

Coupled Mass-Spectrometry-Based Lipidomics Machine Learning Approach for Early Detection of Clear Cell Renal Cell Carcinoma

Malena Manzi, Martín Palazzo, María Elena Knott, Pierre Beausery, Patricio Yankilevich, María Isabel Giménez, and María Eugenia Monge*



Cite This: <https://dx.doi.org/10.1021/acs.jproteome.0c00663>



Read Online

ACCESS |



Metrics & More



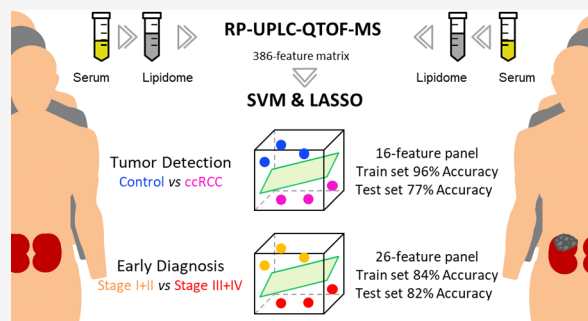
Article Recommendations



Supporting Information

ABSTRACT: A discovery-based lipid profiling study of serum samples from a cohort that included patients with clear cell renal cell carcinoma (ccRCC) stages I, II, III, and IV ($n = 112$) and controls ($n = 52$) was performed using ultraperformance liquid chromatography coupled to quadrupole-time-of-flight mass spectrometry and machine learning techniques. Multivariate models based on support vector machines and the LASSO variable selection method yielded two discriminant lipid panels for ccRCC detection and early diagnosis. A 16-lipid panel allowed discriminating ccRCC patients from controls with 95.7% accuracy in a training set under cross-validation and 77.1% accuracy in an independent test set. A second model trained to discriminate early (I and II) from late (III and IV) stage ccRCC yielded a panel of 26 compounds that classified stage I patients from an independent test set with 82.1% accuracy. Thirteen species, including cholic acid, undecylenic acid, lauric acid, LPC(16:0/0:0), and PC(18:2/18:2), identified with level 1 exhibited significantly lower levels in samples from ccRCC patients compared to controls. Moreover, 3α -hydroxy- 5α -androstane-17-one 3-sulfate, *cis*-5-dodecenoic acid, arachidonic acid, *cis*-13-docosenoic acid, PI(16:0/18:1), PC(16:0/18:2), and PC(O-16:0/20:4) contributed to discriminate early from late ccRCC stage patients. The results are auspicious for early ccRCC diagnosis after validation of the panels in larger and different cohorts.

KEYWORDS: lipidomics, clear cell renal cell carcinoma, biomarkers, ultraperformance liquid chromatography, mass spectrometry, machine learning, support vector machines, LASSO



INTRODUCTION

Renal cell carcinoma (RCC) is the main type of kidney cancer (80%)¹ and accounts for 5 and 3% of all oncological diagnoses in men and women, respectively.^{2,3} The RCC global incidence trend is continuously increasing³ and is one of the most lethal urological malignancies.⁴ More than 50% of cases are incidentally detected by imaging procedures, and 20% of the patients exhibit locally advanced or metastatic RCC at the time of diagnosis with a poor prognosis (5-year survival rate <35%).^{1,3} These tumors are inherently resistant to chemotherapy⁵ and radiotherapy,⁶ though if localized and detected at early stages, the disease is potentially curable by surgery.¹ Actually, the discovery of early detection biomarkers is the most promising approach to reduce RCC mortality.⁷

Clear cell RCC (ccRCC) is the most common RCC subtype, accounting for 75% of RCC cases,^{8,9} and it is classified into stages I, II, III, and IV, the latter being often a metastatic cancer.¹ At a molecular level, the four most frequently mutated genes in ccRCC are VHL (50–80%), PBRM1 (29–46%), BAP1 (6–19%), and SETD2 (8–30%),¹⁰ leading to an extremely heterogeneous disease that cannot be diagnosed based only on genetic information.^{11,12} A gene expression

profile has shown that ccRCC behaves metabolically different with respect to other tumor types, with most metabolic genes being downregulated.¹³ Moreover, discrepancies were observed between results from transcriptomics and metabolic profiling of tissue samples,¹⁴ highlighting that multiple pathways contribute to ccRCC progression.^{14,15} Therefore, a combination of multiple biomarkers may provide larger sensitivity and specificity for clinical diagnostic purposes than a single biochemical marker.

Cellular metabolism reprogramming is an accepted hallmark of cancer disease.¹⁶ The metabolome integrates the responses to internal and external factors,¹⁷ revealing an accurate readout for the study of cancer development and progression. Over the last years, many efforts aimed at discovering RCC biomarkers, principally focused on RCC detection.¹⁸ Metabolic profiling

Received: August 26, 2020

Table 1. Patient Cohort Used for the Study^a

sample class	samples (<i>n</i>)	age (mean ± SD)	gender (<i>n</i> , %)		hospital (<i>n</i> , %)	
			male	female	IAOHR	HIBA
healthy individuals (controls)	52	58 ± 10	35 (67%)	17 (33%)	33 (63.5%)	19 (36.5%)
ccRCC patients	112	60 ± 9	77 (69%)	35 (31%)	28 (25%)	84 (75%)
SI	56	58 ± 10	35 (62.5%)	21 (37.5%)	5 (9%)	51 (91%)
SII	14	59 ± 8	12 (86%)	2 (14%)	5 (36%)	9 (64%)
SIII	18	64 ± 9	12 (67%)	6 (33%)	6 (33%)	12 (67%)
SIV	24	62 ± 8	18 (75%)	6 (25%)	12 (50%)	12 (50%)
chrRCC patients	16	53 ± 14	10 (63%)	6 (38%)	1 (6%)	15 (94%)
papRCC patients	8	61 ± 9	6 (75%)	2 (25%)	2 (25%)	6 (75%)

^a*n*: number of samples; SD: standard deviation; IAOHR: Instituto de Oncología A. H. Roffo; HIBA: Hospital Italiano de Buenos Aires; ccRCC: clear cell renal cell carcinoma; S: stage; chrRCC: chromophobe-RCC; papRCC: papillary-RCC.

studies applying mass spectrometry (MS) and/or nuclear magnetic resonance (NMR) spectroscopy have been mainly conducted to interrogate urine^{19–29} and tissue^{12,14,30–37} samples. Blood analysis offers a fast readout of the body health status, its collection procedures are less invasive than biopsies, and data normalization strategies are less controversial for blood than urine.³⁸ However, only a few studies have focused on the analysis of serum^{39–41} or plasma^{42–44} samples.¹⁸ Currently, there is a lack of robust tissue and/or biofluid-based biomarkers for ccRCC early diagnosis in the clinic.^{18,45,46}

Lipids are the main components of biological membranes and exert many essential biological roles, such as cellular signaling, energy storage, and cell–cell interactions. In addition, lipids are involved in biological processes like cell survival, proliferation, and cell death, acting as mediators of several oncogenic processes.⁴⁷ Therefore, the dysregulation of lipid metabolism and its association with cancer progression have prompted many lipidomics studies for biomarker discovery.⁴⁸ Kidney cancer is considered to be a metabolic disease.⁴⁹ In particular, ccRCC is characterized for being lipid- and glycogen-laden, exhibiting abnormal cholesterol metabolism and fat storage.^{50,51} Several studies have reported altered lipid levels, though lipid profiling for RCC diagnosis has been mainly conducted on tissue samples.^{18,48}

Considering that lipids are promising candidate molecules to be explored in an untargeted fashion for ccRCC detection, we hypothesize that lipid panels may overcome single biomarker limitations for ccRCC diagnosis. Therefore, in this work, we have profiled serum samples from ccRCC patients with different stages of the disease and healthy individuals using a discovery-based lipidomics approach based on ultraperformance liquid chromatography coupled to quadrupole-time-of-flight mass spectrometry (UPLC-QTOF-MS) in combination with machine learning techniques. Multivariate statistical analysis was conducted on a 386-feature matrix using the least absolute shrinkage and selection operator (LASSO)⁵² variable selection method coupled with support vector machines (SVMs).⁵³ The strategy yielded a reduced lipid panel that allowed discriminating ccRCC patients from healthy individuals with high accuracy in training and independent test sets. A second multivariate model trained to discriminate early-stage (I and II) from late-stage (III and IV) ccRCC patients provided another panel that allowed sample classification with high accuracy in the training set and in an independent test set composed of stage I ccRCC patients. Lipids in the discriminant panels were identified with confidence levels 1 and 2 according to the metabolomics standards initiative (MSI).^{54,55} In

addition, discrimination of patients with chromophobe- and papillary-RCC from ccRCC patients and healthy individuals was evaluated.

MATERIALS AND METHODS

Chemicals

Ultrapure water with 18.2 MΩ·cm resistivity (Thermo Scientific Barnstead Micropure UV ultrapure water system, Sunnyvale, CA); Optima liquid chromatography-mass spectrometry (LC-MS)-grade acetonitrile, methanol, and isopropanol, analytical grade acetic acid; and LC-MS-grade formic acid and ammonium acetate (Fisher Chemical, Raleigh, NC) were used to prepare chromatographic mobile phases, samples, and chemical standard solutions. Leucine enkephalin was purchased from Waters Corp. (Milford, MA). Sodium hydroxide was purchased from EMSURE[®]ISO (Merck Millipore, Burlington, MA). Commercial serum from the North American population (H4522), cholic acid, cortisol, lactic acid, lauric acid (C12:0), linoleic acid (C18:2), undecylenic acid (C11:1), PI(16:0/18:1), PC(16:0/18:1), PC(18:1/18:1), PE(18:1/18:1), SM(18:1/16:0), and lysophosphatidylcholine (LPC) (USP reference standard) were purchased from Sigma-Aldrich (St. Luis). *cis*-5-Dodecenoic acid (C12:1), arachidonic acid (C20:4), MG(0:0/16:0/0:0), dehydroepiandrosterone sulfate (sodium salt), 2-hydroxymyristic acid, *cis*-13-docosenoic acid (C22:1), pentacosanoic acid (C25:0), PC(16:0/16:0), PC(18:2/18:2), PC(16:0/18:2), and PC(O-16:0/20:4) were purchased from Cayman Chemical (MI). Decanoic acid (C10:0), undecanoic acid (C11:0), heptadecanoic acid (C17:0), octadecanoic acid (C18:0), eicosanoic acid (C20:0), and docosanoic acid (C22:0) were purchased from Loradan (Solna, Sweden).

Serum Samples

Serum samples were provided by the public oncologic serum biobank “Biobanco Público de Muestras Séricas Oncológicas” (BPMSO) from “Instituto de Oncología A. H. Roffo” (IAOHR). Samples were drawn from healthy individuals and patients from IAOHR and from “Hospital Italiano de Buenos Aires” (HIBA), Argentina, following the guidelines approved by the IAOHR Institutional Review Board, and after the signature of the corresponding informed consent. The cohort consisted of 52 healthy individuals (controls) and 112 ccRCC patients with stage (S) I, II, III, or IV (Table S1). Additional samples were collected from patients with chromophobe-RCC (chrRCC, *n* = 16) and papillary-RCC (papRCC, *n* = 8). The male/female ratio in samples was approximately 2:1 (Table S1), following the worldwide incidence trend.⁴ The age

population was 58 ± 10 years (mean \pm standard deviation (SD)) in controls, 60 ± 9 years in ccRCC patients, 53 ± 14 years in chRCC patients, and 61 ± 9 years in papRCC patients. The means of the age populations for the binary comparisons were not significantly different using the Mann–Whitney *U* test at the 0.05 level (Table 1 and Figure S1). Blood samples were drawn before surgery from untreated patients and from healthy individuals with 8 h of fasting. Samples from all subjects were subjected to a serological analysis to discard those with a positive result for HIV and hepatitis B and C. According to the BPMSO standard operating procedure, 20 mL of blood was collected in tubes without any anticoagulant and left 15 min at 25 °C to allow the clot formation and centrifuged at 600g for 10 min. Serum samples were subsequently split into aliquots and stored at -80 °C. Aliquots were used only once after thawing.

Sample Preparation

Frozen serum samples were thawed at 0 °C on a water-ice bath. Cold (4 °C) isopropanol was added to serum samples in a 3:1 solvent/sample volume ratio for protein precipitation and lipid extraction.^{56,57} Samples were subsequently vortex-mixed for 10 s and centrifuged at 16 000g for 20 min at 4 °C. Different types of quality control (QC) samples were prepared and jointly analyzed with the study samples.⁵⁸ Ten intrastudy QC samples (QC_{ss}) consisting of pools of equal volume aliquots of subsets of the study samples were generated and followed protein precipitation and further sample preparation steps. Each subset included randomly selected serum samples from controls and patients with different RCC stages and histology. A process blank QC sample (QC_{blank}) was prepared by replacing serum with ultrapure water. Commercial serum from the male North American population was used as inter and intralab QC samples, and as a long-term reference sample. Supernatants from the study samples and QC samples were randomly lyophilized in 10 batches at -80 °C and 50 mTorr for 48 h using a Telstar LYOQuest-85 freeze dryer (Telstar, Madrid, Spain). After lyophilization, samples were stored at -80 °C until analysis. Dried samples were reconstituted with a cold mixture solution of water/methanol/acetonitrile/isopropanol (40:10:25:25 v/v) to provide similar initial UPLC gradient conditions. After reconstitution, samples were further vortex-mixed and centrifuged at 16 000g for 10 min at 4 °C. Supernatants were subsequently transferred to LC vials for analysis. A second type of intrastudy QC sample that consisted of pooled reconstituted serum samples from an independent set of healthy individuals ($n = 5$) from the same population under study (QC_{HI}) was included for analysis. The QC_{HI} sample was split into several aliquots and stored at -80 °C until analysis. A chemical standard mixture that included saturated fatty acids (FAs) (C10:0, C11:0, C12:0, C17:0, C18:0, C20:0, C22:0) and different phospholipids (PLs) (PC(16:0/16:0), PC(16:0/18:1), PC(18:1/18:1), PE(18:1/18:1), SM(18:1/16:0) and lysophosphatidylcholine (LPC16:0 and LPC18:0)) dissolved in water/methanol/acetonitrile/isopropanol (40:10:25:25 v/v) was also analyzed to evaluate adduct type formation, retention time, and fragmentation patterns in tandem MS experiments and, therefore, to aid assessing the lipid identification process.

Ultraperformance Liquid Chromatography-Mass Spectrometry

Ultraperformance liquid chromatography-mass spectrometry (UPLC-MS) lipid profiling analyses were performed on a

Waters ACQUITY UPLC I-Class system fitted with a Waters ACQUITY UPLC BEH C₁₈ VanGuard precolumn (2.1 mm \times 5 mm, 1.7 μ m particle size) connected directly to the inlet of a Waters ACQUITY UPLC BEH C₁₈ column (2.1 mm \times 100 mm, 1.7 μ m particle size, Waters Corporation, Milford, MA) and coupled to a Xevo G2S QTOF mass spectrometer (Waters Corporation, Manchester, U.K.) with an electrospray ionization (ESI) source. The typical resolving power and mass accuracy of the Xevo G2S QTOF mass spectrometer were 32 000 full width at half-maximum (FWHM) and 0.2 ppm at m/z 554.2615, respectively. The chromatographic separation method for lipid analysis was adapted from Sarafian et al.⁵⁶ Gradient elution was conducted with acetonitrile/water (60:40 v/v) mixed with 10 mM ammonium acetate and 0.1% acetic acid (mobile phase A) and isopropanol/acetonitrile (90:10 v/v) mixed with 10 mM ammonium acetate and 0.1% acetic acid (mobile phase B), with the following program: 0–2 min 40–43% B; 2–2.1 min 43–50% B; 2.1–12.0 min 50–54% B; 12.0–12.1 min 54–70% B; 12.1–18.0 min 70–99% B; 18.0–19.0 min 99% B. The flow rate was constant at 0.4 mL min⁻¹ for 12 min and was decreased to 0.30 mL min⁻¹ between 12 and 19 min. Subsequently, the gradient was returned to its initial conditions in 11 min. The injection volume was 2 μ L. The temperature of the column and the autosampler tray was set to 55 and 5 °C, respectively. The mass spectrometer was operated in negative ion mode with a probe capillary voltage of 2.3 kV and a sampling cone voltage of 30 V. The source temperature was set to 120 °C, and the desolvation gas temperature was set to 300 °C. The nitrogen gas desolvation flow rate and the cone desolvation flow rate were 600 and 10 L h⁻¹, respectively. Every day, a 0.5 mM sodium formate solution prepared in isopropanol/water (90:10 v/v) was used for calibrating the mass spectrometer across the range of m/z 50–1200. A reference solution of leucine enkephalin (m/z 554.2615) was infused at 2 μ L min⁻¹ every 45 s and was used for data correction during acquisition. Data were acquired in the range of m/z 50–1200 in MS^E continuum mode,⁵⁹ and the scan time was set to 0.5 s. The UPLC-QTOF-MS^E data are available under the data set identifier MTBLS1839 in MetaboLights public repository⁶⁰ (www.ebi.ac.uk/metabolights/MTBLS1839). Base peak intensity (BPI) chromatograms obtained for serum samples of a healthy individual and SI-SIV-ccRCC patients are shown in Figure S2.

For UPLC-MS/MS experiments, the product ion mass spectra were recorded within the range of m/z 30–1200 with collision cell voltages between 10 and 40 V, depending on the analyte. Ultrahigh-purity argon ($\geq 99.999\%$) was used as the collision gas. Data acquisition and processing were carried out using MassLynx version 4.1 (Waters Corp., Milford, MA). Additional experiments were conducted in positive ion mode to assist in feature identification, using a probe capillary voltage of 2.5 kV.

To further assist in compound identification, a hydrophilic interaction liquid chromatography (HILIC) QTOF-MS-based method was used with a Waters ACQUITY UPLC BEH HILIC column (2.1 mm \times 100 mm, 1.7 μ m particle size, Waters Corporation, Milford, MA). Gradient elution was conducted with a 10 mM ammonium acetate solution (pH 5.8) with 5% of acetonitrile (mobile phase A) and acetonitrile (mobile phase B) using the following program: 0–10 min 95–65% B; 10–12 min 65–35% B; 12–16 min 35% B. The flow rate was constant at 0.3 mL min⁻¹ for 12 min and was increased to 0.4 mL min⁻¹ for 4 min. Subsequently, the

gradient was returned to its initial conditions in 12 min. The injection volume was 2 μ L. The temperature of the column and the autosampler tray was set to 35 and 5 $^{\circ}$ C, respectively. The mass spectrometer was operated with the same parameters as described above for the reverse-phase chromatographic method.

Data Acquisition

Quality assurance procedures were followed throughout the analytical workflow. System suitability procedures were performed before data acquisition of the study samples to determine that the method and associated instrumentation were fully functioning. The intrastudy QC_{HI} sample was injected 26 times for conditioning the UPLC-MS system before sample analysis. Table S2 shows a typical analysis order applied in this study. Samples were randomly analyzed in 10 batches by UPLC-QTOF-MS together with a system suitability blank (reconstitution solvent) and QC samples along 10 days. To ensure data quality, the coefficients of variation were daily analyzed for selected features in the intralab QC and intrastudy QC_{HI} samples to ensure that peak area values were below 20%⁶¹ within every batch and the accumulated batches. Moreover, unsupervised principal component analysis (PCA) was performed daily to inspect data quality before analyzing a new batch. These strategies allowed evaluation of the instrument performance and method reproducibility.

Feature Matrix Extraction, Data Preprocessing, and Curation

UPLC-MS spectral features (retention time (R_t), m/z pairs) were extracted using the Progenesis Q1 software package version 2.1 (Nonlinear Dynamics, Waters Corp., Milford, MA). The different steps involved in this procedure included retention time alignment, peak picking, deisotoping, integration, and deconvolution by grouping together adduct ions derived from the same compound. The following adduct ions were considered for deconvolution: $[M - H]^-$, $[2M - H]^-$, $[3M - H]^-$, $[M + Cl]^-$, $[M + Na - 2H]^-$, $[M + K - 2H]^-$, $[M + HCOO]^-$, $[M + CH_3COO]^-$, and $[M - H_2O - H]^-$. Neutral masses provided by Progenesis Q1 as a result of deconvolution were manually reviewed based on the isotopic patterns and the mass difference between detected ions. When incorrect assignments were identified, the m/z values associated with the most intense ionic species from the group of adducts originally informed by the Progenesis Q1 software were retained. Different filters were applied for data cleaning. Blank deduction was applied to features with a peak area of at least 3-fold or more in samples than in process blanks; otherwise, the area was set to 0.⁶² After process blank subtraction, only features present in at least 50% of one class (controls and SI-, SII-, SIII-, and SIV-ccRCC patients) were retained. Only features having monoisotopic peaks with signal intensity larger than 10^3 in the continuum mass spectra, exhibiting a mass difference lower than 10 mDa compared to the intrastudy QC_{ss} sample, and exhibiting the isotopic pattern were retained. The matrix obtained after this procedure, consisting of 386 features, was normalized to the total peak area for each sample (Data Set S1). This matrix was used to develop and train supervised multivariate models based on SVM and the LASSO⁵² variable selection method. The data were autoscaled before statistical analysis. A PCA model was built with the 386-feature matrix to evaluate data quality and confirm intrastudy QC_{ss} sample clustering (Figure S3).

Data Processing, Classification, and Prediction

Two classification problems were considered for training multivariate models based on (i) ccRCC detection and (ii) early ccRCC diagnosis. The sample cohort was split into different training, validation, and test sets for each binary comparison (i.e., ccRCC patients versus controls, and early versus late ccRCC stages) (Figure 1). Training and validation,

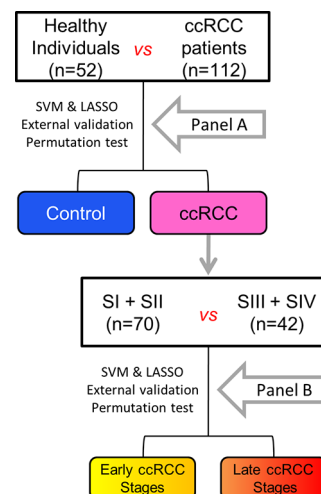


Figure 1. Scheme illustrating the two-model strategy for ccRCC detection and discrimination of early stages (I and II) from late stages (III and IV) of the disease.

sets were used for model building, and the test set was used to evaluate the performance of the models with independent samples. For the first classification problem, the data set consisted of 164 samples categorized as control (healthy individuals) or tumor (ccRCC patients including all four disease stages). Samples were randomly split into three sets; 48.8% were used for training, 8.5% for validation, and 42.7% for testing the models (Table 2A and Figure S4A). The

Table 2. Number of Samples Used for Training, Validating, and Testing the Multivariate SVM and LASSO Models for the Binary Comparison of (A) ccRCC Patients versus Controls and (B) Early-Stage (I and II) versus Late-Stage (III and IV) ccRCC Patients^a

(A)						
set	total (<i>n</i>)	ccRCC				control
		SI	SII	SIII	SIV	
training	80	10	10	10	10	40
validation	14	2	2	2	2	6
test	70	44	2	6	12	6
total	164	56	14	18	24	52
(B)						
set	total (<i>n</i>)	early stages		late stages		
		SI	SII	SIII	SIV	
training	70	24	11	15	20	
validation	14	4	3	3	4	
test	28	28	0	0	0	
total	112	56	14	18	24	

^aTraining and validation sets were balanced between classes.

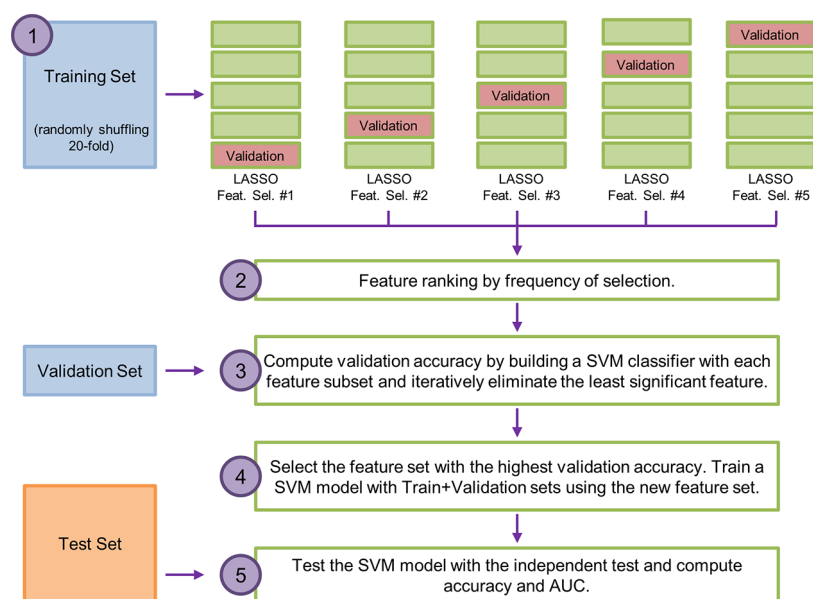


Figure 2. Machine learning pipeline designed for feature selection based on support vector machines (SVMs) coupled with the least absolute shrinkage and selection operator (LASSO) variable selection method. The pipeline consisted of three blocks: (i) features were ranked using the training set, which was 20-fold shuffled (steps 1 and 2), (ii) ranked features were selected with the validation set, and the model was trained on the selected feature panel with the training + validation sets (steps 3 and 4); and (iii) classification and model testing were accomplished with the test set (step 5).

training set included 80 samples and was balanced using the same number of age-matched samples from ccRCC patients and controls. The validation set included 14 samples, of which 42.9% corresponded to control and 57.1% corresponded to tumor, including equal contribution from each ccRCC stage. The test set comprised 70 samples. This design was aimed at having a large enough number of samples for training and validating the models but also for evaluating the model performance with a large independent test set that was not used to build the models. The second classification problem involved discriminating serum samples of patients with early ccRCC stages (SI and SII) from those with late ccRCC stages (SIII and SIV). In this case, the data set consisted of 112 samples that were randomly split in 62.5% for training, 12.5% for validation, and 25.0% for testing the models (Table 2B and Figure S4B). The training and validation sets were balanced according to both classes,⁴⁶ whereas the test set consisted of samples from SI ccRCC patients. Feature autoscaling was performed with the ensemble of the training and validation sets using their mean and standard deviation. The independent test set was autoscaled using the training and validation mean and standard deviation. A Mann–Whitney U test was applied to evaluate the age distribution between classes for the training and validation sets under the null hypothesis that age mean between classes was the same. A resulting p -value >0.05 between classes suggested that there was not enough evidence to reject the null hypothesis.

The machine learning pipeline was divided in three blocks (Figure 2). First, features were ranked using the training set (steps 1 and 2 of Figure 2). Subsequently, a group of ranked features was selected with the validation set (step 3 of Figure 2), and finally, classification and model testing was accomplished with the test set (steps 4 and 5 of Figure 2). The training set was shuffled 20 times, and at each shuffle, samples were fivefold split using 4/5 to select a reduced feature subset by a LASSO method and the remaining 1/5 to validate

the selected features. After fivefold feature selection turns nested within 20 random shuffle iterations, 100 feature subsets were independently obtained. Subsequently, a feature ranking was built based on the selection frequency within the 100 subsets obtained by the LASSO method.⁵²

A variable feature subset, which decreased the number of features for each iteration, was used to build an SVM⁵³ classifier with the training set. The classifier was applied to the validation set while iteratively removing the least important feature until only one remained. Validation accuracy was computed at each iteration. The selected subset consisted of the minimum number of features that maximized the validation accuracy within the 30 most frequently selected features. Once the feature subset was defined, an SVM classifier was built with fivefold cross-validation using the training and validation sets, and the resulting model was tested with the independent test set. The computed performance metrics were accuracy, sensitivity, specificity and the area under the receiver operating characteristic (ROC) curve. In addition, permutation tests were implemented as a complementary way to evaluate possible model overfitting. For that purpose, 1000 iterations were performed considering all of the samples.

The support vector classifier builds a nonlinear decision function (eq 1) for a binary classification problem using Kernel functions by constructing a hyperplane (eq 2) as follows

$$D(x) = \text{sign}[x^T \beta + \beta_0] \quad (1)$$

$$x: f(x) = x^T \beta + \beta_0 \quad (2)$$

where β and β_0 are the hyperplane feature coefficients and bias, respectively. The goal of the binary classification is to estimate a function $f: R \rightarrow \{+1, -1\}$ from the training data samples x_i with label Y_i . More details about the decision function parameters are described in the Materials and Methods section of the Supporting Information.

A one-class SVM classifier was also developed to classify samples from chrRCC and papRCC patients despite the limitation associated with the sample cohort size of these RCC histological subtypes. The one-class SVM model was trained with a selected class instead of two classes; thus, the classification approach determined whether a sample belonged to that class or not.

All machine learning steps including data preprocessing, model building, feature selection, classification, and independent testing were performed with the Scikit Learn Machine Learning library in Python 3.6.

Data were inspected through PCA before and after variable selection (i.e., on the curated spectral feature matrix and on the discriminant feature panel). Fold changes for discriminant features were calculated as the ratio of median peak areas between compared classes. Mann–Whitney *U* tests were used to calculate statistical significance, and *p*-values within each panel were corrected using the Benjamini–Hochberg⁶³ procedure for multiple comparisons with a false discovery rate (FDR) of 0.1. Pairs of features that exhibited a Spearman's correlation coefficient larger than 0.7 were considered highly correlated.

Identification Procedure of Discriminant Features

Lipid identification was attempted for all of the discriminant features. Elemental formulas were generated based on accurate masses and isotopic patterns and were used to assess spectral adduct ions. Lipid identification was further confirmed with tandem MS experiments and validated through retention time and fragmentation pattern matching with chemical standards when possible. Neutral masses were searched in the LIPID MAPS⁶⁴ database, and elemental formulas were searched in the Human Metabolome Database (HMDB).⁶⁵ For some cases, neutral masses or *m/z* values were also searched in the HMDB and the METLIN⁶⁶ database. The mass window tolerance was set at 10 mDa. Experimental tandem MS spectra were compared to the aforementioned databases when the spectra were available, and fragmentation patterns were manually interpreted in all cases. Chemical standard solutions were prepared in water/methanol/acetonitrile/isopropanol (40:10:25:25 v/v) and were analyzed under identical conditions as serum samples to validate putative lipid identification. Different collision cell voltages were used to assist in fragmentation pattern interpretation. Two different quadrupole-mass windows of 6 and 1 Da were used in tandem MS experiments to address spectral overlap due to quadrupole co-selection. The narrower mass window provided higher selectivity at the expense of lower sensitivity. When the precursor ion intensity was too low to conduct tandem MS experiments or when quadrupole co-selection could not be mitigated, tentative identification was not informed. To avoid misidentification, tandem MS experiments were conducted not only for the discriminant feature but also for the different adduct ions corresponding to the same compound both in negative and positive ion modes. Diagnostic and characteristic fragments and neutral losses, as well as the position of substituents in phospholipids (PLs), were assigned based on the literature.^{67–69} The terminology used for lipids and fragment ion annotation follows the lipidomics standards initiative (LSI).^{70–72}

Spiking experiments were also conducted with the standards on serum samples to address retention time differences caused by serum matrix effects. When endogenous compound levels

were too high, samples were diluted 10-fold before spiking with chemical standards. Moreover, for putatively annotated compounds that did not interact with the stationary phase of the C₁₈ column and exhibited a retention time close to the dead time, complementary experiments were conducted using HILIC-QTOF-MS. Serum samples, spiked serum samples with chemical standards, and chemical standard solutions were analyzed under the same conditions to assess retention time and mass spectra matching. Confidence levels for metabolite identification are reported based on the established criteria suggested by the MSI.^{54,55}

RESULTS AND DISCUSSION

Lipid Profiling and Serum Sample Classification

Serum samples from ccRCC patients (*n* = 112) and controls (*n* = 52) were analyzed by UPLC-QTOF-MS to identify discriminant lipid panels that would allow ccRCC detection and early diagnosis. A total of 2962 features (*R*_{*v*}, *m/z* pairs) were extracted from the UPLC-MS negative ion mode data. This matrix was reduced to 2260 features after deconvolution, and it was subjected to consecutive filters to retain only robust-enough features for data analysis and interpretation.^{73,74} Features with the highest mean abundance in the system suitability blank or process blank were eliminated. Following process blank deduction, only features that were present in at least 50% of one sample class were retained. This data set consisted of 1773 features that were manually analyzed to retain only those with signal intensity, peak shape, and mass variance within the established thresholds and that exhibited the isotopic pattern to facilitate identification. This final matrix consisting of 386 spectral features was subsequently normalized (Data Set S1) and used to build unsupervised and supervised multivariate statistical models for sample classification and prediction. Figure 3A,B shows the PCA score plots for samples corresponding to controls and ccRCC patients (stages I–IV) and samples from patients with early and late ccRCC stages, respectively. No outliers were detected by this analysis, and no sample clustering was visualized for any of the binary classification problems using the 386-feature matrix (Figure 3A,B). Therefore, sample classification and prediction were further attempted for each classification problem based on supervised SVM models coupled with the LASSO feature selection method.

The first model was built and trained to discriminate tumors from control samples, and the second model was trained to distinguish early- from late-stage ccRCC samples. The SVM decision function was executed 10 times using different sample subsets for training and testing the models, to generate 10 different models for each binary comparison (Tables S3 and S4). The best models for ccRCC detection and early ccRCC diagnosis were selected based on (i) the classification accuracy, sensitivity, and specificity obtained with the training and the independent test sets; (ii) the number of selected features in the discriminant panels; and (iii) the feature overlapping degree among the panels for each binary comparison. The selected model for discriminating ccRCC from control samples yielded 96.8% accuracy, 100% sensitivity, and 93.5% specificity in the training set under cross-validation and 81.4% accuracy, 79.7% sensitivity, and 100% specificity in the test set, which exhibited an area under the curve (AUC) of 0.89. This model was based on 18 discriminant features (panel A), with two pairs of features identified as the same compound (see Lipid

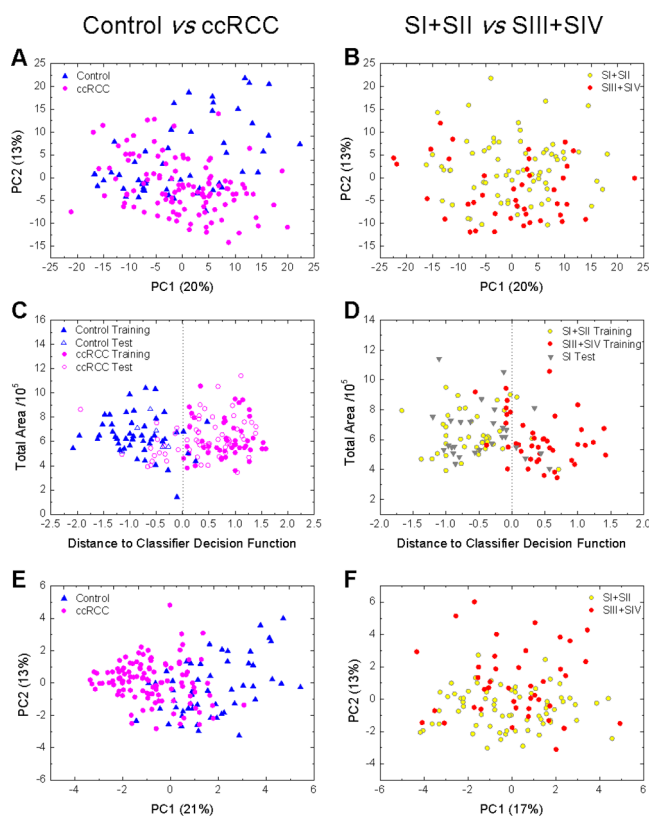


Figure 3. Principal component analysis (PCA) score plot of serum samples from controls and ccRCC patients (left) and from ccRCC patients with early (SI + SII) and late (SIII + SIV) stages (right), using the 386-feature matrix. The model consisted of (A) two PCs with 33% total captured variance and (B) two PCs with 33% total captured variance. Classification plot based on the SVM coupled with LASSO models for (C) ccRCC versus control, which yielded 95.7% accuracy, 91.3% sensitivity, and 91.3% specificity in the training set under cross-validation and 77.1% accuracy, 100% sensitivity, and 75.0% specificity in the test set and exhibited an area under the curve (AUC) of 0.90, and for (D) SI + SII (early) versus SIII + SIV (late) ccRCC stages, which yielded 84.5% of the total accuracy in the training set and 82.1% accuracy in classifying the test set consisting of samples from SI ccRCC patients. PCA score plot of samples from ccRCC patients and controls (E) using the 16-discriminant feature panel and consisted of two PCs with 34% total captured variance. PCA score plot of samples from patients with SI + SII (early) and SIII + SIV (late) ccRCC stages (F), using the 26-discriminant feature panel, and consisting of two PCs with 30% total captured variance. Controls (blue triangles); ccRCC samples (magenta circles); full symbols were used for the training set and empty symbols for the test set; samples from stage I + stage II ccRCC patients (yellow circles); stage III + stage IV ccRCC patients (red circles); and stage I ccRCC patients in the test set (gray triangles).

Identification). Thus, the most intense discriminant features from each pair were retained, leading to a 16-feature panel for further analysis. This subpanel exhibited 95.7% accuracy, 100% sensitivity, and 91.3% specificity in the training set under cross-validation and 77.1% accuracy, 75.0% sensitivity, and 100% specificity in the test set, with an AUC of 0.90 (Figure 3C).

The selected model for the early- versus late-stage ccRCC binary comparison comprised 26 features (panel B). Interestingly, two of these features were common to discriminant panels A and B. The latter provided 84.5% total classification accuracy in the training set, with 92.9 and 76.2% classification accuracy for early and late ccRCC stages,

respectively. Noteworthy, this model provided 82.1% classification accuracy in the test set, composed only of SI ccRCC samples (Figure 3D). To evaluate possible overfitting, PCA models were built for each binary comparison using the discriminant panels with 16 and 26 features. The PCA score plots displayed in Figure 3E,F exhibited an improved sample class discrimination compared to the original 386-feature matrix (Figure 3A,B). In addition, permutation test results supported robustness of each model's performance, discarding overfitting. The permutation test accuracy scores were 0.95 and 0.74 for the control versus tumor and the early versus late ccRCC stage binary comparisons, respectively ($p < 0.001$).

Panel A, selected to optimize the discrimination of ccRCC patients from controls, was also used to evaluate sample classification for other RCC subtypes despite the potential limitation of this strategy. That is, samples from chRCC and papRCC patients were tested for classification with the decision function obtained for the first binary classification problem. As a result, samples were randomly classified as ccRCC or controls, thus revealing that the developed SVM model built and trained with the ccRCC lipid fingerprint cannot effectively classify other RCC histological subtypes. Although this result was a possible outcome, it simultaneously highlights the specificity of the model for the most frequent RCC subtype. Additionally, the one-class SVM model, trained with ccRCC or control samples, provided similar results as the binary classification-based model, since approximately 50% of chRCC and papRCC samples were classified as either ccRCC or control, suggesting that the signature defined by panel A does not represent the phenotype of either chRCC or papRCC.

Lipid Identification

Structural elucidation of lipids is crucial to understand the disease biochemistry since different isomers can exhibit specific roles. Compound identification was attempted for all of the discriminant features in panels A and B (Tables 3 and S5–S7). Adduct ion analysis was performed considering the following ionic species: $[M - H]^-$, $[M + Cl]^-$, $[M + CH_3COO]^-$, $[M + HCOO]^-$, $[M + Na - 2H]^-$, $[M + K - 2H]^-$, $[M - H_2O - H]^-$, $[M - CO_2 - H]^-$, $[M - CH_3]^-$, $[M + nNaCl + Cl]^-$, $[M + nNaCl - H]^-$, $[M + NaCH_3COO - H]^-$, and $[M + NaHCOO - H]^-$. Tandem MS experiments also suggested the presence of $[M + nNaCH_3COO - H]^-$ and $[M + NaCH_3COO + CH_3COO]^-$ ions being associated with mobile phase composition, since adduct ions with CH_3COO^- , CH_3CN , $NaCH_3COO$, and a combination of them were proven to be generated by LC-ESI-MS in negative ion mode.⁷⁵ Considering the huge overlap of isobaric species provided by the lipidome,⁷⁶ the identification process was based on exact mass, isotopic pattern analysis, diagnostic and characteristic product ions, and neutral losses identified in MS/MS experiments, as well as retention time and fragmentation pattern matching with chemical standards for all possible cases (Figures 4 and S5). Comparison with chemical standards also contributed to discard four putatively annotated compounds (Figure S5). When poor MS/MS information was obtained, and molecular formulas were not generated with a high score, the detected features were reported as unknown (Tables S6 and S7). Similarly, when tentative identifications did not match with experimental or in silico MS/MS spectra from databases, only molecular formulas were informed (Tables S6 and S7).

Table 3. Discriminant Features from Panels A and B Identified with a High Confidence Level^a

feature no.	panel	R _t (min)	m/z	ion type	elemental formula	Δm (mDa)	T/C	L/E	p-value	compound identification ^b
200	B	0.62	369.1736	[M - H] ⁻	C ₁₉ H ₃₀ O ₅ S	0.0	+1.11	-1.31	4.2 × 10 ⁻²	3α-hydroxy-5α-androstan-17-one 3-sulfate
795	A	0.74	407.2787	[M - H] ⁻	C ₂₄ H ₄₀ O ₅	1.0	-3.05	+1.05	1.4 × 10 ⁻³	cholic acid
928	A	0.92	243.1955	[M - H] ⁻	C ₁₄ H ₂₈ O ₃	0.5	-1.70	-5.22 ^c	2.9 × 10 ⁻²	2-hydroxymyristic acid
940	A	0.94	183.1385	[M - H] ⁻	C ₁₁ H ₂₀ O ₂	0.0	-1.21	-1.19 ^c	3.6 × 10 ⁻²	undecylenic acid
978	B	1.04	197.1542	[M - H] ⁻	C ₁₂ H ₂₂ O ₂	0.0	-1.03	-1.19	2.2 × 10 ⁻²	cis-5-dodecenoic acid
1021	B	1.15	500.2779	[M - H] ⁻	C ₂₃ H ₄₄ NO ₇ P	0.2	-1.09	-1.01	NS ^d	LPE(20:4/0:0)
1042	A	1.23	199.1698	[M - H] ⁻	C ₁₂ H ₂₄ O ₂	0.0	-1.38	-1.10	1.6 × 10 ⁻⁵	lauric acid
1068	A	1.35	554.3472	[M + CH ₃ COO] ⁻	C ₂₄ H ₅₀ NO ₇ P	-1.4	-1.55	-1.21 ^c	3.4 × 10 ⁻¹³	LPC(16:0/0:0)
1072	A	1.35	636.3474	[M + NaCH ₃ COO + CH ₃ COO] ⁻		1.5	-1.32	-1.10	8.5 × 10 ⁻⁸	
1153	B	1.74	385.2358	[M + NaCH ₃ COO - H] ⁻	C ₂₀ H ₃₂ O ₂	0.3	-1.14	+1.05	NS	arachidonic acid
1185	A	1.86	443.2398	[M + 2(NaCH ₃ COO) - H] ⁻	C ₁₈ H ₃₂ O ₂	-1.2	+1.32	-1.01	9.7 × 10 ⁻⁴	linoleic acid
1480	B	4.09	337.3099	[M - H] ⁻	C ₂₂ H ₄₂ O ₂	-0.8	-1.05	+1.19	NS	cis-13-docosenoic acid
1552	B	5.30	864.5742	[M + CH ₃ COO] ⁻	C ₄₆ H ₈₀ NO ₈ P	-1.3	-1.24 ^c	-1.16	1.1 × 10 ⁻²	PC(18:2/20:4)
1597	A	5.57	766.5408	[M - CH ₃] ⁻	C ₄₄ H ₈₀ NO ₈ P	-2.1	-1.36	-1.14	5.6 × 10 ⁻⁴	PC(18:2/18:2)
1604	A	5.60	922.5782	[M + NaCH ₃ COO + CH ₃ COO] ⁻		0.3	-1.35	-1.15	5.9 × 10 ⁻⁴	
1651	B	6.10	835.5348	[M - H] ⁻	C ₄₃ H ₈₁ O ₁₃ P	1.1	-1.10	-1.21	NS	PI(16:0/18:1)
1733	B	6.81	898.5778	[M + NaCH ₃ COO + CH ₃ COO] ⁻	C ₄₂ H ₈₀ NO ₈ P	-0.7	+1.04	1.13	9.2 × 10 ⁻³	PC(16:0/18:2)
1804	B	7.54	826.5970	[M + CH ₃ COO] ⁻	C ₄₄ H ₈₂ NO ₇ P	0.8	-1.02	-1.05	NS	PC(O-16:0/20:4)
1820	B	7.71	722.5110	[M - H] ⁻	C ₄₁ H ₇₄ NO ₇ P	-1.5	+1.04	-1.34	6.0 × 10 ⁻³	PE(P-16:0/20:4)

^aΔm values were calculated as the mass difference between the m/z values obtained from Progenesis Q1 and the theoretical mass for the monoisotopic peaks. Fold changes were calculated as the ratio of median peak areas between samples from ccRCC patients and controls (T/C), and between samples from ccRCC patients with late- and early-stage tumors (L/E). The p-values reported for the binary comparison associated with each panel were calculated with the Mann–Whitney U test. Statistical significance was evaluated after correcting p-values using the Benjamini–Hochberg procedure for multiple comparisons with an FDR of 0.1. ^bCompounds identified with MSI level 1 are highlighted in bold. The diagnostic ions that assisted in lipid identification and the corresponding structure codes (SMILE) are listed in Table S8. ^cIndicates that p-values <0.05 were obtained with the Mann–Whitney U test for both binary comparisons. ^dNS: non-significant.

For most fatty acids (FAs) and their derivatives, tandem MS spectra provided limited specific information for structural interpretation.⁶⁹ The [M - H]⁻ ion was identified as the base peak of their mass spectra⁶⁹ and corresponded to the monoisotopic peak of discriminant features for the following compounds validated with chemical standards: undecylenic acid (C11:1), cis-5-dodecenoic acid (C12:1), lauric acid (C12:0), and cis-13-docosenoic acid (C22:1). The discriminant feature #1153 (1.74 min m/z 385.2358) was identified as the [M + NaCH₃COO - H]⁻ adduct ion of arachidonic acid (C20:4), which was highly correlated with feature #1150 (1.74 min m/z 303.2330) that was assigned as the [M - H]⁻ ion (Table S5). Diagnostic fragments of this lipid included ions at m/z 259 and 205, from the loss of CO₂ and from the cleavage of the double bond between C5 and C6, respectively.⁶⁹ Linoleic acid (C18:2) was detected as the [M + 2(NaCH₃COO) - H]⁻ ion, which was highly correlated with two other features identified as [M + NaCH₃COO - H]⁻ and [M - H]⁻ (Table S5). Structural information obtained for singly hydroxylated fatty acids from tandem MS experiments was poor, since they yield the same mass loss for any position of the hydroxyl group, except for 3-hydroxy fatty acids.⁶⁹ Discriminant feature #928 (0.92 min m/z 243.1955) was assigned as 2-hydroxymyristic acid based on the retention time match with the chemical standard. Nevertheless, the presence of other n-hydroxy fatty acids cannot be excluded. Feature #961 (0.97 min m/z 355.2263) was putatively annotated as (4Z,7Z,10E,12E,16Z)-18-(3-ethylcycloprop-1-en-1-yl)-14-hydroxyoctadeca-4,7,10,12,16-pentaenoate (14-HDoHE), a hydroxylated docohexanoic acid (DHA, C22:6)-derivative, based on the HMDB search, since no endogenous compounds

matched with the molecular formula or with the exact mass in LIPID MAPS.

Some discriminant features corresponding to compounds eluting at low retention times were detected as [M + NaCl - H]⁻ adduct ions.⁷⁷ This was the case for 3-hydroxydodecane-dioic acid, which was putatively annotated, and for compounds with the tentative molecular formulas of C₄₇H₈₂O₂ and C₃H₆O₃. A complementary HILIC-MS analysis contributed to discard lactic acid as a tentative ID for the latter and assisted in validating the identity of other molecules that were not retained in the reverse-phase chromatographic method. This was the case for cholic acid, which was detected as the [M - H]⁻ ion, as expected for most bile acids and bile acid conjugates in negative ESI mode.⁶⁹ The HILIC-MS method also contributed to identify two discriminant compounds belonging to the steroid family. One of them corresponded to the androgen steroid derivative 3α-hydroxy-5α-androstan-17-one 3-sulfate. Since the chemical standard was not commercially available, the compound was synthesized via a hydrogenation reaction of dehydroepiandrosterone sulfate for confirmation (see Materials and Methods in the Supporting Information). As expected, the only observed product ion in the MS/MS spectrum corresponded to HSO₄⁻ at m/z 96.9596. The second steroid, detected as [M + CH₃COO]⁻, was tentatively identified as 18-hydroxycorticosterone. The HILIC-MS method contributed to discard cortisol as an alternative candidate for this discriminant feature; both compounds are found in biological fluids, have a similar structure, and lead to common product ions in the MS/MS spectra.

Phospholipids (PLs) were also identified as discriminant compounds through tandem MS experiments in negative and

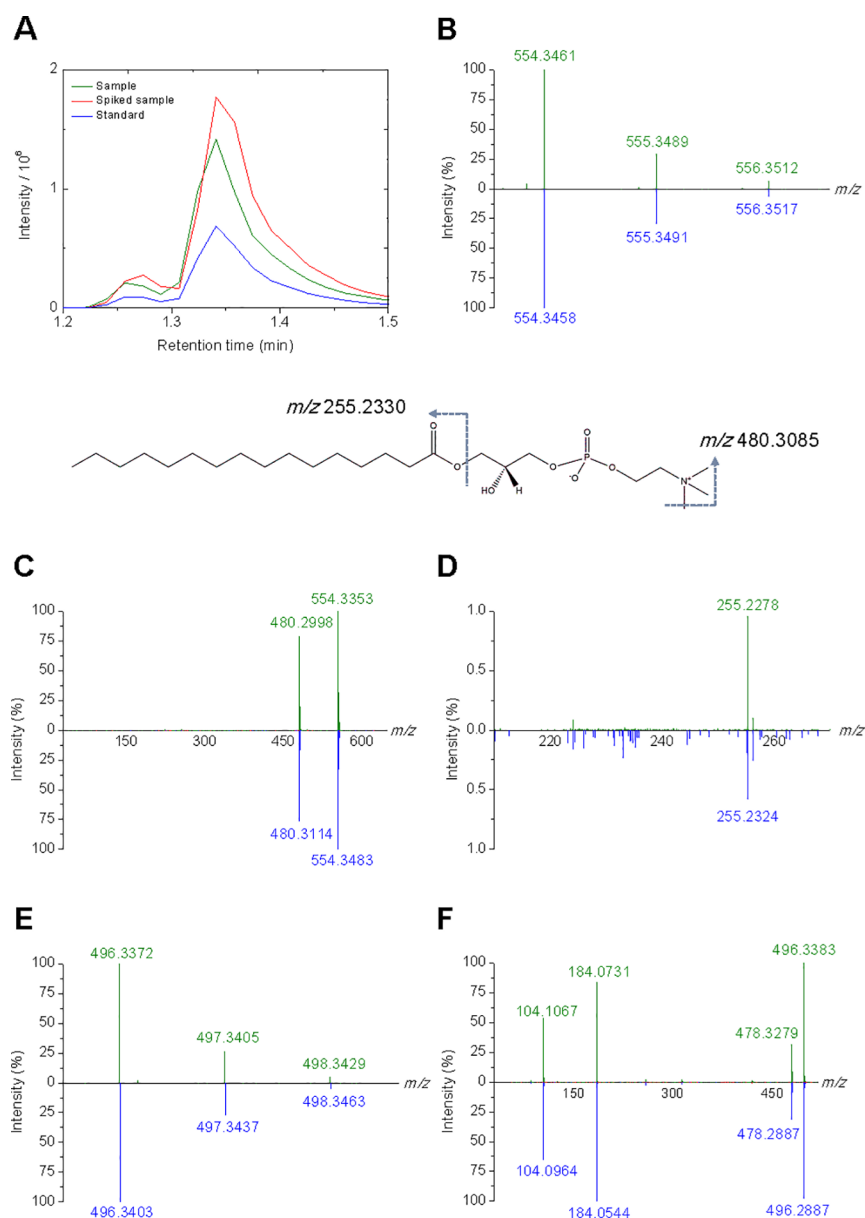


Figure 4. Identification of LPC(16:0/0:0). (A) Extracted ion chromatograms for the $[M + \text{CH}_3\text{COO}]^-$ ion at $m/z 554.3472 \pm 0.0500$ generated from a pooled sample (green), a $15 \mu\text{M}$ spiked pooled sample (red), and a $15 \mu\text{M}$ LPC(16:0/0:0) standard solution (blue). (B) Mass spectrum of $[\text{LPC}(16:0/0:0) + \text{CH}_3\text{COO}]^-$ ion with $m/z 554.3472$ in a pooled sample (green), and its simulated isotopic pattern (blue). (C) Product ion mass spectra of $[\text{LPC}(16:0/0:0) + \text{CH}_3\text{COO}]^-$ precursor ion for a pooled sample (green) and for a $6 \mu\text{M}$ LPC(16:0/0:0) standard solution (blue) using collision cell voltages of 20 V and 10 V, respectively. (D) Mass spectrum showing the range of m/z 210–270. (E) Mass spectrum of $[\text{LPC}(16:0/0:0) + \text{H}]^+$ ion with $m/z 496.3403$ in a pooled sample (green), and its simulated isotopic pattern (blue). (F) product ion mass spectra of $[\text{LPC}(16:0/0:0) + \text{H}]^+$ precursor ion for a pooled sample (green) and for a $6 \mu\text{M}$ LPC(16:0/0:0) standard solution (blue) using a collision cell voltage of 20 V. The 100% intensity in the MS/MS spectrum of the sample corresponds to absolute intensity values of 5.03×10^5 and 3.79×10^5 for (C) and (E), respectively. The diagnostic ions for LPC(16:0) detected in negative ion mode are: $m/z 255.2330$ (FA - H, 16:0) and $m/z 480.3085$ (M - CH_3COO - CH_3) and in positive ion mode are $m/z 184.0733$ (protonated phosphocholine, $\text{C}_5\text{H}_{15}\text{O}_4\text{P}$), $m/z 104.1075$ (protonated choline), and $m/z 478.3292$ (LPC16:0 - H_2O + H).

positive ESI modes. Two quadrupole-mass windows were used to gain confidence in annotation, especially for those cases in which validation with the corresponding chemical standard was not possible. As expected, the most abundant ionic species in negative mode were the $[M - \text{H}]^-$ ion for phosphoethanolamines (PEs) and phosphoinositols (PIs) and the $[M + \text{CH}_3\text{COO}]^-$ ion for phosphocholines (PCs), the latter being identified by the loss of ($\text{CH}_3 + \text{CH}_3\text{COO}^-$; $\Delta m = 74$) corresponding to PC demethylation. Alternative ion types detected for PLs were $[M - \text{CH}_3]^-$, $[M + \text{NaCH}_3\text{COO} +$

$\text{CH}_3\text{COO}]^-$, $[M + \text{H}]^+$, and $[M + \text{Na}]^+$,⁶⁹ the former being produced through in-source fragmentation of PCs. Adduct ion analysis and tandem MS experiments contributed to avoid incorrect assignment of PCs as PEs.^{78,79} For candidates with no chemical standards, tentative identification was provided based on the polar head attached to the sn-3 carbon position that defined the PL class and considering the abundance ratio of the lyso-form ions arising from the neutral loss of the sn-1 or sn-2 moiety as a free fatty acid as $[M - \text{H} - \text{R}_x\text{COOH}]^-$ or as a ketone $[M - \text{H} - \text{R}_x\text{CHCO}]^-$. Hou et al. showed that these

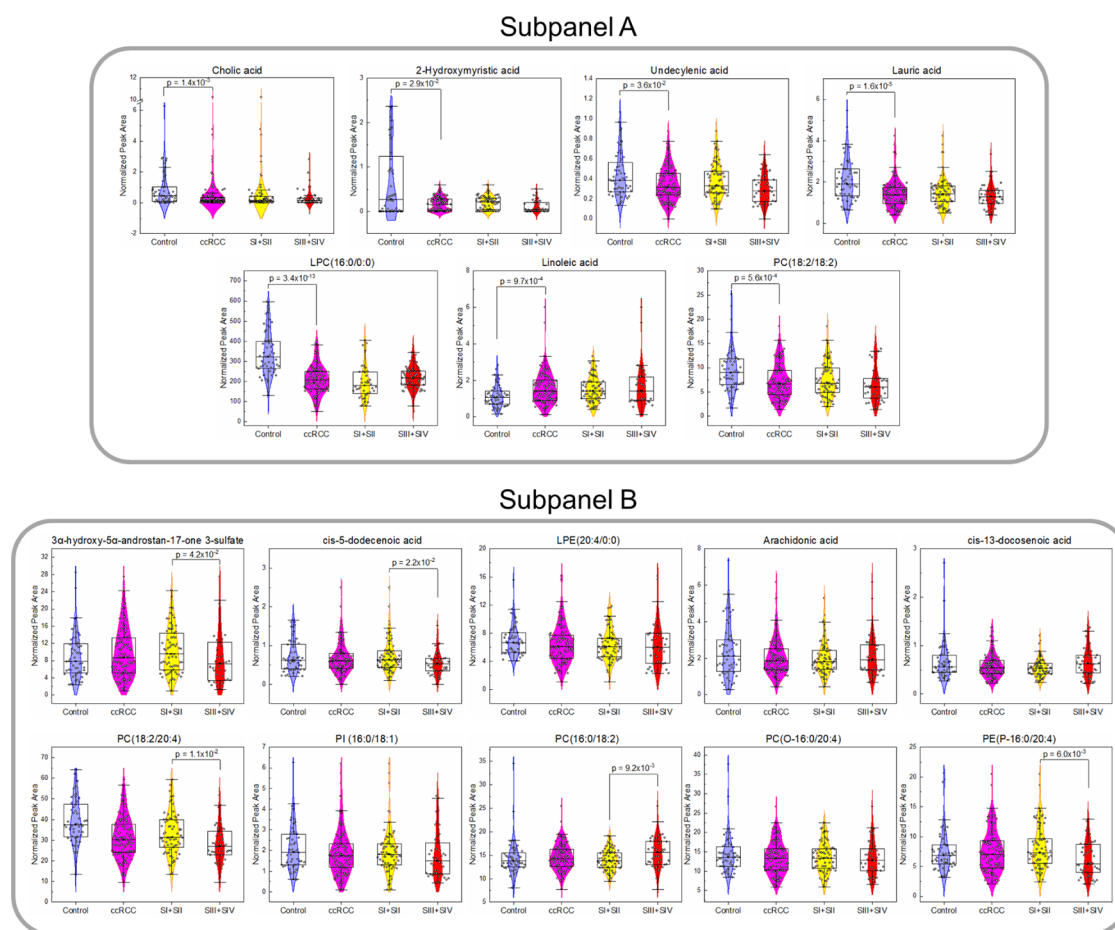


Figure 5. Box plots and overlapped data distribution curves for discriminant lipids identified with a high confidence level. Comparisons of controls versus ccRCC patients, and SI + SII (early) versus SIII + SIV (late) ccRCC stages are shown. Data are represented by a gray circle and overlap with the boxes. Mean values are represented by a filled square in the box; median values are represented by a line in the box; the upper and lower edges of the box are the 25th and 75th percentiles, respectively; and the whisker extends to the most extreme values in the data, not including outliers defined as 1.5 of the interquartile range. A symmetrical Kernel smooth curve type was selected for representing the data distribution. Fold changes were calculated as the ratio of median peak areas between the classes (i) ccRCC and control and (ii) early and late ccRCC stages. The p -values reported for binary comparisons were calculated with the Mann–Whitney U test. Statistical significance was evaluated after correcting p -values with the Benjamini–Hochberg procedure for multiple comparisons with an FDR of 0.1.

fragments exhibited higher abundances when the loss corresponds to the fatty acyl moieties attached to the sn-2 position.⁶⁸ Also, it has been shown that for most abundant PCs in biological systems, the loss of the sn-2 carboxylic acid is the most favorable, with the same trend being observed for PEs.^{69,80,81}

Two highly correlated features from panel A (Tables 3 and S5) were identified as PC(18:2/18:2) (Figure S5). Also, PC(16:0/18:2) was identified with MSI level 1, while PC(18:2/20:4) was putatively annotated based on several characteristic ions that suggested this lipid as the unique candidate. Other annotated PCs included PC(38:5), putatively annotated as PC(18:1/20:4) and PC(16:0/22:5), and PC(37:4), where arachidonate-related product ions were detected, suggesting the presence of an acyl substituent with 17 carbon atoms at the sn-1 position, such as n-methylpalmitate or heptadecanoate. For a tentatively identified PC in panel B, the adduct type remains unidentified (feature #1892: 8.60 min_{m/z} 926.6041, Figure S5). However, the characteristic product ions detected at m/z 283.2643 and 303.2324 corresponding to the acyl substituents and the product ion detected at m/z 508.3398 associated with the

LPC(18:0)–CH₃ fragment suggest PC(20:4_18:0) as the discriminant lipid (Figure S5). Another discriminant PL in panel B was identified based on the lyso-form ions⁶⁸ as PI(16:0/18:1), which was confirmed with a chemical standard.

Two features were identified as compounds belonging to other PL subclasses, mostly found in PC and PE, and known as plasmanyl- and plasmenyl-PLs, in which the sn-1 position is linked to alkyl or vinyl ethers, respectively, and a fatty acyl group is linked to the sn-2 position by an ester bond, even if counterparts were also reported.⁸² The discriminant compound PC(O-16:0/20:4) that was further confirmed with a chemical standard exhibited, as expected,^{69,83} a fragmentation pattern for the $[M + H]^+$ precursor ion identical to that observed for the PC. The other feature that corresponded to these PL subclasses was putatively annotated as PE(P-16:0/20:4) based on the detected product ions in the tandem MS spectrum and the loss of the sn-2 fatty acid as a ketene, as described for ether PEs.⁸⁴ The product ions were identical to those expected from a PE, except for the singly generated carboxylate anion.

An oxidized PL was also identified in the discriminant panel B through manual interpretation of MS/MS spectra (Table

S7). Feature #1540 (5.11 min m/z 832.5702) was tentatively annotated as PC(16:0(OH)_18:2) based on two product ions detected at m/z 279.2307 and 271.2247; the former corresponded to the linoleate acyl anion and the latter to the hydroxylated palmitate acyl anion. Recognition and annotation of oxidized PLs is a challenge itself due to the limited tools available for data analysis.^{85–88} Annotation of oxidized LPCs was recently addressed through the comparison of native and oxidized LPCs, showing that the only difference in the MS/MS spectra relies on the presence of an –OH group in the acyl chain,⁸⁹ as it was observed for feature #1540 (5.11 min m/z 832.5702).

Lyso-glycerophospholipids (LPLs) were also identified in the discriminant panels. LPLs are characterized by having only one acyl chain linked to the sn-1 or the sn-2 position of glycerol and a polar head at the sn-3 position. The enzymes phospholipase A (PLA) 1 and PLA 2 are capable of cleaving diacyl-PLs, producing 2-acyl-1-lysophospholipids (2-acyl-1-LPLs) and 1-acyl-2-LPLs, respectively. It has been reported that acyl chains at the sn-2 position of 2-acyl-1-LPLs readily move to the sn-1 position by a nonenzymatic intramolecular acyl migration; thus, most of the LPLs detected in serum are 1-acyl-2-lyso isomers.⁹⁰ In high-energy positive mode fragmentation experiments, the intensity ratio between product ions at m/z 184.1 and 104.1 allows discriminating between sn-1 and sn-2 isomers, the ratio being >1 for the sn-1 isomer,⁹¹ and the ions corresponding to the aliphatic chain are typically identified in negative ion mode. Figure 4 illustrates the identification process for LPC(16:0/0:0), which was selected by the SVM and LASSO model as two discriminant features in panel A, corresponding to the highly correlated $[M + CH_3COO]^-$ and $[M + NaCH_3COO + CH_3COO]^-$ ions (Table S5). The product ions detected in the MS/MS spectrum followed the expected intensity ratio for the sn-1 isomer, which was confirmed by the analysis of a chemical standard that eluted after its counterpart (2-acyl-1-LPC).⁹² Based on results from tandem MS experiments in both ion modes, another LPL was annotated as LPE(20:4/0:0).

Overall, a total of nine features were successfully identified as seven lipids with MSI level 1 in the discriminant panel A aimed at ccRCC detection. In addition, five features were putatively annotated, and four features remained with no ID (Table S6). Considering only the seven compounds that were identified with high confidence (Tables 3 and S8), the SVM model yielded AUC values of 0.95 and 0.82 for the training and test sets, respectively, highlighting the good performance of this subpanel for discriminating ccRCC patients from controls (Table S9A). Regarding panel B, which was selected to discriminate early- from late-stage ccRCC patients, 7 out of the 26 discriminant features were identified with MSI level 1. In addition, three features were tentatively identified with high confidence based on tandem MS experiments conducted in both ion modes; eight features were putatively annotated, and eight features remained with no ID (Table S7). Thus, the subpanel consisting of 10 compounds identified with high confidence (Tables 3 and S8) was evaluated for classification of early versus advanced ccRCC stages. This new model yielded accuracy values of 72.6% for the training set and 75.0% for the SI ccRCC samples in the test set (Table S9B). These results are promising considering that this subpanel included only 10 out of the 26 discriminant features. It is worth mentioning that this work is the first serum study on ccRCC that challenges the performance of multivariate models with large independent

test sets, supporting the power of SVM coupled with LASSO for biomarker discovery.

Biological Relevance of Identified Discriminant Lipids

The trends obtained for the discriminant lipid levels according to the binary classification problems contribute to understand the altered mechanisms in ccRCC (Figure 5; Tables 3, S6, and S7). Even if each discriminant compound may not be considered a unique biomarker for the disease, the lipid signature provided by each panel can be used for ccRCC detection and early diagnosis.

Most discriminant features in panel A, i.e., 13 out of 16, exhibited significantly decreased levels in samples from ccRCC patients compared to controls, whereas two of them were detected with significantly increased levels in ccRCC (Figure 5 and Table S6). In agreement with our results, other authors reported decreased lipid levels in RCC.^{40,43,51} Moreover, most features of panel A presented decreased levels along the disease progression (Figure 5 and Table S6). Regarding the analysis of samples from chrRCC and papRCC patients, serum levels of the discriminant features in panel A were, in general, lower than those from controls but higher compared to ccRCC patients (Table S10).

Among the 13 compounds that exhibited decreased levels in the samples from ccRCC patients, 4 belonged to the fatty acyl class (Table 3), i.e., undecylenic acid, lauric acid, 2-hydroxymiristic acid, and 3-hydroxydodecanedioic acid. The metabolism of FAs is mainly regulated by their synthesis via the fatty acid synthase (FAS) and through their β -oxidation. Levels of most enzymes involved in the β -oxidation pathway were reported to be decreased in high-grade RCC tissue, suggesting the lack of acyl-CoAs oxidation.^{19,30} In addition, overexpression of FAS has been associated with ccRCC aggressiveness and poor patient survival.^{93,94} The pattern observed in this study may result from an imbalanced synthesis and impaired β -oxidation of FAs. Interestingly, the latter has also been described in prostate cancer.^{95,96} Short-chain free FAs, including lauric acid, have been reported to exhibit decreased levels in RCC tissue in a grade-dependent fashion and coupled with increased levels of carnitines and acylcarnitines.³⁰ The decreased expression reported for the very long-chain acyl-CoA dehydrogenase (VLCAD) enzyme in RCC tissue¹⁵ may explain the lower levels of short-chain FAs detected in serum samples in this study.

Increased levels of linoleic acid, an essential FA, were detected in serum samples from ccRCC patients, with similar levels in samples from different ccRCC stages, chrRCC and papRCC. This trend suggests that the alteration of linoleic acid metabolism may occur in the early stage of the disease and may be common to the different RCC tumors. Linoleic acid levels were also found increased in ccRCC tissue⁹⁷ and decreased with disease progression.⁹⁸ Recently, increased plasma levels of linoleic acid were found in RCC and bladder cancer patients, suggesting a common alteration in both genitourinary tumors.⁴⁴ A mechanism of lipid deposition was suggested in ccRCC based on carnitine palmitoyltransferase 1A (CPT1A), acting as a rate-limiting enzyme of the FA transport into the mitochondria, which was found as a direct target gene of the hypoxia-inducible factor 1 α (HIF1 α) and HIF2 α complexes.⁹⁹ Future studies focused on the analysis of paired data from tissue and serum samples collected from the same patients would probably contribute to understand linoleic acid metabolism in ccRCC.

Two identified and confirmed PLs in panel A, i.e., PC(18:2/18:2) and LPC(16:0/0:0), exhibited decreased levels in samples from ccRCC patients compared to controls (Figure 5 and Table 3). No correlation was observed between PC(18:2/18:2) and linoleic acid in tumor or control samples despite the fatty acyl chains present in this lipid. Previous studies have also detected decreased levels of LPC(16:0/0:0) in RCC serum samples.^{39,40} These works reported reduced levels in the early stages of the disease, while the opposite trend was observed for the ccRCC cohort investigated in this work. This discrepancy may be associated with differences in the sample cohort and the staging system applied for ccRCC classification.¹ Regarding the comparison with other RCC histological subtypes, ccRCC exhibited the lowest LPC(16:0/0:0) serum levels (Table S10).

Cholic acid, a primary bile acid, also presented decreased levels in serum samples from ccRCC patients compared to controls (Figure 5 and Table 3). However, no differences were observed between ccRCC stages. Interestingly, samples from chRCC patients exhibited the lowest levels compared to controls. Bile acids play an important role in lipid absorption and cholesterol catabolism; thus, altered levels may be associated with the impaired cholesterol metabolism characteristic of ccRCC.⁵¹ Bile acids are considered to be markers of liver injury,¹⁰⁰ and high levels have also been associated with hepatocellular carcinoma.¹⁰¹ Also, a few reports have focused on secondary bile acid alterations associated with gastric and colon cancer and the microbiota role;^{102,103} however, their role in RCC remains unclear.

A similar analysis was performed for the discriminant features aimed at ccRCC staging. Twelve out of twenty-six features in panel B exhibited significantly different levels, seven being increased in the early stages and five being increased in advanced stages of the disease (Figure 5 and Table S7). Only eight features were significantly different between controls and ccRCC patients, six being decreased in ccRCC samples. Therefore, the best panel found for discriminating ccRCC stages does not follow the same behavior in terms of lipid level changes as the one selected for discriminating samples of ccRCC patients from controls. That is, most lipids in panel A exhibited lower levels in samples from advanced ccRCC patients compared to early stages, but the fold changes of panel B features presented a random pattern toward ccRCC staging (Table S7). Therefore, the two-model scheme developed in this work for disease detection and early ccRCC diagnosis may not be simplified with a single-step strategy since the combination of features that best distinguished early from late ccRCC stages follows a different behavior than the panel that best resembles the ccRCC phenotype for discriminating ccRCC patients from controls.

Four FAs were identified in the discriminant panel for ccRCC staging, including monounsaturated FAs (MUFAs) and polyunsaturated FAs (PUFAs), i.e., *cis*-5-dodecenoic acid, *cis*-13-docosenoic acid, arachidonic acid, and 14-HDoHE, the latter being an autoxidation product of docosahexaenoic acid. Interestingly, only the short-chain MUFAs exhibited significantly lower levels in advanced ccRCC stages, following the same trend as those FAs identified in panel A. Levels of arachidonic acid and 14-HDoHE exhibited a nonsignificant decrease trend in ccRCC patients compared to controls (Figure 5 and Table 3). In this regard, decreased levels of arachidonic acid were reported in ccRCC tissue by means of

MS imaging⁹⁷ and GC-MS with lower levels in the late stages of the disease.³³

PLs can act as reservoirs of FAs. Interestingly, all identified or annotated PLs in panel B contained arachidonic or palmitic acid as substituents. Arachidonic acid is an essential precursor of eicosanoids, which are regulatory lipids associated with inflammatory processes,¹⁰⁴ and palmitic acid has been related with cell proliferation.^{105,106} Thus, both FAs are linked to tumor progression. The relative abundance of PLs followed different trends along disease stages (Table S7). Lower levels of PC(18:2/20:4) and PE(P-16:0/20:4) and higher levels of PC(20:4_18:0) and PC(16:0/18:2) and its oxidized form PC(16:0(OH)_18:2) were detected in samples from advanced ccRCC patients (Table S7). As expected, levels of the oxidized form were lower than those observed for the native form in each sample class.⁸⁵ Serum levels of PC(18:2/20:4) were significantly decreased in samples from ccRCC patients compared to controls, whereas levels of PC(16:0(OH)_18:2) were increased in tumor samples (Table 3). Higher levels of oxidized PLs have been associated with cancer, supporting this result.¹⁰⁷

FAs with an odd number of carbon atoms are not endogenous in humans. However, PC(17:0_20:4) was a tentative identification for one of the discriminant lipids in panel B (Table S7), without excluding the possibility of an alternative compound with branched fatty acyl chains in the molecule. Previous metabolomics studies using DI-MS and LC-MS have suggested LPC(17:0) and LPC(15:0) as potential serum biomarkers of RCC.^{39,40} In addition, Wettersten et al. found increased levels of medium-chain FAs, including C15:0 and C17:0, in RCC tissues compared to adjacent normal tissues,³⁰ and higher levels of C17:0 were reported in metastasized tumor tissues.³³ Altogether, this evidence and the observed trends for linoleic acid would suggest an altered metabolism of FAs incorporated through the diet. In this regard, several studies have addressed the role of fat diet in cancer development.¹⁰⁸

Two compounds belonging to the steroid lipid class also contributed to ccRCC staging (Figure 5 and Table 3). Patients with early ccRCC stages presented higher levels of 3 α -hydroxy-5 α -androstane-17-one 3-sulfate than controls, whereas advanced ccRCC patients exhibited lower levels. The putatively identified steroid 18-hydroxycorticosterone exhibited decreased levels in serum samples from ccRCC patients. Similarly, tetrahydrocorticosterone levels were found to be decreased in serum samples from RCC patients.^{39,40} In addition, metabolites involved in the steroid hormone biosynthesis were reported to be increased in urine samples from RCC patients, and a panel of three compounds, including cortolone and testosterone, allowed for the discrimination of patients with RCC from those with benign tumors.¹⁰⁹ These results suggest an impaired steroid metabolism in RCC that may be further investigated considering the altered cholesterol metabolism.⁵¹

Overall, results from this analysis suggest a marked downregulation of serum lipid levels associated with the disease. The alterations observed between ccRCC stages cannot be generalized with a common pattern. Nevertheless, it is worth noting that the best panel selected for early diagnosis included up- and downregulated lipids from different classes. Results are promissory for translation to the clinical setting after validation in larger and different cohorts, entailing

specific considerations for assuring the stability of each single lipid class.¹¹⁰

CONCLUSIONS

This study highlights the power of coupling untargeted serum lipidomic strategies with machine learning techniques to identify reduced lipid panels that provide an accurate signature for ccRCC diagnosis and to identify those patients in the early stage of the disease. High classification performance was achieved with training sets under cross-validation and with independent test sets. In addition, the discriminant subpanels that only included identified lipids with a high confidence level still yielded good classification accuracy. Most lipid levels were significantly decreased in serum samples from RCC patients, with ccRCC being the most impaired RCC histology compared to controls. All lipids exhibiting significant changes at different stages of disease were discussed in relation to the altered metabolic pathways to improve the current understanding of lipid-based ccRCC development and progression and to identify new targets for future drug treatment design. Results from this study are promissory for translating the diagnostic lipid fingerprints to the clinical setting after developing targeted assays to evaluate the panels in larger cohorts, including individuals with different ethnicities, diets, and lifestyles as well as patients with other tumor types and renal pathologies.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00663>.

Materials and methods; machine learning pipeline; data processing, classification and prediction; organic syntheses of 3 α -hydroxy-5 α -androstano-17-one 3-sulfate; age distribution and male/female frequency in the patient cohort (Figure S1); representative base peak intensity (BPI) chromatograms obtained for serum samples (Figure S2); principal component analysis (PCA) score plot of serum samples using the 386-feature matrix (Figure S3); sample distribution used for training, validating and testing the SVM and LASSO models (Figure S4); identification of discriminant features in panels A and B (Figure S5); sample cohort description (Table S1); daily run order using the UPLC-QTOF-MS analytical platform (Table S2); performance of the generated SVM and LASSO models for the binary comparison of ccRCC versus controls using different sample subsets for training and testing (Table S3); performance of the generated SVM and LASSO models for the binary comparison of early versus late ccRCC using different sample subsets for training and testing (Table S4); Spearman's correlation coefficients obtained for those features identified as the same compound (Table S5); panel A features (Table S6); panel B features (Table S7); diagnostic ions and SMILE codes for the discriminant features (Table S8); performance of the discriminant panels and subpanels for each binary comparison (Table S9); median values of the 18 discriminant features from panel A for controls and chRCC, papRCC, and ccRCC samples (Table S10) (PDF)

Metabolic feature matrix (Data Set 1) (XLSX)

AUTHOR INFORMATION

Corresponding Author

María Eugenia Monge – *Centro de Investigaciones en Bionanociencias (CIBION), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), C1425FQD CABA, Argentina;* orcid.org/0000-0001-6517-5301; Phone: +54 11 4899 5500; Email: maria.monge@cibion.conicet.gov.ar

Authors

Malena Manzi – *Centro de Investigaciones en Bionanociencias (CIBION), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), C1425FQD CABA, Argentina;* *Departamento de Química Biológica, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, C1113AAD Buenos Aires, Argentina*

Martín Palazzo – *LM2S, Université de Technologie de Troyes, CS42060 Troyes, France; Instituto de Investigación en Biomedicina de Buenos Aires (IBioBA), CONICET, Instituto Partner de la Sociedad Max Planck, C1425FQD CABA, Argentina*

María Elena Knott – *Centro de Investigaciones en Bionanociencias (CIBION), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), C1425FQD CABA, Argentina*

Pierre Beuseroy – *LM2S, Université de Technologie de Troyes, CS42060 Troyes, France*

Patricio Yankilevich – *Instituto de Investigación en Biomedicina de Buenos Aires (IBioBA), CONICET, Instituto Partner de la Sociedad Max Planck, C1425FQD CABA, Argentina*

María Isabel Giménez – *Departamento de Diagnóstico y Tratamiento, Hospital Italiano de Buenos Aires, C1199ABB CABA, Argentina*

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00663>

Author Contributions

M.E.M. conceived and designed the study. M.I.G. coordinated serum sample collection at HIBA. M.E.K. processed the serum samples and organized the metadata. M.E.M. and M.E.K. developed the UPLC-MS method. M.M. conducted LC-MS and MS/MS experiments. M.P., P.B., P.Y., and M.E.M. designed the SVM analysis. M.P., P.B., and P.Y. conducted the SVM analysis. M.M. and M.E.M. analyzed and interpreted the data. M.M. and M.E.M. wrote the paper. All authors revised the paper.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

M.E.M. acknowledges Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina, PUE 055 project) and the National Agency of Scientific and Technological Promotion (ANPCyT, PRH-PICT-2015-0022, and PICT-2018-02137 projects) for providing the funding. P.Y. acknowledges CONICET, ANPCyT, and FOCEM-Mercosur for providing the funding. M.P. acknowledges Universidad Tecnológica Nacional in Argentina and Université de Technologie de Troyes in France for providing the funding. We acknowledge the “Bióbanco Público de Muestras Séricas

Oncológicas” (BPMSO) from Instituto de Oncología A. H. Roffo and Hospital Italiano de Buenos Aires for providing the samples. M.E.M., M.M., and P.Y. are research staff members from CONICET. We also would like to thank Dr. Lydia I. Puricelli for helpful discussions, Dr. Mariela Bollini for conducting the synthesis of 3 α -hydroxy-5 α -androstan-17-one 3-sulfate, and Dr. Vanesa Herlax for providing lipid chemical standards.

REFERENCES

- (1) Escudier, B.; Porta, C.; Schmidinger, M.; Rioux-Leclercq, N.; Bex, A.; Khoo, V.; Grünwald, V.; Gillessen, S.; Horwich, A.; Committee, E. G. Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up[†]. *Ann. Oncol.* **2019**, *30*, 706–720.
- (2) Siegel, R. L.; Miller, K. D.; Jemal, A. Cancer statistics, 2018. *CA—Cancer J. Clin.* **2018**, *68*, 7–30.
- (3) Capitanio, U.; Bensalah, K.; Bex, A.; Boorjian, S. A.; Bray, F.; Coleman, J.; Gore, J. L.; Sun, M.; Wood, C.; Russo, P. Epidemiology of Renal Cell Carcinoma. *Eur. Urol.* **2019**, *75*, 74–84.
- (4) Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R. L.; Torre, L. A.; 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*, 394–424.
- (5) Diamond, E.; Molina, A. M.; Carbonaro, M.; Akhtar, N. H.; Giannakakou, P.; Tagawa, S. T.; Nanus, D. M. Cytotoxic chemotherapy in the treatment of advanced renal cell carcinoma in the era of targeted therapy. *Crit. Rev. Oncol./Hematol.* **2015**, *96*, 518–526.
- (6) De Meerleer, G.; Khoo, V.; Escudier, B.; Joniau, S.; Bossi, A.; Ost, P.; Briganti, A.; Fonteyne, V.; Van Vulpen, M.; Lumen, N.; Spahn, M.; Mareel, M. Radiotherapy for renal-cell carcinoma. *Lancet Oncol.* **2014**, *15*, e170–e177.
- (7) Gao, H.; Dong, B.; Jia, J.; Zhu, H.; Diao, C.; Yan, Z.; Huang, Y.; Li, X. Application of ex vivo (1)H NMR metabonomics to the characterization and possible detection of renal cell carcinoma metastases. *J. Cancer Res. Clin. Oncol.* **2012**, *138*, 753–761.
- (8) Hsieh, J. J.; Purdue, M. P.; Signoretti, S.; Swanton, C.; Albiges, L.; Schmidinger, M.; Heng, D. Y.; Larkin, J.; Ficarra, V. Renal cell carcinoma. *Nat. Rev. Dis. Primers* **2017**, *3*, No. 17009.
- (9) Hakimi, A. A.; Pham, C. G.; Hsieh, J. J. A clear picture of renal cell carcinoma. *Nat. Genet.* **2013**, *45*, 849–850.
- (10) Hsieh, J. J.; Le, V. H.; Oyama, T.; Ricketts, C. J.; Ho, T. H.; Cheng, E. H. Chromosome 3p Loss-Orchestrated VHL, HIF, and Epigenetic Deregulation in Clear Cell Renal Cell Carcinoma. *J. Clin. Oncol.* **2018**, *36*, 3533–3539.
- (11) Diamandis, E. P. The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem? *BMC Med.* **2012**, *10*, No. 87.
- (12) Leuthold, P.; Schaeffeler, E.; Winter, S.; Buttner, F.; Hofmann, U.; Murdter, T. E.; Rausch, S.; Sonntag, D.; Wahrheit, J.; Fend, F.; Hennenlotter, J.; Bedke, J.; Schwab, M.; Haag, M. Comprehensive Metabolomic and Lipidomic Profiling of Human Kidney Tissue: A Platform Comparison. *J. Proteome Res.* **2017**, *16*, 933–944.
- (13) Gatto, F.; Nookaew, I.; Nielsen, J. Chromosome 3p loss of heterozygosity is associated with a unique metabolic network in clear cell renal carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, E866–E875.
- (14) Hakimi, A. A.; Reznik, E.; Lee, C. H.; Creighton, C. J.; Brannon, A. R.; Luna, A.; Aksoy, B. A.; Liu, E. M.; Shen, R.; Lee, W.; Chen, Y.; Stirdivant, S. M.; Russo, P.; Chen, Y. B.; Tickoo, S. K.; Reuter, V. E.; Cheng, E. H.; Sander, C.; Hsieh, J. J. An Integrated Metabolic Atlas of Clear Cell Renal Cell Carcinoma. *Cancer Cell* **2016**, *29*, 104–116.
- (15) Wettersten, H. I.; Aboud, O. A.; Lara, P. N., Jr.; Weiss, R. H. Metabolic reprogramming in clear cell renal cell carcinoma. *Nat. Rev. Nephrol.* **2017**, *13*, 410–419.
- (16) Hanahan, D.; Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **2011**, *144*, 646–674.
- (17) Patel, S.; Ahmed, S. Emerging field of metabolomics: big promise for cancer biomarker identification and drug discovery. *J. Pharm. Biomed. Anal.* **2015**, *107*, 63–74.
- (18) Manzi, M.; Riquelme, G.; Zabalegui, N.; Monge, M. E. Improving diagnosis of genitourinary cancers: Biomarker discovery strategies through mass spectrometry-based metabolomics. *J. Pharm. Biomed. Anal.* **2019**, *178*, No. 112905.
- (19) Perroud, B.; Lee, J.; Valkova, N.; Dhirapong, A.; Lin, P. Y.; Fiehn, O.; Kultz, D.; Weiss, R. H. Pathway analysis of kidney cancer using proteomics and metabolic profiling. *Mol. Cancer* **2006**, *5*, No. 64.
- (20) Kim, K.; Aronov, P.; Zakharkin, S. O.; Anderson, D.; Perroud, B.; Thompson, I. M.; Weiss, R. H. Urine metabolomics analysis for kidney cancer detection and biomarker discovery. *Mol. Cell. Proteomics* **2009**, *8*, 558–570.
- (21) Kim, K.; Taylor, S. L.; Ganti, S.; Guo, L.; Osier, M. V.; Weiss, R. H. Urine metabolomic analysis identifies potential biomarkers and pathogenic pathways in kidney cancer. *OMICS: J. Integr. Biol.* **2011**, *15*, 293–303.
- (22) Ganti, S.; Taylor, S. L.; Kim, K.; Hoppel, C. L.; Guo, L.; Yang, J.; Evans, C.; Weiss, R. H. Urinary acylcarnitines are altered in human kidney cancer. *Int. J. Cancer* **2012**, *130*, 2791–2800.
- (23) Kind, T.; Tolstikov, V.; Fiehn, O.; Weiss, R. H. A comprehensive urinary metabolomic approach for identifying kidney cancer. *Anal. Biochem.* **2007**, *363*, 185–195.
- (24) Monteiro, M.; Carvalho, M.; Henrique, R.; Jeronimo, C.; Moreira, N.; de Lourdes Bastos, M.; de Pinho, P. G. Analysis of volatile human urinary metabolome by solid-phase microextraction in combination with gas chromatography-mass spectrometry for biomarker discovery: application in a pilot study to discriminate patients with renal cell carcinoma. *Eur. J. Cancer* **2014**, *50*, 1993–2002.
- (25) Monteiro, M.; Moreira, N.; Pinto, J.; Pires-Luis, A. S.; Henrique, R.; Jeronimo, C.; Bastos, M. L.; Gil, A. M.; Carvalho, M.; Guedes de Pinho, P. GC-MS metabolomics-based approach for the identification of a potential VOC-biomarker panel in the urine of renal cell carcinoma patients. *J. Cell. Mol. Med.* **2017**, *21*, 2092–2105.
- (26) Monteiro, M. S.; Barros, A. S.; Pinto, J.; Carvalho, M.; Pires-Luis, A. S.; Henrique, R.; Jeronimo, C.; Bastos, M. L.; Gil, A. M.; Guedes de Pinho, P. Nuclear Magnetic Resonance metabolomics reveals an excretory metabolic signature of renal cell carcinoma. *Sci. Rep.* **2016**, *6*, No. 37275.
- (27) Ragone, R.; Sallustio, F.; Piccinonna, S.; Rutigliano, M.; Vanessa, G.; Palazzo, S.; Lucarelli, G.; Ditonno, P.; Battaglia, M.; Fanizzi, F. P.; Schena, F. P. Renal Cell Carcinoma: A Study through NMR-Based Metabolomics Combined with Transcriptomics. *Diseases* **2016**, *4*, No. 7.
- (28) Falegan, O. S.; Arnold Egloff, S. A.; Zijlstra, A.; Hyndman, M. E.; Vogel, H. J. Urinary Metabolomics Validates Metabolic Differentiation Between Renal Cell Carcinoma Stages and Reveals a Unique Metabolic Profile for Oncocytomas. *Metabolites* **2019**, *9*, No. 155.
- (29) Falegan, O. S.; Ball, M. W.; Shaykhtudinov, R. A.; Pieroraio, P. M.; Farshidfar, F.; Vogel, H. J.; Allaf, M. E.; Hyndman, M. E. Urine and Serum Metabolomics Analyses May Distinguish between Stages of Renal Cell Carcinoma. *Metabolites* **2017**, *7*, No. 6.
- (30) Wettersten, H. I.; Hakimi, A. A.; Morin, D.; Bianchi, C.; Johnstone, M. E.; Donohoe, D. R.; Trott, J. F.; Aboud, O. A.; Stirdivant, S.; Neri, B.; Wolfert, R.; Stewart, B.; Perego, R.; Hsieh, J. J.; Weiss, R. H. Grade-Dependent Metabolic Reprogramming in Kidney Cancer Revealed by Combined Proteomics and Metabolomics Analysis. *Cancer Res.* **2015**, *75*, 2541–2552.
- (31) Schaeffeler, E.; Buttner, F.; Reustle, A.; Klumpp, V.; Winter, S.; Rausch, S.; Fisel, P.; Hennenlotter, J.; Kruck, S.; Stenzl, A.; Wahrheit, J.; Sonntag, D.; Scharpf, M.; Fend, F.; Agaimy, A.; Hartmann, A.; Bedke, J.; Schwab, M. Metabolic and Lipidomic Reprogramming in Renal Cell Carcinoma Subtypes Reflects Regions of Tumor Origin. *Eur. Urol. Focus* **2019**, *5*, 608–618.
- (32) Buck, A.; Ly, A.; Balluff, B.; Sun, N.; Gorzalka, K.; Feuchtinger, A.; Janssen, K. P.; Kuppen, P. J.; van de Velde, C. J.; Weirich, G.

- Erlmeier, F.; Langer, R.; Aubele, M.; Zitzelsberger, H.; Aichler, M.; Walch, A. High-resolution MALDI-FT-ICR MS imaging for the analysis of metabolites from formalin-fixed, paraffin-embedded clinical tissue samples. *J. Pathol.* **2015**, *237*, 123–132.
- (33) Catchpole, G.; Platzer, A.; Weikert, C.; Kempkensteffen, C.; Johannsen, M.; Krause, H.; Jung, K.; Miller, K.; Willmitzer, L.; Selbig, J.; Weikert, S. Metabolic profiling reveals key metabolic features of renal cell carcinoma. *J. Cell Mol. Med.* **2011**, *15*, 109–118.
- (34) Sato, T.; Kawasaki, Y.; Maekawa, M.; Takasaki, S.; Saigusa, D.; Ota, H.; Shimada, S.; Yamashita, S.; Mitsuzuka, K.; Yamaguchi, H.; Ito, A.; Kinoshita, K.; Koshiba, S.; Mano, N.; Arai, Y. Value of global metabolomics in association with diagnosis and clinicopathological factors of renal cell carcinoma. *Int. J. Cancer* **2019**, *145*, 484–493.
- (35) Nizioł, J.; Bonifay, V.; Ossolinski, K.; Ossolinski, T.; Ossolinska, A.; Sunner, J.; Beech, I.; Arendowski, A.; Ruman, T. Metabolomic study of human tissue and urine in clear cell renal carcinoma by LC-HRMS and PLS-DA. *Anal. Bioanal. Chem.* **2018**, *410*, 3859–3869.
- (36) Cifková, E.; Holcapek, M.; Lisa, M.; Vrana, D.; Melichar, B.; Student, V. Lipidomic differentiation between human kidney tumors and surrounding normal tissues using HILIC-HPLC/ESI-MS and multivariate data analysis. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2015**, *1000*, 14–21.
- (37) Saito, K.; Arai, E.; Maekawa, K.; Ishikawa, M.; Fujimoto, H.; Taguchi, R.; Matsumoto, K.; Kanai, Y.; Saito, Y. Lipidomic Signatures and Associated Transcriptomic Profiles of Clear Cell Renal Cell Carcinoma. *Sci. Rep.* **2016**, *6*, No. 28932.
- (38) Warrack, B. M.; Hnatyshyn, S.; Ott, K. H.; Reily, M. D.; Sanders, M.; Zhang, H.; Drexler, D. M. Normalization strategies for metabolomic analysis of urine samples. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2009**, *877*, 547–552.
- (39) Lin, L.; Yu, Q.; Yan, X.; Hang, W.; Zheng, J.; Xing, J.; Huang, B. Direct infusion mass spectrometry or liquid chromatography mass spectrometry for human metabolomics? A serum metabolomic study of kidney cancer. *Analyst* **2010**, *135*, 2970–2978.
- (40) Lin, L.; Huang, Z.; Gao, Y.; Yan, X.; Xing, J.; Hang, W. LC-MS based serum metabolomic analysis for renal cell carcinoma diagnosis, staging, and biomarker discovery. *J. Proteome Res.* **2011**, *10*, 1396–1405.
- (41) Zheng, H.; Ji, J.; Zhao, L.; Chen, M.; Shi, A.; Pan, L.; Huang, Y.; Zhang, H.; Dong, B.; Gao, H. Prediction and diagnosis of renal cell carcinoma using nuclear magnetic resonance-based serum metabolomics and self-organizing maps. *Oncotarget* **2016**, *7*, 59189–59198.
- (42) Zira, A. N.; Theocharis, S. E.; Mitropoulos, D.; Migdalis, V.; Mikros, E. (1)H NMR metabolomic analysis in renal cell carcinoma: a possible diagnostic tool. *J. Proteome Res.* **2010**, *9*, 4038–4044.
- (43) Sülentrop, F.; Moka, D.; Neubauer, S.; Haupt, G.; Engelmann, U.; Hahn, J.; Schicha, H. 31P NMR spectroscopy of blood plasma: determination and quantification of phospholipid classes in patients with renal cell carcinoma. *NMR Biomed.* **2002**, *15*, 60–68.
- (44) Liu, X.; Zhang, M.; Cheng, X.; Liu, X.; Sun, H.; Guo, Z.; Li, J.; Tang, X.; Wang, Z.; Sun, W.; Zhang, Y.; Ji, Z. LC-MS-Based Plasma Metabolomics and Lipidomics Analyses for Differential Diagnosis of Bladder Cancer and Renal Cell Carcinoma. *Front. Oncol.* **2020**, *10*, No. 717.
- (45) Kennedy, A. D.; Wittmann, B. M.; Evans, A. M.; Miller, L. A. D.; Toal, D. R.; Lonergan, S.; Elsea, S. H.; Pappan, K. L. Metabolomics in the clinic: A review of the shared and unique features of untargeted metabolomics for clinical research and clinical testing. *J. Mass Spectrom.* **2018**, *53*, 1143–1154.
- (46) Xia, J.; Broadhurst, D. I.; Wilson, M.; Wishart, D. S. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics* **2013**, *9*, 280–299.
- (47) van Meer, G.; de Kroon, A. I. Lipid map of the mammalian cell. *J. Cell Sci.* **2011**, *124*, 5–8.
- (48) Perrotti, F.; Rosa, C.; Cicalini, I.; Sacchetta, P.; Del Boccio, P.; Genovesi, D.; Pieragostino, D. Advances in Lipidomics for Cancer Biomarkers Discovery. *Int. J. Mol. Sci.* **2016**, *17*, No. 1992.
- (49) Linehan, W. M.; Srinivasan, R.; Schmidt, L. S. The genetic basis of kidney cancer: a metabolic disease. *Nat. Rev. Urol.* **2010**, *7*, 277–285.
- (50) Bylesjö, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemom.* **2006**, *20*, 341–351.
- (51) Gebhard, R. L.; Clayman, R. V.; Prigge, W. F.; Figenshau, R.; Staley, N. A.; Reese, C.; Bear, A. Abnormal cholesterol metabolism in renal clear cell carcinoma. *J. Lipid Res.* **1987**, *28*, 1177–1184.
- (52) Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. B* **1996**, *58*, 267–288.
- (53) Friedman, J.; Hastie, T.; Tibshirani, R. *The Elements of Statistical Learning*, 2nd ed.; Springer-Verlag: New York, NY, 2009; pp 417–455.
- (54) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reily, M. D.; Thaden, J. J.; Viant, M. R. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **2007**, *3*, 211–221.
- (55) Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites* **2018**, *8*, No. 31.
- (56) Sarafian, M. H.; Gaudin, M.; Lewis, M. R.; Martin, F. P.; Holmes, E.; Nicholson, J. K.; Dumas, M. E. Objective set of criteria for optimization of sample preparation procedures for ultra-high throughput untargeted blood plasma lipid profiling by ultra performance liquid chromatography-mass spectrometry. *Anal. Chem.* **2014**, *86*, 5766–5774.
- (57) Polson, C.; Sarkar, P.; Incedon, B.; Raguvaran, V.; Grant, R. Optimization of protein precipitation based upon effectiveness of protein removal and ionization effect in liquid chromatography-tandem mass spectrometry. *J. Chromatogr. B* **2003**, *785*, 263–275.
- (58) Viant, M. R.; Ebbels, T. M. D.; Beger, R. D.; Ekman, D. R.; Epps, D. J. T.; Kamp, H.; Leonards, P. E. G.; Loizou, G. D.; MacRae, J. I.; van Ravenzwaay, B.; Rocca-Serra, P.; Salek, R. M.; Walk, T.; Weber, R. J. M. Use cases, best practice and reporting standards for metabolomics in regulatory toxicology. *Nat. Commun.* **2019**, *10*, No. 3041.
- (59) Bateman, K. P.; Castro-Perez, J.; Wrona, M.; Shockcor, J. P.; Yu, K.; Oballa, R.; Nicoll-Griffith, D. A. MSE with mass defect filtering for in vitro and in vivo metabolite identification. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 1485–1496.
- (60) Haug, K.; Cochrane, K.; Nainala, V. C.; Williams, M.; Chang, J.; Jayaseelan, K. V.; O'Donovan, C. MetaboLights: a resource evolving in response to the needs of its scientific community. *Nucleic Acids Res.* **2020**, *48*, D440–D444.
- (61) Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2011**, *6*, 1060–1083.
- (62) Kirwan, J. A.; Weber, R. J.; Broadhurst, D. I.; Viant, M. R. Direct infusion mass spectrometry metabolomics dataset: a benchmark for data processing and quality control. *Sci. Data* **2014**, *1*, No. 140012.
- (63) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. B* **1995**, *57*, 289–300.
- (64) Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E. A.; Glass, C. K.; Merrill, A. H., Jr.; Murphy, R. C.; Raetz, C. R.; Russell, D. W.; Subramaniam, S. LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* **2007**, *35*, D527–D532.
- (65) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.;

- Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeronic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. HMDB: the Human Metabolome Database. *Nucleic Acids Res.* **2007**, *35*, D521–D526.
- (66) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **2005**, *27*, 747–751.
- (67) Godzien, J.; Ciborowski, M.; Martinez-Alcazar, M. P.; Samczuk, P.; Kretowski, A.; Barbas, C. Rapid and Reliable Identification of Phospholipids for Untargeted Metabolomics with LC-ESI-QTOF-MS/MS. *J. Proteome Res.* **2015**, *14*, 3204–3216.
- (68) Hou, W.; Zhou, H.; Bou Khalil, M.; Seebun, D.; Bennett, S. A.; Figeys, D. Lyso-form fragment ions facilitate the determination of stereospecificity of diacyl glycerophospholipids. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 205–217.
- (69) Murphy, R. C. *Tandem Mass Spectrometry of Lipids: Molecular Analysis of Complex Lipids*; Royal Society of Chemistry: Cambridge, U.K., 2015; pp 1–270.
- (70) Liebisch, G.; Vizcaino, J. A.; Kofeler, H.; Trotschmuller, M.; Griffiths, W. J.; Schmitz, G.; Spener, F.; Wakelam, M. J. Shorthand notation for lipid structures derived from mass spectrometry. *J. Lipid Res.* **2013**, *54*, 1523–1530.
- (71) Pauling, J. K.; Hermansson, M.; Hartler, J.; Christiansen, K.; Gallego, S. F.; Peng, B.; Ahrends, R.; Ejsing, C. S. Proposal for a common nomenclature for fragment ions in mass spectra of lipids. *PLoS One* **2017**, *12*, No. e0188394.
- (72) Liebisch, G.; Ahrends, R.; Arita, M.; Arita, M.; Bowden, J. A.; Ejsing, C. S.; Griffiths, W. J.; Holcápek, M.; Köfeler, H.; Mitchell, T. W.; Wenk, M. R.; Ekroos, K. Lipidomics Standards Initiative, C., Lipidomics needs more standardization. *Nat. Metab.* **2019**, *1*, 745–747.
- (73) Goodacre, R.; Broadhurst, D.; Smilde, A. K.; Kristal, B. S.; Baker, J. D.; Beger, R.; Bessant, C.; Connor, S.; Capuani, G.; Craig, A.; Ebbels, T.; Kell, D. B.; Manetti, C.; Newton, J.; Paternostro, G.; Somorjai, R.; Sjöström, M.; Trygg, J.; Wulfert, F. Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics* **2007**, *3*, 231–241.
- (74) Dudzik, D.; Barbas-Bernardos, C.; Garcia, A.; Barbas, C. Quality assurance procedures for mass spectrometry untargeted metabolomics. a review. *J. Pharm. Biomed. Anal.* **2018**, *147*, 149–173.
- (75) Trötz Müller, M.; Guo, X.; Fauland, A.; Kofeler, H.; Lankmayr, E. Characteristics and origins of common chemical noise ions in negative ESI LC-MS. *J. Mass Spectrom.* **2011**, *46*, 553–560.
- (76) Koelmel, J. P.; Ulmer, C. Z.; Jones, C. M.; Yost, R. A.; Bowden, J. A. Common cases of improper lipid annotation using high-resolution tandem mass spectrometry data and corresponding limitations in biological interpretation. *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2017**, *1862*, 766–770.
- (77) Knott, M. E.; Manzi, M.; Zabalegui, N.; Salazar, M. O.; Puricelli, L. I.; Monge, M. E. Metabolic Footprinting of a Clear Cell Renal Cell Carcinoma in Vitro Model for Human Kidney Cancer Detection. *J. Proteome Res.* **2018**, *17*, 3877–3888.
- (78) Gathungu, R. M.; Larrea, P.; Sniatynski, M. J.; Marur, V. R.; Bowden, J. A.; Koelmel, J. P.; Starke-Reed, P.; Hubbard, V. S.; Kristal, B. S. Optimization of Electrospray Ionization Source Parameters for Lipidomics To Reduce Misannotation of In-Source Fragments as Precursor Ions. *Anal. Chem.* **2018**, *90*, 13523–13532.
- (79) Godzien, J.; Armitage, E. G.; Angulo, S.; Martinez-Alcazar, M. P.; Alonso-Herranz, V.; Otero, A.; Lopez-Gonzalez, A.; Barbas, C. In-source fragmentation and correlation analysis as tools for metabolite identification exemplified with CE-TOF untargeted metabolomics. *Electrophoresis* **2015**, *36*, 2188–2195.
- (80) Hsu, F. F.; Turk, J. Charge-remote and charge-driven fragmentation processes in diacyl glycerophosphoethanolamine upon low-energy collisional activation: a mechanistic proposal. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 892–899.
- (81) Hvattum, E.; Hagelin, G.; Larsen, A. Study of mechanisms involved in the collision-induced dissociation of carboxylate anions from glycerophospholipids using negative ion electrospray tandem quadrupole mass spectrometry. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1405–1409.
- (82) Ricketts, C. J.; De Cubas, A. A.; Fan, H.; Smith, C. C.; Lang, M.; Reznik, E.; Bowlby, R.; Gibb, E. A.; Akbani, R.; Beroukhi, R.; Bottaro, D. P.; Choueiri, T. K.; Gibbs, R. A.; Godwin, A. K.; Haake, S.; Hakimi, A. A.; Henske, E. P.; Hsieh, J. J.; Ho, T. H.; Kanchi, R. S.; Krishnan, B.; Kwiatkowski, D. J.; Lui, W.; Merino, M. J.; Mills, G. B.; Myers, J.; Nickerson, M. L.; Reuter, V. E.; Schmidt, L. S.; Shelley, C. S.; Shen, H.; Shuch, B.; Signoretti, S.; Srinivasan, R.; Tamboli, P.; Thomas, G.; Vincent, B. G.; Vocke, C. D.; Wheeler, D. A.; Yang, L.; Kim, W. Y.; Robertson, A. G.; Spellman, P. T.; Rathmell, W. K.; Linehan, W. M.; et al. The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep.* **2018**, *23*, 313.e5–326.e5.
- (83) Hsu, F. F.; Turk, J.; Thukkani, A. K.; Messner, M. C.; Wildsmith, K. R.; Ford, D. A. Characterization of alkylacyl, alk-1-enylacyl and lyso subclasses of glycerophosphocholine by tandem quadrupole mass spectrometry with electrospray ionization. *J. Mass Spectrom.* **2003**, *38*, 752–763.
- (84) Hsu, F. F.; Turk, J. Differentiation of 1-O-alk-1'-enyl-2-acyl and 1-O-alkyl-2-acyl glycerophospholipids by multiple-stage linear ion-trap mass spectrometry with electrospray ionization. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 2065–2073.
- (85) Godzien, J.; Kalaska, B.; Adamska-Patruno, E.; Siroka, J.; Ciborowski, M.; Kretowski, A.; Barbas, C. Oxidized glycerophosphatidylcholines in diabetes through non-targeted metabolomics: Their annotation and biological meaning. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2019**, *1120*, 62–70.
- (86) Gil de la Fuente, A.; Traldi, F.; Siroka, J.; Kretowski, A.; Ciborowski, M.; Otero, A.; Barbas, C.; Godzien, J. Characterization and annotation of oxidized glycerophosphocholines for non-targeted metabolomics with LC-QTOF-MS data. *Anal. Chim. Acta* **2018**, *1037*, 358–368.
- (87) Gil-de-la-Fuente, A.; Godzien, J.; Saugar, S.; Garcia-Carmona, R.; Badran, H.; Wishart, D. S.; Barbas, C.; Otero, A. CEU Mass Mediator 3.0: A Metabolite Annotation Tool. *J. Proteome Res.* **2019**, *18*, 797–802.
- (88) Ni, Z.; Angelidou, G.; Hoffmann, R.; Fedorova, M. LPPtger software for lipidome-specific prediction and identification of oxidized phospholipids from LC-MS datasets. *Sci. Rep.* **2017**, *7*, No. 15138.
- (89) López-López, Á.; Godzien, J.; Soldevilla, B.; Gradillas, A.; López-González, A.; Lens-Pardo, A.; La Salvia, A.; del Carmen Riesco-Martínez, M.; García-Carbonero, R.; Barbas, C. Oxidized lipids in the metabolic profiling of neuroendocrine tumors – Analytical challenges and biological implications. *J. Chromatogr., A* **2020**, *1625*, No. 461233.
- (90) Okudaira, M.; Inoue, A.; Shuto, A.; Nakanaga, K.; Kano, K.; Makide, K.; Saigusa, D.; Tomioka, Y.; Aoki, J. Separation and quantification of 2-acyl-1-lysophospholipids and 1-acyl-2-lysophospholipids in biological samples by LC-MS/MS. *J. Lipid Res.* **2014**, *55*, 2178–2192.
- (91) Dong, J.; Cai, X.; Zhao, L.; Xue, X.; Zou, L.; Zhang, X.; Liang, X. Lysophosphatidylcholine profiling of plasma: discrimination of isomers and discovery of lung cancer biomarkers. *Metabolomics* **2010**, *6*, 478–488.
- (92) Creer, M. H.; Gross, R. W. Separation of isomeric lysophospholipids by reverse phase HPLC. *Lipids* **1985**, *20*, 922–928.
- (93) von Roemeling, C. A.; Marlow, L. A.; Wei, J. J.; Cooper, S. J.; Caulfield, T. R.; Wu, K.; Tan, W. W.; Tun, H. W.; Copland, J. A. Stearoyl-CoA desaturase 1 is a novel molecular therapeutic target for clear cell renal cell carcinoma. *Clin. Cancer Res.* **2013**, *19*, 2368–2380.
- (94) Horiguchi, A.; Asano, T.; Asano, T.; Ito, K.; Sumitomo, M.; Hayakawa, M. Fatty acid synthase over expression is an indicator of

tumor aggressiveness and poor prognosis in renal cell carcinoma. *J. Urol.* **2008**, *180*, 1137–1140.

(95) Zha, S.; Ferdinandusse, S.; Hicks, J. L.; Denis, S.; Dunn, T. A.; Wanders, R. J.; Luo, J.; De Marzo, A. M.; Isaacs, W. B. Peroxisomal branched chain fatty acid beta-oxidation pathway is upregulated in prostate cancer. *Prostate* **2005**, *63*, 316–323.

(96) Liu, Y. Fatty acid oxidation is a dominant bioenergetic pathway in prostate cancer. *Prostate Cancer Prostatic Dis.* **2006**, *9*, 230–234.

(97) Vijayalakshmi, K.; Shankar, V.; Bain, R. M.; Nolley, R.; Sonn, G. A.; Kao, C. S.; Zhao, H.; Tibshirani, R.; Zare, R. N.; Brooks, J. D. Identification of diagnostic metabolic signatures in clear cell renal cell carcinoma using mass spectrometry imaging. *Int. J. Cancer* **2020**, *147*, 256–265.

(98) Tamura, K.; Horikawa, M.; Sato, S.; Miyake, H.; Setou, M. Discovery of lipid biomarkers correlated with disease progression in clear cell renal cell carcinoma using desorption electrospray ionization imaging mass spectrometry. *Oncotarget* **2019**, *10*, 1688–1703.

(99) Du, W.; Zhang, L.; Brett-Morris, A.; Aguila, B.; Kerner, J.; Hoppel, C. L.; Puchowicz, M.; Serra, D.; Herrero, L.; Rini, B. I.; Campbell, S.; Welford, S. M. HIF drives lipid deposition and cancer in ccRCC via repression of fatty acid metabolism. *Nat. Commun.* **2017**, *8*, No. 1769.

(100) Tang, Y. M.; Wang, J. P.; Bao, W. M.; Yang, J. H.; Ma, L. K.; Yang, J.; Chen, H.; Xu, Y.; Yang, L. H.; Li, W.; Zhu, Y. P.; Cheng, J. B. Urine and serum metabolomic profiling reveals that bile acids and carnitine may be potential biomarkers of primary biliary cirrhosis. *Int. J. Mol. Med.* **2015**, *36*, 377–385.

(101) Chen, T.; Xie, G.; Wang, X.; Fan, J.; Qiu, Y.; Zheng, X.; Qi, X.; Cao, Y.; Su, M.; Xu, L. X.; Yen, Y.; Liu, P.; Jia, W. Serum and urine metabolite profiling reveals potential biomarkers of human hepatocellular carcinoma. *Mol. Cell. Proteomics* **2011**, *10*, No. M110.004945.

(102) Liang, Q.; Wang, C.; Li, B. Metabolomic Analysis Using Liquid Chromatography/Mass Spectrometry for Gastric Cancer. *Appl. Biochem. Biotechnol.* **2015**, *176*, 2170–2184.

(103) Wang, S.; Dong, W.; Liu, L.; Xu, M.; Wang, Y.; Liu, T.; Zhang, Y.; Wang, B.; Cao, H. Interplay between bile acids and the gut microbiota promotes intestinal carcinogenesis. *Mol. Carcinog.* **2019**, *58*, 1155–1167.

(104) Adam, O.; Tesche, A.; Wolfram, G. Impact of linoleic acid intake on arachidonic acid formation and eicosanoid biosynthesis in humans. *Prostaglandins, Leukotrienes Essent. Fatty Acids* **2008**, *79*, 177–181.

(105) Fatima, S.; Hu, X.; Gong, R. H.; Huang, C.; Chen, M.; Wong, H. L. X.; Bian, Z.; Kwan, H. Y. Palmitic acid is an intracellular signaling molecule involved in disease development. *Cell. Mol. Life Sci.* **2019**, *76*, 2547–2557.

(106) Louie, S. M.; Roberts, L. S.; Mulvihill, M. M.; Luo, K.; Nomura, D. K. Cancer cells incorporate and remodel exogenous palmitate into structural and oncogenic signaling lipids. *Biochim. Biophys. Acta, Mol. Cell Res.* **2013**, *1831*, 1566–1572.

(107) Reuter, S.; Gupta, S. C.; Chaturvedi, M. M.; Aggarwal, B. B. Oxidative stress, inflammation, and cancer: how are they linked? *Free Radical Biol. Med.* **2010**, *49*, 1603–1616.

(108) Gerber, M. Background review paper on total fat, fatty acid intake and cancers. *Ann. Nutr. Metab.* **2009**, *55*, 140–161.

(109) Zhang, M.; Liu, X.; Liu, X.; Li, H.; Sun, W.; Zhang, Y. A pilot investigation of a urinary metabolic biomarker discovery in renal cell carcinoma. *Int. Urol. Nephrol.* **2020**, *52*, 437–446.

(110) Ulmer, C. Z.; Koelmel, J. P.; Jones, C. M.; Garrett, T. J.; Aristizabal-Henao, J. J.; Vesper, H. W.; Bowden, J. A. A Review of Efforts to Improve Lipid Stability during Sample Preparation and Standardization Efforts to Ensure Accuracy in the Reporting of Lipid Measurements. *Lipids* **2020**, DOI: 10.1002/lipd.12263.