Plasma Lipidomic Profiles of Kidney, Breast, and Prostate Cancer Patients Differ from Healthy Controls

Denise Wolrab

University of Pardubice

Robert Jirásko

University of Pardubice

Ondřej Peterka

University of Pardubice

Jakub Idkowiak

University of Pardubice

Michaela Chocholoušková

University of Pardubice

Zuzana Vaňková

University of Pardubice

Karel Hořejší

University of Pardubice

Ivana Brabcová

University of Pardubice

David Vrána

Palacký University and University Hospital

Hana Študentová

Palacký University and University Hospital

Bohuslav Melichar

Palacký University and University Hospital

Michal Holčapek (✉ Michal.Holcapek@upce.cz)

University of Pardubice

Research Article

Keywords: Plasma , kidney, breast, healthy controls

Posted Date: July 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-734203/v1>

Abstract

Early cancer screening is one of the unmet needs in clinical medicine. The peripheral blood analysis is a preferred method for efficient population screening as blood collection is well embedded in clinical practice and minimally invasive for patients. Lipids are important biomolecules, and variations in lipid concentrations may reflect pathological disorders. The lipidomic profiling by ultrahigh-performance supercritical fluid chromatography hyphenated to mass spectrometry of human plasma for distinguishing samples obtained from breast, kidney, and prostate cancer patients and healthy controls is investigated. The mean sensitivity, specificity, and accuracy of the new lipidomic profiling approach were 85%, 95%, and 92% for kidney cancer; 91%, 97%, and 94% for breast cancer; and 87%, 95%, and 92% for prostate cancer. No association of statistical models with tumor stage is observed. The statistically most significant lipid species for differentiation of studied cancer types are CE 16:0, Cer 42:1, LPC 18:2, PC 36:2, PC 36:3, SM 32:1, and SM 41:1.

Introduction

Cancer incidence and mortality are increasing worldwide as a result of aging as well as changing patterns of other risk factors¹. Malignant disorders are categorized according to the site where the tumor growth has initiated, regardless of subsequent metastatic spread to other parts of the body². Prostate cancer is the second most commonly diagnosed cancer in men³. Breast cancer is the most commonly diagnosed cancer and one of the principal causes of cancer-related death in women⁴. On the other hand, kidney cancer is the 9th most common cancer in males and the 14th most common cancer in females⁵. However, there is a marked geographical variation of the incidence rate with the highest kidney cancer incidence rates for males in the Czech Republic among all European countries⁶. The first tests to detect prostate cancer include prostate-specific antigen (PSA) levels in peripheral blood and digital rectal examination (DRE). In case of abnormal DRE or elevated PSA levels, transrectal ultrasound-guided biopsy is performed for verification⁷. While mammography represents the principal method for breast cancer screening, the diagnosis is commonly complemented with other imaging methods including magnetic resonance imaging, positron-emission tomography, computed tomography, or single-photon emission computed tomography⁸. Kidney cancer is often discovered by chance during the examination with imaging methods⁹, such as ultrasound scanning, computed tomography, or magnetic resonance imaging¹⁰. Generally, all imaging methods are subject to the limitation that very small tumors are not properly visualized, resulting in the low sensitivity for early stage¹¹. Staging examinations are performed after the cancer diagnosis based on the information regarding the location, spread, and extent of tumor¹². The diagnosis and treatment of patients at the early stage of cancer increases the chance for survival and cure compared to patients with the diagnosis at the late stage. Cancer screening methods aim at the detection of cancer at the early stage for high-risk individuals¹³. Recently, the research interest increases in developing cancer screening methods based on examination of peripheral blood increased, including liquid biopsy. The analysis of circulating cells, platelets, extracellular vesicles, mRNA, miRNA, proteins,

cell-free DNA (cfDNA), and circulating tumor DNA (ctDNA) in blood are currently being investigated as potential approaches in cancer screening¹⁴. Recently, the successful detection of eight cancer types based on the analysis of proteins and mutations in ctDNA in blood was reported¹⁵. Metabolomics also attracts research attention in cancer screening¹⁶. Lipidomics can be considered a part of metabolomics that deals with the comprehensive analysis of lipids as important biomolecules involved in many biological processes, such as signaling molecules or constituents of cell membranes, energy storage, and various other metabolic pathways¹⁷. Clinical lipidomics revealed that plasma lipid concentrations can change for various malignant diseases¹⁸.

Here we aim to quantitatively determine the plasma lipidome of kidney, breast, and prostate cancer patients and compare it with the lipidome of healthy controls. The lipidomic changes are investigated and visualized using multivariate data analysis (MDA) and other statistical tools. The ultrahigh-performance supercritical fluid chromatography - mass spectrometry (UHPSFC/MS) is used as a powerful high-throughput and sensitive method for quantitative lipidomic analysis based on the lipid class separation approach recommended for reliable quantitation together with the use of exogenous internal standards for individual lipid classes^{19,20}.

Results

Study design

Heparin plasma samples from 289 cancer patients and 192 volunteers without the history of previous malignant disease (further referred to as healthy controls) were obtained. The patients were diagnosed with breast, prostate, or kidney cancer based on the standard medical procedures at the University Hospital in Olomouc. The sample set was divided into training and validation sets, where about 25% of each sample type assigned to the validation set. Finally, the training set included 135 samples from healthy controls, 209 samples obtained from cancer patients (77 breast, 82 kidney, and 50 prostate cancers), and the validation set included 57 samples obtained from healthy control donors and 80 samples from cancer patients (26 breast, 37 kidney, and 17 prostate cancers). The overview of all samples and the clinical information are summarized in Fig. 1 and Supplementary Tables S1 and S2. The average age of healthy volunteers was lower than patients and the average body mass index (BMI) was comparable for both sample groups. Cancer patients are categorized according to the TNM system. The majority of samples is assigned as T1 stage, typically obtained from patients with breast (59%) and kidney cancer (47%), whereby T2 stage is predominant for prostate cancer (69%).

Previous studies suggested minor differences in plasma lipidome depending on gender^{21–24}. For the studied sample set, the prediction performance using MDA for both genders and gender-separated models was performed (Supplementary Fig. S1 showing sensitivity, specificity, and accuracy values for the training and validation sets together with OPLS-DA models). The accuracy was slightly higher for gender-separated models, in particular for females. Therefore, the sample set was divided according to

gender. Obviously, prostate cancer occurs only in males, and the overwhelming majority of breast cancer patients are females, so only for kidney cancer samples the gender separation is an important issue with 73% of male samples and 27% of female samples in this study.

Discovery phase measured by UHPSFC/MS

The order of samples was randomized separately for the extraction and UHPSFC/MS measurements to exclude any possible biases. The plasma lipidome analysis in the discovery phase resulted in the quantitation of 138 lipids (Supplementary Table S3a) belonging to glycerolipids, glycerophospholipids, and sphingolipids. Non-supervised principal component analysis (PCA) and supervised orthogonal projection to latent structures discriminant analysis (OPLS-DA) were applied for all training set samples to visualize differences between sample groups (healthy controls and cancer patients) for three studied cancer types (Fig. 2a-d). MDA allows the prediction of samples to belong to a particular sample group. The samples from the validation set were predicted by the corresponding OPLS-DA model built on the training set samples (Supplementary Tables S4 and S5). The specificity, sensitivity, and accuracy of the models for kidney cancer patients were 91%, 73%, and 82% for males and 88%, 71%, and 84% for females, for female breast cancer patients 88%, 63%, and 76%, for prostate cancer patients 90%, 82%, and 87%. The specificity and sensitivity values depending on the cancer stage and accuracy values depending on the cancer type are summarized in Fig. 3 for training and validation sets. The prediction performance is only slightly higher for the training set than the validation set, which indicates the possible use for samples with the unknown classification. The receiver operating characteristics (ROC) curves for the diagnostic ability to classify healthy control samples or cancer samples are illustrated in Fig. 3i-l for individual cancer types. The AUC values for the different cancer types ranged from 0.917 to 0.967 for the training set and from 0.868 to 0.953 for the validation set.

Qualification phase measured by UHPSFC/MS

UHPSFC/MS measurements were repeated after several months to verify the repeatability of the results, and repeated measurements are called as the qualification phase. The same sample extracts were measured using different sample measurement sequences to minimize the risk of method-induced biases. A total of 138 lipids were also quantified in the qualification phase (Supplementary Table S6a), with 126 of 138 lipids (91%) quantified in both the discovery and qualification phase. Lipid identification differences were mainly observed for low abundant short fatty acyl glycerolipids. PCA score plots for UHPSFC/MS measurements were compared in Fig. 4a,b. Both data sets show the quality control (QC) samples cluster in the PCA score plot, indicating satisfactory method stability during the measurement sequence. The partial group separation is observed for samples obtained from healthy controls and cancer patients. PCA score plots confirm a high reproducibility of the lipidomic profiling, as illustrated by the numbers of selected samples in PCA plots in Fig. 4a,b. The first and the second data sets were compared by calculating the relative standard deviation (RSD) for each lipid in each sample (Supplementary Table S7). In total, 65 % of all values have RSD < 20%, and the average of all RSD for each lipid and all samples is 19%. Figure 4c-e further illustrates the reproducibility of quantitative results for selected dysregulated lipids in the discovery and qualification phases, as the medians of the box plots

of the first and second measurements are comparable. Furthermore, the box plots also show that the selected lipid species are downregulated in all cancer types compared to the control group. MDA was also applied for the repeated measurements using the training sample set for building models. Generally, the prediction performance was comparable for the discovery and qualification phase, and the results are summarized in Supplementary Tables S4 and S5. ROC curves are summarized for training and validation sets in Fig. 5a-d for all cancer types using UHPSFC/MS. AUC values ranged from 0.888 to 0.994.

Furthermore, MDA models for the discovery phase were used to predict the sample set of the qualification phase (Supplementary Tables S4 and S8). The specificity, sensitivity, and accuracy of the models for kidney cancer patients were 62%, 91%, and 76% for males and 86%, 72, and 83% for females, for female breast cancer patients 94%, 57%, and 76%, for prostate cancer patients 88%, 75%, and 82%. ROC curves are summarized for all samples in Fig. 5e-h for all cancer types. AUC values ranged from 0.864 to 0.901.

Shotgun MS measurements

To further verify that plasma lipidomics profiling can be used for the prediction of the diagnosis in samples obtained from healthy control, kidney, breast, or prostate cancer patients, shotgun MS was applied as an alternative technique to verify conclusions from UHPSFC/MS measurements. 412 lipid species were quantified by shotgun MS for glycerolipid, glycerophospholipid, and sphingolipid categories (Supplementary Table S6c), which is about 3 times more quantified lipid species in comparison to UHPSFC/MS. The MDA prediction performance was comparable with shotgun MS and UHPSFC/MS (Supplementary Tables S4 and S5), suggesting that the whole lipidomic profile is important for the prediction, not only the selected lipid species. ROC curves for training and validation sets are summarized in Fig. 5i-l for all cancer types. AUC values ranged from 0.841 to 1.00. For two of the most significant regulated lipid species, the box plots are presented in Fig. 5m,n for all pathological states and data sets using UHPSFC/MS and shotgun MS.

Influence of quantified lipid species on the prediction

The influence of the number of quantified lipid species used for MDA on the accuracy to correctly classify samples obtained from various pathological disorders was investigated. First, lipid species common in the discovery and qualification phases for the data sets obtained with UHPSFC/MS and shotgun MS were investigated. In total, 91 common lipid species were identified and used for MDA in each data set. The overall prediction performance was comparable to that obtained when all quantified lipid species were used for MDA, independent of the employed method and diagnosis (Supplementary Tables S4 and S9). The number of lipid species included to build the MDA models for prediction was further reduced by considering the statistical parameters, such as fold change ($\pm 20\%$), p-value (< 0.05), and VIP value (> 1), obtained for individual methods and sample types. Supplementary Table S10 provided supplementary information how variables were reduced. The whole data set was divided into 5 data subsets (healthy control vs. kidney cancer samples for males and females, healthy control vs. breast cancer samples for males and females, and healthy control vs. prostate cancer samples for males) for UHPSFC/MS (1st and 2nd measurements) and shotgun MS, which resulted in 15 data subsets. In total, 29 lipid species were statistically significant after the Bonferroni correction in > 10 from 15 data subsets considering all cancer

types, methods, and measurements. The accuracy slightly decreased with decreasing number of lipids used for MDA independent of the investigated cancer type and method (Supplementary Tables S4 and S11). However, as the decrease of the prediction performance was not so pronounced, the effect of further reduction of lipids used for MDA on the prediction performance to correctly assign the sample type was investigated. CE 16:0, Cer 42:1, LPC 18:2, PC 36:2, PC 36:3, SM 32:1, and SM 41:1 were significant according to the Bonferroni correction in > 14 of the 15 data subsets considering all cancer types, methods and measurements. MDA for these 7 lipid species was performed and the prediction performance was evaluated (Supplementary Tables S4 and S12). The average of sensitivity, specificity, and accuracy values for the different number of lipids considering all methods and genders was calculated (Fig. 6). Generally, the sensitivity and consequently the accuracy decreased with decreasing number of lipids used for MDA. The specificity was not affected by the number of lipid species, independent of the cancer type (Fig. 6). No effect of the cancer stage on the concentrations of the most significant lipid species was observed for all cancer types (Supplementary Fig. S2).

Statistical evaluation of data

The different plasma lipidomic profiles depending on the cancer type were investigated by evaluating statistically significant lipid species after the Bonferroni correction, lipid species with the fold change of $\pm 20\%$, and VIP value > 1. The percentage of lipid species belonging to the lipid class fulfilling the defined criteria were calculated, as illustrated by the pie charts for different cancer types in Fig. 7a-d. The nonpolar lipid species, triacylglycerols and cholesterol esters, are of higher relevance in kidney cancer, while the influence of glycerophospholipids and sphingolipids appears to be dominant in breast and prostate cancer.

The most significant lipid species for all methods and data sets are downregulated in plasma samples of cancer patients, independent of the cancer type, as illustrated in Fig. 7e-g. MDA was employed to investigate the differences between healthy control samples and different cancer types for all quantified lipid species, 91 lipid species common for UHPSFC/MS and shotgun MS, 29, and finally 7 most significant lipid species for all data sets. OPLS-DA models for samples obtained from healthy male controls and male patients suffering from prostate and kidney cancer as well as samples obtained from healthy female controls and female patients suffering from kidney and breast cancer are shown in Fig. 8a-b. The question was whether the differentiation and prediction of the cancer type and healthy control samples is possible using UHPSFC/MS. The specificity ranged from 67 to 97% with the average of 83%, sensitivity for kidney cancer from 49 to 74% with the average of 61%, sensitivity for prostate from 0 to 66% with the average of 43% and the accuracy from 57 to 74% with the average of 65% for the training and validation set and different numbers of lipids (138, 91, 29, and 7 lipid species) included to build the MDA models considering samples obtained from male donors (Fig. 8a, Supplementary Table S13). The specificity ranged from 80 to 96% with the mean of 90%, sensitivity for kidney cancer from 0 to 44% with the mean of 24%, sensitivity for breast cancer from 60 to 83% with the mean of 73% and the accuracy from 66 to 78% with the mean of 74% for the training and validation set and different numbers of lipids (138, 91, and 29 lipid species) included to build the MDA models considering samples obtained from

female donors (Fig. 8b). It was not possible to perform the MDA model using 7 lipid species as variables due to the insufficient number of components for samples obtained from female donors.

The differentiation of the cancer type was also investigated by performing OPLS-DA models classifying kidney cancer vs. prostate cancer for males (Fig. 8c) and kidney cancer vs. breast cancer for females (Fig. 8d and Supplementary Table S14). OPLS-DA models were evaluated using 138 and 91 lipids as variables for males and 138, 91, and 29 for females using UHPSFC/MS data, as for the lower number of lipids a lack of components was observed. The sensitivity for prostate cancer was 71–88% with the mean of 81% and for kidney cancer 57–82% with the mean of 72% for the training and validation set and both UHPSFC/MS data sets considering male samples. The sensitivity for breast cancer ranged between 94 and 100% with the mean of 98%, the sensitivity for kidney cancer ranged between 14 and 86% with the mean of 57%, and the accuracy ranged between 80 and 97% with the mean of 88% for the training and validation set and both UHPSFC/MS data sets considering female samples.

Discussion

Cancer screening as a part of regular health examination may allow early cancer detection and timely treatment, resulting in the improvement of the clinical outcome. The measurement of circulating biomarkers as a minimally invasive and routinely employed method seems to be one of the most attractive and convenient methods for the screening of high-risk individuals. Current approaches focus on the analysis of genetic mutations, ctDNA, or proteins for early cancer diagnosis in plasma or serum. The applicability of the lipidomic analysis for this purpose has not been clearly demonstrated so far.

In the present study, quantitative lipidomics of human plasma samples collected from healthy controls and cancer patients was performed by UHPSFC/MS. MDA revealed the applicability of lipidomics for the diagnosis with high sensitivity, specificity, and accuracy for all studied cancer types. The involvement of lipids and dependence of lipid concentration changes on the pathological condition was previously suggested²⁵, due to the multifunctional character of lipids²⁶. Previous studies investigated the lipidome in plasma or serum of patients with different cancer types¹⁸, including breast cancer^{27–29}, pancreatic cancer^{21,30}, kidney cancer³¹, lung cancer^{32,33}, and prostate cancer^{34,35}. The results of the present study are consistent with previous reports. However, the special focus of this study was put on the accurate molar quantitation of lipid species towards the possibility of future interlaboratory comparison of the results by using the same measurement protocol.

Nonpolar lipids like cholesteryl esters and triacylglycerols are more important for kidney cancer, while for breast or prostate cancer, statistically significant differences are more pronounced for polar lipids belonging to sphingolipids and glycerophospholipids. The reduction of quantified lipid species used for MDA showed only a small loss in specificity, sensitivity, and accuracy. This would reduce the method complexity in comparison to lipidomic profiling and would pave the way for clinical use. However, the investigation of a higher number of samples in a controlled prospective cohort using multiple collection sites is needed for the verification of clinical utility, which should be the next step in the journey towards

the development of clinical test. The investigation of the targeted method for 7 lipid species identified as statistically most significant, CE 16:0, Cer 42:1, LPC 18:2, PC 36:2, PC 36:3, SM 32:1, and SM 41:1, is intended.

In conclusion, the present data indicate the potential of lipidomic profiling in cancer screening at least for breast, kidney, and prostate cancers. The use of individual MDA models to distinguish healthy control samples and the single cancer type results in higher accuracy than the use of MDA models that include multiple cancer types. The use of internal standards for each lipid class allows the quantitation of lipid species and the comparison of lipid concentrations between different laboratories. Subsequent prospective studies are necessary for 7 lipid species identified as potential biomarkers for cancer screening.

Methods

Human samples.

A retrospective study on 481 human plasma samples was performed. A total of 192 control samples and 289 cancer samples from patients suffering from breast, kidney, or prostate cancer were collected. The criteria for healthy controls were included no history of any type of cancer and age over 18 years. For cancer patients, the disease was histologically confirmed by needle biopsy or by examining the surgical resection specimen. Both cancer patients and healthy controls were of the same ethnicity (Caucasian), collected at the same place (University Hospital in Olomouc), and processed in the same way. No other exclusion criteria were applied. The clinical information for all patients and controls is summarized in Fig. 1 and Supplementary Tables S1 and S2. The sample set was divided into training (using to build OPLS-DA models) and validation (indicates the possible use for samples with unknown classification) sets. Each 4th sample was assigned to the validation set, to obtain a distribution of 75% of samples belonging to the training set and 25% of samples to the validation set. Patients had no treatment before the blood collection. Human plasma was collected in 9 mL lithium-heparin collection tubes and then centrifuged. The supernatant was transferred, aliquoted, and stored at -80 °C until further processing for lipidomic analysis.

Ethics declaration.

The study was approved by the ethical committee at the University Hospital Olomouc. All subjects signed an informed consent. All methods were carried out in line with Ethical Principles for Medical Research Involving Human Subjects (Declaration of Helsinki).

Study phases.

The lipidome of 481 plasma samples was measured by UHPSFC/MS in the discovery phase. To guarantee that UHPSFC/MS results are reproducible, the same extracts were measured again several months later corresponding to the qualification phase. The sequence of sample measurements was randomized to exclude any measurement bias. The data set was independently processed and the results were compared to the discovery phase. Furthermore, the extracts were also measured with shotgun MS to exclude any bias caused by the employed method, independently processed, and compared with UHPSFC/MS results.

Chemicals.

Solvents for analysis, such as acetonitrile, 2-propanol, methanol (HPLC/MS grade), water (UHPLC/MS grade), and hexane, were purchased from Honeywell (Riedel-da Haën, CHROMASOLV™ LC-MS Ultra, Hamburg, Germany), distributed by Fisher Scientific (Waltham, Massachusetts, USA). Chloroform stabilized with 0.5-1% ethanol was purchased from either Sigma-Aldrich (St. Louis, MO, USA) or Merck (Darmstadt, Germany), respectively. Ammonium acetate was purchased from Fisher Scientific. Deionized water for liquid-liquid extraction was obtained from a Milli-Q Reference Water Purification System (Molsheim, France). Carbon dioxide of 4.5 grade (99.995%) was purchased from Messer Group (Bad Soden, Germany). Non-endogenous lipids were used as internal standards (IS) for quantitative analysis, *i.e.*, MG 19:1/0:0/0:0, DG 12:1/0:0/12:1, and TG 19:1/19:1/19:1 from Nu-ChekPrep (Elysian, MN, USA); CE 16:0 D7, Cer d18:1/12:0, cholesterol D7, LPC 17:0/0:0, LPE 14:0/0:0, PC 14:0/14:0, PC 22:1/22:1, PE 14:0/14:0, PI 15:0/18:1 D7, SM d18:1/12:0, PS 14:0/14:0, PA 14:0/14:0, PG 14:0/14:0, LPG 14:0/0:0, HexCer d18:1/12:0, Hex2Cer d18:1/12:0, and SHexCer d18:1/12:0 from Avanti Polar Lipids (Alabaster, AL, USA). The concentrations of stock solutions of individual IS and the volumes needed to prepare the IS mixture are summarized in Supplementary Table S15.

Lipidomic analysis.

For the lipid extraction, a modified Folch procedure was employed, which was previously validated³⁶. The same sample extracts were analyzed with UHPSFC/MS and shotgun MS. Human serum (25 µL) and the mixture of IS (17.5 µL) were homogenized in 3 mL of chloroform/methanol (2:1, v/v) for 10 min in an ultrasonic bath (40 °C). When the samples reached ambient temperature, 600 µL of water was added, and the mixture was vortexed for 1 min. After 3 min of centrifugation (3000 rpm), the aqueous layer was removed, and the organic layer was evaporated under a gentle stream of nitrogen. The residue was dissolved in a mixture of 500 µL of chloroform/2-propanol (1:1, v/v), carefully vortexed and filtered (0.2 µm syringe filter). The extract was diluted 1:20 with the mixture of hexane/2-propanol/chloroform (7:1.5:1.5, v/v/v) for UHPSFC/MS analysis and 1:8 with chloroform/methanol/2-propanol (1:2:4, v/v/v) mixture containing 7.5 mM of ammonium acetate and 1% of acetic acid for shotgun MS analysis.

UHPSFC/MS measurements were carried out on an Acquity Ultra Performance Convergence Chromatography (UPC2) system hyphenated to the hybrid quadrupole traveling wave ion mobility time-of-

flight mass spectrometer Synapt G2-Si from Waters by using the commercial interface kit (Waters, Milford, MA, USA). The chromatographic settings were used with minor improvements from the previously published method³⁷. The main difference is that the data were recorded in continuum and sensitivity mode. The peptide leucine enkephalin was used as the lock mass with the scan time of 0.1 s and the interval of 30 s. The lock mass was scanned but the mass correction was not automatically applied. All samples were measured in duplicates. Noise reduction was performed on the raw files using the Waters compression tool. Data files were lock mass corrected and converted into centroid data using the exact mass measure tool from Waters. The MarkerLynx software from Waters was used for data preprocessing. Further data processing was done by LipidQuant 1.0 software³⁸.

Shotgun experiments were performed on a quadrupole linear ion trap mass spectrometer 6500 QTRAP (Sciex, Concord, ON, Canada) equipped with ESI probe using the characteristic precursor ion (PIS) and neutral loss (NL) scan events³⁹. Raw data files were processed with the LipidView Software from Sciex in order to obtain a summary table of m/z vs. intensity for each scan mode (NL and PIS) of all samples. The raw data were prefiltered by applying the following settings in the positive ion mode, a tolerance mass window of 0.5 Da, a minimum intensity threshold of 0.1%, and a minimum signal-to-noise ratio of 3 after smoothing. The summary tables of m/z vs. intensity for all samples were exported as txt files and further processed by the LipidQuant 1.0 software.

Data processing.

LipidQuant 1.0 is a Microsoft Excel based script used for the automated data processing of txt files³⁸ including m/z values vs. intensities for all samples. The experimental m/z values were compared to the theoretical m/z values from the embedded database for lipid identification, depending on the retention time window or scan type defining the lipid class. The lipid quantitation was performed by calculating the ratio of the intensities of the target lipid and the internal standard and multiplying with the known concentration of the internal standard. Isotopic correction type II⁴⁰ was automatically applied and a summary table containing lipid concentrations in all samples was generated. Zero filling for missing values was applied by setting the number for 80 % of the minimum measured concentration for a given lipid species for all samples. If the concentration was not determined for more than 25% of the samples, then the lipid species was excluded from the data set. The data set was divided into training and validation set by assigning each 4th sample to the validation set. The clinical information for samples, like gender and pathological state, were revealed and samples were assigned. The final tables containing the lipid concentrations for all samples and fulfilling all defined criteria were used for MDA and other statistical tools.

Statistical analysis.

MDA was performed by the SIMCA software, version 13.0 (Umetrics, Sweden). The lipid species were defined as variables and the samples as observations. The data set was preprocessed by applying the logarithmic transformation, pareto scaling, and centering. The data preprocessing should facilitate the normal distribution of lipid concentrations and that low abundant lipid species contribute similarly to the MDA as high abundant lipid species. PCA was performed to evaluate for outliers, estimate the measurement quality by checking the clustering of QC samples, and evaluate the sample group clustering depending on the pathological state. OPLS-DA is a statistical tool to visualize differences between sample groups of known classification. OPLS-DA was built using the training set and then used for the sample prediction of the validation set. For both PCA and OPLS-DA, the score scatter plots for the first two components are visualized, even though more components may contribute to the model. The number of components for PCA and OPLS-DA models were determined by selecting the option autofit in the SIMCA software, where only components are considered of significance according to cross-validation rules. The cross-validation is automatically applied following Eastment *et al.* for PCA⁴¹ and Martens *et al.* for OPLS-DA⁴². The data set is divided into 7 groups, omitting one group, building the model, and predicting the excluded group. This is repeated for each group, and the results of predictions reveal the number of significant components, which is provided in Supplementary Table S4. OPLS-DA revealed differences in the lipidome by using gender as a classifier. As a consequence, the data sets for females and males were treated separately for investigation of the prediction performance.

Microsoft Excel was used for the calculation of average lipid concentrations obtained for all sample groups, fold change, T-value, and p-value. For the calculation of p-value, a two-sided two-sample T-test assumed unequal variances (Welch test) for the samples obtained from healthy controls and kidney, breast, or prostate cancer patients. P-values < 0.05 were considered as significant, but p-values were further evaluated according to the Bonferroni correction. All statistical parameters for all lipids are summarized in Supplementary Tables S3 and S6 below the lipid concentrations measured in individual samples and Supplementary Table S10. Another parameter indicating some relevance to differentiate samples from healthy controls and cancer patients, is the variable of importance (VIP) value obtained for each OPLS-DA plot. The most regulated and statistically significant lipid species with a fold change $\pm$ 20%, a p-value < 0.05, and a VIP value > 1 for all methods and phases are summarized in Supplementary Table S10. Box plots were used to better visualize lipid species concentrations depending on the health state. The box plots were constructed in R free software environment (<https://www.r-project.org>) using readxl and ggplot2 packages. In each box plot, the median was presented by a horizontal line, the box represented the 1st and 3rd quartile values, and whiskers stood for 1.5*IQR from the median and each measurement was plotted as a jittered point value. The receiver operating characteristics curves were generated by using the packages readxl and AUC in R. The dendrograms were also constructed in R⁴³. For the circular dendrograms, the Euclidean distances were calculated, and then the upgma function from the phangorn library was used for clustering (the Ward agglomeration method was selected). Circular dendrograms were generated and surrounded by the heatmap (ggtree and gheatmap functions – [ggtree](#) library). For the heatmap presentation, all concentrations were min-max scaled.

Declarations

DATA AVAILABILITY

All data relevant for the presented conclusions are provided in the manuscript or in the supplementary tables. Raw files of all measurements can be provided on request by the corresponding author.

ACKNOWLEDGEMENTS

The project was funded by the project 21-20238S (Czech Science Foundation).

AUTHOR CONTRIBUTIONS

D.W., R.J., and M.H prepared the concept of the study. D.W. and O.P. performed the sample preparation. D.W. analyzed samples by UHPSFC/MS and I.B. by shotgun MS, D.W. processed data, D.W. and J.I. performed statistical analysis. H.S. and B.M. obtained and provided plasma samples and clinical information. D.W. prepared the first draft of the manuscript and Figures. M.H. was responsible for funding and supervision of this study. All co-authors read, reviewed, edited, and approved the manuscript.

COMPETING INTERESTS

M.H., R.J., and D.W. are listed as inventors on the patent A method of diagnosing cancer based on lipidomic analysis of a body fluid (EP18174963.1, filing date 29. 5. 2018) related to this work.

References

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA-Cancer J. Clin*, **68**, 394–424 <https://doi.org/10.3322/caac.21492> (2018).
2. National Cancer Institute. Cancer staging. Cancer.gov. <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>. Updated February 9, 2021.
3. Culp, M. B., Soerjomataram, I., Efstathiou, J. A., Bray, F. & Jemal, A. Recent Global Patterns in Prostate Cancer Incidence and Mortality Rates. *Eur. Urol*, **77**, 38–52 <https://doi.org/10.1016/j.eururo.2019.08.005> (2020).
4. Ahmad, A. *in Breast Cancer Metastasis and Drug Resistance: Challenges and Progress* (Ed Aamir Ahmad) 1–7 (Springer International Publishing, 2019).
5. Znaor, A., Lortet-Tieulent, J., Laversanne, M., Jemal, A. & Bray, F. International Variations and Trends in Renal Cell Carcinoma Incidence and Mortality. *Eur. Urol*, **67**, 519–530 <https://doi.org/10.1016/j.eururo.2014.10.002> (2015).

6. Li, P. *et al.* Regional Geographic Variations in Kidney Cancer Incidence Rates in European Countries. *Eur. Urol*, **67**, 1134–1141 <https://doi.org/10.1016/j.eururo.2014.11.001> (2015).
7. Descotes, J. L. Diagnosis of prostate cancer. *Asian Journal of Urology*, **6**, 129–136 <https://doi.org/10.1016/j.ajur.2018.11.007> (2019).
8. Jafari, S. H. *et al.* Breast cancer diagnosis: Imaging techniques and biochemical markers. *Journal of Cellular Physiology*, **233**, 5200–5213 <https://doi.org/10.1002/jcp.26379> (2018).
9. Capitanio, U. & Montorsi, F. Renal cancer. *The Lancet*, **387**, 894–906 [https://doi.org/10.1016/S0140-6736\(15\)00046-X](https://doi.org/10.1016/S0140-6736(15)00046-X) (2016).
10. Ljungberg, B. *et al.* The 2010 Update. *Eur. Urol*, **58**, 398–406 <https://doi.org/10.1016/j.eururo.2010.06.032> (2010).
11. Lennon, A. M. *et al.* Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention., **369**, eabb9601 <https://doi.org/10.1126/science.abb9601> (2020).
12. Amin, M. B. *et al.* The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. *CA: A Cancer Journal for Clinicians*, **67**, 93–99 <https://doi.org/10.3322/caac.21388> (2017).
13. Tanos, R. & Thierry, A. R. Clinical relevance of liquid biopsy for cancer screening. *Translational Cancer Research*, **S105**, S129 (2018).
14. Chen, M. & Zhao, H. Y. Next-generation sequencing in liquid biopsy: cancer screening and early detection. *Hum. Genomics*, **13**, 10 <https://doi.org/10.1186/s40246-019-0220-8> (2019).
15. Cohen, J. D. *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test., **359**, 926 <https://doi.org/10.1126/science.aar3247> (2018).
16. Zhang, L., Han, X. & Wang, X. Is the clinical lipidomics a potential goldmine? *Cell Biol. Toxicol*, **34**, 421–423 <https://doi.org/10.1007/s10565-018-9441-1> (2018).
17. Yang, K., Han, X. & Lipidomics Techniques, Applications, and Outcomes Related to Biomedical Sciences. *Trends in Biochemical Sciences*, **41**, 954–969 <https://doi.org/10.1016/j.tibs.2016.08.010> (2016).
18. Wolrab, D. *et al.* Mass spectrometric quantitation of lipids in cancer research. *TrAC Trends in Analytical Chemistry*, **120**, 115480 <https://doi.org/10.1016/j.trac.2019.04.012> (2019).
19. Holčapek, M., Liebisch, G., Ekroos, K. & Lipidomic analysis. *Anal. Chem.* **90**, 4249–4257, [doi:https://doi.org/10.1021/acs.analchem.7b05395](https://doi.org/10.1021/acs.analchem.7b05395) (2018).
20. Liebisch, G. *et al.* Lipidomics needs more standarization. *Nat. Metab*, **1**, 745–747 <https://doi.org/10.1038/s42255-019-0094-z> (2019).
21. Wolrab, D. *et al.* Lipidomic profiling of human serum enables detection of pancreatic cancer *Nat. Com.*, revision, preprint available at medRxiv: <https://www.medrxiv.org/content/10.1101/2021.01.22.21249767v1> (2021).
22. Holčapek, M. *et al.* A method of diagnosing pancreatic cancer based on lipidomic analysis of a body fluid, European patent EP3514545 (granted), date of filling 22.1. 2018.

23. Holčapek, M., Wolrab, D., Jirásko, R. & Cífková, E. A method of diagnosing cancer based on lipidomic analysis of a body fluid, EP18174963.1 (pending), filing date 29. 5 2018.
24. Sales, S. *et al.* Gender, Contraceptives and Individual Metabolic Predisposition Shape a Healthy Plasma Lipidome. *Sci. Rep*, **6**, 27710 <https://doi.org/10.1038/srep27710> (2016).
25. Wenk, M. R. The emerging field of lipidomics. *Nature Reviews Drug Discovery*, **4**, 594–610 <https://doi.org/10.1038/nrd1776> (2005).
26. Gross, R. W. & Han, X. Lipidomics at the Interface of Structure and Function in Systems Biology. *Chem. Biol*, **18**, 284–291 <https://doi.org/10.1016/j.chembiol.2011.01.014> (2011).
27. Chen, X. *et al.* Plasma lipidomics profiling identified lipid biomarkers in distinguishing early-stage breast cancer from benign lesions. *Oncotarget* **7** (2016).
28. Cala, M. P. *et al.* Multiplatform plasma metabolic and lipid fingerprinting of breast cancer: A pilot control-case study in Colombian Hispanic women. *PLOS ONE*, **13**, e0190958 <https://doi.org/10.1371/journal.pone.0190958> (2018).
29. Cui, M., Wang, Q. & Chen, G. Serum metabolomics analysis reveals changes in signaling lipids in breast cancer patients. *Biomed. Chromatogr*, **30**, 42–47 <https://doi.org/10.1002/bmc.3556> (2016).
30. Guo, Y. *et al.* Simultaneous Quantification of Serum Multi-Phospholipids as Potential Biomarkers for Differentiating Different Pathophysiological states of lung, stomach, intestine, and pancreas. *Journal of Cancer*, **8**, 2191–2204 <https://doi.org/10.7150/jca.19128> (2017).
31. Lin, L. *et al.* LC-MS-based serum metabolic profiling for genitourinary cancer classification and cancer type-specific biomarker discovery. *PROTEOMICS*, **12**, 2238–2246 <https://doi.org/10.1002/pmic.201200016> (2012).
32. Yu, Z. *et al.* Global lipidomics identified plasma lipids as novel biomarkers for early detection of lung cancer. *Oncotarget* **8** (2017).
33. Guo, Y. *et al.* Probing gender-specific lipid metabolites and diagnostic biomarkers for lung cancer using Fourier transform ion cyclotron resonance mass spectrometry. *Clin. Chim. Acta*, **414**, 135–141 <https://doi.org/10.1016/j.cca.2012.08.010> (2012).
34. Zhou, X. *et al.* Identification of Plasma Lipid Biomarkers for Prostate Cancer by Lipidomics and Bioinformatics. *PLOS ONE*, **7**, e48889 <https://doi.org/10.1371/journal.pone.0048889> (2012).
35. Patel, N., Vogel, R., Chandra-Kuntal, K., Glasgow, W. & Kelavkar, U. A. Novel Three Serum Phospholipid Panel Differentiates Normal Individuals from Those with Prostate Cancer. *PLOS ONE*, **9**, e88841 <https://doi.org/10.1371/journal.pone.0088841> (2014).
36. Wolrab, D., Chocholoušková, M., Jirásko, R., Peterka, O. & Holčapek, M. Validation of lipidomic analysis of human plasma and serum by supercritical fluid chromatography–mass spectrometry and hydrophilic interaction liquid chromatography–mass spectrometry. *Analytical and Bioanalytical Chemistry*, **412**, 2375–2388 <https://doi.org/10.1007/s00216-020-02473-3> (2020).
37. Lída, M. & Holčapek, M. High-Throughput and Comprehensive Lipidomic Analysis Using Ultrahigh-Performance Supercritical Fluid Chromatography–Mass Spectrometry. *Anal. Chem*, **87**, 7187–7195 <https://doi.org/10.1021/acs.analchem.5b01054> (2015).

38. Wolrab, D. *et al.* LipidQuant tool for automated data processing in lipid class separation-mass spectrometry workflows. *Bioinformatics*, submitted, preprint available at Zenodo: <https://zenodo.org/record/4899685#.YNDy10wzUaG> (2021).
39. Lísá, M., Cífková, E., Khalikova, M., Ovčáčíková, M. & Holčapek, M. Lipidomic analysis of biological samples: Comparison of liquid chromatography, supercritical fluid chromatography and direct infusion mass spectrometry methods. *Journal of Chromatography A*, **1525**, 96–108 <https://doi.org/10.1016/j.chroma.2017.10.022> (2017).
40. Wang, M., Wang, C. & Han, X. Selection of internal standards for accurate quantification of complex lipid species in biological extracts by electrospray ionization mass spectrometry—What, how and why? *Mass Spectrometry Rev*, **36**, 693–714 <https://doi.org/10.1002/mas.21492> (2016).
41. Eastment, H. T. & Krzanowski, W. J. Cross-Validatory Choice of the Number of Components from a Principal Component Analysis., **24**, 73–77 <https://doi.org/10.2307/1267581> (1982).
42. Martens, H. & Naes, T. M. v. C., Wiley, Chichester, 1989.
43. Yu, G. Using ggtree to visualize data on tree-like structures. *Current Protocols in Bioinformatics*, **69**, e96 <https://doi.org/10.1002/cpbi.96> (2020).

Figures

Figure 1

Overview of the sample set (n = 481) used for the study in 3 cancer types. Samples were divided into training (75%) and validation sets (25%). Plasma samples obtained from patients suffering from kidney

(n = 119), breast (n = 103), and prostate (n = 67) cancers and control samples from healthy volunteers (n = 192) were included in the study.

Figure 2

OPLS-DA models used for the differentiation of cancer and control plasma samples. The training set from the discovery phase was measured by UHPSFC/MS and then used to build OPLS-DA models: a kidney cancer vs. control samples for males, b kidney cancer vs. control samples for females, c prostate cancer vs. control samples for males, and d breast cancer vs. control samples for females. Annotation: blue - control samples (N), yellow - cancer stage T1 (T1), orange - cancer stage T2 (T2), light red - cancer stage T3 or Tis (T3/Tis), and dark red - unknown cancer stage (Tx).

Figure 3

OPLS-DA models were used to predict the pathological state of human subjects. The training set was used to build OPLS-DA models. The percentage of specificity (blue), sensitivity (yellow, orange, red), and accuracy (green) for the training and validation set using UHPSFC/MS data from the discovery phase are presented. The sensitivity was determined for each cancer stage (yellow - T1, orange - T2, and red - T3), excluding samples with unknown cancer stage. Training set: a kidney cancer vs. control samples for males, b kidney cancer vs. control samples for females, c prostate cancer vs. control samples for males, and d breast cancer vs. control samples for females. Validation set: e kidney cancer vs. control samples for males, f kidney cancer vs. control samples for females, g prostate cancer vs. control samples for males, and h breast cancer vs. control samples for females. The ROC curves with the corresponding AUC values are presented, whereby the continuous lines represent the ROC curve for the training set and dashed lines for the validation set. i kidney cancer vs. control samples for males, j kidney cancer vs. control samples for females, k prostate cancer vs. control samples for males, and l breast cancer vs. control samples for females.

Figure 4

Comparison of UHPSFC/MS results in the discovery and qualification phases. a PCA of all samples (validation and training set) in the discovery phase. b PCA of all samples (validation and training set) in the qualification phase (red - cancer samples (T), blue - control samples (N), and green - QC samples). Selected samples were annotated for comparison reasons. Boxplots comparing the concentrations of samples obtained from patients with different pathological states (blue: control samples, orange: kidney cancer, light blue: prostate cancer, pink: breast cancer) for the discovery phase (1) and qualification phase (2) for c) LPC 18:2, d) PC 36:2, e) Cer 42:1 using UHPSFC/MS.

Figure 5

The ROC curves with the corresponding AUC values are presented, whereby the continuous lines represent the ROC curve for the training set and dashed lines for the validation set using UHPSFC/MS. a kidney cancer vs. control samples for males, b kidney cancer vs. control samples for females, c prostate cancer vs. control samples for males, and d breast cancer vs. control samples for females using data from the qualification phase for MDA, and e kidney cancer vs. control samples for males, f kidney cancer vs.

control samples for females, g prostate cancer vs. control samples for males, and h breast cancer vs. control samples for females predicting the data from the qualification phase using the discovery phase for MDA, i kidney cancer vs. control samples for males, j kidney cancer vs. control samples for females, k prostate cancer vs. control samples for males, and l breast cancer vs. control samples for females using shotgun MS data for MDA. Box plots comparing the concentrations of samples obtained from patients with different pathological states (blue: control samples, orange: kidney cancer, light blue: prostate cancer, pink: breast cancer) for the discovery phase (1) and qualification phase (2) using UHPSFC/MS and shotgun MS (SG): m PC 36:3, and n SM 32:1.

Figure 6

Influence of reducing the number of lipid species used for performing multivariate data analysis using UHPSFC/MS data in the discovery phase. (138 lipids: no exclusion, 91 lipids: common lipids from the discovery and qualification phase using UHPSFC/MS and shotgun MS, 29: only lipid species included, which are significant according to the Bonferroni correction, >10 from 15 models (considering all cancer types, methods, and models). SM 38:1 and SM 42:1 were excluded as those variables were only

significant for UHPSFC/MS data sets (10/10) and therefore a method bias cannot be excluded. 7: only lipid species included, which are significant according to the Bonferroni correction, >14 from 15 models (considering all cancer types, methods, and models). The average of the specificity (blue), sensitivity (red), and accuracy (green) for the validation and training set as well as for both genders were calculated. a kidney cancer, b breast cancer, and c prostate cancer.

Figure 7

Distribution of lipid class percentage, when considering only lipid species from a lipid class, which are significant according to the Bonferroni correction for the discovery phase: a kidney cancer males, kidney cancer females, c prostate cancer males, and d breast cancer females. Dendrograms considering 7 lipid species quantified in the discovery phase using UHPSFC/MS for both genders and training and validation sets for: e kidney cancer, f prostate cancer, and g breast cancer.

Figure 8

OPLS-DA models for the differentiation of the sample type like cancer type and control samples a males and b females using the concentrations of the 138 lipids determined in the discovery phase with UHPSFC/MS for the training set. OPLS-DA models for the differentiation of c prostate and kidney cancer samples for males and d breast and kidney cancer samples for females.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [WolrabSciRepSupplementaryTables.xlsx](#)

- [WolrabSciRepSupplementaryinformation.pdf](#)