Serum Lipidome Profiling Reveals a Distinct Signature of Ovarian Cancer in Korean Women

Samyukta Sah[1,2], Olatomiwa O. Bifarin[1], Samuel G. Moore[2], David A. Gaul[1,2], Hyewon Chung[3], Sun Young Kwon[4], Hanbyoul Cho[5], Chi-Heum Cho[3], Jae-Hoon Kim[6], Jaeyeon Kim[7], and Facundo M. Fernández[1,2]

ABSTRACT

Background: Distinguishing ovarian cancer from other gynecological malignancies is crucial for patient survival yet hindered by non-specific symptoms and limited understanding of ovarian cancer pathogenesis. Accumulating evidence suggests a link between ovarian cancer and deregulated lipid metabolism. Most studies have small sample sizes, especially for early-stage cases, and lack racial/ethnic diversity, necessitating more inclusive research for improved ovarian cancer diagnosis and prevention.

Methods: Here, we profiled the serum lipidome of 208 ovarian cancer, including 93 early-stage patients with ovarian cancer and 117 nonovarian cancer (other gynecological malignancies) patients of Korean descent. Serum samples were analyzed with a high-coverage liquid chromatography high-resolution mass spectrometry platform, and lipidome alterations were investigated via statistical and machine learning (ML) approaches.

Results: We found that lipidome alterations unique to ovarian cancer were present in Korean women as early as when the cancer is localized, and those changes increase in magnitude as the diseases progresses. Analysis of relative lipid abundances revealed specific patterns for various lipid classes, with most classes showing decreased abundance in ovarian cancer in comparison with other gynecological diseases. ML methods selected a panel of 17 lipids that discriminated ovarian cancer from non-ovarian cancer cases with an AUC value of 0.85 for an independent test set.

Conclusions: This study provides a systemic analysis of lipidome alterations in human ovarian cancer, specifically in Korean women.

Impact: Here, we show the potential of circulating lipids in distinguishing ovarian cancer from nonovarian cancer conditions.

Introduction

The high mortality rate of ovarian cancer is largely due to the asymptomatic progression of the disease, accompanied by our lack of understanding of ovarian cancer biology. Ovarian cancer is not a single disease; it consists of several histological subtypes (1) and its heterogenous nature is one major obstacle in understanding disease biology and identifying new biomarkers (2). Presently, there are no screening tests available for detecting ovarian cancer in the general population (3). In many cases, women with ovarian cancer experience non-specific symptoms such as pelvic pain and bloating, and thus are frequently dismissed and misdiagnosed as benign conditions (4). Symptomatic patients with a risk of developing ovarian cancer are screened via transvaginal sonography (TVS) or the measurement of serum protein biomarker CA125 (5). However, these diagnostic strategies lack adequate sensitivity and specificity. For instance, increased levels of CA125 are observed for several nonovarian cancer–related gynecological malignancies whereas it is only elevated in 50% of early ovarian cancer cases (5). Likewise, TVS can misdiagnose ovarian cancer as benign diseases, leading to mismanagement that significantly reduces patient survival rate (6). Patients with ovarian cancer misdiagnosed for benign diseases suffer worse prognosis compared with women who are treated by gynecologic oncologists (7), and currently 30% to 50% of women with ovarian cancer in the United States do not receive appropriate ovarian cancer treatment (4). Thus, although effective screening in the general population remains the goal, accurate diagnosis and triage of women suspected of ovarian cancer is crucial for improved prognosis.

Substantial effort has been put into improving the clinical diagnosis of ovarian cancer and understanding the underlying disease mechanisms. For example, the two-stage approach that uses both CA125 and TVS sequentially showed improved specificity in clinical trials, and the use of protein biomarker HE4 in combination with CA125 has shown better specificity for distinguishing malignant from benign pelvic masses (8). Other potential biomarkers for improved diagnosis include autoantibodies and antigen–autoantibody complexes, sometimes combined with CA125 (9, 10), serum microRNAs (11) or circulating tumor DNA (12). Recently, protein markers detected in Pap test fluid (13) have shown promising results to distinguish ovarian cancer from women with normal cytology. One characteristic feature of malignant tumors is their ability to rewire metabolism in response to high cell proliferation rates (14). Metabolomics, the examination of

[1]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia. [2]Petit Institute of Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, Georgia. [3]Department of Obstetrics and Gynecology, School of Medicine, Keimyung University, Daegu Republic of Korea. [4]Department of Pathology, School of Medicine, Keimyung University, Daegu, Republic of Korea. [5]Department of Obstetrics and Gynecology, Institute of Women's Life Medical Science, Yonsei University College of Medicine, Seoul, Republic of Korea. [6]Department of Obstetrics and Gynecology, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea. [7]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indiana University Melvin and Bren Simon Comprehensive Cancer Center, Indianapolis, Indiana.

S. Sah and O.O. Bifarin contributed equally as co-authors of this article.

Corresponding Authors: Facundo M. Fernández, Georgia Institute of Technology, 901 Atlantic Drive NW, Atlanta, GA 30332. E-mail: facundo.fernandez@chemistry.gatech.edu; and Jaeyeon Kim, 635 Barnhill Drive, MS4053 Indianapolis, IN 46202. E-mail: jaeyeonk@iu.edu

Cancer Epidemiol Biomarkers Prev 2024;33:681–93

doi: 10.1158/1055-9965.EPI-23-1293

metabolites, including lipids, carbohydrates, and amino acids, provides a powerful platform for investigating metabolic alterations and accelerating biomarker discovery.

In recent years, alterations in ovarian cancer lipid metabolism have gained increased attention (15, 16). Lipids function as building blocks for cell membranes, participate in cellular signaling, and are regulators of numerous cellular functions that drive energy-related processes (17). Given the close connection between altered lipid metabolism and oncogenesis, there is accumulating evidence showing specific lipid profiles associated with ovarian cancer growth and metastasis (15, 18, 19). Serum lipidome profiling of patients with ovarian cancer against normal controls and benign malignancies has shown evidence of dysregulation in glycerophospholipids, ceramides, and triglycerides (TG) being associated with malignant ovarian tumors (15, 16, 20). Some studies have suggested that the use of lipid panels in combination with CA125 can achieve enhanced diagnostic power (16, 21). The diagnostic potential of gangliosides, a class of glycolipids involved in immunosuppressive response in tumors, has recently been reported for distinguishing patients with ovarian cancer from nonovarian cancer–related diseases as well as healthy controls (22). Increased ganglioside levels in plasma, tissue, and ascites fluid from patients with ovarian cancer have also been reported (22, 23). These studies, however, are often hindered by small sample sizes, limited number of early-stage samples, lack of external validation datasets, and inconsistencies in sample collection and processing protocols, thereby limiting the statistical significance and robustness of the final results. Most studies focus on cohorts of non-Hispanic white women or women with European ancestry, and a few examples involving cohorts of Chinese women (24, 25). However, studies with patients of Korean descent are largely lacking. Although an integrated serum proteomics and metabolomics study of Korean patients has been previously reported, it only involved 10 patients with ovarian cancer (26).

Here, we present a comprehensive serum lipidomic study of patients with ovarian cancer of Korean descent with various histological types and disease stages ($n = 208$) and of women with other gynecological malignancies, including invasive cervical cancer ($n = 117$). Samples were secured from two independent tissue banks. Ultra-high performance liquid chromatography-high resolution mass spectrometry (UHPLC-MS) combined with machine learning (ML) were used to identify circulating lipidome alterations unique to patients with ovarian cancer. Data were analyzed on a lipid class basis, providing a systems level perspective of ovarian cancer–related lipidome perturbations. Furthermore, a panel of optimal lipid markers was identified with an ML pipeline to distinguish patients with ovarian cancer from nonovarian cancer. This panel included ether phospholipids, sphingolipids, and gangliosides.

Materials and Methods

Patient cohort

Serum samples were obtained from two independent tissue banks in South Korea: Dongsan Hospital Human Tissue Bank and the Human Tissue Bank of Gangnam Severance Hospital, Yonsei University College of Medicine (No. HTB-P2019–13). Samples from both tissue banks were obtained after the approval from their respective IRB and the patient's written informed consent. The study was conducted according to the guidelines of the Declaration of Helsinki. The Severance cohort included 185 samples from patients with ovarian cancer, 47 from women with benign ovarian tumors, 50 from invasive cervical cancers, and 21 samples from patients with benign uterine tumors. Blood was collected from all patients during surgery after anesthesia and at least 8 hours of fasting. In the Dongsan cohort, 88 women had ovarian cancer, 12 had benign ovarian tumors, 10 had benign uterine tumors, and 9 women had cervical cancer. As with the Severance cohort, samples from these patients were collected during surgery after anesthesia and at least 6 hours of fasting. All recruited participants were of Korean descent. Samples from both cohorts were grouped together and patients with ovarian cancer and all other gynecological malignancies (nonovarian cancer) were age-matched. The matched cohort included 208 patients with ovarian cancer (mean age, 51.9 years) and 117 nonovarian cancer (mean age, 49.9 years). Disease stages and histological characteristics of each patient are given in Supplementary Table S1. Among the patients with ovarian cancer, 93 patients had early-stage (I and II) cancers. Ten of the patients with ovarian cancer had recurrent cancer and 9 have had their samples collected more than once, although at different stages of ovarian cancer development. Samples from normal controls (i.e., women with no known gynecological malignancies) were also collected during regular health exams at Severance hospital. In this case, blood was collected after fasting for at least 8 hours. As noted earlier, blood from patients with ovarian cancer or patients with other conditions was collected after the initiation of anesthesia, and thus could not be directly compared with normal controls without major confounding effects. Therefore, control samples were excluded from the main data analysis pipeline. However, we also conducted a lipidome comparison study between healthy controls and patients with ovarian cancer for reference purposes. Results from this study are presented in the Supplementary Information section and the data are shared together with the rest of the cohort.

Chemicals

LC/MS grade 2-propanol, water, formic acid (99.5+%), ammonium formate, and ammonium acetate were purchased from Fisher Chemical (Fisher Scientific International, Inc.) and used for the preparation of chromatographic mobile phases and sample extraction. Isotopically labeled lipid standards (Supplementary Table S2) were purchased from Avanti Polar Lipids.

Sample preparation

Serum samples were thawed on ice, followed by extraction of the non-polar (lipid) metabolome. The extraction solvent was prepared by addition of 725 μL of the isotopically labeled lipid standard mixture (Supplementary Table S2) to 43.5-mL 2-propanol (1:60 ratio) and kept on ice. This cold extraction mixture was added to serum samples in a solvent: Serum 3:1 ratio for protein precipitation, followed by vortex mixing for 15 seconds. Samples were centrifuged at 13,000 rpm for 7 minutes and the resulting supernatant was transferred to LC vials. The supernatant was stored at −80°C until UHPLC-MS analysis, which was performed within one week. A blank sample, prepared with LC/MS grade water, underwent the same sample preparation process as the serum samples. Pooled quality control (QC) samples were prepared by combining 5 to 10 μL aliquots of each serum sample extract. This pooled QC sample was analyzed every 10 LC/MS runs to monitor and correct instrument stability through the course of the experiment. Samples were randomized both for sample preparation and LC/MS analysis.

Ultra-high performance LC/MS serum lipidomics

Reverse phase (RP) chromatography was performed in a Vanquish LC system equipped with a Thermo Accucore C30, 150 × 2.1 mm, 2.6-μm particle size column. An Orbitrap ID-X Tribrid mass spectrometer

(Thermo Fisher Scientific) was used for MS analysis. For negative ion mode, mobile phase A was 10 mmol/L ammonium acetate with water/acetonitrile (40:60 v/v) and mobile phase B was 10 mmol/L ammonium acetate with 2-isopropanol/acetonitrile (90:10 v/v). For positive ion mode, mobile phase A was 10 mmol/L ammonium formate with water/acetonitrile (40:60 v/v) and 0.1% formic acid. Mobile phase B was 10 mmol/L ammonium formate with 2-isopropanol/acetonitrile (90:10 v/v) and 0.1% formic acid. All samples were kept in the autosampler at 4°C during LC/MS runs and an injection volume of 2 μL was used in all cases. MS data were acquired in the 150 to 2,000 *m/z* range with a 120,000 mass resolution setting. The most important MS parameters and the chromatographic gradient used are given in Supplementary Tables S3 and S4, respectively. For MS/MS experiments, the Deep AcquireX data acquisition workflow was applied. Stepped normalized collision energy values of 15, 30, 45 were used for fragmenting precursor ions in the HCD cell followed by Orbitrap analysis at 30,000 mass resolving power. Precursor ions were also fragmented with a collision-induced dissociation (CID) energy of 40 and analyzed in the ion trap.

Data processing

Spectral feature (retention time, *m/z*) pairs were extracted from raw data using Compound Discoverer v3.3 (Thermo Fisher Scientific). This step included chromatographic peak alignment, peak peaking, peak area integration, and adjustment for instrument drift using the pooled QC injections and the systemic error removal using random forest algorithm (27). Chromatographic peaks with less than five times the peak area of the matching sample blank peaks were marked as background noise and removed from the dataset. Further filtering was performed by removing features not present in at least 50% of the QC sample injections or that had a relative standard deviation (RSD) greater than 30% in QC samples. Following feature extraction, all retained features were matched against a curated in-house lipid spectral database. Exact masses, elemental formulas, and MS-MS spectra were used for matching purposes, and results manually curated. All annotated features were subject to ML feature selection.

QC

Data quality was assessed using the set of pooled QC runs. The average intensities of all QC runs were examined with RawMeat (Vast Scientific) software and their RSD for positive and negative ion modes calculated. An average RSD below 15% was obtained for positive ion mode data. A slight drift in the negative ion mode dataset was observed, with a %RSD of 35%. For this dataset, features that could not be aligned after data processing in Compounds Discoverer were removed. Principal component analysis (PCA) showed excellent clustering of QC samples (Supplementary Fig. S1), confirming good reproducibility. In addition, the quality of all sample and QC runs was evaluated across each batch using the internal standard (IS) peak areas (Supplementary Table S5) and found to be excellent.

Selection of discriminant lipids

Exploratory analysis of the overall lipidome alterations in ovarian cancer was conducted using the following pipeline: First, one of a pair of two highly correlated features was removed using a Pearson's correlation coefficient cutoff value of 0.85. Fold changes for all remaining features were then calculated as the base 2 logarithm of the average lipid abundance ratios between patients with ovarian and nonovarian cancers. Following this step, the statistical significance of each detected lipid was calculated using the Welch's *t* test with a Benjamini–Hochberg correction. Lipids with a *q* value of <0.05 were considered statistically significant. Altered lipid features were then autoscaled followed by feature selection using the SelectFromModel function with a random forest classifier in the Python sci-kit-learn library (v1.1.2). Features were ranked by their Gini index; lipids with a Gini index equal to or greater than the Gini index mean were selected. The sci-kit-learn default parameters were used and the number of trees for the random forest classifier was set to 100. To study differences between early-stage ovarian cancer, nonovarian cancer, and late-stage ovarian cancer, the dataset was stratified by ovarian cancer stages. The early-stage class consisted of stages I and II and the late-stage ovarian cancer class was built using stages III and IV. Lipids statistically different between ovarian and nonovarian cancer groups (*P* value <0.05) were retained, followed by the random forests feature selection process described above.

ML pipeline for biomarker panel selection

A small panel of lipids differentiating ovarian cancer from nonovarian cancer samples was selected using the following ML pipeline: (i) One of two highly correlated features was removed using a Pearson's correlation coefficient cutoff value of 0.85. (ii) The dataset was split into training (70%, ovarian cancer $n = 144$ and nonovarian cancer $n = 83$) and test set (30%, ovarian cancer $n = 64$ and nonovarian cancer $n = 34$). (iii) Next, the Welch's *t* test ($P < 0.05$) was applied to ovarian and nonovarian cancer samples in the training set. (iv) Because the dataset contained fewer number of nonovarian cancer samples than ovarian cancer samples, the training and test sets were imbalanced and consisted of almost twice as many ovarian cancer samples than nonovarian cancer. Imbalanced datasets often lead to poor classification performance as the classification classes are not equally represented (28). To achieve improved classification power, the training set was balanced *via* the Synthetic Minority Over-sampling Technique (SMOTE; ref. 28), which can be used to create synthetic minority class samples for imbalanced datasets. Python's imbalanced-learn library (v. 0.9.1) was used to implement SMOTE. (v) Feature selection was carried out on the balanced training set. In this case, random forests were used, and features were ranked by their Gini index feature importance score. The top 10 lipids provided the best discriminating power for the training set and were selected for classification purposes. The sci-kit-learn default parameters were used whereas the number of trees was set to 100. The classification power of ganglioside lipids was also evaluated. The best discriminating gangliosides were selected following the same feature selection process as described above. The top seven ganglioside features provided the best discriminating power for the training set and were selected for further analysis. Jupyter notebooks was used as the integrated development environment for all statistical and ML analysis. All code is available via GitHub: https://github.com/facundof2016/Sah_Ovarian-cancer_2023.

ML classification

Classification tasks were performed by training ML models to differentiate ovarian cancer from nonovarian cancer samples. A balanced training dataset ($n = 144$ ovarian cancer, $n = 144$ nonovarian cancer) was used to train the models, with a 10-fold cross validation conditions. Random forests, logistic regression, k-nearest neighbor (k-NN), linear kernel support vector machine (SVM-Lin), and a voting ensemble classifier were used. The estimators for the voting classifier included all four ML methods: Random forests, logistic regression, k-NN, and SVM-Lin. The default parameters of Python's sci-kit-learn library (v1.1.2) were used. All classifiers were evaluated on the basis of sensitivity, specificity, accuracy, and their area under the receiver operating characteristic curve (AUC-ROC).

AutoML technique for classification tasks

For AutoML using auto-sklearn (version 0.15.0), the AutoSklearnClassifier was allocated 3,600 seconds to identify the optimal ML pipelines. We used a cross-validation resampling strategy with standard parameters, adopting the hold-out method. This method further segmented the training data into an internal training and validation set with a 67:33 ratio. The ensemble approach was implemented, with consideration given to up to 50 models for inclusion in the ensemble. ROC-AUC was used as the evaluation metrics. During optimization, each model was restricted to a maximum runtime of 1,440 seconds. After establishing the AutoML pipelines using autosklearn, the final ensemble was trained on the complete training dataset through 5-fold cross-validation. Subsequently, its performance was assessed on unseen datasets.

Data availability

Data generated in this study are available through the NIH Metabolomics Workbench (http://www.metabolomicsworkbench.org/) with project ID PR001623 (Study ID ST002521) http://dx.doi.org/10.21228/M8X42D. Code is provided at https://github.com/facundof2016/Sah_Ovarian-cancer_2023

Results

Patient cohort

A high-coverage serum lipidomic LC/MS profiling workflow was applied to the combined patient cohorts. Each cohort was composed of patients with ovarian cancer, benign ovarian tumors, benign uterine tumors, or cervical cancer. Patients with benign conditions and cervical cancers were grouped together as nonovarian cancer samples, whereas patients with ovarian cancer of various stages and histological types were all combined into the ovarian cancer group. To remove potential confounders from the dataset, ovarian and nonovarian cancer groups were age matched. The statistical significance between the ages of patients with ovarian and nonovarian cancers was assessed with the Welch's *t* test. The age-matched cohort consisted of 208 patients with ovarian cancer and 117 patients with nonovarian cancer (Supplementary Table S1; **Fig. 1**).Among patients with ovarian cancer, 93 women had early-stage (I and II) cancers, of which 30% (28/93) were of serous histology, the most aggressive subtype (1). In contrast, serous tumors accounted for 86% of advanced-stage (III and IV) cases whereas the remaining 14% of the cases included clear cell, transitional, mucinous, and carcinosarcoma subtypes.

Serum lipidome differences between patients with ovarian and nonovarian cancers

Lipidomics data were acquired for all ovarian and nonovarian cancer samples using RP UHPLC-MS. A total of 24,297 and 5,485 de-isotoped and de-adducted spectral features (retention time, *m/z* pairs) were extracted from positive and negative ion mode datasets, respectively. From these, a total of 994 lipid species assigned to 22 lipid subclasses were successfully annotated using our in-house MS/MS spectral library. Detected lipid classes included fatty acids, glycerophospholipids, glycerolipids, sphingolipids, and sterol lipids. TG and

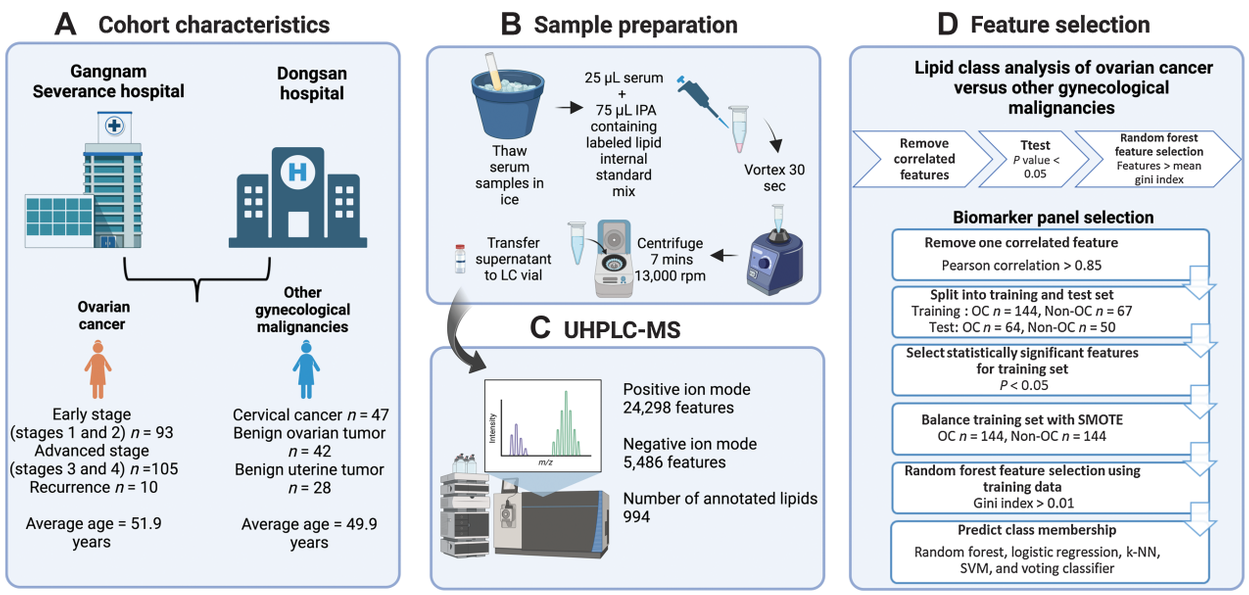

Figure 1.

Study design overview, UHPLC-MS workflow, and machine learning pipeline. **A,** Serum samples from patients with ovarian cancer ($n = 208$) and nonovarian cancer ($n = 117$) obtained from Gangnam Severance and Dongsan Hospitals were studied. **B,** Serum sample preparation workflow before UHPLC-MS analysis. **C,** UHPLC-MS data collection in both positive and negative ion modes. Data were processed with Compound Discoverer v.3.3 (Thermo Fisher Scientific) and features annotated using in-house MS-MS libraries. **D,** Machine learning workflow for selecting the most relevant lipids to differentiate ovarian and nonovarian cancer conditions. For the lipid class analysis of ovarian cancer versus other gynecological malignancies, selection of the best differential lipids was performed using random forests feature selection. Lipids with a Gini index greater than the mean of all Gini index values were selected. For the biomarker panel selection, selection of a 10-lipid panel for differentiating ovarian and nonovarian cancer serum samples was performed using random forests algorithm. Lipids with a Gini index of >0.01 were selected as the best differential features in the training set. Five different machine learning models—random forests, logistic regression, k-Nearest Neighbor (KNN), support vector machines (SVM) and voting classifier, which is an ensemble of the four listed classifiers—were used for classification. SMOTE, Synthetic Minority Oversampling Technique. (Created with BioRender.com.)

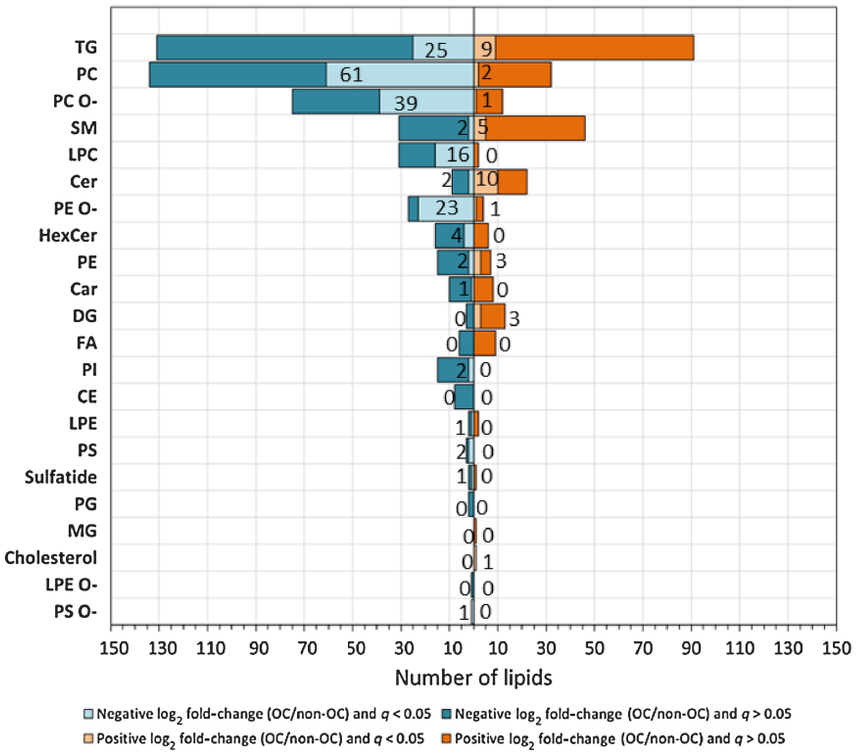

Figure 2.

Annotated lipids, grouped by lipid class, showing number of lipids with positive and negative fold changes between ovarian and nonovarian cancer groups. Fold changes were calculated as the base 2 logarithm of the average lipid abundance ratios for ovarian cancer versus nonovarian cancer. A positive fold-change value indicates higher levels in ovarian cancer samples. Negative values indicate lower levels in ovarian cancer samples. The number of lipids with negative fold-change values and positive fold-change values for each lipid class are shown as blue and orange bars, respectively. The number of statistically significant lipids (FDR corrected P <0.05) with negative and positive fold-change values are labeled as light blue and orange bars, respectively. TG, Triacylglycerols; PC, Phosphatidylcholines; PC O-, Ether phosphatidylcholines; SM, Sphingomyelins; LPC, Lysophosphatidylcholines; Cer, Ceramides; PE O-, Ether phosphatidylethanolamines; Car, Carnitines; HexCer, Hexosylceramides; PE, Phosphatidylethanolamines; DG, Diacylglycerols; FA, Fatty acids; PI, Phosphatidylinositols; CE, Cholesterol esters; LPE, Lysophosphatidylethanolamines; PS, Phosphatidylserines; PG, Phosphatidylglycerols; MG, Monoradylglycerols; LPE O-, Ether Lysophosphatidylethanolamines; PS O-, Ether phosphatidylserines.

phosphatidylcholines (PC) accounted for 29.6% and 18.2% of the dataset, respectively. Representative raw data showing separation of the various lipid classes detected are shown in Supplementary Information, Supplementary Fig. S2, showcasing the excellent lipid coverage obtained in our method. Exploratory analysis to investigate lipidome differences among different ovarian cancer histological types was first conducted. In addition to common lipids, 22 gangliosides were annotated by matching their elemental formulas and exact masses to online databases, including LipidMaps (29) and HMDB (30). MS/MS information was available for 16 of the 22 annotated gangliosides and was also matched against online databases (Supplementary Table S6). PCA using the combined set annotated lipids did not show any clear clustering based on various ovarian cancer histological types (Supplementary Fig. S3B). Unsurprisingly, the one sample with low-grade serous carcinoma (LGSC) histology deviated the most from the rest of the samples. This could be due to the low number of LGCS samples present in the dataset ($n = 1$). Compared with high-grade serous carcinoma (HGSC), LGSC has a different mode of carcinogenesis with distinct molecular-genetic features (31).

To investigate differences between patients with ovarian and nonovarian cancers at the lipidome level, unsupervised and supervised multivariate analysis was conducted. PCA using the combined set of (+) and (−) RP UHPLC-MS 994 annotated lipids showed minimal separation between patients with ovarian and nonovarian cancers (Supplementary Fig. S4A). This was expected as differences between different types of malignancies are likely to be relatively subtle. The dataset was further explored with supervised multivariate analysis. Orthogonal partial least squares-discriminate analysis (oPLS-DA) for the same dataset indicated better clustering between the two groups (Supplementary Fig. S4B), which could further be improved *via* ML-based feature selection processes. Performance characteristics of this oPLS-DA model with 10-fold cross validation were a modest 0.73 sensitivity and 0.65 specificity. To better visualize lipidome differences between patients with ovarian and nonovarian cancers, the dataset was analyzed on a lipid class basis. Fold changes and statistical significance between ovarian and nonovarian cancer groups were calculated for all annotated lipids (**Fig. 2**). Among the 994 annotated lipids, 218 had FDR corrected P values of <0.05 ($q < 0.05$). Fold changes calculated between ovarian and nonovarian cancer groups showed an overall decrease in the serum lipid abundances of patients with ovarian cancer. Seventy-four percent of the detected ether phosphatidylethanolamines (PE O-), for example, had significantly decreased abundance in patients with ovarian cancer. Lipid classes, including PC, ether PCs (PC O-), lysophosphatidylcholines (LPC), and lysophosphatidylethanolamines (LPE), were also significantly reduced in women with ovarian cancer (**Fig. 2**). Although most lipids were reduced in the ovarian cancer samples, some lipids, including ceramides (Cer) and TG, were increased. Interestingly, the only two ceramides species that were lower in patients with ovarian cancer were Cer(d18:0/22:0) and its oxidated form Cer(t18:0/22:0).

An ML workflow, described in the methods section and **Fig. 1**, was developed to select features that better describe lipidome differences between ovarian and nonovarian cancer conditions. The one hundred best differential lipids were selected (**Fig. 3**), including ceramides, LPC, PC, PC O-, PE O-, and TG, indicating diverse changes in the circulating lipid abundances in serum of patients with ovarian cancer. Unsupervised PCA and supervised oPLS-DA using these selected 100 lipids (**Fig. 3**) showed better clustering than the previously built models using all annotated lipids (Supplementary Fig. S4). Performance characteristics of the oPLS-DA model on the first 2 latent variables were 0.75 and 0.71 cross-validated sensitivity and specificity, respectively. Comparison of the relative lipid abundances in ovarian and nonovarian cancer conditions showed patterns of alterations based on the lipid class, with majority of the lipids showing decreased

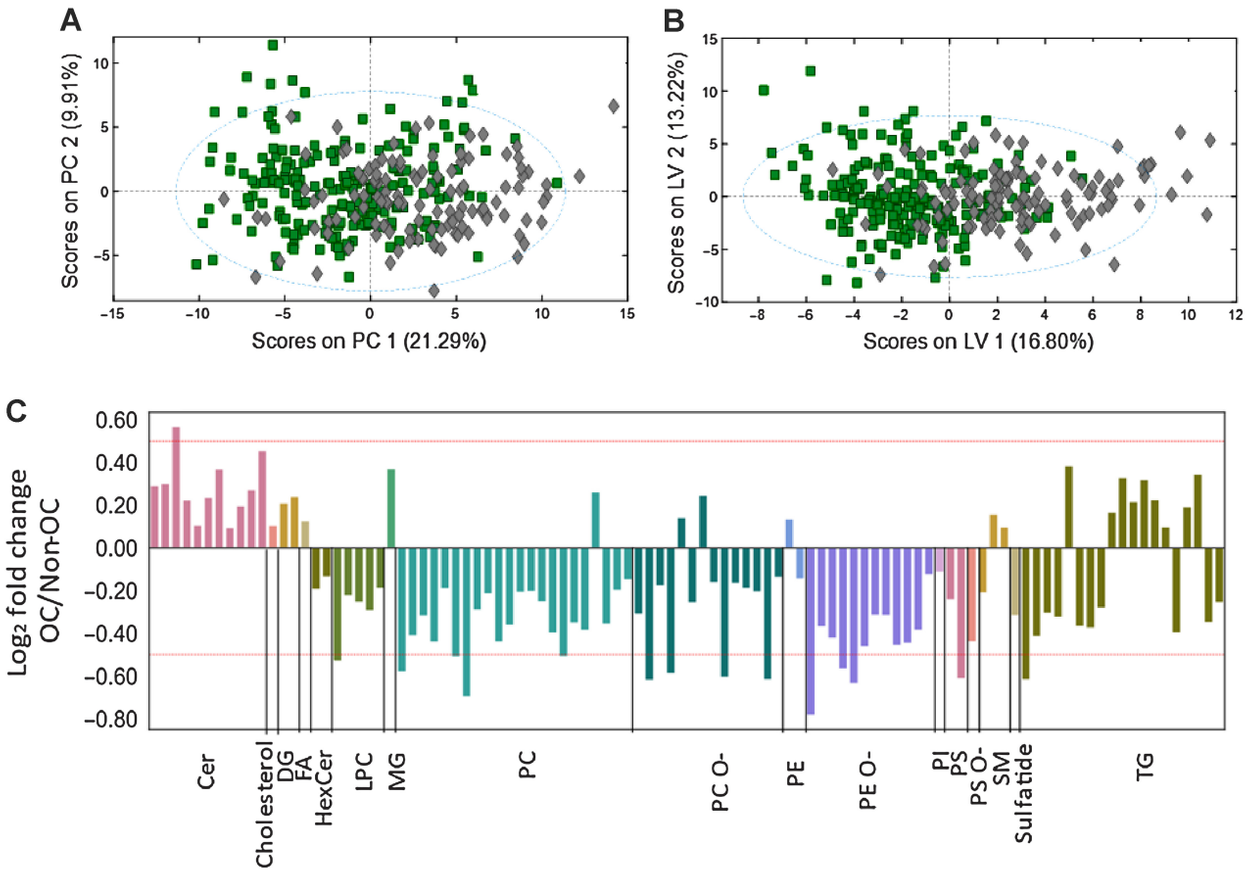

Figure 3.

Serum lipidome analysis of ovarian and nonovarian cancer samples using only the abundances of select lipids. Lipids were selected with the following feature selection workflow: One of the two highly correlated lipids was filtered out using a Pearson correlation coefficient cutoff value of 0.85. Next, lipids with *P* values lower than 0.05 were selected, followed by random forest feature selection in which lipids with a Gini index greater than the mean of all Gini indices were selected. **A,** PCA score plot showing clustering of ovarian and nonovarian cancer samples using the selected lipids. **B,** o-PLS-DA score plot for the same dataset. Ovarian cancer samples are depicted as green squares and nonovarian cancer samples are shown with gray diamonds. **C,** Fold changes for the selected lipids. Fold changes were calculated as the base 2 logarithm of the average lipid abundance ratios for ovarian cancer versus nonovarian cancer. A positive fold-change value indicates higher levels in ovarian cancer samples. Negative values indicate lower levels in ovarian cancer samples. TG, Triacylglycerols; PC, Phosphatidylcholines; PC O-, Ether phosphatidylcholines; SM, Sphingomyelins; LPC, Lysophosphatidylcholines; Cer, Ceramides; PE O-, Ether phosphatidylethanolamines; Car, Carnitines; HexCer, Hexosylceramides; PE, Phosphatidylethanolamines; DG, Diacylglycerols; FA, Fatty acids; PI, Phosphatidylinositols; CE, Cholesterol esters; LPE, Lysophosphatidylethanolamines; PS, Phosphatidylserines; PG, Phosphatidylglycerols; MG, Monoradylglycerols; LPE O-, Ether Lysophosphatidylethanolamines; PS O-, Ether phosphatidylserines.

serum abundance in patients with ovarian cancer. Only ceramides, some SM, cholesterol, FA(20:1), diglycerides, monoglyceride (MG), and some TG species were increased in the patients with ovarian cancer. TG species showed variable trends based on the FA side chain: Most TG species with long fatty acyl side chain composition were increased whereas TG with short fatty acyl side chains were decreased. In addition, 15 of the 100 selected lipids had significantly different fold changes between ovarian and nonovarian cancer groups ($\log_2$ fold change < -0.5 or $\log_2$ fold change > 0.5), confirming that lipids exhibit profound changes in ovarian cancer. These mainly consisted of PE O-, PC O-, and PC.

Lipidome analysis of HGSC

The heterogeneity of ovarian cancer is one major obstacle in biomarker discovery (22). HGSC accounts for only 30% of early-stage ovarian cancer, but 86% of advanced-stage ovarian cancer. The overwhelming majority of early-stage ovarian cancers (~69%) are of non-HGSCs, the ovarian cancer types that are generally slow-growing, less aggressive, and therefore commonly detected at early stages (1). In contrast, most advanced-stage ovarian cancers (86%) are HGSC, which are aggressive and typically not detected until advanced stage. These findings indicate that, potentially, HGSC has biological mechanisms of development and progression distinct from non-HGSCs. We investigated whether HGSC exhibits a distinctive lipid profile compared with other ovarian cancer subtypes. The Welch's *t* test was performed to select statistically different features between HGSC and other ovarian cancer types. Seventy-four lipids were selected as statistically significant ($P < 0.05$); however, when *P* values were corrected for false positives (FDR corrected) no lipids remained statistically significant, indicating that lipidome differences between different histological types are likely small, or not present in the subset of annotated lipids. Nevertheless, we still examined the lipidome of HGSC with the 74 selected features with a *P* value of < 0.05. Comparison of the serum lipid abundance of HGSC versus non-HGSC

ovarian cancer types showed subtle differences (Supplementary Fig. S5). Only one TG species showed a $\log_2$ fold change greater than 0.5. Overall, the serum lipidome of HGSC showed increased lipid abundance compared with non-HGSC types. Most lipid classes, including all ceramides, and most glycerophospholipids, and glycerolipids, were increased in HGSC, whereas only ether phospholipids, including PE O- and PC O- and a few SM, PC and PE species, were reduced. To further investigate the serum lipid profile of HGSC, we compared the serum lipidome of HGSC with nonovarian cancer malignancies. The Welch's *t* test selected 202 statistically significant features with FDR corrected *P* values of <0.05. Fold changes between HGSC and nonovarian cancer showed that, unlike HGSC versus other ovarian cancer subtypes, in this case most lipid classes were reduced in HGSC (Supplementary Fig. S6). These results bear similarities with our previous analysis of all ovarian cancer subtypes versus nonovarian cancer (**Fig. 3**), reinforcing the finding that lipid alterations between ovarian and nonovarian cancer conditions are much more profound than they are between different histological ovarian cancer types.

Stage-stratified analysis

Next, we evaluated the lipidome profile associated with early-stage ovarian cancer (I and II) and advanced-stage ovarian cancer (III and IV) versus nonovarian cancer conditions using a stage-stratified feature selection approach, as described in the Materials and Methods section. The early-stage ovarian cancer cohort consisted of 93 patients with ovarian cancer and 117 patients with nonovarian cancer, whereas the late-stage ovarian cancer cohort had 115 patients with ovarian cancer and 117 patients with nonovarian cancer. One-hundred twenty lipids were selected as the best differential lipids between early-stage ovarian and nonovarian cancers whereas 102 lipids were selected for late-stage ovarian versus nonovarian cancer conditions. Although fewer lipids were selected for late-stage ovarian cancer versus nonovarian cancer than early-stage ovarian cancer versus nonovarian cancer, a greater diversity of lipid classes was altered for late-stage ovarian cancer (Supplementary Tables S7 and S8). Both early- and late-stage ovarian cancer exhibited changes in glycerophospholipids, glycerolipids, and sphingolipids, whereas some lipid classes—including diglycerides, fatty acids, and cholesterol—differed only for late-stage patients with ovarian cancer. Most of the selected differential lipids for both early-stage and late-stage ovarian cancer included PC, TG, PC O-, PE O-, and Cer (**Fig. 4**). Compared with late-stage ovarian cancer, a higher number of LPC and SM species were selected for early-stage ovarian cancer. Furthermore, as expected, lipidome alterations in early disease stages were less prominent than in advanced disease stage. Volcano plots showing the fold changes and statistical significance for the selected lipids show that ceramides, some TG, and SM were increased in both early- and late-stage ovarian cancer. Lipid classes identified as significantly altered (*P* value <0.05 and fold change greater than $\log_2$ 0.5) in early-stage ovarian cancer included PE O-, PC O-, PC, and one PS species (Supplementary Table S9). Late-stage ovarian cancer exhibited significant changes in Cer, PE O-, PC O-, PC, LPC, PS, and TG (Supplementary Table S9).

Among the differentially abundant lipids selected for early-stage or late-stage ovarian cancer versus nonovarian cancer conditions, 42 lipids were altered in both subgroups (Supplementary Fig. S7). These were mainly sphingolipids, including Cer and SM, glycerophospholipids, including PC, PS and PI, ether phospholipids (PC O-and PE O-), and TG species, Lipids with fatty acid alkyl chains of C18:0, C18:1, C18:2, C20:4, C22:5, and C22:6 were frequently observed in this panel. In addition, all selected ceramide species were composed of d18:1 or d18:2 fatty acyl backbone, whereas the fatty acyl side chains were very long fatty acids, including C24, C25, or C26. Similarly, the sn-1 fatty acid alkyl chain for most PE O-, PE, and PS species were composed of C18:1 or C18:2, whereas sn-2 fatty acid alkyl chains were composed of very long chain fatty acyl chains of C22 and C24. This panel of lipids also consisted of glycerophospholipids with dietary odd fatty acyl chain composition (C17:0, C17:1, and C17:2). Next, fold changes for these 42 lipids between early- and late-stage patients with ovarian versus nonovarian cancers were analyzed. Subtle changes in serum lipid abundance were observed for early-stage ovarian cancer versus nonovarian cancer. Among the selected lipids, PE O-(17:1/20:4) had the greatest fold change. Lipid alterations for advanced stage ovarian cancer versus nonovarian cancer were much more profound than in early-stage patients with ovarian cancer, and 42 lipids showed alterations that were directionally concordant with the disease stage. These results indicate that changes in the serum lipid abundance of ovarian cancer can be detected when the cancer is localized, and that these changes are amplified as the diseases progresses (Supplementary Fig. S7).

In addition, we evaluated the serum lipidome of early-stage HGSC samples using the 120 lipids selected as differential features for early-stage ovarian cancer versus nonovarian cancer (Supplementary Fig. S8). In line with our lipidome analysis of HGSC shown earlier (Supplementary Fig. S6), the serum lipidome of early-stage serous patients exhibited alterations that in most cases were like the combined early-stage ovarian cancer cohort of various histological types (**Fig. 4**). Only a few PC species showed increased abundance in early-stage serous samples. In addition, five lipids, including TG(56:6), PC(O-35:4), PC(O-32:2), PE(O-18:2/20:1), and PC(37:3), showed significant alteration (Supplementary Fig. S8), suggesting that circulating lipids can aid in early diagnosis of serous carcinoma.

Biomarker panel

Considering the importance of biomarkers in ovarian cancer diagnosis and management, we selected a panel of lipids to discriminate ovarian cancer and from nonovarian cancer conditions with the maximum possible accuracy. All 994 annotated lipids from (+) and (−) RP UHPLC-MS datasets were combined and subjected to the ML workflow detailed in the Materials and Methods section and **Fig. 1**. Although the AUC values for the imbalanced dataset (ovarian cancer $n = 144$, nonovarian cancer $n = 83$) were acceptable, the specificity for these classification models was very low (Supplementary Table S10). For the imbalanced dataset, a logistic regression model provided the best AUC value of 0.76 whereas the specificity for this model was only 0.50. Thus, to achieve better classification power, the training set was balanced using the SMOTE (28), yielding a 10-lipid panel that consisted mainly of ceramides and ether-linked glycerophospholipids (**Table 1**). The best classification performance for the training set was achieved with a random forest classier: The AUC, sensitivity, and specificity values under 10-fold cross validation were 0.85, 0.78, and 0.76, respectively (Supplementary Table S11; **Table 2**). For the test set, the random forest classifier provided the AUC value of 0.82 whereas the sensitivity and specificity values were 0.78 and 0.76, respectively. The selected panel of lipids provided good classification performance for the test set, highlighting the potential of lipid markers to distinguish the serum lipid profile of Korean women with ovarian cancer and different types of malignant or benign gynecological diseases.

As accurate early diagnosis of ovarian cancer, especially of early-stage HGSC is crucial to improve clinical outcomes, we evaluated the classification power of this 10-lipid panel for early-stage samples ($n = 31$) versus nonovarian cancer samples ($n = 30$) from the test set, as well as for all 28 early-stage HGSC samples (samples used in training and

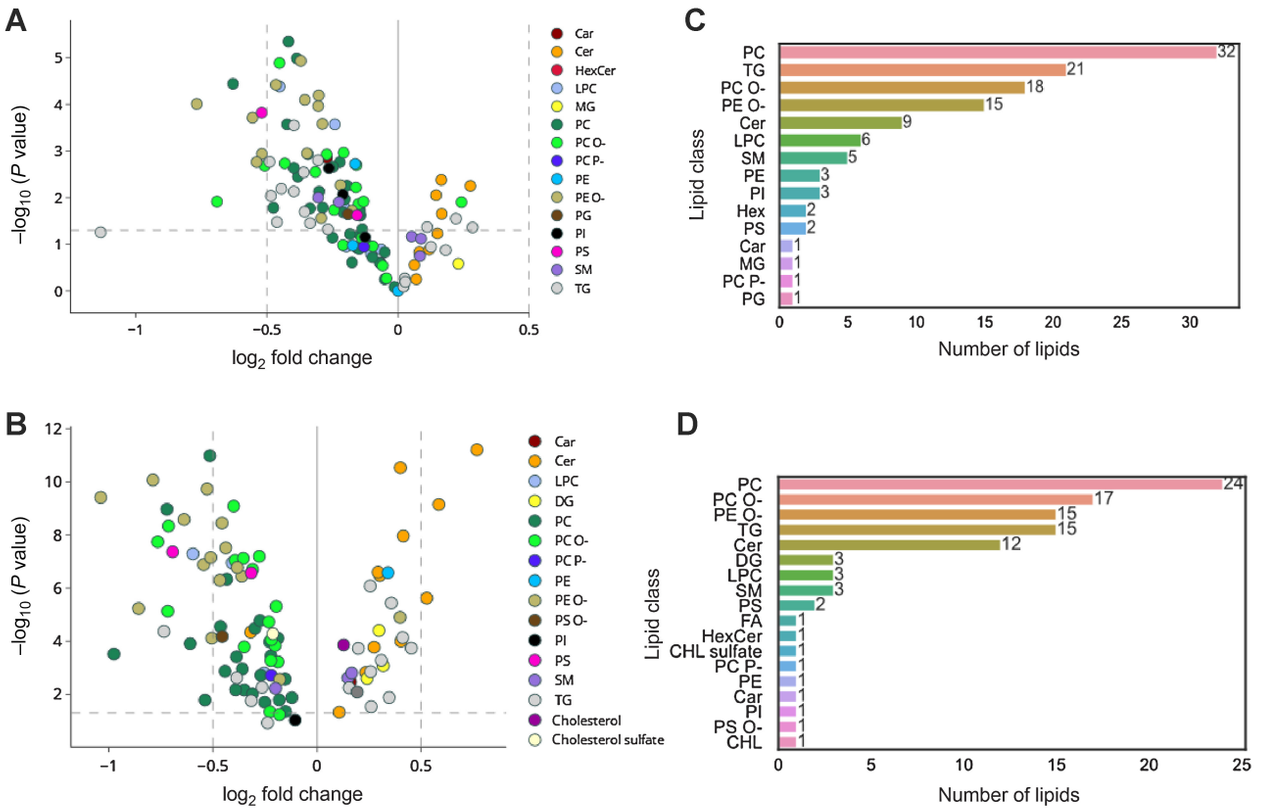

Figure 4.

Serum lipidome differences in early-stage (I and II) or advanced-stage (III and IV) patients with ovarian versus nonovarian cancers using only select lipids. Most relevant lipid species for differentiating early-stage ovarian cancer from nonovarian cancer, and advanced-stage ovarian cancer from nonovarian cancer were selected using random forests feature selection. Volcano plots showing lipidome differences with 120 best discriminating lipids selected for early-stage ovarian cancer versus nonovarian cancer (**A**) and 102 best discriminating lipids selected for advanced stage ovarian cancer versus nonovarian cancer samples (**B**). Lipid species are color-coded by lipid class, as indicated on the plots. Fold changes were calculated as the base 2 logarithm of the average lipid abundance ratios for ovarian cancer versus nonovarian cancer. A positive fold-change value indicates higher levels in ovarian cancer samples. Negative values indicate lower levels in ovarian cancer samples. *P* values were calculated using the Welch's *t* test. Bar graphs showing number of lipids selected, grouped by lipid class, for early-stage ovarian cancer versus nonovarian cancer (**C**) and advanced-stage ovarian cancer versus nonovarian cancer (**D**). TG, Triacylglycerols; PC, Phosphatidylcholines; PC O-, Ether phosphatidylcholines; SM, Sphingomyelins; LPC, Lysophosphatidylcholines; Cer, Ceramides; PE O-, Ether phosphatidylethanolamines; Car, Carnitines; HexCer, Hexosylceramides; PE, Phosphatidylethanolamines; DG, Diacylglycerols; FA, Fatty acids; PI, Phosphatidylinositols; CE, Cholesterol esters; LPE, Lysophosphatidylethanolamines; PS, Phosphatidylserines; PG, Phosphatidylglycerols; MG, Monoradylglycerols; LPE O-, Ether Lysophosphatidylethanolamines; PS O-, Ether phosphatidylserines; CHL, cholesterol.

test sets) versus nonovarian cancer samples ($n = 30$). Random forest classification model for early-stage versus nonovarian cancer samples gave the AUC value of 0.75 whereas the specificity value was only of 0.45 (Supplementary Table S11). Better specificity could be achieved by selecting a panel of lipids specifically for discriminating early-stage ovarian cancer from nonovarian cancer conditions. In addition, because the nonovarian cancer dataset consisted of invasive cervical cancer samples, it is not surprising to obtain low classification performance as we are discriminating early-stage ovarian cancer samples from other cancers. Improved classification performance was achieved for a sample set without cervical cancer samples (Supplementary Table S12). The 10-lipid biomarker panel discriminated early-stage ovarian cancer from benign conditions (without cervical cancer) with the AUC, sensitivity, and specificity values of 0.86, 0.81, and 0.69, respectively.

Comparison of the relative abundances of these 10 selected lipid markers between early-stage patients with ovarian cancer, late-stage ovarian cancer, and nonovarian cancer showed trends following the course of the disease (Supplementary Fig. S9). Ceramides showed an increasing trend whereas phospholipids decreased with the disease progression. Although we were not able to compare diseased patients against normal controls, the directionally concordant changes in lipid abundances suggest that the lipidome profile of women with gynecological malignancies can be monitored for improved clinical triage.

Serum ganglioside alterations with ovarian cancer

Previous studies have indicated increased gangliosides levels in ovarian cancer cell lines, tissues, serum, as well as ascites fluid from patients with ovarian cancer (22, 23), suggesting that gangliosides could play a role as ovarian cancer markers. However, existing data are still sparse regarding the importance of these glycolipids in ovarian cancer development (22). A total of 22 ganglioside species were detected in serum by LC/MS (Supplementary Table S6), and their structures confirmed by tandem MS experiments. Examination of the average ganglioside abundances for patients with early-stage ovarian cancer, advanced-stage ovarian cancer, and nonovarian cancer showed

Table 1. Discriminant (ovarian vs. nonovarian cancers) 10-lipid panel selected using a random forests algorithm.

Annotation	Adduct	Experimental *m/z*	Mass error (ppm)	MS Confidence level	Log_2 fold change		FDR corrected *P* value	
					Early OC/non-OC	Advanced OC/non-OC	Early OC/non-OC	Advanced OC/non-OC
Cer(d18:1_16:0)	$(M+H)^+$	538.5199	0.87	2	0.14	0.40	0.02	1E−09
Cer(d18:1_25:1)	$(M+H)^+$	662.6451	0.77	2	0.15	0.23	0.08	0.002
Cer(d18:2_23:0)	$(M+H)^+$	634.6138	0.78	2	0.12	0.30	0.19	1E−06
LPC(14:0)	$(M+H)^+$	468.3014	0.42	2	−0.45	−0.59	0.0007	3E−07
PC(O-36:5)	$(M+H)^+$	766.5756	1.39	2	−0.43	−0.76	0.007	1E−07
PC(40:7)	$(M+CH_3COOH-H)^-$	876.6123	−0.14	2	−0.51	−0.71	0.03	4E−08
PC(O-32:2)	$(M+CH_3COOH-H)^-$	716.5596	0.98	2	−0.15	−0.20	0.029	0.00026
PE(O-18:1_22:4)	$(M+H)^+$	780.5889	−1.66	2	−0.24	−0.38	−0.38	7E−07
PE(O-38:4)	$(M+H)^+$	754.5753	1.12	2	−0.46	−0.78	0.00071	2E−09
PS(18:0_20:4)	$(M+H)^+$	812.5432	−0.48	2	−0.15	−0.31	0.040	1E−8

Note: Proposed lipid annotation, main adduct type detected, experimental monoisotopic *m/z* value, mass error (ppm), MS annotation confidence level, *P* values, and abundance log-transformed fold changes are shown. MS annotation level was assigned on the basis of the following criteria: (i) MS1 and MS/MS spectrum of standard matched to the feature; (ii) MS1 and MS/MS spectrum of the feature matched with library spectra; and (iii) Tentative ID assignment based on elemental formula match with literature; (iv) unknowns.

an increase in overall ganglioside abundances concurrent with disease progression (Supplementary Fig. S10). However, the changes in ganglioside abundances between patients with nonovarian and ovarian cancers were not significant if tested univariately. Therefore, the diagnostic power of gangliosides was explored by combining them in panels. Classification models were built using all annotated gangliosides, which showed acceptable performance. The best random forest model provided AUC values for training and test sets of 0.78 and 0.68, respectively. To investigate whether gangliosides could further enhance the diagnostic power of the previously selected 10-lipid panel (Supplementary Table S11; **Table 1**), the top 7 discriminating gangliosides features in these models were added to the panel, resulting in a new 17-lipid panel (Supplementary Table S13). Improved AUC values were obtained for training and test sets with this set of lipids: The best AUC values for the training and test set were 0.88 and 0.83, respectively (Supplementary Table S14; **Table 2**). As with previous comparisons, the training set was balanced using SMOTE and was cross-validated 10-fold.

AutoML classification

In addition to traditional ML approaches, an AutoML method, Auto-sklearn (32), that automatically identifies optimal ML pipelines for a given task, was applied. The Auto-sklearn technique uses Bayesian optimization, meta-learning, and ensemble techniques to automatically select ML pipelines. Through Bayesian optimization, it assesses hyperparameter settings for optimal results, whereas meta-learning helps to learn from past experiences to enhance performance on new tasks. The framework also constructs ensembles by selecting models that minimize training data errors. Here, the AutoML technique (detailed in Materials and Methods section) provided improved AUC, accuracy and specificity for the test set using the 17-lipid biomarker panel. Classification metrics using auto-sklearn were 0.85, 0.78, 0.75, and 0.82 AUC, accuracy, sensitivity, and specificity, respectively (**Table 2**).

Serum lipidomics of normal controls versus patients with ovarian cancer

Blood samples from normal controls were collected for 150 women during regular checkups. Although these controls were excluded from the feature selection process, we investigated whether lipids selected for differentiating patients with ovarian cancer from nonovarian cancer could also distinguish normal controls. Clustering between ovarian cancer and normal controls was investigated using the full set of annotated lipids (Supplementary Fig. S11). Fold changes between ovarian cancer and normal controls for these 100 select lipids also showed lower serum lipidome abundances in patients with ovarian cancer (Supplementary Fig. S12). Ceramides and TG showed variable trends with most ceramides showing increased abundances in patients with ovarian cancer. Compared with our previous analysis of ovarian cancer versus nonovarian cancer conditions (**Fig. 3**), lipid alterations between ovarian cancer and normal controls were much more profound. Several lipid species showed significant differences: PC

Table 2. Machine learning performance of various lipid panels.

	10-Lipid panel training set *OC: n = 144, Non-OC: n = 144*	10-Lipid panel test set *OC: n = 64, Non-OC: n = 34*	17-Lipid panel training set *OC: n = 144, Non-OC: n = 144*	17-Lipid panel test set *OC: n = 64, Non-OC: n = 34*	AutoML 17-lipid panel test set *OC: n = 64, Non-OC: n = 34*
AUC	0.85 ± 0.07	0.82	0.88 ± 0.07	0.83	0.85
Accuracy	0.77 ± 0.09	0.78	0.81 ± 0.09	0.76	0.78
Sensitivity	0.78 ± 0.15	0.78	0.83 ± 0.09	0.75	0.75
Specificity	0.76 ± 0.12	0.76	0.79 ± 0.12	0.76	0.82

Note: Classification performance using a random forest classier and the AutoML model. The top 7 discriminant gangliosides were added to the 10-lipid panel to create the 17-lipid biomarker panel. Training data were balanced with the Synthetic Minority Oversampling Technique (SMOTE).

(18:0/22:6) and LPC(18:2) were reduced approximately 10- and 6-fold in patients with ovarian cancer, respectively. The classification performance of the selected 10-lipid maker panel was also evaluated for age-matched ovarian cancer versus normal controls (Supplementary Table S11). Because the dataset had a larger number of ovarian cancer samples than normal controls, SMOTE (28) was applied to balance the number of samples. Classification performance for the balanced set of patients with ovarian cancer versus normal controls was evaluated using five ML methods: Random forest, SVM, logistic regression, k-NN, and voting method. The best results were obtained with a random forest model. Performance characteristics of this model were 0.94, 0.88, 0.91, and 0.84 for the AUC, accuracy, sensitivity, and specificity, respectively.

Discussion

Accurate distinction between ovarian cancer and other benign or cancerous gynecological malignancies remains an unmet clinical challenge with significant impact on patient survival (7). Significantly improved survival rates are observed when women with ovarian cancer are referred to tertiary care specialists rather than general gynecologists, and although treatment of women with benign diseases by specialists imposes no harm, benign disease misdiagnosed for cancer causes unnecessary burden and patient anxiety (6, 33–36).

In recent years, dysregulation in lipid metabolism has been established as a crucial feature of ovarian cancer progression, reflecting the increased energy demands of highly proliferating cancer cells and cell membrane remodeling (37). However, previous ovarian cancer lipidomic studies have been limited by low numbers of early-stage patients, low lipidome coverage, and biases originating in the sample collection process (15, 16, 21). Although alteration in lipid metabolism is a characteristic feature of ovarian cancer, findings from different studies still show numerous inconsistencies (37). Previous data show racial/ethnic differences in risk, incidence, and survival of patients with ovarian cancer (38–40), yet most studies consist mainly of non-Hispanic white women with European ancestry. Compared with other racial groups, Asian women have a higher incidence rate of mucinous ovarian tumors (41) and it has been shown that the protein biomarker HE4 may be more useful for that cancer in this ethnic group (40). It is generally agreed that studies in specific racial/ethnic groups are needed to better understand ovarian cancer diversity and accelerate preventive and diagnostic strategies.

The study showcased here presents results for the comprehensive analysis of the serum lipidome of patients with ovarian cancer of Korean descent, recruited from two different clinical sites. A brief outline of the comparisons conducted in this study is presented in Supplementary Fig. S13. Among other findings, the comparison yielded a 17-lipid panel that distinguishes between patients with ovarian and nonovarian cancers (**Table 2**; Supplementary Fig. S14 and Supplementary Table S15). Furthermore, a consistent decrease in the overall lipid abundance across various ovarian cancer stages and histological types was observed in our dataset when compared against nonovarian cancer conditions. These results agree with previous studies showing an overall abundance decrease in serum glycerophospholipid abundances in patients with ovarian cancer of non-Hispanic white and European decent (15, 16, 21). This overall decrease in lipid abundance is likely due to altered lipoprotein level, as suggested in previous studies (42, 43). High-density lipoprotein (HDL) particles are rich in phospholipids; decreased HDL-cholesterol levels have been reported in patients with ovarian cancer (42). Furthermore, stage-stratified analysis comparing early- or advanced-stage patients with ovarian versus nonovarian cancers (**Fig. 4**; Supplementary Fig. S7) suggests that lipid alterations are detectable even in the early stages, amplifying as the disease progresses. These results further underscore the potential role of circulating lipids in monitoring ovarian cancer development and progression. Similar results were reported by Buas and colleagues (21) when comparing benign diseases with ovarian cancer (21) and recent studies suggest changes in circulating lipid abundances in pre-diagnostic specimens (44), indicating that lipid profiles exhibit changes years before ovarian cancer diagnosis. Mounting evidence suggests that monitoring lipid profiles could aid in identifying women who are at higher risk of developing ovarian cancer.

Although most lipids exhibited wide-scale abundance reductions in patients with ovarian cancer, some lipid classes, including Cer, some TG species, and DG, showed a consistent increase across the disease spectrum. Increase in blood ceramides levels have been frequently reported for ovarian cancer (16, 20, 45, 46), with significant increase in specific ceramides, including C16, C20, C22, and C24 derivatives. In our study, a consistent increase in almost all statistically significant ceramide species was observed for patients with ovarian cancer. Ceramides identified as differentially abundant for ovarian cancer included those with d18:0, d18:1, and d18:2 fatty acyl backbones with C16, C23, C24, C25, and C26 fatty acid chains. Previously, Niemi and colleagues (15) identified distinctive Cer alterations based on fatty acid chain length: Ceramide species with d16:1, d18:0, d18:1, and d18:2 backbones of 16:0, 18:0, 20:0, and 24:1 fatty acyl side chain were increased in patients with ovarian cancer, whereas those with 23:0 and 24:4 fatty acid side chain were decreased. In our study, however, no such trend was observed, and all key Cer species showed increased abundance in ovarian cancer. Although understanding the biological basis of these lipid alterations in ovarian cancer is an active area of investigation, literature suggests that increased Cer levels may be associated with the increased activity of ceramide synthases that promote tumor growth (47). The frequently observed dysregulation of ceramides in ovarian cancer along with their biological importance relative to tumor growth calls for a deeper understanding of ovarian cancer sphingolipid metabolism.

Along with increased Cer levels, reduced abundance of ether phospholipids such as PE O- and PC O- is also frequently reported in ovarian cancer studies (15). Here, we observed 74% of the detected PE O- species to be significantly decreased in patients with ovarian cancer, suggesting that PE O- are of particular importance in ovarian cancer pathogenesis. In particular, in agreement with previous studies, we observed decreased abundance of ether phospholipid species, including PE (O-18:1_22:4), PE(O-18:1_24:5; ref. 46), and PC(O-34:2; ref. 15). In addition, ratios between Cer and ether phospholipids abundances have frequently been reported as potential markers for ovarian cancer (46, 48), suggesting an important association between sphingolipid and ether phospholipid metabolism (49). Salminen and colleagues (48) identified the ratio between Cer(d18:1/18:0) and PC(O-38:4) as a prognostic marker for ovarian cancer. Four of the best 10 discriminating lipids were ether phospholipids and 3 were ceramides (**Table 1**), highlighting their diagnostic value and reinforcing the notion that a further understanding of the interplay between these lipid families is much needed in the context of ovarian cancer.

A recent study has suggested gangliosides GD2 and GD3 as potential diagnostics for ovarian cancer (22). Gangliosides are a class of glycolipids present in the plasma membranes of almost all vertebrate cells and are crucial in biological processes that include cell recognition and adhesion, transmembrane signaling, cell growth, and differentiation (50). Gangliosides are known to be involved in tumor–host immune system interactions (51), and studies have reported

pronounced local defects in ovarian cancer immune responses (52, 53). Gangliosides are shed into the extracellular environment, particularly during the malignant transformation of cells (54, 55), making them ideal biomarker candidates. Past studies have reported lower survival rates of patients with cancer with increased circulating ganglioside levels (23). Despite the potentially important role of gangliosides in ovarian cancer, however, reports on ganglioside expression in patients with ovarian cancer are quite sparse. In our study, we conducted in depth profiling of gangliosides, qualitatively observing an increased average abundance in patients with ovarian cancer compared with nonovarian cancer conditions. Abundance differences, however, were not statistically significant. A generally increasing trend in abundance with disease malignancy was observed, potentially reflecting increases in ganglioside secretion associated with tumor development (23). These results are in line with previous studies reporting significantly higher ganglioside levels in ovarian cancer cells compared with nonovarian cancer malignancies (23). In addition, our lipid panel results showed a moderate AUC improvement with the addition of 7 select gangliosides (Supplementary Table S14; **Table 2**). Further improvement was achieved with the use of the developed AutoML method (**Table 2**). Future quantitative studies with enhanced sensitivity and specificity for gangliosides are likely to facilitate biomarker discovery and solidify the development of therapeutic targets for ovarian cancer.

One salient characteristic of our study was the relatively large sample size for patients with ovarian cancer ($n = 208$), including 93 early-stage patients with ovarian cancer (stage I and II). Securing samples from early-stage patients with ovarian cancer is a major challenge and thus most studies reported only on a very limited number of early-stage cases. In addition, as mentioned previously, most studies are limited to samples from non-Hispanic white women. Here, a stage-stratified analysis for 93 early-stage ovarian cancer samples versus 117 nonovarian cancer samples selected a variety of lipid classes as differential lipids (**Fig. 4**; Supplementary Fig. S7). Among the selected lipids, ether glycerophospholipids exhibited significant changes in early disease stage emphasizing their potential for distinguishing ovarian cancer from other gynecological malignancies when the cancer is localized. Of all early-stage samples, the cohort included 23 early-stage patients with HGSC—the deadliest subtype of ovarian cancer (56). Only 13% of serous ovarian carcinomas are diagnosed when this cancer is localized, leading to poor prognosis and very low survival rates. Our lipidome analysis of early-stage patients of serous histology showed systematic alterations in lipid abundance based on lipid class. Although, most lipid alterations for early-stage serous and early-stage ovarian cancer samples of various histological types showed the same overall trend, an increased abundance of some specific PC species was observed only in the early-stage serous subtype (**Fig. 4**; Supplementary Fig. S8), indicating that some lipid alterations may be unique to HGSC. Alterations in the majority of lipid classes, however, followed a pattern similar to the other histological types. As indicated in Supplementary Fig. S5, as well as the exploratory PCA analysis of different ovarian cancer subtypes (Supplementary Fig. S3B), very subtle differences were observed between different ovarian cancer histological types. Meanwhile, lipidome analysis between HGSC versus nonovarian cancer conditions (Supplementary Fig. S6) and all ovarian cancer subtypes versus nonovarian cancer (**Figs. 3** and **4**) exhibited more prominent changes, suggesting that changes in lipid profile associated with ovarian cancer versus nonovarian cancer cases surpass most of the differences caused by the histological diversity of ovarian cancer. These findings underscore the daunting challenge of identifying ovarian cancer subtype-specific biomarkers. Our results are in line with previous metabolomics studies (15), but further analysis with larger samples of especially of early-stage patients with serous carcinoma will be required to confirm these observations as well as to identify makers unique to HGSC.

The dataset presented here also provides an opportunity to contrast lipid changes associated with ovarian cancer against cervical cancer. Data suggest that the ovarian cancer lipidome profile is distinguishable from that of cervical cancer, opening the possibility of developing screening tools to simultaneously test for both ovarian and cervical cancers. Because the testing for cervical cancer *via* Pap tests has been routinely performed for over 50 years, developing a lipid-based Pap tests screening strategy for both cervical and ovarian cancers could lead to improved ovarian cancer diagnosis and disease prognosis (13). In the future, we expect that further diagnostic gains could be achieved by integrating lipid/glycolipid markers with existing protein biomarkers such as CA125 and/or HE4 (21).

Our study had some limitations. We note that our study does not account for factors such as the use of oral contraceptives, postmenopausal hormone use, and body mass index. These factors may influence the lipid profile and introduce confounding effects in the dataset. In addition, our study does not examine the performance of our selected lipid markers against, or in conjunction with CA125 and HE4. As suggested previously (21), a combination of CA125 and circulating lipid markers may provide enhanced accuracy in distinguishing patients with ovarian cancer from nonovarian cancer. Further experiments using chemical standards and targeted metabolomics methods are also needed to validate our findings.

In conclusion, an in-depth metabolomic analysis of the serum lipidome of several benign and malignant gynecological diseases in Korean women revealed lipids unique to ovarian cancer. In comparison with benign ovarian and uterine tumors and invasive cervical cancer, the serum lipidome of ovarian cancer shows reduced abundance of most lipid classes, except for a few specific cases. Although ovarian cancer metabolomics still show some level of disagreement, especially regarding PC and LPC abundances, most studies, including our own, have consistently provided evidence of increased Cer levels. Alterations in specific ceramide species have frequently been reported, and thus a detailed biological understanding of sphingolipid metabolism in ovarian cancer warrants further efforts. Reductions in PE O- and PC O- have also been frequently reported, and thus the link between ether phospholipid metabolism and ovarian cancer calls for further investigation. Our data show the ovarian cancer lipid profile can not only be distinguished from benign conditions but also from other gynecological cancers, suggesting that future work focusing on identifying lipid markers unique to different gynecological malignancies may lead to lipid panels that can simultaneously test for multiple diseases. Developing clinical tests that combine lipid markers with already existing proteins markers, including CA125, may provide a route for enhanced diagnosis, detection, and accurate triage of women with gynecological malignancies.

Authors' Disclosures

S. Sah reports a patent for Methods and Markers for Ovarian Cancer Diagnosis and Stratification pending. O.O. Bifarin reports a patent for Methods and Marks of Ovarian Cancer Diagnosis and Stratification pending. D.A. Gaul reports a patent for 63/587,180 pending. J.-H. Kim reports grants from NIH R01CA218664 and U.S. Department of Defense W81XWH2210390 during the conduct of the study; as well as reports a US patent 63/587,180 pending. F.M. Fernández reports grants from NIH R01CA218664 and DoD W81XW2210390 during the conduct of the study; as well as reports a US patent 63/587,180 pending. No disclosures were reported by the other authors.

Authors' Contributions

S. Sah: Data curation, software, formal analysis, investigation, visualization, methodology, writing–original draft, writing–review and editing. **O.O. Bifarin:** Software, formal analysis, validation, investigation, visualization, methodology, writing–review and editing. **S.G. Moore:** Investigation. **D.A. Gaul:** Resources, investigation. **H. Chung:** Resources. **S.Y. Kwon:** Resources. **H. Cho:** Resources. **C.-H. Cho:** Resources. **J.-H. Kim:** Conceptualization, resources, supervision, funding acquisition, investigation. **J. Kim:** Conceptualization, resources, supervision, funding acquisition, investigation, writing–review and editing. **F.M. Fernández:** Conceptualization, resources, supervision, funding acquisition, writing–review and editing.

Acknowledgments

F.M. Fernandez and J. Kim acknowledge support from NIH 1R01CA218664–01 and Department of Defense (W81XWH2210390). J. Kim acknowledges the support from Indiana University Health–Indiana University School of Medicine Strategic Research Initiative. We acknowledge the Systems Mass Spectrometry Core at the Georgia Institute of Technology for UHPLC-MS analysis. The authors acknowledge support from NSF MRI CHE-1726528 grant for the acquisition of an ultra-high resolution Fourier transform ion cyclotron resonance (FTICR) mass spectrometer for the Georgia Institute of Technology core facilities.

Note

Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Received October 19, 2023; revised January 11, 2024; accepted February 23, 2024; published first February 27, 2024.

References

1. Kim J, Park EY, Kim O, Schilder JM, Coffey DM, Cho CH, et al. Cell origins of high-grade serous ovarian cancer. Cancers 2018;10:433.
2. Atallah GA, Abd Aziz NH, Teik CK, Shafiee MN, Kampan NC. New predictive biomarkers for ovarian cancer. Diagnostics 2021;11:465.
3. Montagnana M, Benati M, Danese E. Circulating biomarkers in epithelial ovarian cancer diagnosis: from present to future perspective. Ann Transl Med 2017;5:276.
4. Mercado C, Zingmond D, Karlan BY, Sekaris E, Gross J, Maggard-Gibbons M, et al. Quality of care in advanced ovarian cancer: the importance of provider specialty. Gynecol Oncol 2010;117:18–22.
5. Kobayashi E, Ueda Y, Matsuzaki S, Yokoyama T, Kimura T, Yoshino K, et al. Biomarkers for screening, diagnosis, and monitoring of ovarian cancer. Cancer Epidemiol Biomarkers Prev 2012;21:1902–12.
6. Moore RG, Miller MC, Disilvestro P, Landrum LM, Gajewski W, Ball JJ, et al. Evaluation of the diagnostic accuracy of the risk of ovarian malignancy algorithm in women with a pelvic mass. Obstet Gynecol 2011;118:280–8.
7. Goff BA, Miller JW, Matthews B, Trivers KF, Andrilla CH, Lishner DM, et al. Involvement of gynecologic oncologists in the treatment of patients with a suspicious ovarian mass. Obstet Gynecol 2011;118:854–62.
8. Bast RC Jr, Lu Z, Han CY, Lu KH, Anderson KS, Drescher CW, et al. Biomarkers and strategies for early detection of ovarian cancer. Cancer Epidemiol Biomarkers Prev 2020;29:2504–12.
9. Guo J, Yang WL, Pak D, Celestino J, Lu KH, Ning J, et al. Osteopontin, macrophage migration inhibitory factor and anti–interleukin-8 autoantibodies complement CA125 for detection of early-stage ovarian cancer. Cancers 2019;11:596.
10. Anderson KS, Cramer DW, Sibani S, Wallstrom G, Wong J, Park J, et al. Autoantibody signature for the serologic detection of ovarian cancer. J Proteome Res 2015;14:578–86.
11. Elias KM, Fendler W, Stawiski K, Fiascone SJ, Vitonis AF, Berkowitz RS, et al. Diagnostic potential for a serum miRNA neural network for detection of ovarian cancer. eLife 2017;6:e28932.
12. Bast RC Jr, Matulonis UA, Sood AK, Ahmed AA, Amobi AE, Balkwill FR, et al. Critical questions in ovarian cancer research and treatment: report of an American association for cancer research special conference. Cancer 2019;125:1963–72.
13. Boylan KLM, Afiuni-Zadeh S, Geller MA, Argenta PA, Griffin TJ, Skubitz APN. Evaluation of the potential of Pap test fluid and cervical swabs to serve as clinical diagnostic biospecimens for the detection of ovarian cancer by mass spectrometry-based proteomics. Clin Proteomics 2021;18:4.
14. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646–74.
15. Niemi RJ, Braicu EI, Kulbe H, Koistinen KM, Sehouli J, Puistola U, et al. Ovarian tumours of different histologic type and clinical stage induce similar changes in lipid metabolism. Br J Cancer 2018;119:847–54.
16. Braicu EI, Darb-Esfahani S, Schmitt WD, Koistinen KM, Heiskanen L, Pöhö P, et al. High-grade ovarian serous carcinoma patients exhibit profound alterations in lipid metabolism. Oncotarget 2017;8:102912–22.
17. Butler LM, Perone Y, Dehairs J, Lupien LE, de Laat V, Talebi A, et al. Lipids and cancer: emerging roles in pathogenesis, diagnosis and therapeutic intervention. Adv Drug Deliv Rev 2020;159:245–93.
18. Li J, Xie H, Li A, Cheng J, Yang K, Wang J, et al. Distinct plasma lipids profiles of recurrent ovarian cancer by liquid chromatography-mass spectrometry. Oncotarget 2017;8:46834–45.
19. Onwuka JU, Okekunle AP, Olutola OM, Akpa OM, Feng R. Lipid profile and risk of ovarian tumours: a meta-analysis. BMC Cancer 2020;20:200.
20. Fan L, Zhang W, Yin M, Zhang T, Wu X, Zhang H, et al. Identification of metabolic biomarkers to diagnose epithelial ovarian cancer using a UPLC/QTOF/MS platform. Acta Oncol 2012;51:473–9.
21. Buas MF, Drescher CW, Urban N, Li CI, Bettcher L, Hait NC, et al. Quantitative global lipidomics analysis of patients with ovarian cancer versus benign adnexal mass. Sci Rep 2021;11:18156.
22. Galan A, Papaluca A, Nejatie A, Matanes E, Brahimi F, Tong W, et al. GD2 and GD3 gangliosides as diagnostic biomarkers for all stages and subtypes of epithelial ovarian cancer. Front Oncol 2023;13:1134763.
23. Santin AD, Ravindranath MH, Bellone S, Muthugounder S, Palmieri M, O'Brien TJ, et al. Increased levels of gangliosides in the plasma and ascitic fluid of patients with advanced ovarian cancer. BJOG 2004;111:613–8.
24. Xie H, Hou Y, Cheng J, Openkova MS, Xia B, Wang W, et al. Metabolic profiling and novel plasma biomarkers for predicting survival in epithelial ovarian cancer. Oncotarget 2017;8:32134–46.
25. Zhang Y, Liu Y, Li L, Wei J, Xiong S, Zhao Z. High resolution mass spectrometry coupled with multivariate data analysis revealing plasma lipidomic alteration in ovarian cancer in Asian women. Talanta 2016;150:88–96.
26. Ahn H-S, Yeom J, Yu J, Kwon Y-I, Kim J-H, Kim K. Convergence of plasma metabolomics and proteomics analysis to discover signatures of high-grade serous ovarian cancer. Cancers 2020;12:3447.
27. Fan S, Kind T, Cajka T, Hazen SL, Tang WHW, Kaddurah-Daouk R, et al. Systematic error removal using random forest for normalizing large-scale untargeted lipidomics data. Anal Chem 2019;91:3590–6.
28. Chawla N, Bowyer K, Hall L, Kegelmeyer W. SMOTE: synthetic minority oversampling technique. J Artif Intell Res 2002;16:321–57.
29. Fahy E, Sud M, Cotter D, Subramaniam S. LIPID MAPS online tools for lipid research. Nucleic Acids Res 2007;35:W606–W12.
30. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, et al. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res 2018;46:D608–d17.
31. Lu Z, Chen J. [Introduction of WHO classification of tumours of female reproductive organs, fourth edition]. Zhonghua Bing Li Xue Za Zhi 2014;43:649–50.
32. Feurer M, Klein A, Eggensperger K, Springenberg JT, Blum M, Hutter F. Autosklearn: efficient and robust automated machine learning. In: Hutter F, Kotthoff L, Vanschoren J, editors. Automated machine learning: methods, systems, challenges. Cham, Switzerland: Springer International Publishing; 2019. p. 113–34.
33. Bristow RE, Tomacruz RS, Armstrong DK, Trimble EL, Montz FJ. Survival effect of maximal cytoreductive surgery for advanced ovarian carcinoma during the platinum era: a meta-analysis. J Clin Oncol 2002;20:1248–59.
34. Guidelines for referral to a gynecologic oncologist: rationale and benefits. The society of gynecologic oncologists. Gynecol Oncol 2000;78:S1–13.
35. ACOG committee opinion: number 280, December 2002. The role of the generalist obstetrician–gynecologist in the early detection of ovarian cancer. Obstet Gynecol 2002;100:1413–6.

36. Engelen MJ, Kos HE, Willemse PH, Aalders JG, de Vries EG, Schaapveld M, et al. Surgery by consultant gynecologic oncologists improves survival in patients with ovarian carcinoma. Cancer 2006;106:589–98.
37. Saorin A, Di Gregorio E, Miolo G, Steffan A, Corona G. Emerging role of metabolomics in ovarian cancer diagnosis. Metabolites 2020;10:419.
38. Torre LA, Trabert B, DeSantis CE, Miller KD, Samimi G, Runowicz CD, et al. Ovarian cancer statistics, 2018. CA Cancer J Clin 2018;68:284–96.
39. Peres LC, Schildkraut JM. Chapter One—Racial/ethnic disparities in ovarian cancer research. In: Ford ME, Esnaola NF, Salley JD, editors. Advances in Cancer Research. Volume 146. Cambridge, MA: Academic Press; 2020. p. 1–21.
40. Chan KK, Chen CA, Nam JH, Ochiai K, Wilailak S, Choon AT, et al. The use of HE4 in the prediction of ovarian cancer in Asian women with a pelvic mass. Gynecol Oncol 2013;128:239–44.
41. Tung KH, Goodman MT, Wu AH, McDuffie K, Wilkens LR, Kolonel LN, et al. Reproductive factors and epithelial ovarian cancer risk by histologic type: a multiethnic case–control study. Am J Epidemiol 2003;158:629–38.
42. Qadir MI, Malik SA. Plasma lipid profile in gynecologic cancers. Eur J Gynaecol Oncol 2008;29:158–61.
43. Kozak KR, Su F, Whitelegge JP, Faull K, Reddy S, Farias-Eisner R. Characterization of serum biomarkers for detection of early-stage ovarian cancer. Proteomics 2005;5:4589–96.
44. Hada M, Edin ML, Hartge P, Lih FB, Wentzensen N, Zeldin DC, et al. Prediagnostic serum levels of fatty acid metabolites and risk of ovarian cancer in the prostate, lung, colorectal, and ovarian (PLCO) cancer screening trial. Cancer Epidemiol Biomarkers Prev 2019;28:189–97.
45. Gaul DA, Mezencev R, Long TQ, Jones CM, Benigno BB, Gray A, et al. Highly-accurate metabolomic detection of early-stage ovarian cancer. Sci Rep 2015;5:16351.
46. Sah S, Ma X, Botros A, Gaul DA, Yun SR, Park EY, et al. Space- and time-resolved metabolomics of a high-grade serous ovarian cancer mouse model. Cancers 2022;14:2262.
47. Kozar N, Kruusmaa K, Bitenc M, Argamasilla R, Adsuar A, Goswami N, et al. Metabolomic profiling suggests long chain ceramides and sphingomyelins as a possible diagnostic biomarker of epithelial ovarian cancer. Clin Chim Acta 2018; 481:108–14.
48. Salminen L, Braicu EI, Lääperi M, Jylhä A, Oksa S, Hietanen S, et al. A novel two-lipid signature is a strong and independent prognostic factor in ovarian cancer. Cancers 2021;13:1764.
49. Jiménez-Rojo N, Leonetti MD, Zoni V, Colom A, Feng S, Iyengar NR, et al. Conserved functions of ether lipids and sphingolipids in the early secretory pathway. Curr Biol 2020;30:3775–87.
50. Hakomori S. Tumor malignancy defined by aberrant glycosylation and sphingo (glyco)lipid metabolism. Cancer Res 1996;56:5309–18.
51. Daniotti JL, Lardone RD, Vilcaes AA. Dysregulated expression of glycolipids in tumor cells: from negative modulator of anti-tumor immunity to promising targets for developing therapeutic agents. Front Oncol 2015;5: 300.
52. Feng W, Dean DC, Hornicek FJ, Shi H, Duan Z. Exosomes promote pre-metastatic niche formation in ovarian cancer. Mol Cancer 2019;18:124.
53. Merogi AJ, Marrogi AJ, Ramesh R, Robinson WR, Fermin CD, Freeman SM. Tumor-host interaction: analysis of cytokines, growth factors, and tumor-infiltrating lymphocytes in ovarian carcinomas. Hum Pathol 1997;28: 321–31.
54. Portoukalian J, David MJ, Shen X, Richard M, Dubreuil C. Tumor size-dependent elevations of serum gangliosides in patients with head and neck carcinomas. Biochem Int 1989;18:759–65.
55. Kong Y, Li R, Ladisch S. Natural forms of shed tumor gangliosides. Biochim Biophys Acta 1998;1394:43–56.
56. Kim J, Coffey DM, Ma L, Matzuk MM. The ovary is an alternative site of origin for high-grade serous ovarian cancer in mice. Endocrinology 2015; 156:1975–81.