clear cell cancer (5.3%), mucinous cancer (1.8%) or mixed/unknown cancer (7.1%), which was consistent with the training cohort.

Methylation markers selection
We first selected OC-specific methylation markers by performing the targeted methylation sequencing in
tissues from OC, BOD and HC. By comparing the methylation signatures between OC and BOD/HC, we identified 1,272 OC-specific DMRs (Figure 2a). Of these DMRs, 629 (49.4%) showed higher methylation levels in cancer tissues. GO enrichment analysis showed that the genes with those DMRs were enriched in the pathways related to cell adhesion, cell differentiation, cell development, and regulation of GTPase activity (Figure 2b), and the rest (50.6%) with lower methylation levels in tumour tissues, enriched in the pathways related to cell adhesion, cell differentiation, cell development, and regulation of GTPase activity (Figure 2b).

The methylation levels for the 1,272 OC-specific DMRs are depicted in the heatmap (Figure 2d), exhibiting different methylation patterns between OC and BOD/HC. Moreover, the unsupervised clustering in OC and BOD/HC was performed, and cancer samples showed similar methylation patterns, which were different from BOD/HC, suggesting these markers were cancer specific (Supplementary Figure 1a). The principal component analysis further demonstrated the distinct component between OC and BOD/HC (Supplementary Figure 1b). Altogether, these results indicate a robust discrimination between OC and BOD/HC based on these selected DMRs. Hence, those DMRs were used to develop models for detection and prognostic estimation of OC using cfDNA methylation.

Development of the OC-D model

To develop a cfDNA detection model based on the OC-specific DMRs, plasma samples from OC and BOD/HC were sequenced and analyzed (Figure 1). We identified 18 methylation markers in the cfDNA sequenced by ELSA-seq and constructed an OC-D model by LASSO regression (Figure 1a). The methylation markers identified in the training set are presented in Figure 1c, showing the different methylation patterns between cases and controls. The coefficients, detail annotations and reference genes relating to these markers are displayed in Supplementary Table S2. There was no linear correlation among these markers (Spearman's correlation, P > 0.05) (Supplementary Figure 2b).

Using a best cutoff value (0.4226) as determined via the Youden's index, the OC-D model demonstrated a sensitivity of 95.7% (95% CI: 85.3%–99.5%) and a specificity of 94.5% (95% CI: 83.5%–98.7%), respectively, to discriminate OC from BOD/HC in the training set (Supplementary Table S3). Compared with baseline CA125 of the participants in the training cohort, the OC-D model had a better detection performance (AUC = 0.987, 95% CI: 0.971–1.00 vs. AUC = 0.940, 95% CI: 0.895–0.985; DeLong method, P = 0.028, Figure 3d).

The predicted risk score of OC-D model increased with tumor stage and was significantly higher in cancer than in BOD and HC (Kruskal-Wallis's test, P < 0.001, Figure 3e). The sensitivity of OC-D model was 80.0% (95% CI: 44.4%–97.5%) in stage I, 100.0% (95% CI: 54.1%–100.0%) in stage II, 100.0% (95% CI: 82.4%–100.0%) in stage III, and 100% (95% CI: 73.5%–100.0%) in stage IV (Supplementary Table S3). We further constructed a model combining of the OC-D model and CA125. As a result, the combined model outperformed CA125 alone (AUC: 0.988, 95% CI: 0.971–1.00 vs. AUC: 0.940, 95% CI: 0.895–0.985, DeLong method, P = 0.015, Figure 3f), but not the OC-D model (DeLong method, P = 0.62). The specificity slightly increased to 98.0% (95% CI: 89.4%–100.0%), compared with OC-D model (specificity: 94.0%, 95% CI: 83.5%–98.7%), while the overall sensitivity remained stable (sensitivity: 95.7%, 95% CI: 85.5%–99.55%, Figure 3g, h, Supplementary Table S3).

Independent validation of the OC-D model

To further validate the performance of the OC-D model, we included an external independent validation cohort with plasma samples. The methylation patterns of the selected 18 DMRs were different between cases and controls (Figure 4a). The sensitivity and specificity of the OC-D model were 94.7% (85.4%–100.0%) for discriminating OC from BOD/HC (Supplementary Table S3). The performance of the OC-D model also outperformed CA125 (AUC: 0.967, 95% CI: 0.940–0.994 vs. AUC: 0.905, 95% CI: 0.841–0.969; DeLong method, P = 0.03, Figure 4b). Consistently, the predicted risk score of the OC-D model increased with tumour stage and was significantly higher in cancer than in BOD and HC in the validation cohort (Kruskal-Wallis’s test, P < 0.001, Figure 4c). The sensitivities of the OC-D model were 83.3% (95% CI: 35.9%–99.6%) in stage I, 85.7% (95% CI: 42.1%–99.6%) in stage II, 96.9% (95% CI: 83.8%–99.9%) in stage III, and 100% (95% CI: 39.8%–100.0%) in stage IV at a specificity of 88.7% (95% CI: 78.7%–94.9%) (Figure 4d, Supplementary Table S3).

We then performed subgroup analysis in the total set including training and validation sets to demonstrate the robustness of the OC-D model. The sensitivities showed no significant differences stratified by age (Fisher's exact test, P = 0.50) or histology subtypes (Fisher's exact test, P = 0.50) (Figure 4e, Supplementary Table S4), indicating that the performance of the OC-D model was not influenced by these clinical covariates.
The OC-D model also performed better in distinguishing OC from HC/BOC than CA125 (AUC: 0.983 vs. 0.939, DeLong method, P = 0.001, Figure 4f). In addition, the OC-D model showed good discriminability in the CA125-negative OC (AUC: 0.881, 95% CI, 0.785–0.977). Conversely, CA125 performed inferior in OC-D model-negative OC (AUC: 0.650, 95% CI: 0.369–0.931, DeLong method, P = 0.12, Figure 4g). Altogether, these results suggested that the performance of the OC-D model was robust and better than CA125.

Eight borderline OC were exploratively analyzed, and all the borderline OCs were epithelial tumours (three serous, four mucinous, one endometrioid and one mixed epithelial cell). Only one patient was positive in the total set, which yielded a sensitivity of 12.5% (95% CI: 0.32–52.6%).

**Prognostic model training and validation**

In total, 51 patients with HG-SOC who underwent surgery and had survival data were analysed in the NCC cohort. The cfDNA from plasma was collected and sequenced by ELSA-seq from the eligible participants. The detailed characteristics of participants are depicted in the Supplementary Table S5. Most patients were in
Figure 3. Development of OC-D model-based OC-specific DMRs. (a) Tuning parameter ($\lambda$) selection in the LASSO model using five-fold cross-validation via minimum criteria (the 1-SE criteria). (b) A coefficient profile plot was produced against the log ($\lambda$).
stage III (72.5%). A majority of patients (74.7%) had no residual tumour or minimal residual tumour (<0.5 cm) after surgery. The median follow-up was 16.8 months (range: 1 to 29.2 months).

We firstly investigated whether the OC-D model was associated with survival in the HG-SOC. The OC-D model rather than CA125 was a prognostic indicator for HG-SOC (log-rank test, \(P = 0.013\); HR, 0.35; 95% CI, 0.15–0.84, Supplementary Figure 3a-b), suggesting the possible role of cfDNA methylation markers in prognostic prediction. Since the OC-D model was constructed for the detection of OC, we further constructed an OC-P model based on the OC-specific DMRs to better predict prognosis. The OC-P model was constructed by LASSO Cox regression and included 15 methylation markers (Supplementary Figure 3c-d). No linear correlation was among these markers (Supplementary Figure 3e). The coefficients, gene types and reference genes relating to these markers are displayed in Supplementary Table S6. Patients were divided into high-risk and low-risk groups according to the median value (1.10) of the risk score of OC-P model. The median PFS in the high-risk group was significantly shorter than that in the low-risk group (log-rank test, \(P = 0.01\), Table 1). Moreover, more patients in the high-risk group were platinum-resistant than in the platinum-resistant group compared with the platinum-sensitive group (Mann-Whitney test, \(P < 0.05\), Supplementary Figure 3f).

We further tested the rationality of the OC-P model in the tissue samples in the ICGC cohort since no cfDNA methylation data is available. Eight of 15 genes were matched in the ICGC cohort, and the CpG sites closest to promoter region of these genes were selected and validated using the same coefficient as in training set (Supplementary Table S6). Similar results were observed that patients in high-risk group had shorter OS than those in low-risk group in both the univariable (HR: 0.58, 95% CI: 0.35–0.96, Cox regression, \(P = 0.034\); Figure 5F) and multivariable Cox regression (HR: 0.56, 95% CI: 0.32–0.98, Cox regression, \(P = 0.04\); Supplementary Table S7). Altogether, these results suggest that OC-P can be an independent prognostic factor of OC.

Mechanistic investigation based on the prognostic methylation markers

The methylation markers identified in the prognostic model and their implications in cancer development, immune contexture, and homologous recombination repair (HRR) signalling were investigated in the TCGA-OC cohort with tissue methylome data. Eight of 15 genes were matched in the TCGA cohort, and the CpG sites closest to promoter region of these genes were selected (Supplementary Table S6).

Consistent with the training results, patients in high-risk group had shorter OS (log-rank test, \(P = 0.01\); HR: 0.75, 95% CI: 0.60–0.93, Figure 6a) and PFS (log-rank test, \(P = 0.043\); HR: 0.82, 95% CI: 0.67–0.99, Figure 6b) than those in low-risk group. Moreover, in the TCGA cohort, the low-risk group was associated with a higher homologous recombination deficiency (HRD) score (median: 62 vs. 58; Mann-Whitney, \(P = 0.04\)) and higher large-scale state transition (LST) score (median: 22 vs. 19; Mann-Whitney, \(P = 0.03\)), but no significant differences in the scores of loss of heterozygosity (LOH: median: 16 vs. 14; NTAI: median: 24 vs. 24, Mann-Whitney, \(P > 0.05\)) or telomeric allelic imbalance were observed between high and low-risk groups.
Figure 4. OC-D model in the independent validation cohort. (a) Heatmap illustrating the DMRs between OC (n = 57) and benign ovarian disease (n = 31) and healthy controls (n = 31) including hypomethylated and hypermethylated genes in the validation cohort. Borderline OC (n = 8) was exploratory observation. (b) Comparison of the ROC curves delineating the association between the predictive probability of OC-D model and CA125 and cancer (DeLong’s test, P = 0.03) in the validation cohort (n = 127). (c) Predicted risk scores of healthy control (n = 31), benign diseases (n = 31), and OC (n = 57) with different clinical stages generated from OC-D model, CA125 and the combined model in the validation cohort (n = 127). (d) The sensitivity stratified by stages in the validation cohort (n = 127). (e) The sensitivity and specificity stratified by clinical covariates in the total cohort (n = 224). (f) Comparison of ROC curves delineating the association between the predictive probability of the OC-D model and CA125 and cancer (DeLong’s test, P < 0.001) in the total cohort (n = 224). (g) Comparison of ROC curves delineating the association between the predictive probability of the OC-D model and cancer in CA125-negative patients (<35 IU/ml) (n = 100) and the association between the predictive probability of CA125 and cancer in the OC-D model-negative patients (n = 98) (DeLong’s test, P = 0.12). Abbreviation: OC, ovarian cancer; ROC, Receiver operating characteristic.
Figure 5. The training and validation of the prognostic model. (a) Kaplan-Meier survival curves comparing PFS between the high-risk group (n = 26) and low-risk group (n = 25) in the training set (n = 51). (Log-rank test, P < 0.001). (b) ROCs and corresponding AUCs with 95% CI for 18-month survival predicted by the prognostic model and CA125. (c) Subgroup analysis presenting the PFS for low-risk group (n = 25) and high-risk group (n = 26) stratified by the clinical covariates. (d) Kaplan-Meier survival curves comparing PFS between high-risk group (n = 26) and low-risk group (n = 25) stratified by stages in NCC cohort (n = 51). P values were compared by log-rank test. (e) Histogram depicting the ratio of sensitive or resistant to platinum in the high-risk group (n = 26) and low-risk
indicating high levels of chromosomal instability in the low-risk group (Figure 6c, d). However, the frequencies of mutated genes in the HRR pathway including BRCA1, BRCA2, ATM, PALB2, and RAD50/51 were relatively balanced between the high-risk and low-risk groups (BRCA1: 5.18% vs. 4.25%, BRCA2: 6.37% vs. 3.09%, ATM: 4.38% vs. 4.25%, PALB2: 2.39% vs. 0.77%, RAD50: 1.20% vs. 2.32%, RAD51: 1.2% vs. 4.25%, Chi-square test, P > 0.05; Figure 6c), suggesting that the variability of HRD may not be mediated by gene mutation in HRR pathway. Further investigation into immune characteristics in the low-risk group revealed a lower MSIsensor Score (median: 0.7 vs. 0.91; Mann-Whitney, P = 0.02) and a lower aneuploidy score (median: 12 vs. 14; Mann-Whitney, P = 0.06; Figure 6c, e), which represented higher immunogenicity and decreased immune evasion, respectively. Moreover, increased immune cell infiltrations were also observed in the low-risk group, such as macrophage (Mann-Whitney, P = 0.05; Figure 6c, e), mast cell (Mann-Whitney, P = 0.06; Figure 6c, e), and type 1 T helper cell (Mann-Whitney, P = 0.03, Supplementary Figure S4). Furthermore, individual mRNA expression of key genes in the TCGA cohort that relate to corresponding methylation sites, such as ANKRD33, IRF2, PGCP, and SNHG18, were correlated with the pathway score of diversity of B/T cell receptors (BCR/TCR, the density of tumour-infiltrating immune cells (Supplementary Figure S5a, Supplementary Table S8), and the activation of immune-related pathways Supplementary Figure S5a, Supplementary Table S8) than defined by Puleo et al. previously.24 The above results suggest that the low-risk group was associated with higher HRD scores, a high immunogenicity, and immune cell infiltrations. Core genes involving the mechanisms warrant further study.

**Discussion**

In this study, we aimed to build cfDNA methylation models for the detection and prognostic stratification of OC patients. The cfDNA OC-D model could effectively separate patients with OC from BOC/HC in the independent validation cohort using plasma samples, and outperformed CA125. Additionally, we developed an OC-P model with 15-methylation markers using plasma cfDNA, which could effectively stratify OC patients into high-risk or low-risk groups and performed better than CA125. Altogether, our results provided the potential utility of cfDNA methylation as detection and prognostic markers in OC.

Currently, there is no standard screening paradigm for OC. Although serum CA125 is the most widely used protein biomarker in OC clinically, it remains controversial for its poor performance. In a meta-analysis by Ferraro et al.,19 the specificity of CA125 for detecting ovarian cancer was only 78% (95% CI: 76%–80%), and the unsatisfactory specificity was also observed in our study (80.6%). A few studies have explored the rationality and accuracy of cfDNA methylation in the detection of OC. Compared with previous studies exploring cfDNA methylation in OC,18,20 the development of the OC-D
Figure 6. Mutational landscape in HRR signaling and HRD score in TCGA cohort. (a) Kaplan-Meier survival curves comparing OS between the high-risk group (n = 279) and low-risk group (n = 279) in the TCGC-OC cohort (log-rank test, P = 0.009). (b) Kaplan-
Articles

model was well designed, which has gone through marker selection, model development, and validation to ensure the robustness of the model. Our study included an independent validation set, which may lower the possibility of overestimation of the performance. Compared with CA125, the OC-D model exhibited excellent performance in detecting early-stage OC (sensitivity: stage I, 83.3%; stage II, 85.7%) at the specificity of 88.7% in the external validation cohort, which would be meaningful in clinical application considering the remarkably poor outcomes of advanced OC. Moreover, the OC-D model detected 62.5% (5 out of 8) OC in the CA125-negative OC. However, considering the relatively low cost and comparable early-stage sensitivity of CA125 testing, cfDNA methylation may be a complement but not an alternative to CA125 for the detection of OC.

Notably, our study was limited by a relatively small size of patients with non-HG-SOC, such as clear cell, endometrioid and mucinous ovarian cancer. Although these histological subtypes are biologically distinct cancers,28 the DMRs identified in this study shared overlapping signatures in different subtypes. These shared methylation signatures suggest that those candidate DMRs were representatives of these pathological subtypes. Borderline OC was also included for exploratory analysis; however, its sensitivity was relatively low. In the current study, all eight borderline OC were epithelial tumours with no infiltrative growth pattern (three serous, four mucinous, one endometrioid and one mixed epithelial cell). Owning to the stromatic non-invasiveness or microinvasiveness of borderline OC, few cfDNA is released into the blood, which may be the main reason to limit analytical sensitivity.29

Beyond detection of OC, we also explored the possible clinical utility of cfDNA methylation in prognostic stratification of OC. CA125 is a serum marker with a persistent increase in many recurrent OC patients, but it has been demonstrated to be an unideal surrogate to predict recurrence and progression.30 Previous studies have demonstrated that methylation markers in tissues can predict the survival of patients with HG-SOC.31,32 However, these models heavily rely on the sequencing of tissues. Since the methylation alteration can be detected in cfDNA, we hypothesized that the cfDNA methylation could also be used as prognostic stratification. Our results showed that cfDNA methylation markers could differentiate the prognosis of OC patients and it was an independent risk factor for disease progression. Prognostic stratification analysis may help identify patients who may benefit from aggressive treatment and frequent surveillance. In our study, 15-methylation-marker classifier could be utilized to evaluate the risk of relapse/progression in patients with HG-SOC, independent of clinical risk factors such as stage and treatment and performed better than CA125.

Since the prognostic markers are most likely associated with key signalling of cancer development and evolution, which could provide insights for therapeutic implications of OC in the future. The methylation markers identified in the prognostic model and their implications in cancer development, immune contexture, and HRR signalling were investigated in the TCGA-OC cohort. High levels of HRD score and chromosomal instability were shown in the low-risk group, and the variability of HRD may not be mediated by gene mutations in the HRR pathway. It may partly explain why patients in the low-risk group are more sensitive to platinum-containing regimens. Further investigation into immune characteristics revealed higher immunogenicity and immune cell infiltrations in the low-risk group, respectively. Indubitably, elucidating the mechanisms underlying the prognosis of OC at the molecular level is of significance to facilitate the treatment of ovarian cancer and improve the survival of patients. However, the treatment responsiveness regarding target therapy or immunotherapy warrants further investigation.

Several limitations of this study should not be ignored. First, our study was limited by a relatively short clinical follow-up (median follow-up time, 16.8 months), and further investigations with longer clinical surveillance are needed to adequately assess the reliability of this OC-P model in clinical decision-making for patients. Second, the robustness of the OC-P model will be further confirmed if more demographic data such as ethnicity and comorbidities were included in the multivariable models. Third, the rationality of the OC-P model was tested with methylome data from the ICGC and TCGA cohort, which instead of cfDNA, is the methylation sequencing data of tissues. The patient cohort with the identical sequence approach and methylation sites is needed to validate the prognostic performance of
the OC-P model in the future. Finally, whether the OC-specific methylation markers could be utilized to monitor the treatment responsiveness in tumor real-time needs further investigation.

Collectively, our study demonstrated the rationality and accuracy of cfDNA methylation markers for the detection and prognostic prediction of OC. The validation in a larger population is warranted in the future.

Contributors
Conception and design: Lingying Wu, Ning Li, Chengcheng Li, Yu Zhang. Collection and assembly of data: Yifan Li, Guangwen Yuan, Yangchun Sun, Rong Zhang, Xiaoguang Li, Jing Zhao, Yuzi Zhang, Xiaofang Wen, Weiqi Nian, Xin Zhu, Shangli Cai. Data analysis and interpretation: Leilei Liang, Chengcheng Li, Yuchen Liao, Guoqiang Wang, Jiayue Xu, Yifan Li. Manuscript writing: All authors. Accessed and verified the data: Lingying Wu and Ning Li. All authors read and approved the final version of the manuscript.

Data sharing statement
The authors declare that relevant data supporting the findings of this study are available within the paper and its Supplementary files. The ICGC and TCGA cohorts in this study were publicly available as described in the Method section. The script for the classification model was submitted as a supplementary file. Due to ethical and privacy concerns, we are unable to publish the patient-level data in our study, of which readers may find the online version at doi:10.1016/j.ebiom.2022.104222. None.

FundingNone.

Acknowledgements
We acknowledge all the donors who supported our study. We thank all the patients and family members for their contribution to this study. We thank the clinical research teams at all our centres for their contribution to this study.

Declaration of interests
Chengcheng Li, Yuchen Liao, Guoqiang Wang, Jiayue Xu, Yuzi Zhang, Jing Zhao, Xiaofang Wen, Xin Zhu, and Shangli Cai are employees of Burning Rock Biotech. The other authors declare no competing interests.

References
1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018. GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
2. Armstrong DK, Alvarez RD, Bakkkum-Ganzes JN, et al. Ovarian Cancer, Version 2.2020. NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Cancer Network. 2021;19(2):219–246.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7–33.
4. Menon U, Gentry-Maharaj A, Burnell M, 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. 2021;397(10290):2182–2193.
5. Kalsi J, Gentry-Maharaj A, Ryan A, et al. Performance characteristics of the ultrasound strategy during incidence screening in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). Cancer. 2021;117(4):838.
6. Ferraro S, Braga F, Lanzoni M, Boracchi P, Bignozzi EM, Panethegini M. Serum human epididymis protein 4 vs carbohydrate antigen 125 for ovarian cancer diagnosis: a systematic review. J Clin Pathol. 2015;68(4):273–281.
7. Pinsky PF, Yu K, Kramer BS, Black A, Prorok PC. Extended mortality results for ovarian cancer screening in the PLCO trial with median 15 years follow-up. Gynecol Oncol. 2016;144(2):270–275.
8. Henderson JT, Webber EM, Sawaya GF. Screening for ovarian cancer: updated evidence report and systematic review for the US Preventive services task force. JAMA. 2018;319(6):555–566.
9. Zhang M, Cheng S, Jin Y, Zhao Y, Wang Y. Roles of CA125 in diagnosis, prediction, and oncogenesis of ovarian cancer. Biochim Biophys Acta Rev Cancer. 2022;1875(4):889–903.
10. Wan JCM, Massie C, Garcia-Corbacho J, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nat Rev Cancer. 2017;17(7):223–238.
11. Gao Q, Zeng Q, Wang Z, et al. Circulating cell-free DNA for cancer early detection. Innate. 2022;24(4):452–466.
12. Heitzer E, Haque IS, Roberts CE, Speicher MR. Current and future perspectives of liquid biopsies in genomics-driven oncology. Nat Rev Genet. 2019;20(2):71–88.
13. Dor Y, Cedar H. Principles of DNA methylation and their implications for biology and medicine. Lancet North Am Ed. 2018;392(10149):777–786.
14. Liu L, Young JM, Jassowicz AF, et al. Targeted methylation sequencing of plasma cell-free DNA for cancer detection and classification. Ann Oncol. 2018;29(6):1445–1453.
15. Xu RH, Wei W, Krawczyk M, et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. Nat Mater. 2017;16(11):1155–1161.
16. Luo H, Zhao Q, Wei W, et al. Circulating tumour DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer. Sci Transl Med. 2020;12(524):eabc435.
17. Liu M, Oxnard G, Klein E, et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann Oncol. 2020;31(6):743–759.
18. Widischwender M, Zikan M, Wahl B, et al. The potential of circulating tumor DNA methylation analysis for the early detection and management of ovarian cancer. Genome Med. 2017;9(1):116.
19. Luo B, Ma F, Liu H, et al. Cell-free DNA methylation markers for differential diagnosis of hepatocellular carcinoma. BMC Med. 2022;20(6):8.
20. Zhang Q, Hu G, Yang Q, et al. A multiplex methylation-specific PCR assay for the detection of early-stage ovarian cancer using cell-free serum DNA. Gynecol Oncol. 2021;219(5):132–139.
21. Gao Q, Li B, Cai S, et al. Early detection and localization of multiple cancers using a blood-based methylation assay (ELSA-seq). J Clin Oncol. 2021;39(17_suppl):e59.
22. Liang N, Li B, Jia Z, et al. Ultrasensitive detection of circulating tumour DNA via deep methylation sequencing aided by machine learning. Nat Biomed Eng. 2021;5(6):586–599.
23. Patch AM, Christie EL, Etemadmoghadam D, et al. Whole-genome characterization of chemoresistant ovarian cancer. Nature. 2015;521(7553):489–494.
24. Thorsson V, Gibbs DL, Brown SD, et al. The immune landscape of cancer. Immunity. 2018;48(4):812–830.e14.
25 Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf*. 2013;14:7.
26 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143 (1):29–36.
27 Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983;148(3):839–843.
28 Lheureux S, Braunstein M, Oza AM. Epithelial ovarian cancer: evolution of management in the era of precision medicine. *CA Cancer J Clin*. 2019;69(4):280–304.
29 du Bois A, Trillsch F, Mahner S, Heitz F, Harter P. Management of borderline ovarian tumors. *Ann Oncol*. 2016;27 suppl 1:i120–i12.
30 Karam AK, Karlan BY. Ovarian cancer: the duplicity of CA125 measurement. *Nat Rev Clin Oncol*. 2010;7(6):335–339.
31 Guo W, Zhu L, Yu M, Zhu R, Chen Q, Wang Q. A five-DNA methylation signature act as a novel prognostic biomarker in patients with ovarian serous cystadenocarcinoma. *Clin Epigenet*. 2018;10 (1):142.
32 Wei SH, Balch C, Paik HH, et al. Prognostic DNA methylation biomarkers in ovarian cancer. *Clin Cancer Res*. 2006;12(9):2788–2794.