


RESEARCH

Open Access



# Data-driven identification of plasma metabolite clusters and metabolites of interest for potential detection of early-stage non-small cell lung cancer cases versus cancer-free controls

Julian O. Kim<sup>1,2\*</sup> , Robert Balshaw<sup>3</sup>, Connel Trevena<sup>4</sup>, Shantanu Banerji<sup>2,5</sup>, Leigh Murphy<sup>2,6</sup>, David Dawe<sup>2,5</sup>, Lawrence Tan<sup>7</sup>, Sadeesh Srinathan<sup>7</sup>, Gordon Buduhan<sup>7</sup>, Biniyam Kidane<sup>7</sup>, Gefei Qing<sup>8</sup>, Michael Domaratzki<sup>4</sup> and Michel Aliani<sup>9</sup>

## Abstract

**Background:** Metabolomics is a potential means for biofluid-based lung cancer detection. We conducted a non-targeted, data-driven assessment of plasma from early-stage non-small cell lung cancer (ES-NSCLC) cases versus cancer-free controls (CFC) to explore and identify the classes of metabolites for further targeted metabolomics biomarker development.

**Methods:** Plasma from 250 ES-NSCLC cases and 250 CFCs underwent ultra-high-performance liquid chromatography/quadrupole time-of-flight mass spectrometry (UHPLC-QTOF-MS) in positive and negative electrospray ionization (ESI) modes. Molecular feature extraction, formula generation, and find-by-ion tools annotated metabolic entities. Analysis was restricted to endogenous metabolites present in  $\geq 80\%$  of samples. Unsupervised hierarchical cluster analysis identified clusters of metabolites. The metabolites with the strongest correlation with the principal component of each cluster were included in logistic regression modeling to assess discriminatory performance with and without adjustment for clinical covariates.

**Results:** A total of 1900 UHPLC-QTOF-MS assessments identified 1667 and 2032 endogenous metabolites in the ESI-positive and ESI-negative modes, respectively. After data filtration, 676 metabolites remained, and 12 clusters of metabolites were identified from each ESI mode. Multivariable logistic regression using the representative metabolite from each cluster revealed effective classification of cases from controls with overall diagnostic accuracy of 91% (ESI positive) and 94% (ESI negative). Metabolites of interest identified for further targeted analysis include the following: 1b, 3a, 12a-trihydroxy-5b-cholanoic acid, pyridoxamine 5'-phosphate, sphinganine 1-phosphate, gamma-CEHC, 20-carboxy-leukotriene B4, isodesmosine, and 18-hydroxycortisol.

\*Correspondence: Jkim7@cancercare.mb.ca

<sup>2</sup> CancerCare Manitoba Research Institute, Winnipeg, Manitoba, Canada  
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusions:** Plasma-based metabolomic detection of early-stage NSCLC appears feasible. Further metabolomics studies targeting phospholipid, steroid, and fatty acid metabolism are warranted to further develop noninvasive metabolomics-based detection of early-stage NSCLC.

**Keywords:** Early-stage non-small cell lung cancer, Plasma metabolomics, Early detection, Non-targeted metabolomics

## Background

Lung cancer is the leading cause of cancer-related mortality worldwide with an estimated 1.59 million persons succumbing to lung cancer annually [1]. Presently, most lung cancers are detected at either locally advanced or metastatic stages for which chances of cure are limited. Earlier detection of lung cancer represents a strategy to reduce mortality by allowing more patient opportunities for curative therapies including surgery or stereotactic ablative radiotherapy before disease progression to incurable disease. As proof of this principle, several randomized trials [2] have demonstrated that early detection of lung cancer using low-dose CT screening scans reduces lung cancer mortality. Although CT screening is being adopted in various jurisdictions, concerns remain regarding financial burden [3], high false-positive rate which can trigger unnecessary ancillary test procedures [4, 5], and risks of radiation-induced malignancy [6, 7]. Thus, there remains a bona fide need for the development of a cost-effective, noninvasive, accurate test for early NSCLC detection.

Metabolomics is an omics field which consists of measuring endogenous and exogenous low-molecular-weight metabolites in an organism at a specified time under specific environmental conditions [8]. Alterations in the metabolomic phenotype of cancer cells were first reported in 1956 by Warburg et al. who observed a higher rate of glycolysis and lactic acid production [9]. Technological advancements over the subsequent decades have led to the use of metabolomics methods including gas chromatography (GC), mass spectrometry (MS), liquid chromatography (LC), and  $^1\text{H}$ -nuclear magnetic resonance ( $^1\text{H}$ -NMR) as means for the diagnosis and prediction of outcomes in oncology. Metabolomic profiling has been used to assess the impact of malignancy on the metabolome using tumor tissue [10], urine [11], blood [12, 13], stool [14], or other biofluids [15]. To date, several groups [12, 13, 16–22] have investigated metabolomics profiles of NSCLC cases versus controls using combinations of different analytic platforms, biofluids, patient populations, and statistical modeling approaches.

With variations in methodologies employed in these studies, there is minimal overlap of metabolomic profiles that distinguish non-small cell lung cancer (NSCLC) cases from controls. In addition to the differences in

analytical approaches, the lack of assessment for the impact of potential confounding clinical variables such as age, sex, or smoking status may have contributed to this lack of similarity. Furthermore, these studies utilized relatively modest numbers of NSCLC cases for analysis and profile training, and few have focused primarily on the detection of early-stage NSCLC cases (ES-NSCLC), the group for which early detection is most relevant and technically challenging.

This metabolomics biomarker discovery study assessed the differences in global metabolomics profiles of plasma from a large group of clinical ES-NSCLC patients versus cancer-free controls to reduce the risk of model overfitting. We aimed to determine which classes of metabolites appear promising for future use in targeted metabolomics studies to further refine metabolomics profiles associated with ES-NSCLC. Our secondary aim was to assess the impact of clinical covariates, namely age, smoking history, and sex, on the classification performance of metabolomic entities for differentiating ES-NSCLC cases from cancer-free controls.

## Methods

### Patient population

From 2004 to 2014, 250 patients with clinical early-stage lung cancer (based on preoperative CT imaging scans) had blood samples collected by venipuncture prior to surgical resection. Final pathological staging was assigned following pathological analysis of pulmonary resection specimens using AJCC 6th edition. Blood samples were centrifuged into component blood products, and plasma aliquots were frozen at  $-80^\circ\text{C}$  and stored at the Manitoba Tumour Bank (MTB), a provincial biorepository certified by the Canadian Tissue Repository Network. Surgery for NSCLC cases consisted of wedge resection, segmentectomy, or lobectomy with lymph node sampling or lymph node dissection as clinically indicated. During the same time period, plasma samples from cancer-free controls were collected by venipuncture and stored using identical laboratory procedures in the same biorepository. Cancer-free controls consisted of two groups: (1) patients with suspected lung cancer (based on preoperative imaging) who underwent surgical resection with final pathology showing only benign pulmonary disease such as tuberculosis, granulomatous

disease, or pulmonary fibrosis and no evidence of malignancy and (2) family members of cancer patients with no prior personal history of malignancy (with exception of completely excised non-melanomatous skin cancers). Controls were excluded if cancer was diagnosed within 2 years following the date of the blood collection. Fasting was not mandatory prior to blood collection, and so fasting and non-fasting specimens were included in the analysis.

#### **Metabolomics analysis**

All analytical work was carried out in duplicate by blinded laboratory personnel who are unaware of the case/control status of plasma samples so to reduce unintentional bias. The laboratory workflow and procedures employed have been previously published [23, 24] and are summarized briefly below.

#### **Plasma extraction**

Frozen aliquots of plasma were thawed to 20 °C, and four 100- $\mu$ L aliquots were extracted per sample. For each aliquot, 200  $\mu$ L of acetonitrile was added and vortexed for 30 s and centrifuged at 10,000g for 10 min at 4 °C, and then, 250  $\mu$ L of supernatant was pipetted into an Eppendorf tube and dried under a gentle stream of N<sub>2</sub> and stored at -80 °C. The dried samples were reconstituted in 100  $\mu$ L of 4:1 acetonitrile and deionized H<sub>2</sub>O prior to analysis.

#### **Ultra-high-performance liquid chromatography/quadrupole time-of-flight mass spectrometry (UHPLC-QTOF-MS) analysis**

A 1260 Rapid Resolution system (Agilent Technologies, Santa Clara, USA) containing a binary pump and degasser, well-plate autosampler, and thermostatted column compartment (maintained at 55 °C) was used for analyses. Chromatographic separations were performed in duplicate on an Agilent ZORBAX SB-Poroshell column 2.1 mm  $\times$  50 mm, 2.7  $\mu$ m. MS analysis was performed on an Agilent 6538 QTOF mass spectrometer equipped with dual electrospray ionization (ESI) source in positive and negative modes.

#### **Identification of metabolic entities**

Agilent Profinder software, version B.08 (Agilent Technologies, Santa Clara, USA) including molecular feature extraction (MFE), formula generation, and find-by-ion tools, was used to prepare raw data and identify individual metabolic entities. Log<sub>2</sub> normalization of the concentrations of individual metabolites was performed using the Mass Profiler Professional software, version 12.6 (Agilent Technologies, Santa Clara, USA), and exported as raw data files for model building.

#### **Quality control**

Five quality control mixtures were made by pooling 100  $\mu$ L of plasma (randomly chosen from 10 samples in each group) and were analyzed in a random manner amongst all other samples.

#### **Statistical analysis**

Patient characteristics recorded at the time of sample collection were tabulated by case versus control status and compared using the Welch two-sample *t*-test or Pearson's chi-squared test.

#### **Data filtration and data cleaning**

Data analysis was restricted to known endogenous human metabolites identified by the Metlin and Human Metabolome databases. Candidate endogenous metabolites not detected in 80% or more of all samples were judged to be unlikely to possess useful classification values and were dropped from further analysis. Missing values of individual metabolites were replaced with one-half of the smallest positive measured quantity for each metabolite. All data analysis was conducted using the R statistical software package.

#### **Cluster analysis**

Given the highly dimensional nature of metabolomics data and the apparent risk of numerous collinear entities to mask overall global alterations present in the data, an unsupervised hierarchical cluster analysis was used to abridge the data into groups of 12 subsets of similar metabolites based on their distance in multi-vector space (1-correlation) using the complete linkage method of the Log<sub>2</sub> normalized concentrations for each metabolite, and samples were clustered using the Mahalanobis distance using complete linkage method. For the cluster analyses, the averaged value of the Log<sub>2</sub> normalized concentrations of metabolites arising from duplicated UHPLC-QTOF-MS assessments was used, and separate cluster analyses were done for the ESI positive and negative ionization modes. Hierarchical cluster analysis dendrogram heat maps were generated to display clusters of similar metabolites with individual metabolites on the *x*-axis and individual samples on the *y*-axis. A principal component analysis (PCA) of the cohort was performed to visually and qualitatively assess the separation of cases from controls in multi-vector space by ESI mode using 12 cluster-representative metabolic entities per ESI mode. For this PCA, the metabolite within each cluster with the strongest correlation to the first principle component was designated a "cluster-representative metabolite," which was

tabulated by ESI mode to provide qualitative examples of entities within each cluster.

### **Logistical regression modeling**

To assess the classification potential of metabolomic entities in determining case versus control status, the 12 cluster-representative metabolites identified from ESI-positive and ESI-negative cluster analyses were utilized as explanatory variables in a multivariable logistic regression model for the endpoint of NSCLC case status, with a separate logistic regression model for the ESI-positive and ESI-negative modes. With the aim of exploring the impact of potential clinical confounding variables (known lung cancer risk factors) on model fit, two additional multivariable logistic regression models were built which included clinical explanatory variables (age, sex, and smoking history) in addition to the 12 cluster-representative metabolites for each ESI mode. Forest plots were generated to visualize the relative strength of association and distribution of odds ratios associated with each cluster-representative metabolite with and without adjustment for clinical explanatory variables. Volcano plots were generated of cluster-representative metabolites from each ESI mode to visualize the negative log  $p$ -values ( $-\log_{10}$  (corrected  $p$ -value)) from Welch's  $t$ -test (unadjusted  $p$ -values) versus the  $\text{Log}_2$  fold change in the mean concentration between cases and controls for each metabolite.

### **Classification performance**

Classification performance of cluster-representative metabolites was assessed by multiplying regression coefficients for every 12 representative metabolites by the  $\text{Log}_2$  normalized concentrations of the same metabolites for each patient and comparing logistic regression predicted case versus control status versus known case versus control identity. Sensitivity, specificity, and overall classification accuracy by ESI mode both with and without the inclusion of clinical covariates were calculated, and receiver operator characteristic (ROC) curves were generated plotting sensitivity as a function of (1-specificity). The workflow used for this study is summarized in supplemental Table A1.

### **Ethical considerations**

This study was conducted with written approval from the University of Manitoba Health Research Ethics Board (HS19421) and the St. Boniface Hospital Research Review Committee (RRC/2016/1553).

## **Results**

### **Patient characteristics**

Data from 500 patients consisting of 2000 individual UHPLC-QTOF-MS assessments were available (two ESI-positive and two ESI-negative analyses from each study participant), which was merged with the database containing patient clinical covariates obtained from the Manitoba Tumour Bank. Four patients were excluded from the analysis due to the lack of consent to disclosure of clinical variables annotated to their samples; 11 patients were excluded from the analysis as they were identified as having provided more than one sample to the Manitoba Tumour Bank during the study period (duplicate patients). Loss of metabolomic data fidelity was detected in 10 patients and were excluded from the analysis.

Thus, data from 475 unique patients consisting of 241 lung cancer cases, and 234 cancer-free controls from which 1900 individual UHPLC-QTOF-MS assessments were conducted, were included in the final analysis. From these, a total of 2032 metabolic entities were detected in the ESI-positive mode, 1667 were detected in the ESI-negative mode, and 1529 entities were detected in both ESI modes. After filtering low prevalence entities, 676 metabolites remained of which 353 were detected in the ESI-positive mode and 323 from the ESI-negative mode which were used for further classification assessments.

The baseline clinical characteristics of the 475 included patients are shown in Table 1. Amongst the NSCLC cases, 177 (73%) had adenocarcinoma, and 64 (27%) had squamous cell carcinoma. For NSCLC cases, the final postoperative pathological staging (AJCC 6th ed.) was as follows: stage 1 (60%), stage 2 (21%), stage 3 (17%), and stage 4 (2%). NSCLC cases had a median age of 69 (range 49–88) versus 55 (range 20–89) for cancer-free controls ( $p < 0.001$ ). Males comprised 46% of cases versus 29% of cancer-free controls, and the median body mass index was similar between cases (27.2, range 14.8 to 49.5) and controls (27.0, range 16.4 to 49.6). NSCLC cases had higher proportions with significant comorbidities including diabetes, cardiovascular disease, dyslipidemia, and hypertension. NSCLC cases had significant smoking history such that 27% were current smokers, 65% were previous smokers, and 8% were never smokers. By contrast, cancer-free controls had lower levels of smoking exposure such that 6% were current smokers, 22% were previous smokers, and 48% were never smokers.

### **Cluster analysis**

Metabolic entities were clustered based on their correlations to one another by their distance in vector space (1-correlation) using the complete linkage method, and

**Table 1** Baseline characteristics of the cohort

Variable	NSCLC cases (n = 241)	Cancer-free controls (n = 234)	p-value
<b>Age</b>			
Mean (range)	69 (49–88)	55 (20–89)	p < 0.001
<b>Sex</b>			
Male (%)	112 (46%)	69 (29%)	p < 0.001
Female (%)	129 (54%)	165 (71%)	
<b>Stage (AJCC 7th ed.)</b>			
I (%)	145 (60%)	N/A	-
II (%)	50 (21%)		
III (%)	41 (17%)		
IV (%)	5 (2%)		
<b>NSCLC type</b>			
Adenocarcinoma (%)	177 (73%)	N/A	-
Squamous cell carcinoma (%)	64 (27%)		
<b>Body mass index</b>			
Mean (range)	27.2 (14.8–49.5)	27.0 (16.4–49.6)	0.7
<b>Comorbidities</b>			
Diabetes (%)	55 (23%)	16 (7%)	< 0.001
COPD (%)	68 (28%)	17 (7.3%)	< 0.0001
Hypertension (%)	129 (54%)	44 (19%)	< 0.001
Dyslipidemia (%)	91 (38%)	35 (15%)	< 0.001
Cardiovascular disease (%)	72 (30%)	14 (6%)	< 0.001
<b>Smoking history</b>			
Current smoker (%)	65 (27%)	15 (6%)	< 0.001
Ex-smoker (%)	156 (65%)	51 (22%)	
Never smoker (%)	20 (8%)	48 (21%)	
Unknown (%)	0	120 (51%)	

samples were clustered based on Mahalanobis distance using the complete linkage in order to generate the cluster analysis and heat map visualization for ESI-positive (Fig. 1A) and ESI-negative (Fig. 1B) metabolites.

#### Cluster-representative metabolites

The representative metabolite from each of the 12 identified clusters of metabolic entities is shown in Table 2. The complete list of metabolites found in each cluster is tabulated in a downloadable file (supplemental file 1). Amongst the ESI-positive entities, lipid metabolism (including phosphosphingolipids, glycerolipids), fatty acid metabolism, and steroids predominated as cluster-representative metabolites. Amongst ESI-negative entities, there was a greater variety of classes of metabolites which were observed as cluster-representative metabolites including the following: organic acids (ketoacids, carboxylic acids), steroids, fatty acyls, and hydroxyindol.

#### Principal component analysis

Principal component analysis of the cohort using the 12 cluster-representative metabolites from the ESI-positive

and ESI-negative modes demonstrated useful separation of the cohort by the first two principal components in each ESI mode (Fig. 2).

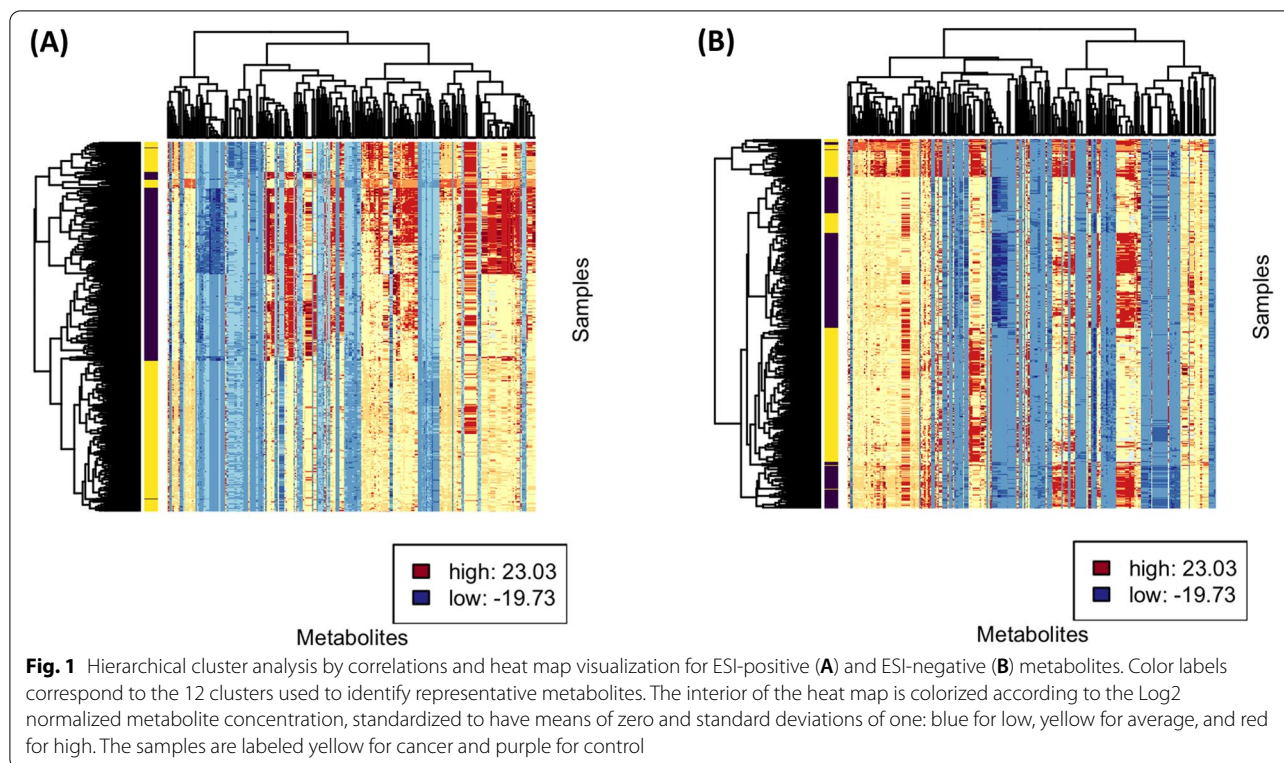
#### Assessment of volcano plots

Assessment of the volcano plots (Fig. A1) reveals that most of the cluster-representative metabolites identified from cluster analyses are not outliers. A number of cluster-representative metabolites, however, were situated at or below the threshold of 0.5, indicative of smaller mean differential concentrations of the metabolites between cases and controls.

#### Logistic regression modeling

Multivariable logistic regression analysis for the endpoint of NSCLC case status using the 12 cluster-representative metabolites revealed that a number of cluster-representative metabolites functioned as statistically significant predictors of NSCLC case status, while others did not (summarized as forest plots in Fig. 3A, B).

The ESI-positive class representative metabolites significantly associated with NSCLC case status in



the logistic model of the metabolites only included the following: MG(0:0/18:1/0:0) (*OR* 1.33, 95% *CI* 1.15–1.55), calcidiol (*OR* 1.51, 95% *CI* 1.25–1.85), 3-methoxybenzenepropanoic acid (*OR* 1.38, 95% *CI* 1.21–1.59), glycocholic acid (*OR* 1.21, 95% *CI* 1.01–1.46), pyridoxamine 5'-phosphate (*OR* 0.90, 95% *CI* 0.85–0.95), sphinganine 1-phosphate (*OR* 0.84, 95% *CI* 0.75–0.92), gamma-CEHC (*OR* 0.40, 95% *CI* 0.31–0.51), and 1b,3a,12a-trihydroxy-5b-cholanoic acid (*OR* 1.12, 95% *CI* 1.03–1.23). The impact of adjustment for clinical covariates was substantial, whereby several cluster-representative metabolites lost statistical significance. The following ESI-positive cluster-representative metabolites conserved statistical significance after adjustment for clinical covariates: MG(0:0/18:1/0:0) (*OR* 1.33, 95% *CI* 1.10–1.60), pyridoxamine 5'-phosphate (*OR* 0.86, 95% *CI* 0.80–0.93), sphinganine 1-phosphate (*OR* 0.87, 95% *CI* 0.76–0.99), gamma-CEHC (*OR* 0.48, 95% *CI* 0.34–0.66), and 1b,3a,12a-trihydroxy-5b-cholanoic acid (*OR* 1.19, 95% *CI* 1.05–1.35).

The ESI-negative cluster-representative metabolites significantly associated with NSCLC case status in the multivariable logistic model of the metabolites only included the following: 20-carboxy-leukotriene B4 (*OR* 1.51, 95% *CI* 1.30–1.77), 11-beta-hydroxyandrostosterone-3-glucuronide (*OR* 1.36, 95% *CI* 1.14–1.63), lithocholic acid glycine conjugate (*OR* 0.75, 95% *CI* 0.63–0.89),

18-hydroxycortisol (*OR* 0.48, 95% *CI* 0.35–0.63), formaldehyde (*OR* 1.11, 95% *CI* 1.01–1.21), isodesmosine (*OR* 1.16, 95% *CI* 1.09–1.24), 3-methyl-2-oxovaleric acid (*OR* 0.92, 95% *CI* 0.89–0.95), and deoxycholic acid 3-glucuronide (*OR* 0.93, 95% *CI* 0.89–0.97). The following ESI-negative cluster-representative metabolites maintained statistical significance after adjustment for clinical covariates: lithocholic acid glycine conjugate (*OR* 0.65, 95% *CI* 0.51–0.82), formaldehyde (*OR* 1.12, 95% *CI* 1.00–1.25), isodesmosine (*OR* 1.16, 95% *CI* 1.07–1.27), 18-hydroxycortisol (*OR* 0.63, 95% *CI* 0.42–0.91), 3-methyl-2-oxovaleric acid (*OR* 0.93, 95% *CI* 0.88–0.98), and deoxycholic acid 3-glucuronide (*OR* 0.91, 95% *CI* 0.85–0.98). The complete logistic regression analysis results are viewable in supplemental Tables 2 (ESI positive) and 3 (ESI negative).

#### Classification performance of cluster-representative metabolites

The classification performance of the ESI-positive and ESI-negative models both with and without covariates is detailed in supplemental Table A4. Using the metabolites alone, the diagnostic accuracy of between 75 (ESI positive) and 82% (ESI negative) was observed. Diagnostic accuracy notably improved with the addition of clinical covariate variables (age, sex, smoking history) to 90% (ESI positive)

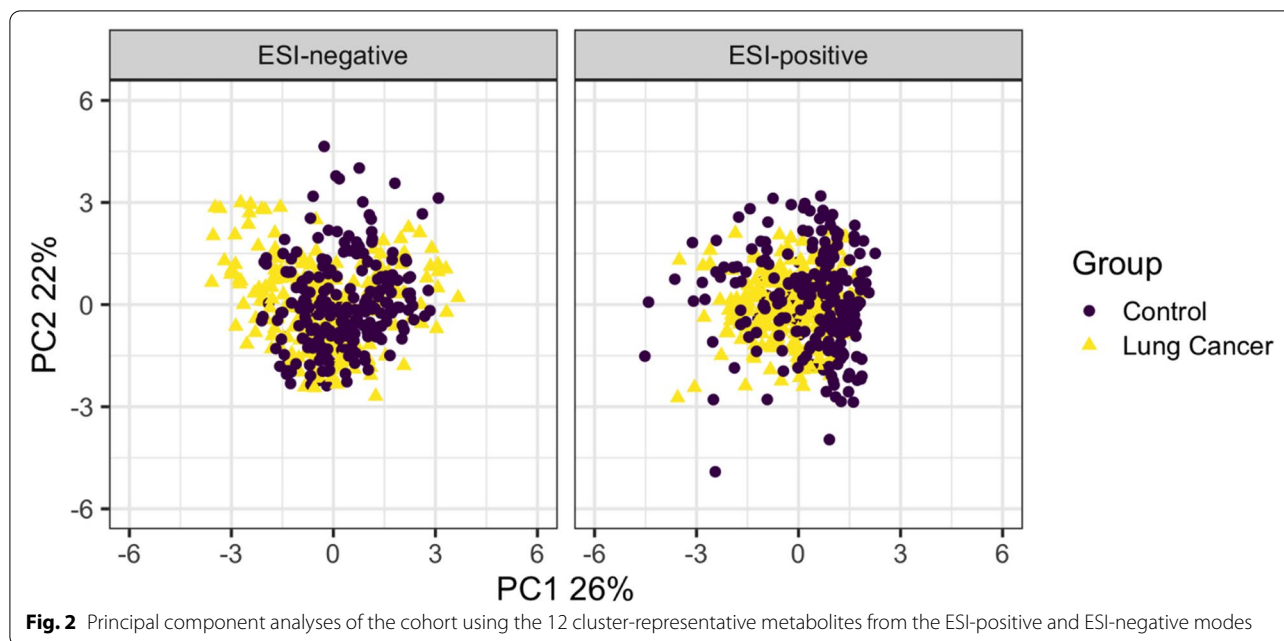
**Table 2** Summary of cluster analysis representative metabolites

Cluster representative	ESI mode	Formula	m/z	Metabolite class	Function	Disease associations
Sphingosine 1-phosphate	+	C <sub>18</sub> H <sub>38</sub> NO <sub>5</sub> P	380.2533	Phosphosphingolipid	Cell survival, inflammatory response, lipid metabolism	Hepatocellular carcinoma [25], other cancers [26]
Pyridoxamine 5'-phosphate (vitamin B6)	+	C <sub>8</sub> H <sub>13</sub> N <sub>2</sub> O <sub>5</sub> P	249.0618	Pyridoxamines	Amino acid metabolism neurotransmitter biosynthesis, lipid metabolism	Ovarian cancer [27]
Sphinganine 1-phosphate	+	C <sub>18</sub> H <sub>40</sub> NO <sub>5</sub> P	382.2718	Phosphosphingolipid	Membrane stabilization	-
Calcidiol (25-hydroxyvitamin D)	+	C <sub>27</sub> H <sub>44</sub> O <sub>2</sub>	401.3432	Vitamin D and derivatives	Vitamin D precursor	Prostate, breast, and colorectal cancer survival [28]; conflicting data for lung cancer incidence [29, 30]
3-Methoxybenzenepropanoic acid	+	C <sub>10</sub> H <sub>12</sub> O <sub>3</sub>	181.0858	Phenylpropanoic acids	-	-
8-Hydroxyguanine	+	C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>6</sub>	168.0559	Purine derivative	Mutagenic base, marker of DNA damage	Lung and stomach cancer [31]
1b,3a,12a-Trihydroxy-5b-cholanoic acid	+	C <sub>24</sub> H <sub>40</sub> O <sub>5</sub>	409.3037	Steroids	Fat absorption and transport	-
Glycocholic acid	+	C <sub>26</sub> H <sub>43</sub> NO <sub>6</sub>	466.3250	Steroids	Fat emulsification, bile acid	Hepatocellular carcinoma [32], cholangiocarcinoma [33, 34], prostate cancer [35]
MG(0:0/18:1/0:0)	+	C <sub>21</sub> H <sub>40</sub> O <sub>4</sub>	357.2989	Glycerolipids	Lipid metabolism, lipid transport	-
2-Hydroxydecanedioic acid	+	C <sub>10</sub> H <sub>18</sub> O <sub>5</sub>	219.1263	Hydroxy acids	Cell membrane stabilizer, energy storage	Zellweger syndrome [36]
Gamma-carboxyethyl hydroxychroman (gamma-CEHC)	+	C <sub>15</sub> H <sub>20</sub> O <sub>4</sub>	249.1541	Benzopyrans	Vitamin E metabolism	Colorectal cancer [37]
Cholic acid glucuronide	-	C <sub>30</sub> H <sub>48</sub> O <sub>11</sub>	583.3146	Steroids	Cholesterol metabolism	-
Formaldehyde	-	CH <sub>2</sub> O	59.0137	Carbonyl compounds	Protein and nucleic acid metabolism	Leukemia [38], nasopharyngeal cancer [38]
17-Hydroxypregnenolone sulfate	-	C <sub>21</sub> H <sub>32</sub> O <sub>6</sub> S	411.1838	Steroids	Lipid metabolism, cell signaling	-
N <sub>1</sub> -Acetylspermine	-	C <sub>9</sub> H <sub>21</sub> N <sub>3</sub> O	303.2322	Carboxylic acids	Cellular metabolism	-
Isodesmosine	-	C <sub>24</sub> H <sub>40</sub> N <sub>5</sub> O <sub>8</sub>	525.2811	Carboxylic acids	Elastin degradation	Liver cirrhosis [34, 39], cystic fibrosis [40, 41]
11-Beta-hydroxyandrost-3-one-3-glucuronide	-	C <sub>25</sub> H <sub>38</sub> O <sub>9</sub>	481.2450	Hydroxyindoles	Lipid metabolism	-
Lithocholic acid glycine conjugate	-	C <sub>26</sub> H <sub>43</sub> NO <sub>4</sub>	432.3120	Steroids	Fat excretion, absorption, and transport	-
3-Methyl-2-oxovaleric acid	-	C <sub>6</sub> H <sub>10</sub> O <sub>3</sub>	129.0549	Ketoacids	Amino acid metabolism	Maple syrup urine disease [42, 43]; colorectal cancer [44, 45]
18-Hydroxycortisol	-	C <sub>21</sub> H <sub>30</sub> O <sub>6</sub>	377.1993	Steroids	-	Primary aldosteronism [46, 47]
N(6)-Methyllysine	-	C <sub>7</sub> H <sub>16</sub> N <sub>2</sub> O <sub>2</sub>	159.1177	Carboxylic acid	Amino acid metabolism	-
Deoxycholic acid 3-glucuronide	-	C <sub>30</sub> H <sub>48</sub> O <sub>10</sub>	567.3370	Steroids	Fat emulsification	-
20-Carboxy-leukotriene B4	-	C <sub>20</sub> H <sub>30</sub> O <sub>6</sub>	411.1899	Fatty acyls	Lipid and drug metabolism	-
Pyroglutamic acid	-	C <sub>5</sub> H <sub>7</sub> NO <sub>3</sub>	128.0347	Carboxylic acid	Amino acid metabolism	NSCLC [48]

Abbreviations: m/z mass-to-charge ratio

and 94% (ESI negative). Receiver operator characteristic curves of the logistic regression models (Fig. 4) demonstrated an area under the curve (AUC) of

0.94 for both the ESI-positive and ESI-negative metabolites when clinical covariates were included in the model.



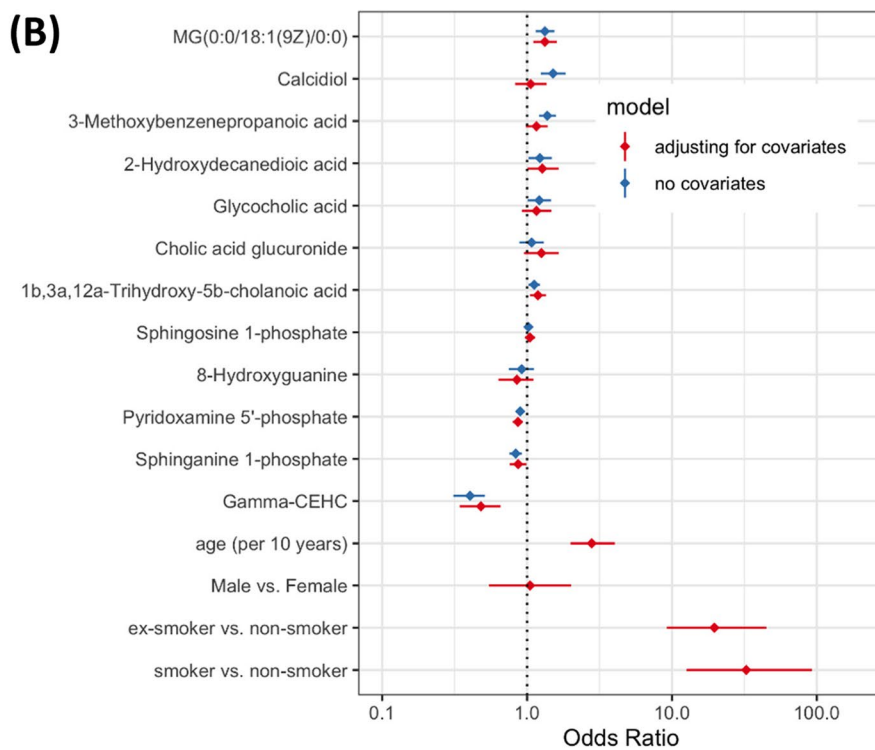
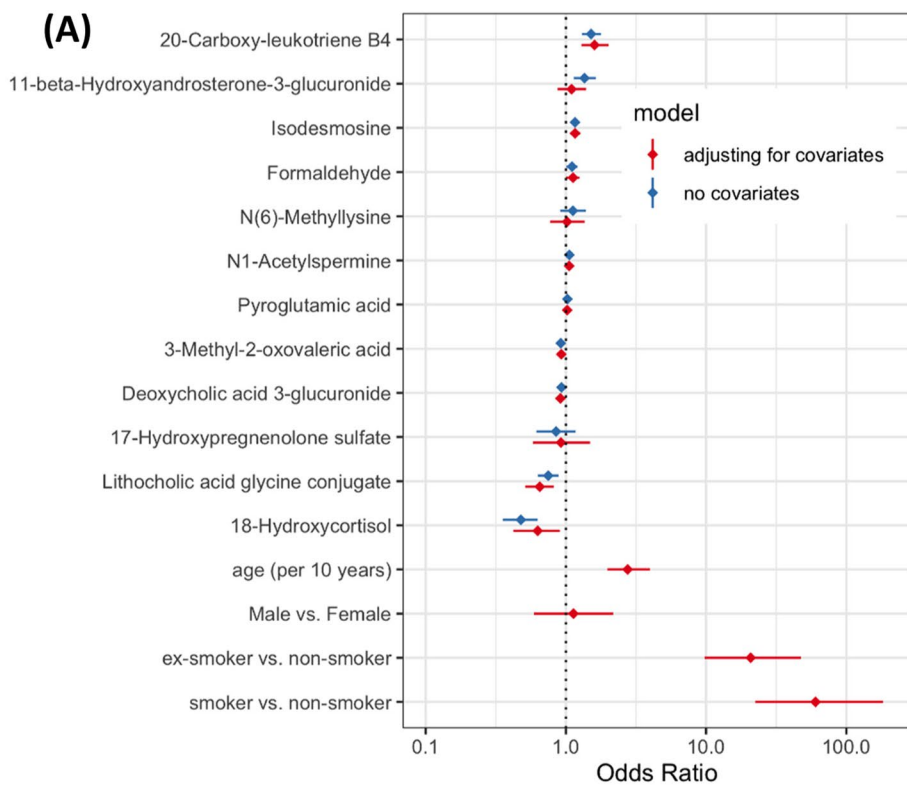
## Discussion

This study demonstrates that metabolomic analysis of plasma using UHPLC-QTOF-MS can potentially differentiate patients with clinical early-stage NSCLC from those who are lung cancer-free based on a wide variety of metabolites. To our knowledge, this study represents the largest cohort of patients with clinical early-stage NSCLC who have undergone non-targeted metabolomic analysis. We observed that plasma contains a number of classes of metabolites with potential classification utility, including phospholipids, fatty acids, steroids, fatty acyls, and amino acids. Cluster-representative metabolites of interest were identified including MG(0:0/18:1/0:0), 3-pyridoxamine 5'-phosphate, sphinganine 1-phosphate, gamma-CEHC, 1b,3a,12a-trihydroxy-5b-cholanoic acid from the ESI-positive mode, and lithocholic acid glycine conjugate, formaldehyde, isodesmosine, 18-hydroxycortisol, 20-carboxy-leukotriene B4, 3-methy-2-oxovaleric acid, and deoxycholic acid 3-glucuronide in the ESI-negative mode. Our findings suggest that these classes of metabolites and selected cluster-representative metabolites are worthy of further evaluation with targeted, quantitative metabolomic analyses for the development for the detection of early-stage NSCLC.

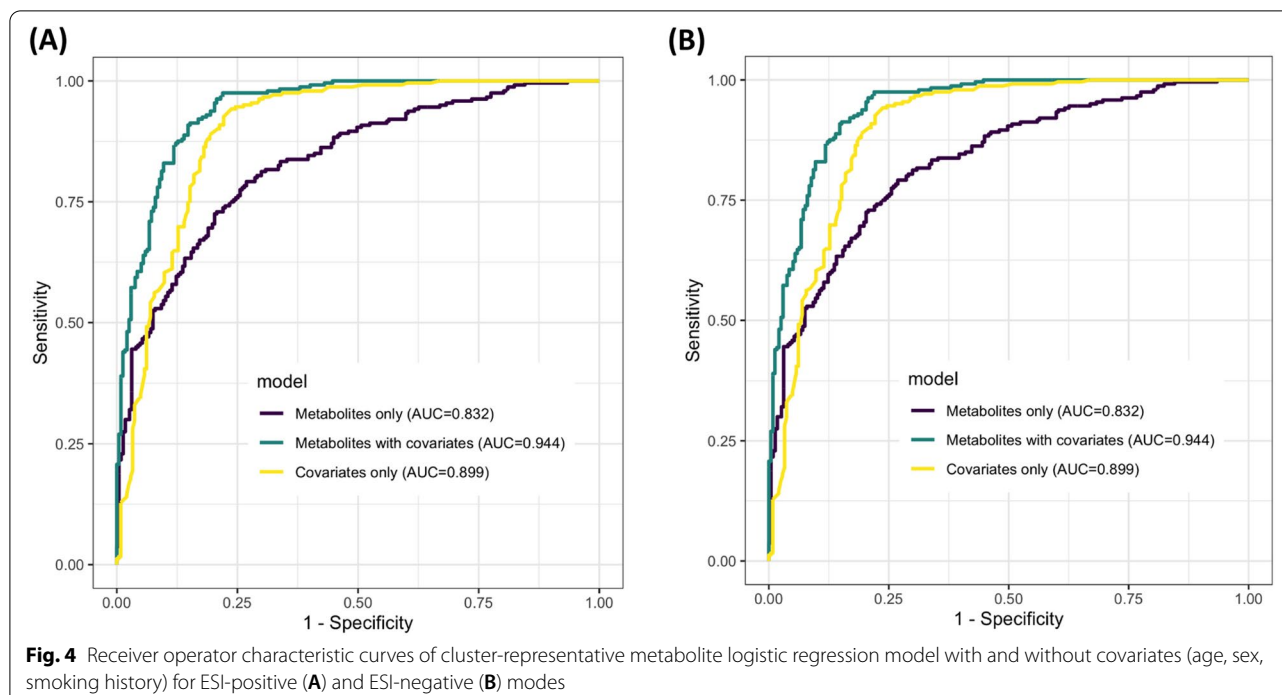
A number of the cluster-representative metabolites reported herein have been previously associated with oncogenesis. Alterations in fatty acid and lipid metabolism have been reported amongst studies assessing the serum or plasma [12, 13, 25] of lung cancer cases compared to controls. Sphinganine and sphingosine [12] overlap with several of the cluster-representative

metabolites arising from this study, notably sphingosine-1-phosphate and sphinganine-1-phosphate. Phosphosphingolipids, including sphingosine 1-phosphate (S1P), are cell membrane-derived and have important roles in cell signaling, cell survival, inflammatory response, angiogenesis, and tumor growth [26]. S1P is implicated as a pro-tumorigenic factor which activates signaling pathways including Ras/ERK, PI3K/RAC, STAT3, and PLC which are associated with various cancers [27–30]. Steroid hormones including calcidiol (25-hydroxyvitamin D), the primary circulating form of vitamin D, and calcitriol (1,25-dihydroxyvitamin D), the primary active form of vitamin D, are implicated in the incidence [31] and prognosis [32] of a number of human malignancies including breast, colorectal, and prostate cancer. The association between vitamin D and NSCLC incidence and prognosis, however, remains controversial with conflicting directions of association reported [31] which may be partially explained by the complex interaction of effect modifiers including age, sex, and smoking status [33]. Curiously, in this study, we observed calcidiol to have utility as a classifier of NSCLC status, but its statistical significance in the logistic regression model disappeared when the clinical covariates were added to the model, not dissimilar to the aforementioned pathogenesis and pharmacoepidemiology studies. Of interest, a cluster-representative metabolite, 8-hydroxyguanine, is a mutagenic base which denotes hydroxyl-mediated DNA damage which is implicated in lung and gastric cancer oncogenesis in preclinical models [35]. This metabolite did not perform as well as other cluster-representative metabolites at predicting





**Fig. 3** Forest plot of the distribution of odds ratios from the multivariable logistic regression analysis for cluster-representative metabolites with and without adjustment for covariates of age, sex, and smoking status metabolites for **A** ESI-positive and **B** ESI-negative analyses



NSCLC status in this study but may merit further targeted evaluations in the future due to its potential link to oncogenesis.

Low-molecular-weight organic acids, including lactic acid, are known to be produced from cancers as a result of their altered glucose metabolism [36]. In this study, in the ESI-negative mode, a low-molecular-weight organic acid, pyroglutamic acid, was found to be a cluster-representative metabolite. Interestingly, pyroglutamic acid has been previously scrutinized for its discriminatory performance of serum from NSCLC versus matched controls in a panel of ten other candidate organic acids [37] and was found to have the best discrimination, with an AUC of 0.76. However, when compared with other cluster-representative metabolites from this study, pyroglutamic acid was outmatched in discriminatory capacity. This highlights the importance of starting metabolomic biomarker development processes with wide non-targeted approaches prior to narrowing down to further quantitative targeted approaches rather than the opposite.

The small overlap of NSCLC-specific altered metabolites reported in the literature as compared to our list of cluster-representative metabolites is likely a consequence of the heterogeneity of biofluids assessed, variations in pre/post-analytical laboratory procedures, and sample sizes utilized. Further differences may be explained by the choice of statistical methodology, driven by differences in

analytical objectives, as well as preferred local best practices that have arisen in this rapidly growing field.

Exploratory analysis of the classification performance using representative metabolites from within each cluster of metabolites demonstrated classification accuracies in the range of 90%. With our experimental design planned to have a nearly 1:1 ratio of cases to controls, these results suggest further development work for this application of metabolomics is warranted. In this large cohort of patients with clinical early-stage NSCLC, we observed that age, sex, and smoking history, which are known risk factors for the development of NSCLC, significantly affected the performance of fitted predictive models. Of these, smoking status and age had particularly strong effects in the multivariable logistic regression models. The degree to which the clinical covariates influenced the classification characteristics of the logistic regression model mirrored the findings of Miyamoto et al., who found that the inclusion of age and sex improved the sensitivity and specificity of blood-based lung cancer metabolomic models [13]. These findings underline the need for future studies of metabolomic profiling for purposes of lung cancer detection to utilize comprehensively annotated biospecimens, including, at minimum, smoking history, age, and sex variables so to allow effective study design (e.g., stratified or matched sampling of cases and controls) as well

as analytical methods such as multivariable adjustment [13] or propensity matching [16]. Furthermore, there are likely many other biologically plausible confounding clinical variables including ethnicity, body mass index, dietary, or lifestyle choices. Given the current paucity of published data assessing the impact of these additional covariates, any future metabolomics studies planned for this milieu should assess any potential impact of these variables on metabolomic profiles in order to improve the calibration of any future diagnostic models.

A limitation of this study is that specimen collection was not explicitly controlled for the prandial state. However, since many of our specimens were collected alongside other routine bloodwork which require fasting prior to sample collection, it is conceivable that a considerable proportion of samples were indeed collected in the fasting state. Unfortunately, since prandial status data were not recorded in the biorepository databases, it could not be verified for each patient. This issue may be common to biorepository specimens like those used in our study, as most biorepositories were conceived with genomic or transcriptomic studies in mind where the prandial state is of little importance as compared to metabolomic analyses. As a result, this study was unable to control for the effects of the prandial state. A literature search did not find any published data which explored the impact of the prandial state on lung cancer metabolomic profiles. Thus, future work in this milieu should be performed amongst patients who have undergone standardized pre-sampling preparations which would reduce the impact of the digestion of various foods on their metabolomic profiles used for lung cancer classification.

Metabolomic data sets are highly multidimensional, leading to an increased risk that key associations may go unnoticed as “noise” as opposed to “signal.” With this in mind, the next logical step in the development of a metabolomics profile aimed at the detection of early-stage lung cancer would be targeted metabolomic analyses using shorter, pre-specified lists of metabolites from biospecimens that are completely annotated with clinical covariate data and preferably collected in the fasting state. Based on the findings of this study, analyses targeting lipid, fatty acid, steroid, and amino acid entities such as some of the cluster-representative metabolites would be reasonable. By targeting specific regions of the metabolome, associations of individual metabolites with ES-NSCLC would be more accurately assessed and evaluated. These associations may be developed in clinical trials and in the setting of case-control trial design, for example, but will ultimately need to be studied and calibrated using cohort study designs where the prevalence of ES-NSCLC is a known truth. A potential source of such specimens and data

may be found in image-based lung cancer screening programs which are already operational or are running in the context of a clinical trial.

## Conclusions

Global alterations in metabolism were found in the plasma of early-stage NSCLC cases compared to cancer-free controls. Further targeted analyses of specific metabolites and classes of metabolites using clinical covariate annotated biospecimens are warranted to refine this non-invasive approach to lung cancer detection.

## Abbreviations

<sup>1</sup>H-NMR: <sup>1</sup>H-nuclear magnetic resonance; CFC: Cancer-free controls; CIHR: Canadian Institutes of Health Research; CT: Computed tomography; ES-NSCLC: Early-stage non-small cell lung cancer; ESI: Electrospray ionization; MTB: Manitoba Tumour Bank; m/z: Mass-to-charge ratio; PCA: Principal component analysis; ROC: Receiver operator characteristic; S1P: Sphingosine 1-phosphate; UHPLC-QTOF-MS: Ultra-high-performance liquid chromatography/quadrupole time-of-flight mass spectrometry.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40170-022-00294-9>.

**Additional file 1: Supplemental Figure A1.** Volcano plots of the cluster-representative metabolites detected from the ESI positive and negative modes.

**Additional file 2: Supplemental Table A1.** Summary of Metabolomic Analysis Workflow. **Supplemental Table A2.** Estimated odds-ratios from multi-predictor logistic regression models based on representative ESI-positive metabolites, with and without adjustment for covariates. **Supplemental Table A3.** Estimated odds-ratios from multi-predictor logistic regression models based on representative ESI-negative metabolites, with and without adjustment for covariates. **Supplemental Table A4.** Classification performance of logistic regression models of cluster-representative metabolites both with and without adjustment for clinical confounding variables (Age, Sex, Smoking History) highlighting the importance of clinical covariates in metabolomic based assessments for lung cancer diagnosis Classification.

**Additional file 3: Supplemental File 1.** Table of all metabolites by cluster.

## Acknowledgements

The authors thank the volunteers who provided biofluid specimens to the Manitoba Tumour Bank. We thank the personnel from the Manitoba Tumour Bank for the provision of archived samples and associated annotated data used for the conduct of this study. We thank Dr. Eilean McKenzie-Matwiy from the CancerCare Manitoba Research Institute who assisted with grant applications. We thank Shiva Shariati for her analytical work for the metabolomic analysis in Dr. Aliani's laboratory. We thank Cheryl Clague for her assistance with the regulatory and ethics compliance work for this study.

## Authors' contributions

JK, SB, DD, RB, and MA were responsible for the study conceptualization, methodology development, and study analysis. MA and MD were responsible for the laboratory analysis, data curation, and supervision of graduate students. JK and RB drafted the original manuscript. JK, RB, MA, DD, MD, CT, and SB revised the manuscript. JK, MA, SB, and DD applied for and administered the grant funding for the study. RB, JK, and MA were responsible for the statistical analysis and preparation of the data visualizations. LM, LT, SS, GB, BK,

and GQ were responsible for the provision of study materials and editing of the manuscript. The authors read and approved the final manuscript.

### Funding

Funding for this study was provided by a grant from the Canadian Institutes of Health Research (CIHR grant no. 151566) to JOK, MA, DD, and SB. The Manitoba Tumor Bank is supported by funding from the CancerCare Manitoba Foundation. Neither the CIHR nor the CancerCare Manitoba Foundation had any role in the design, data collection, analysis, or drafting of the manuscript.

### Availability of data and materials

All data generated and analyzed in this study are available by reasonable request of the corresponding author.

### Declarations

#### Ethics approval and consent to participate

This study was conducted with written approval from the University of Manitoba Health Research Ethics Board (HS19421) and the St. Boniface Hospital Research Review Committee (RRC/2016/1553).

#### Consent for publication

N/A

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Section of Radiation Oncology, Department of Radiology, Max Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada.

<sup>2</sup>CancerCare Manitoba Research Institute, Winnipeg, Manitoba, Canada.

<sup>3</sup>George and Fay Yee Center for Healthcare Innovation, University of Manitoba, Winnipeg, Manitoba, Canada.

<sup>4</sup>Department of Computer Science, University of Manitoba, Winnipeg, Manitoba, Canada. <sup>5</sup>Section of Medical Oncology, Department of Internal Medicine, Max Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada. <sup>6</sup>Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, Manitoba, Canada.

<sup>7</sup>Section of Thoracic Surgery, Department of Surgery, Max Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada. <sup>8</sup>Department of Human Pathology, Max Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada. <sup>9</sup>Food and Human Nutritional Sciences, University of Manitoba, Winnipeg, Manitoba, Canada.

Received: 27 September 2021 Accepted: 22 September 2022

Published online: 12 October 2022

### References

- Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer*. 2019;144:1941–53. <https://doi.org/10.1002/ijc.31937>.
- N.L.S.T.R. Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365:395–409. <https://doi.org/10.1056/NEJMoa1102873>.
- Black WC, Gareen IF, Soneji SS, Sicks JD, Keeler EB, Aberle DR, et al. Cost-effectiveness of CT screening in the National Lung Screening Trial. *N Engl J Med*. 2014;371:1793–802. <https://doi.org/10.1056/NEJMoa1312547>.
- de Koning HJ, Meza R, Plevritis SK, ten Haaf K, Munshi VN, Jeon J, et al. Benefits and harms of computed tomography lung cancer screening programs for high-risk populations. *Rockv AHRQ*. 2013;2.
- Moyer VA, U.S.P.S.T. Force. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;160:330–8. <https://doi.org/10.7326/M13-2771>.
- Rampinelli C, De Marco P, Origgi D, Maisonneuve P, Casiraghi M, Veronesi G, et al. Exposure to low dose computed tomography for lung cancer screening and risk of cancer: secondary analysis of trial data and risk-benefit analysis. *BMJ*. 2017;356:j347. <https://doi.org/10.1136/bmj.j347>.
- Mascalchi M, Mazzoni LN, Falchini M, Belli G, Picozzi G, Merlini V, et al. Dose exposure in the ITALUNG trial of lung cancer screening with low-dose CT. *Br J Radiol*. 2012;85:1134–9. <https://doi.org/10.1259/bjr/20711289>.
- De Preter V, Verbeke K. Metabolomics as a diagnostic tool in gastroenterology. *World J Gastrointest Pharmacol Ther*. 2013;4:97–107. <https://doi.org/10.4292/wjgpt.v4.i4.97>.
- Warburg O. On the origin of cancer cells. *Science*. 1956;123:309–14.
- Brown MV, McDunn JE, Gunst PR, Smith EM, Millburn MV, Troyer DA, et al. Cancer detection and biopsy classification using concurrent histopathological and metabolomic analysis of core biopsies. *Genome Med*. 2012;4:33.
- Wu Q, Wang Y, Gu X, Zhou J, Zhang H, Lv W, et al. Urinary metabolomic study of non-small cell lung carcinoma based on ultra high performance liquid chromatography coupled with quadrupole time-of-flight mass spectrometry. *J Sep Sci*. 2014;37:1728–35. <https://doi.org/10.1002/jssc.201400222>.
- Chen Y, Ma Z, Li A, Li H, Wang B, Zhong J, et al. Metabolomic profiling of human serum in lung cancer patients using liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry and gas chromatography/mass spectrometry. *J Cancer Res Clin Oncol*. 2015;141:705–18. <https://doi.org/10.1007/s00432-014-1846-5>.
- Miyamoto S, Taylor SL, Barupal DK, Taguchi A, Wohlgemuth G, Wikoff WR, et al. Systemic metabolomic changes in blood samples of lung cancer patients identified by gas chromatography time-of-flight mass spectrometry. *Metabolites*. 2015;5:192–210. <https://doi.org/10.3390/metabo5020192>.
- Brown DG, Rao S, Weir TL, O'Malia J, Bazan M, Brown RJ, et al. Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. *Cancer Metab*. 2016;4:11-016-0151-y. <https://doi.org/10.1186/s40170-016-0151-y> eCollection 2016.
- Hassanein M, Callison JC, Callaway-Lane C, Aldrich MC, Grogan EL, Massion PP. The state of molecular biomarkers for the early detection of lung cancer. *Cancer Prev Res (Phila)*. 2012;5:992–1006. <https://doi.org/10.1158/1940-6207.CAPR-11-0441>.
- Mazzone PJ, Wang XF, Beukemann M, Zhang Q, Seeley M, Mohny R, et al. Metabolite profiles of the serum of patients with non-small cell carcinoma. *J Thorac Oncol*. 2016;11:72–8. <https://doi.org/10.1016/j.jtho.2015.09.002>.
- Louis E, Adriaenssens P, Guedens W, Bigirimurame T, Baeten K, Vanhove K, et al. Detection of lung cancer through metabolic changes measured in blood plasma. *J Thorac Oncol*. 2016;11:516–23. <https://doi.org/10.1016/j.jtho.2016.01.011>.
- Puchades-Carrasco L, Jantus-Lewintre E, Perez-Rambla C, Garcia-Garcia F, Lucas R, Calabuig S, et al. Serum metabolomic profiling facilitates the non-invasive identification of metabolic biomarkers associated with the onset and progression of non-small cell lung cancer. *Oncotarget*. 2016;7:12904–16. <https://doi.org/10.18632/oncotarget.7354>.
- Chen Y, Ma Z, Zhong J, Li L, Min L, Xu L, et al. Simultaneous quantification of serum monounsaturated and polyunsaturated phosphatidylcholines as potential biomarkers for diagnosing non-small cell lung cancer. *Sci Rep*. 2018;8:7137-018-25552-z. <https://doi.org/10.1038/s41598-018-25552-z>.
- Cameron SJ, Lewis KE, Beckmann M, Allison GG, Ghosal R, Lewis PD, et al. The metabolomic detection of lung cancer biomarkers in sputum. *Lung Cancer*. 2016;94:88–95. <https://doi.org/10.1016/j.lungcan.2016.02.006>.
- Callejon-Leblic B, Garcia-Barrera T, Gravalos-Guzman J, Pereira-Vega A, Gomez-Ariza JL. Metabolic profiling of potential lung cancer biomarkers using bronchoalveolar lavage fluid and the integrated direct infusion/ gas chromatography mass spectrometry platform. *J Proteome*. 2016;145:197–206.
- Mathe EA, Patterson AD, Haznadar M, Manna SK, Krausz KW, Bowman ED, et al. Noninvasive urinary metabolomic profiling identifies diagnostic and prognostic markers in lung cancer. *Cancer Res*. 2014;74:3259–70. <https://doi.org/10.1158/0008-5472.CAN-14-0109>.
- Mayengbam S, House JD, Aliani M. Investigation of vitamin B6 inadequacy, induced by exposure to the anti-B6 factor 1-amino D-proline, on plasma lipophilic metabolites of rats: a metabolomics approach. *Eur J Nutr*. 2016;55:1213–23. <https://doi.org/10.1007/s00394-015-0934-x>.

24. Surendran A, Aliani M, Ravandi A. Metabolomic characterization of myocardial ischemia-reperfusion injury in ST-segment elevation myocardial infarction patients undergoing percutaneous coronary intervention. *Sci Rep.* 2019;9:11742–13. <https://doi.org/10.1038/s41598-019-48227-9>.
25. Li Y, Song X, Zhao X, Zou L, Xu G. Serum metabolic profiling study of lung cancer using ultra high performance liquid chromatography/quadrupole time-of-flight mass spectrometry. *J Chromatogr B.* 2014;966:147–53.
26. Zeng Y, Yao X, Chen L, Yan Z, Liu J, Zhang Y, et al. Sphingosine-1-phosphate induced epithelial-mesenchymal transition of hepatocellular carcinoma via an MMP-7/ syndecan-1/TGF-beta autocrine loop. *Oncotarget.* 2016;7:63324–37. <https://doi.org/10.18632/oncotarget.11450>.
27. Wang P, Yuan Y, Lin W, Zhong H, Xu K, Qi X. Roles of sphingosine-1-phosphate signaling in cancer. *Cancer Cell Int.* 2019;19:295-019-1014-8. <https://doi.org/10.1186/s12935-019-1014-8> eCollection 2019.
28. Shida D, Takabe K, Kapitov D, Milstien S, Spiegel S. Targeting SphK1 as a new strategy against cancer. *Curr Drug Targets.* 2008;9:662–73. <https://doi.org/10.2174/138945008785132402>.
29. Riboni L, Hadi LA, Navone SE, Guarnaccia L, Campanella R, Marfia G. Sphingosine-1-phosphate in the tumor microenvironment: a signaling hub regulating cancer hallmarks. *Cells.* 2020;9. <https://doi.org/10.3390/cells9020337>.
30. Pyne NJ, Pyne S. Sphingosine 1-phosphate and cancer. *Nat Rev Cancer.* 2010;10:489–503.
31. Vojdeman FJ, Madsen CM, Frederiksen K, Durup D, Olsen A, Hansen L, et al. Vitamin D levels and cancer incidence in 217,244 individuals from primary health care in Denmark. *Int J Cancer.* 2019;145:338–46. <https://doi.org/10.1002/ijc.32105>.
32. Vaughan-Shaw PG, O'Sullivan F, Farrington SM, Theodoratou E, Campbell H, Dunlop MG, et al. The impact of vitamin D pathway genetic variation and circulating 25-hydroxyvitamin D on cancer outcome: systematic review and meta-analysis. *Br J Cancer.* 2017;116:1095–110. <https://doi.org/10.1038/bjc.2017.44>.
33. Wang W, Hu W, Xue S, Chen Q, Jiang Y, Zhang H, et al. Vitamin D and lung cancer; association, prevention, and treatment. *Nutr Cancer.* 2020;0:1–13. <https://doi.org/10.1080/01635581.2020.1844245>.
34. Afdhal NH, Keaveny AP, Cohen SB, Nunes DP, Maldonado N, O'Brien M, et al. Urinary assays for desmosine and hydroxylslypyridinoline in the detection of cirrhosis. *J Hepatol.* 1997;27:993–1002. [https://doi.org/10.1016/S0168-8278\(97\)80142-0](https://doi.org/10.1016/S0168-8278(97)80142-0).
35. Russo MT, De Luca G, Degan P, Parlanti E, Dogliotti E, Barnes DE, et al. Accumulation of the oxidative base lesion 8-hydroxyguanine in DNA of tumor-prone mice defective in both the Myh and Ogg1 DNA glycosylases; 2004.
36. Hirschhaeuser F, Sattler UGA, Mueller-Klieser W. Lactate: a metabolic key player in cancer. *Cancer Res.* 2011;71:6921–5. <https://doi.org/10.1158/0008-5472.CAN-11-1457>.
37. Klupczynska A, Plewa S, Dyszkiewicz W, Kasprzyk M, Sytek N, Kokot ZJ. Determination of low-molecular-weight organic acids in non-small cell lung cancer with a new liquid chromatography-tandem mass spectrometry method. *J Pharm Biomed Anal.* 2016;129:299–309. <https://doi.org/10.1016/j.jpba.2016.07.028>.
38. Zhang L, Zhou D, Guan W, Ren W, Sun W, Shi J, et al. Pyridoxine 5'-phosphate oxidase is a novel therapeutic target and regulated by the TGF-beta signalling pathway in epithelial ovarian cancer. *Cell Death Dis.* 2017;8:3214. <https://doi.org/10.1038/s41419-017-0050-3>.
39. Zhang A, Sun H, Yan G, Han Y, Ye Y, Wang X. Urinary metabolic profiling identifies a key role for glycocholic acid in human liver cancer by ultra-performance liquid-chromatography coupled with high-definition mass spectrometry. *Clin Chim Acta.* 2013;418:86–90. <https://doi.org/10.1016/j.cca.2012.12.024>.
40. Song WS, Park HM, Ha JM, Shin SG, Park HG, Kim J, et al. Discovery of glycocholic acid and taurochenodeoxycholic acid as phenotypic biomarkers in cholangiocarcinoma. *Sci Rep.* 2018;8:1–8. <https://doi.org/10.1038/s41598-018-29445-z>.
41. Ma S, Geraghty P, Dabo A, McCarthy C, McElvaney NG, Turino GM. Cystic fibrosis disease severity correlates with plasma levels of desmosine and isodesmosine, biomarkers of elastin degradation. *ERJ Open Res.* 2019;5:00250–2018. <https://doi.org/10.1183/23120541.00250-2018>.
42. Liang Q, Liu H, Xiang L, Li X, Zhang AH. High-throughput metabolomics enables biomarker discovery in prostate cancer. *RSC Adv.* 2017;7:2587–93. <https://doi.org/10.1039/C6RA25007F>.
43. Podebrad F, Heil M, Leib S, Geier B, Beck T, Mosandl A, et al. Analytical approach in diagnosis of inherited metabolic diseases: maple syrup urine disease (MSUD) - simultaneous analysis of metabolites in urine by enantioselective multidimensional capillary gas chromatography-mass spectrometry (Enantio-MDGC-MS). *HRC J High Resolut Chromatogr.* 1997;20:355–62. <https://doi.org/10.1002/jhrc.1240200703>.
44. Matsumoto M, Kuhara T, Inoue Y, Shinka T, Matsumoto I. Abnormal fatty acid metabolism in patients in hopanenate therapy during clinical episodes. *J Chromatogr B Biomed Sci Appl.* 1991;562:139–45. [https://doi.org/10.1016/0378-4347\(91\)80572-T](https://doi.org/10.1016/0378-4347(91)80572-T).
45. R. Sinha, J. Ahn, J.N. Sampson, J. Shi, G. Yu, X. Xiong, R.B. Hayes, J.J. Goedert, Fecal microbiota, fecal metabolome, and colorectal cancer interrelations, *PLoS One* 11 (2015) 1–13. <https://doi.org/10.1371/journal.pone.0152126>.
46. Goedert JJ, Sampson JN, Moore SC, Xiao Q, Xiong X, Hayes RB, et al. Fecal metabolomics: assay performance and association with colorectal cancer. *Carcinogenesis.* 2014;35:2089–96. <https://doi.org/10.1093/carcin/bgu131>.
47. Mulatero P, Di Cella SM, Monticone S, Schiavone D, Manzo M, Mengozzi G, et al. 18-hydroxycorticosterone, 18-hydroxycortisol, and 18-oxocortisol in the diagnosis of primary aldosteronism and its subtypes. *J Clin Endocrinol Metab.* 2012;97:881–9. <https://doi.org/10.1210/jc.2011-2384>.
48. Swenberg JA, Moeller BC, Lu K, Rager JE, Fry RC, Starr TB. Formaldehyde carcinogenicity research: 30 years and counting for mode of action, epidemiology, and cancer risk assessment. *Toxicol Pathol.* 2013;41:181–9. <https://doi.org/10.1177/0192623312466459>.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

