Metabolomics (2022) 18:31
https://doi.org/10.1007/s11306-022-01891-x

ORIGINAL ARTICLE

Lung cancer metabolomic data from tumor core biopsies enables risk-score calculation for progression-free and overall survival

Hunter A. Miller[1] · Shesh N. Rai[1,2,3] · Xinmin Yin[4] · Xiang Zhang[4] · Jason A. Chesney[1,3,5,6] · Victor H. van Berkel[3,7] · Hermann B. Frieboes[1,3,8,9]

Received: 10 January 2022 / Accepted: 19 April 2022 / Published online: 14 May 2022

Abstract

Introduction Metabolomics has emerged as a powerful method to provide insight into cancer progression, including separating patients into low- and high-risk groups for overall (OS) and progression-free survival (PFS). However, survival prediction based mainly on metabolites obtained from biofluids remains elusive.

Objectives This proof-of-concept study evaluates metabolites as biomarkers obtained directly from tumor core biopsies along with covariates age, sex, pathological stage at diagnosis (I/II vs. III/VI), histological subtype, and treatment vs. no treatment to risk stratify lung cancer patients in terms of OS and PFS.

Methods Tumor core biopsy samples obtained during routine lung cancer patient care at the University of Louisville Hospital and Norton Hospital were evaluated with high-resolution 2DLC-MS/MS, and the data were analyzed by Kaplan–Meier survival analysis and Cox proportional hazards regression. A linear equation was developed to stratify patients into low and high risk groups based on log-transformed intensities of key metabolites. Sparse partial least squares discriminant analysis (SPLS-DA) was performed to predict OS and PFS events.

Results Univariable Cox proportional hazards regression model coefficients divided by the standard errors were used as weight coefficients multiplied by log-transformed metabolite intensity, then summed to generate a risk score for each patient. Risk scores based on 10 metabolites for OS and 5 metabolites for PFS were significant predictors of survival. Risk scores were validated with SPLS-DA classification model (AUROC 0.868 for OS and AUROC 0.755 for PFS, when combined with covariates).

Conclusion Metabolomic analysis of lung tumor core biopsies has the potential to differentiate patients into low- and high-risk groups based on OS and PFS events and probability.

Keywords Lung cancer · Metabolomics · Overall survival · Progression free survival · Risk score calculator · Tumor core biopsy

Abbreviations

HR	Hazard ratio
NSCLC	Non-small cell lung cancer
OS	Overall survival
PC1	First principal component
PC2	Second principal component
PC3	Third principal component
PFS	Progression free survival

✉ Hermann B. Frieboes
hbfrie01@louisville.edu

1 Department of Pharmacology and Toxicology, University of Louisville, Louisville, USA

2 Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, USA

3 James Graham Brown Cancer Center, University of Louisville, Louisville, USA

4 Department of Chemistry, University of Louisville, Louisville, USA

5 Division of Medical Oncology and Hematology, Department of Medicine, University of Louisville, Louisville, USA

6 Department of Biochemistry and Molecular Genetics, University of Louisville, Louisville, USA

7 Department of Cardiovascular and Thoracic Surgery, University of Louisville, Louisville, USA

8 Department of Bioengineering, University of Louisville, Lutz Hall 419, Louisville, KY 40292, USA

9 Center for Predictive Medicine, University of Louisville, Louisville, USA

PLS-DA Partial least squares discriminant analysis
SPLS-DA Sparse partial least squares discriminant analysis

1 Introduction

Accounting for an estimated 12–13% of new cancer diagnoses in 2021, lung cancer is one of the most prevalent cancer types in the U.S. Of ~ 6,00,000 deaths attributed to cancer by the end of 2021, 22% are estimated to be lung cancer (Siegel et al., 2021). Survival has increased since the early 1990s, likely due to improvements in early diagnosis and treatment options. Yet, the outlook for any particular patient is dire, with a 5-year survival rate < 5% for all stages combined (Howlader et al., 2021). Reasons for this poor survival are multifactorial, with two dominant issues being presentation at an advanced stage and poor treatment success rates for patients at an advanced stage (Goldstraw et al., 2007; Puchades-Carrasco et al., 2016). Once patients have been diagnosed, identifying metabolic markers associated with high risk of early death has potential clinical utility for risk stratification and management of patient care.

Metabolomics aims to quantify a broad spectrum of small molecule metabolites within a biological sample and has emerged as a powerful tool for phenotyping a variety of diseases, including cancer (Collino et al., 2013). Mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR) are two major techniques commonly employed in metabolomics studies. Specifically for lung cancer, metabolomics has proven useful in a variety of diagnostic and prognostic roles (Bamji-Stocke et al., 2018), helping to resolve the link between genotype and phenotype. Several studies have investigated the metabolic profiles of lung cancer patients compared to healthy controls (Fan et al., 2009; Hori et al., 2011; Kami et al., 2013), showing profound metabolic dysregulation in lung cancer. Metabolic alterations have also been found in different histology types (Hori et al., 2011; Wikoff et al., 2015) and pathological stages (Hori et al., 2011; Puchades-Carrasco et al., 2016) indicating that metabolomics is useful for determining disease progression.

Recent studies have investigated the use of metabolic profiles from serum, plasma, urine, and tissue samples for predicting survival in a variety of cancers (Berker et al., 2019; Huang et al., 2019; Jin et al., 2014; Xie et al., 2017). In particular, a study using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) metabolomics data derived from paired serum-tissue samples discovered 9 spectral regions with significant differentiation between short and prolonged survival groups of early stage lung cancer patients in a Kaplan–Meier survival analysis ($p = 0.0100$) (Berker et al., 2019). Recently, four metabolites in were included in a weighted risk score calculation and significant differences in survival probability were found between low- and high-risk advanced non-small cell lung cancer (NSCLC) patients receiving first-line chemotherapy (Shen et al., 2017). Based on the literature, there is a promising outlook on the use of metabolomics for survival prediction in lung cancer patients, and the area is relatively unexplored.

In contrast to recent cancer patient serum metabolomic (Peng et al., 2018; Tian et al., 2018) or radiomic (Dercle et al., 2020) profiling, we report patient survival based on metabolomic analysis of tumor core biopsies routinely obtained during lung cancer patient care. In (Berker et al., 2019), tissue samples were found to result in better prediction accuracy for differentiating lung cancer subtypes ($AUROC_{Test}$ 0.82) compared to serum samples ($AUROC_{Test}$ 0.73). In our previous study (Miller et al., 2021), metabolomic profiles of patient tissue samples successfully predicted lung cancer disease control and progressive disease groups as response to first-line therapy. In the current study, tissue-derived metabolites along with the covariates pathological staging at diagnosis, age, sex, histology subtype, and treatment vs. non-treatment were evaluated as markers for determining hazard ratios and survival probability in lung cancer patients. Our hypothesis is that metabolomic analysis of the tumor tissue itself has the potential to reveal biomarkers for distinguishing patients based on survival. These metabolic biomarkers would have clinical utility for calculating a weighted risk score to separate patients into low- and high-risk groups for OS and PFS.

2 Methods

The data for this study were obtained from an on-going study (Miller et al., 2021). Experimental workflow and data wrangling pipeline are summarized below and in Fig. 1.

2.1 Tissue collection

Patients were recruited from an ongoing study, as described in (Miller et al., 2021). Informed consent was obtained to participate. All specimens were collected following approved Internal Review Board protocols at University of Louisville Hospital (IRB 05.0523) and Norton Hospital (IRB 18.0264) from patients with known or suspected NSCLC. Demographic information, including age, sex, race, smoking history, personal history of malignancy, and relevant family history were recorded. Samples were collected by the clinical team, blinded to the research analysis.

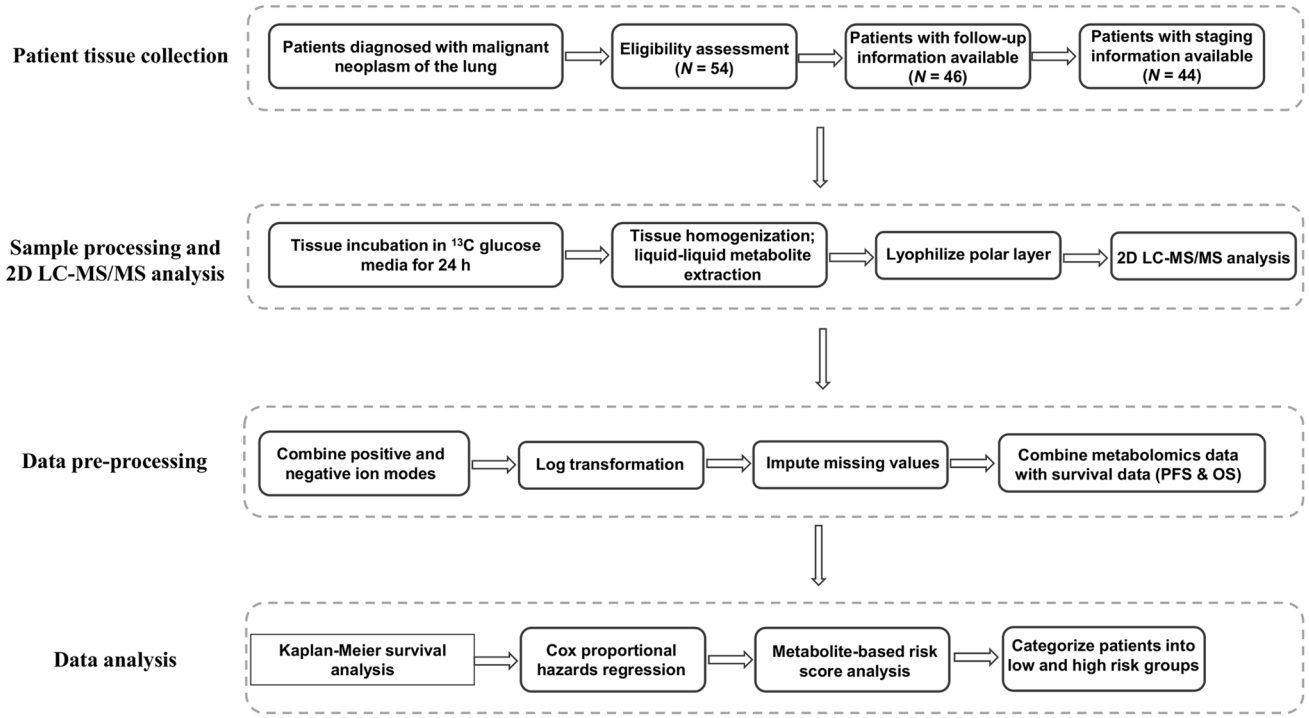

Fig. 1 Diagram of study workflow

2.2 Tissue processing and metabolite extraction

Tumor core biopsies were immediately placed into 1 mL ^{13}C labeled glucose medium in 24-well cell culture plates and incubated at 5% CO_2 and 37 °C on a shaker for 24 h. Following incubation, a liquid–liquid metabolite extraction was performed. Briefly, 500 mL acetonitrile was added, and tissue was homogenized with a pellet mixer. After 2–3 min homogenization, 376 mL of DNase/RNase free water and 250 mL chloroform were added. Contents were vortexed and centrifuged 180×*g* for 20 min. The top (polar) layer was aspirated and frozen at − 80 °C. A quality control sample was prepared from an NSCLC tissue biopsy incubated in unlabeled glucose media for 24 h and processed likewise. Sample polar layers were flash frozen in liquid N_2, then lyophilized for 24–48 h until dried and transported on ice to CREAM core facility for 2DLC-MS/MS analysis.

2.3 2DLC-MS/MS analysis and data pre-processing

All samples were analyzed in random order on a Thermo Q Exactive HF Hybrid Quadrupole-Orbitrap Mass Spectrometer coupled with a Thermo DIONEX UltiMate 3000 HPLC system (Thermo Fisher Scientific, Waltham, MA, USA) equipped with a reversed-phase chromatography (RPC) column and a hydrophilic interaction chromatography (HILIC) column configured to form a parallel 2DLC-MS system(Klavins et al., 2014). To obtain full MS data, every sample was analyzed by parallel 2DLC-MS in positive (+) and negative (−) modes. One unlabeled sample in each group was analyzed by 2DLC-MS/MS in positive and negative modes to acquire MS/MS spectra at 20, 40, and 60 eV collision energies as a quality control for metabolite identification. Data of unlabeled samples were first matched to an in-house database that contains parent ion m/z, MS/MS spectra, and retention time of authentic standards (MSI Level 1 identification). Data without a match (MSI Level 2 identification) were analyzed using Compound Discoverer software v2.0 (Thermo Fisher Scientific, Germany).

2.4 Organization of MS peak intensity data

2DLC-MS data was presented as an alignment table for each batch with retention time, m/z, signal intensity, stable isotope labeling, name of identified metabolite, and database used for metabolite identification. Once metabolites from positive and negative modes were combined, a preliminary step to handle missing values was performed by removing features that contained more than 50% missing values and resulted in a dataset of 52 metabolites with approximately 25.7% missing values. Based on previous analyses, some data appeared to be missing not at random (MNAR) while some was missing at random (MAR). It is clear that certain metabolites were not detected in the quality control samples when particular batches of samples were analyzed, but there were other instances of seemingly random missing values.

Therefore, we chose to handle missing values by imputation rather than limit of detection (LOD) substitution. Data were log-transformed (van den Berg et al., 2006), which is a common method to normalize biological data by centering it and correcting for heteroscedasticity (van den Berg et al., 2006), and imputed by probabilistic principal component analysis (PPCA) (Tipping & Bishop, 1999). A complete subset of the original data was taken and simulated missing values were introduced. A variety of substitution and imputation methods were tested, where imputation by PPCA resulted in the lowest RMSE (Miller et al., 2021). The R package *pcaMethods* was used to perform PPCA. The function was seeded with an integer of '1234' to allow for reproducibility and "maxIterations" was set to 1000.

2.5 Patient clinical data

Out of 54 eligible patients, 46 had follow-up and survival time information, 44 of which had staging information. Data collected included subject (age, sex, primary ethnicity, primary race, status, age at death, overall survival, cancer description, histology subtypes, cancer stage and substage, and progression-free survival in days), treatment (therapy type, therapy details, days of therapy, chemotherapy agent type, chemotherapy agent repeat units, number of chemotherapy cycles completed, surgery results, overall response to surgery/chemotherapy, days since diagnosis and response assessment types), and other information regarding the specimens.

2.6 Data analysis

All statistical analyses were conducted in the R programming language version 4.1.0. Hazard ratios were calculated by Cox proportional hazards regression model (*coxph* function in *survival* package). Survival curves were generated by Kaplan–Meier analysis with p-values found by log-rank test using high and low log transformed metabolite intensity values to stratify patient groups, defined as metabolite levels above or below the median (*survfit* function in *survival* package; *ggsurvplot* function in *survminer* package). Transformed metabolite intensity values were visualized by *heatmap.2* function in *gplots* package. Metabolite correlations were calculated by the *cor* function in the base R *stats* package. All possible combinations of three significant metabolites from the univariable Cox regression models were investigated in addition to covariates in the multivariable models. The optimal model was chosen based on the lowest average p-value across all three statistical tests [likelihood ratio test, Wald test, score (logrank) test]. Dimensionality reduction was achieved with partial least squares discriminant analysis (PLS-DA) to ensure the discrete dependent responses were taken into account. For PLS-DA component analysis, the dependent response variable was found by stratifying patients according to event (death or progression) vs. non-event. PLS-DA was implemented with *plsda* function in *mdatools* package. Pathological stage at diagnosis, sex, histological subtype, and treatment vs. non-treatment included as covariates, were one-hot encoded and classified as stage I/II (1) vs. stage III/IV (2), female (1) vs. male (2), adenocarcinoma (1) vs. squamous cell carcinoma (2), and non-treatment (1) vs. treatment (2).

2.7 Metabolite-based risk score analysis

A linear equation was developed to stratify patients into low and high risk groups based on significant metabolites. Univariable Cox proportional hazards regression model coefficients divided by the standard errors were used as weight coefficients multiplied by log-transformed metabolite intensity and summed to generate a risk score for each patient (Eq. 1). Patients categorized as high risk are expected to experience increased risk of event (death or progression), as defined by the median risk score as the cut-off:

$$Score_i = \sum_{k=1}^{num} w_k M_{i,k} \tag{1}$$

where $Score_i$ is the weighted risk score of each patient i, w_k is the weight coefficient assigned to metabolite k, $M_{i,k}$ is the log-transformed metabolite intensity for sample i and metabolite k, and num is the total number of significant metabolites identified by univariable Cox regression.

2.8 Validation of metabolite-based risk scores

A sparse partial least squares discriminant analysis (SPLS-DA) classification model was created using the covariates only and covariates combined with the metabolite-based risk scores for predicting event vs. non-event OS and PFS with fivefold cross validation and 100 resampling iterations. Classification model was trained using *train* function in *caret* package with method "spls". P-values between covariates only and covariates + risk scores AUROC values were calculated using unpaired T-test (assuming equal variances).

3 Results

3.1 Patient characteristics

Patient clinical characteristics are summarized in Table 1. Out of 46 patients with follow-up information, 14 expired and 32 were alive at the end of the respective follow-up period. The median overall survival was 20.50 months, and median progression-free survival was 17.41 months. Patients

Table 1 Clinical and pathological patient population characteristics

All patients (n = 46)	
Age at diagnosis	
Median age (range)	67 (42–95)
Gender	
Male	19
Female	27
Stage	
I	11
II	9
III	12
IV	12
Not available	2
Primary histology	
Non-small cell	46
Histology subtype	
Adenocarcinoma	38
Squamous cell carcinoma	8
Treatment	
Chemotherapy	19
Immunotherapy	6
Chemotherapy + immunotherapy	5
Surgical resection	17
Radiation	23
Status	
Alive	32
Expired	14
Survival time	
Median: overall (range)	20.50 months (0.69–101.94)
Median: progression-free (range)	17.41 months (0.69–76.40)

were evenly distributed in terms of pathological stage, while 44 of the 46 patients had pathological staging information, 30 of 46 patients received systemic therapy as some form of chemotherapy or immunotherapy, 17 had a surgical resection and 23 received radiation treatment. Primary histology of all 46 patients was non-small cell. Histology subtypes included 38 with adenocarcinoma and 8 with squamous cell carcinoma. All patients with pathological staging information (n = 44) were considered for analysis.

3.2 Metabolomic data and correlation analysis

A heatmap of the dataset with missing values present is in Supplementary Fig. 1. The maximum, minimum, and number of missing values in each patient category for the log-transformed metabolite intensity values are shown in Supplementary Table 1. Correlations were found between all metabolites in the dataset, displayed as a heat map (Supplementary Fig. 2A). Several pairs of strongly correlated (≥ 0.75) metabolites were identified (Supplementary Fig. 2B) and these pairs were filtered from univariable and multivariable Cox regression models. Correlations between metabolites identified as significant in the univariable Cox proportional hazards model analysis are reported for OS and PFS (Supplementary Table 2).

3.3 Univariable and multivariable analysis of metabolomic profiles

Metabolites were considered as separate variables during Cox proportional hazards model analysis, independent of patient characteristics. Covariates significant in the univariable models were adjusted for in the multivariable models. Cox regression coefficients and hazard ratios for metabolites related to OS and PFS are in Table 2.

In univariable analysis, 6 of 52 metabolites were identified as significant ($0.01 \leq p \leq 0.05$) and 4 were identified as highly significant ($p \leq 0.01$) for OS, while 4 metabolites were identified as significant and 1 was identified as highly significant for PFS. The first 3 components of PLS-DA analysis (PC1, PC2, PC3) were significant covariates for OS, while only PC2 was significant for PFS. The covariates pathological stage and sex were significant in the OS model, while none of the covariates were significant in the PFS model. Male patients and those having pathological stage III/IV had an increased risk of death (HR 4.75; HR 6.65). A similar trend (not significant) was observed for PFS. Significant metabolites from the univariable models were included in multivariable analysis along with the significant covariates (Table 2). All possible combinations of three significant metabolites from the univariable Cox regression models were investigated in addition to covariates in the multivariable models. The optimal model was chosen based on the lowest average p-value across all three statistical tests (likelihood ratio test, Wald test, score (logrank) test). Multivariable models achieved significance for OS (likelihood ratio test $p = 7E{-}05$, Wald test $p = 0.002$, log-rank test $p = 4E{-}05$) and for PFS (likelihood ratio test $p = 0.008$, Wald test $p = 0.002$, log-rank test $p = 4E{-}04$).

Kaplan–Meier survival analysis (Fig. 2) revealed that several metabolites contributed to OS and PFS probability over time, where the maximum OS was 101.6 months. Metabolites contributing to lower OS probability included hypoxanthine, L-pyroglutamic acid, N6,N6,N6-trimethyl-L-lysine, proline, adipic acid, benzoic acid, and inosine (Fig. 2A). For PFS, low levels of adenine and histamine and high levels of inosine contributed to shorter survival (Fig. 2B).

3.4 Risk score analysis

Metabolite based risk scores were calculated using Eq. 1 for OS and PFS (Figs. 3 and 4, respectively). For both OS

Table 2 Cox proportional hazards regression models of metabolites, PLS-DA components (PC's) and covariates

Univariable				
Overall survival	*P* value	Coefficient	HR	HR 95% CI
Acetyl-L-carnitine	0.0175	1.55	4.73	1.314–17.053
Creatine	0.0020	1.52	4.57	1.745–11.957
Cystine	1.89E−04	−1.56	0.21	0.093–0.477
Guanosine	0.0202	−0.90	0.41	0.190–0.869
L-Pyroglutamic acid	0.0265	1.36	3.90	1.173–12.976
N6,N6,N6-Trimethyl-L-lysine	2.59E−04	2.90	18.21	3.838–86.425
Pipecolinic acid	0.0371	−0.69	0.50	0.263–0.960
Proline	0.0124	1.29	3.63	1.322–9.954
Benzoic acid	0.0257	1.71	5.55	1.232–25.019
Inosine	0.0021	1.57	4.79	1.768–12.998
PC1 (PLS-DA)	4.51E−05	9.16	9495.70	116.57–773511.75
PC2 (PLS-DA)	0.0375	3.82	45.48	1.248–1657.547
PC3 (PLS-DA)	0.0328	3.18	24.08	1.297–447.066
Sex	0.0125	1.56	4.75	1.397–16.118
Age at diagnosis	0.0867	−0.05	0.95	0.901–1.007
Pathological stage (I/II vs. III/IV)	0.0145	1.89	6.65	1.456–30.322
Histology subtype	0.9985	−19.27	4.27E−09	0–Inf
Treatment vs. non-treatment	0.0883	−1.38	0.25	0.052–1.229
Progression-free survival	*P* value	Coefficient	HR	HR 95% CI
Glutamic acid	0.0082	−0.86	0.42	0.22–0.80
Benzoic acid	0.0190	1.82	6.17	1.35–28.24
Inosine	0.0410	0.87	2.39	1.04–5.53
Itaconic acid	0.0257	1.08	2.95	1.14–7.64
Pyruvate	0.0468	1.25	3.47	1.02–11.85
PC1 (PLS-DA)	0.2190	2.82	16.77	0.19–1502.87
PC2 (PLS-DA)	0.0031	5.49	242.09	6.40–9154.02
PC3 (PLS-DA)	0.6594	0.86	2.37	0.05–108.85
Sex	0.3473	0.47	1.61	0.60–4.32
Age at diagnosis	0.4542	−0.02	0.98	0.94–1.03
Pathological stage (I/II vs. III/IV)	0.1947	0.66	1.94	0.71–5.30
Histology subtype	0.2426	−1.21	0.30	0.04–2.27
Treatment vs. non-treatment	0.9979	17.11	2.69E+07	0–Inf
Multivariable				
Overall survival	*P* value	Coefficient	HR	HR 95% CI
Proline	0.1151	1.244	3.469	0.74–16.29
Inosine	0.1537	0.901	2.462	0.71–8.49
Guanosine	0.0466	−0.829	0.436	0.19–0.99
Sex	0.1197	1.287	3.620	0.71–18.31
Pathological stage (I/II vs. III/IV)	0.1469	1.174	3.234	0.66–15.79
Likelihood ratio test		Wald test		Score (logrank) test
7.00E−05		0.002		4.00E−05
Progression-free survival	*P* value	Coefficient	HR	HR 95% CI
Glutamic acid	0.0277	−0.677	0.508	0.28–0.93
Benzoic acid	0.3809	0.704	2.022	0.42–9.77
Inosine	0.2475	0.545	1.725	0.68–4.35

Table 2 (continued)

Likelihood ratio test	Wald test	Score (logrank) test
0.008	0.002	4.00E−04

Pathological stage at diagnosis, sex, histological subtype, and treatment vs. non-treatment included as covariates, were one-hot encoded and classified as stage I/II (1) vs. stage III/IV (2), female (1) vs. male (2), adenocarcinoma (1) vs. squamous cell carcinoma (2), and non-treatment (1) vs. treatment (2). Univariable analysis revealed 10 significant metabolites, 3 significant PC's and 2 significant covariates for Overall Survival (OS) while 5 significant metabolites and one significant PC were found for Progression Free Survival (PFS). Multivariable analysis resulted in a significant model after adjusting for sex and pathological stage, where guanosine remained as the only significant metabolite for OS, while glutamic acid remained as the only significant metabolite for PFS

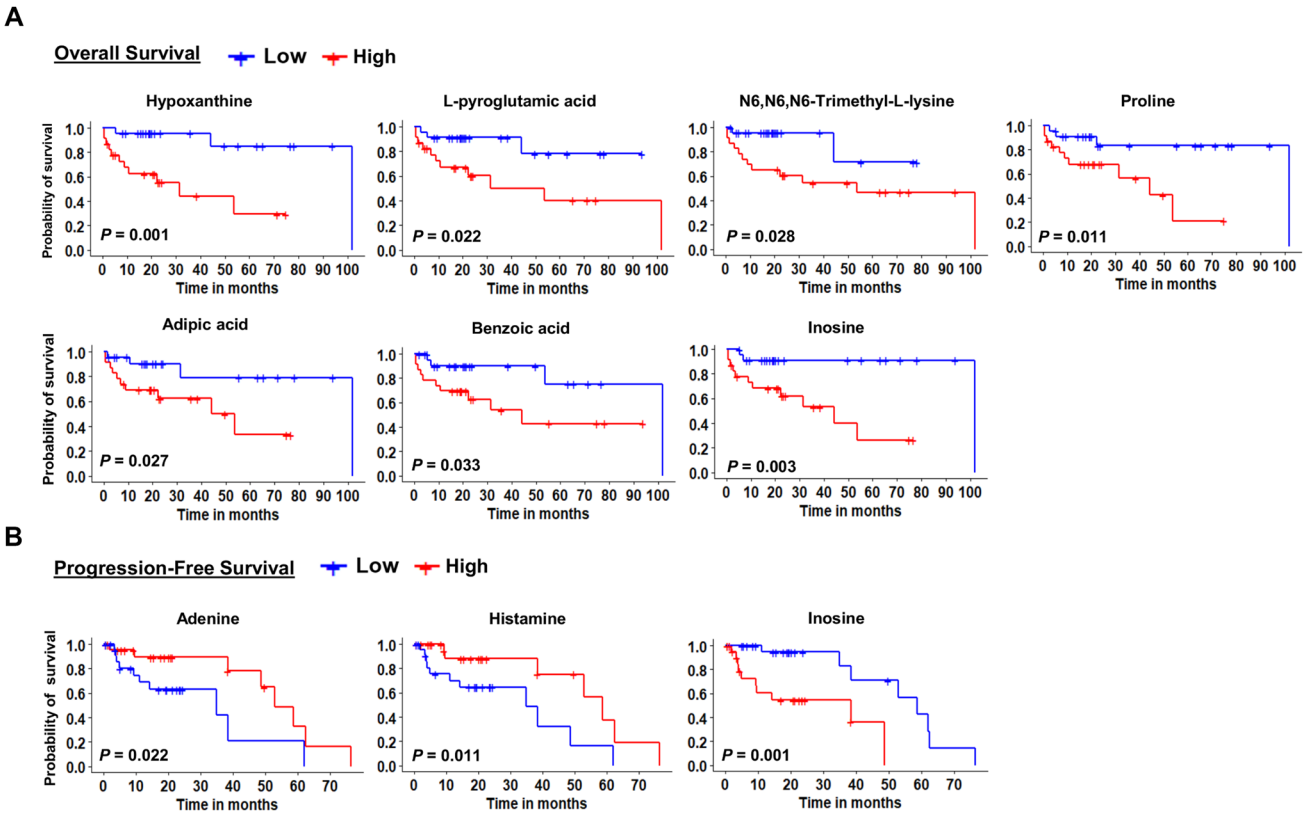

Fig. 2 Kaplan–Meier survival analysis. Survival curves of lung cancer patients with high metabolite levels (red; > median) and low metabolite levels (blue; ≤ median) for overall survival (**A**) and progression-free survival (**B**). P-values from log-rank test

and PFS, key metabolites were selected by significance in the univariable Cox proportional hazards regression model and correspond to the transformed intensity for each sample. In each case, the cut-off was defined as the median risk score to stratify high and low risk patients (Figs. 3A, 4A). Probability of survival for low and high risk patients was determined by Kaplan–Meier analysis, resulting in a marginally significant survival difference for OS ($p = 0.0764$) and highly significant difference for PFS ($p = 0.0012$) (Figs. 3B, 4B). Transformed intensity values of key metabolites among low and high risk patients are visualized as a heatmap (Figs. 3C, 4C). A Cox proportional hazards regression model was generated to determine hazard ratios of metabolite-based risk score and pathological stage at diagnosis (Figs. 3D, 4D). In univariable models, metabolite-based risk scores for OS [HR 1.15 (1.08–1.23), $p = 6.36E{-}06$] and for PFS [HR 1.29 (1.14–1.46), $p = 7.97E{-}05$] were significant predictors of survival. In OS multivariable model, metabolite-based risk score remained significant [HR 1.12 (1.06–1.19), $p = 2.46E{-}04$] after adjusting for sex and pathological stage. The multivariable model overall was significant for OS (likelihood ratio test $p = 2E{-}06$; Wald test $p = 2E{-}05$; log-rank test $p = 8E{-}08$).

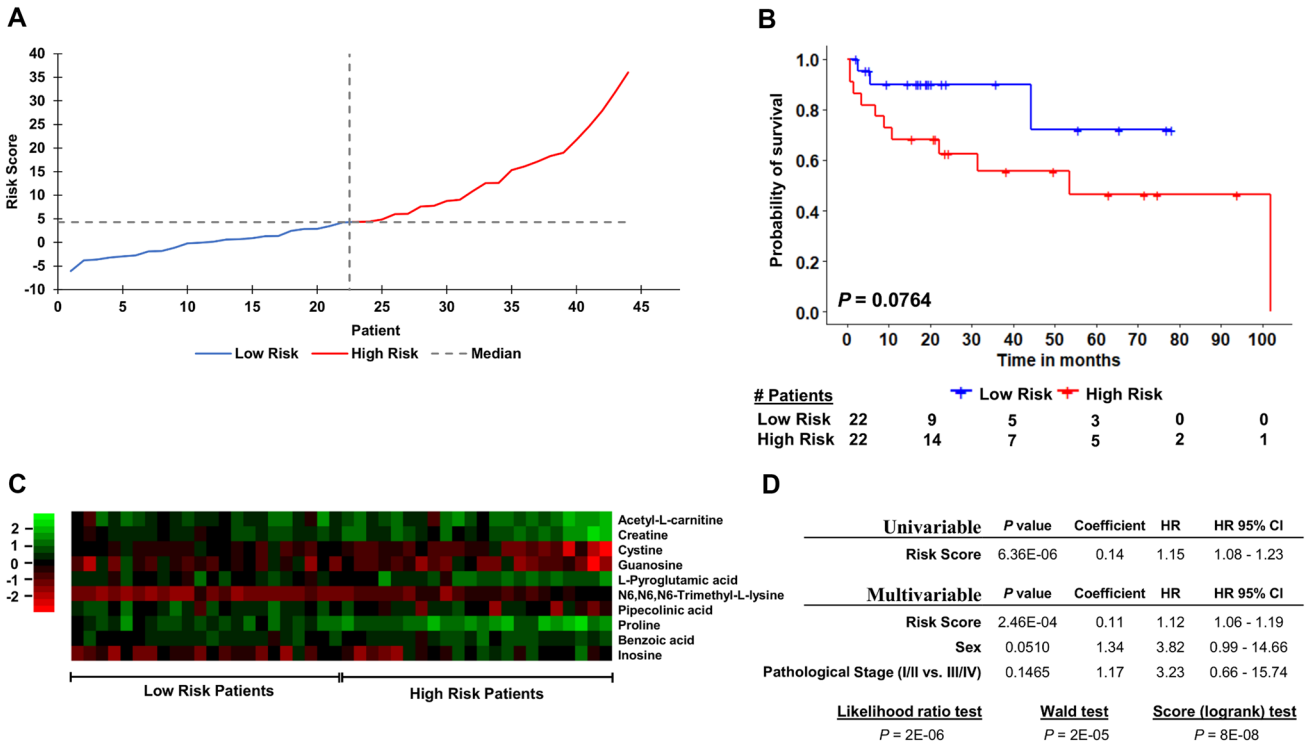

Fig. 3 Metabolite-based risk score analysis of lung cancer patient overall survival (OS). Patients were stratified into low and high risk based on the median risk score as the cut-off (**A**). Kaplan–Meier survival analysis revealed that low and high risk patients had a significant difference in survival probability over time (**B**). Heatmap of log transformed intensity values of key metabolites for low risk (left) and high risk (right) patients (**C**). Cox proportional hazards regression results with metabolite-based risk score, sex and pathological stage at diagnosis as predictors (**D**). Multivariable analysis indicates that the metabolite-based risk score was a more significant predictor of overall survival compared to sex or pathological stage at diagnosis. Pathological stage at diagnosis, sex, histological subtype, and treatment vs. non-treatment included as covariates, were one-hot encoded and classified as stage I/II (1) vs. stage III/IV (2), female (1) vs. male (2), adenocarcinoma (1) vs. squamous cell carcinoma (2), and non-treatment (1) vs. treatment (2)

3.5 Risk score validation

SPLS-DA classification model was employed for validation of metabolite-based risk scores in predicting event vs. non-event. With covariates only (sex, age, pathological stage, histological subtype, treatment vs. non-treatment), OS and PFS predictions achieved AUROC 0.819 and AUROC 0.652, respectively (Supplementary Fig. 3A). Classification model performance when covariates were combined with metabolite-based risk scores, resulted in statistically significantly ($p < 1E{-}10$) higher AUROC 0.868 and AUROC 0.755 for OS and PFS, respectively (Supplementary Fig. 3B).

4 Discussion

This proof-of-concept study tests the hypothesis that metabolomic analysis of tumor core biopsy samples routinely obtained during lung cancer patient care can reveal biomarkers for distinguishing patients based on OS and PFS. The study adhered to REMARK guidelines (McShane et al., 2005). The univariable analysis revealed metabolite-based risk scores derived from 10 key metabolites for OS and 5 key metabolites for PFS as significant predictors of survival. For OS, the multivariable model was significant after adjusting for sex and pathological stage at diagnosis, and the metabolite-based risk score was a more significant predictor than either of these covariates. In the multivariable model, guanosine remained as a significant metabolite for OS and glutamic acid remained significant for PFS. Five metabolites were identified in common for OS by both Kaplan–Meier survival analysis and Cox proportional hazards regression models (L-pyroglutamic acid, proline, benzoic acid, inosine, and N6,N6,N6-Trimethyl-L-lysine) and one metabolite was identified in common by both methods for PFS (inosine). The metabolite-based risk scores proved useful for predicting event vs. non-event in OS and PFS; when the risk scores were combined with covariates age, sex, pathological stage, histological subtype, and treatment vs. non-treatment, the AUROC improved compared to the AUROC obtained through the covariates alone. To our knowledge, this is the first study to show that tumor core biopsy-derived

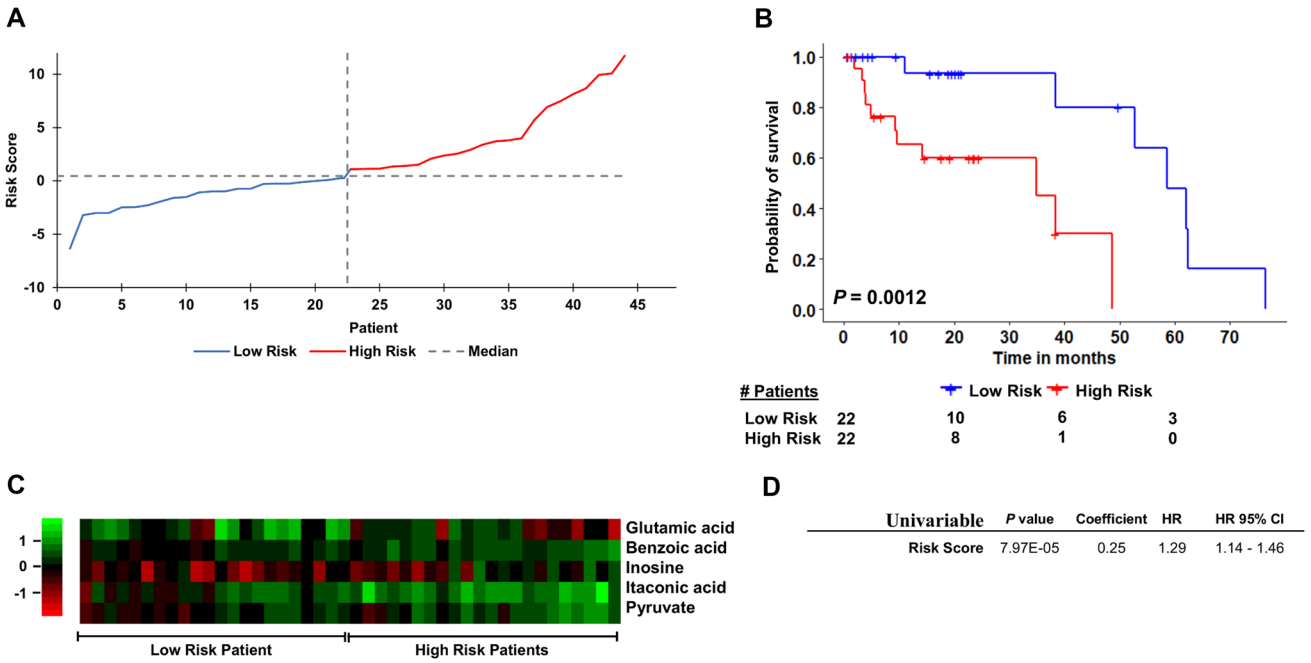

Fig. 4 Metabolite-based risk score analysis of lung cancer patient progression-free survival (PFS). Patients were stratified into low and high risk based on the median risk score as the cut-off (**A**). Kaplan–Meier survival analysis revealed that low and high risk patients had a significant difference in survival probability over time (**B**). Heatmap of log transformed intensity values of key metabolites for low risk (left) and high risk (right) patients (**C**). Cox proportional hazards regression results with metabolite-based risk score as predictor (**D**). Univariable analysis indicates that the metabolite-based risk score was a significant predictor of progression-free survival. Pathological stage at diagnosis, sex, histological subtype, and treatment vs. non-treatment included as covariates, were one-hot encoded and classified as stage I/II (1) vs. stage III/IV (2), female (1) vs. male (2), adenocarcinoma (1) vs. squamous cell carcinoma (2), and non-treatment (1) vs. treatment (2)

metabolomic profiles analyzed by 2DLC-MS/MS can risk stratify lung cancer patients based on survival probability.

Several metabolites related to nucleic acids were identified as significant by the models in this study, including the nucleobase adenine, nucleosides guanosine and inosine, and the purine derivative hypoxanthine. A recent study which investigated serum/plasma metabolomics for survival prediction in lung cancer patients receiving first-line chemotherapy found 4 metabolites which differed significantly ($p < 0.05$) between a discovery and validation cohort, including caffeine, paraxanthine, stachydrine, and methyl glucopyranoside (Shen et al., 2017). Metabolites related to caffeine metabolism generally had higher levels in patients with poor survival. In the current study, hypoxanthine was significantly associated with OS (Fig. 2A). Interestingly, hypoxanthine is also related to caffeine metabolism (Suzuki & Takahashi, 1975). However, the link to caffeine metabolism may be coincidental. Hypoxanthine guanine phosphoribosyltransferase (HPRT) was found to be expressed in cell membranes of NSCLC cell lines H460 and A549 (Townsend et al., 2017). H460 cells have nearly double HPRT expression than A549 and have roughly double the growth rate, indicating that HPRT expression may be directly related to cell proliferation (Townsend et al., 2017). HPRT is an enzyme that functions by transferring phosphoribose from phosphoribosyl pyrophosphate (PRPP) to hypoxanthine or guanine bases to form inosine monophosphate (IMP) and guanine monophosphate (GMP) (Stout & Caskey, 1985; Wilson et al., 1983). Therefore, it is unsurprising that high levels of hypoxanthine and inosine were strongly associated with shorter survival (Fig. 2A) as well as progression and death (Table 2). Interestingly, the opposite trend was observed with guanosine and adenine. High levels of adenine resulted in greater PFS probability over time (Fig. 2B), while increased guanosine was associated with OS (HR 0.41, $p = 0.0202$) (Table 2).

Several amino acids and amino acid derivatives were identified as significant, including glutamic acid, proline, L-pyroglutamic acid, N6,N6,N6-trimethyl-L-lysine, and pipecolinic acid. Hori et al. discovered that glutamic acid levels were lower in serum and higher in lung tissue samples of lung cancer patients compared to healthy controls (Hori et al., 2011). However, another study found that glutamic acid was significantly increased in serum of lung and breast cancer patients compared to healthy controls (Cascino et al., 1995). In the current study, increased glutamic acid in lung tumor tissue was associated with decreased risk of progression (HR 0.42, $p = 0.0082$) (Table 2). This suggests a

complex relationship between lung cancer survival and glutamic acid levels in tumor tissue compared to serum, and a more targeted experimental approach is required to elucidate the mechanisms involved.

The results further show that elevated proline was associated with death (HR 3.63, $p = 0.0124$) (Table 2), consistent with its role supporting cancer cell survival and proliferation (Tanner et al., 2018). High levels of pyroglutamic acid (PGA) were also associated with death (HR 3.90, $p = 0.0265$) (Table 2). PGA is known to accumulate during oxidative stress (Gueta et al., 2020) due to aberrant cancer cell redox homeostasis. Elevated N6,N6,N6-Trimethyl-L-lysine strongly associated with death (HR 18.21, $p = 2.59E{-}04$), highlighting its known role in cancer progression (Maas et al., 2020). Pyruvate was associated with decreased PFS (HR 3.47, $p = 0.0468$), likely due to its promotion of angiogenesis (Jung et al., 2011), tumor proliferation (Sellers et al., 2015), and immune downstaging (Abusalamah et al., 2020). Itaconic acid was also associated with decreased PFS (HR 2.95, $p = 0.0257$).

Higher abundance of cystine, composed of two cysteine molecules, was strongly associated with OS (HR 0.21, $p = 1.89E{-}04$) (Table 2). Cysteine is critical for cancer metabolic remodeling (Serpa, 2020), and abundance potentially due to reduced uptake may indicate decreased tumor metabolic dysregulation—consistent with previous findings that enhanced cystine processing is associated with more aggressive cancers (Ji et al., 2018). Similarly, elevated histamine (Medina & Rivera, 2010) was also associated with improved survival (PFS) probability (Fig. 2B), On the other hand, elevated acetyl-L-carnitine was associated with death (HR 4.73, $p = 0.0175$) (Table 2), consistent with its role in promoting angiogenesis (Cooke & Ghebremariam, 2008) and cancer cell proliferation (Yu et al., 2017). Increased creatine was further associated with death (HR 44.57, $p = 0.0020$) (Table 2). Creatine can be converted into phosphocreatine, which acts as an energy reservoir and is involved in the production of ATP from ADP (Loo et al., 2015). Increased creatine in tumor tissue may lead to an increase in ATP production, which is associated with high cancer metabolic activity (Puchades-Carrasco et al., 2016). In another study, creatine was significantly increased in lung cancer tissue compared to normal tissue, measured by NMR (Wald et al., 1997). In contrast, although several benzoic acid derivatives have shown anti-tumor activity (Anantharaju et al., 2017; Girouard et al., 2020; Yue et al., 2016; Zhang et al., 2021), little is known about how benzoic acid is involved in lung cancer metabolism. Previously, increased benzoic acid was found in late stage lung cancer patients relative to early stage and healthy controls (Hori et al., 2011), consistent with our findings that increased benzoic acid levels were associated with progression (HR 6.17, $p = 0.0190$) and death (HR 5.55, $p = 0.0257$) (Table 2).

PLS-DA component analysis was performed as an alternative to the metabolite based risk score calculated in Eq. 1, where the PC1 and PC2 scores were used for OS and PFS, respectively. The PLS-DA component was ineffective for separating low risk and high risk patients based on PFS in Kaplan–Meier analysis ($p = 0.165$) (data not shown). To increase separation, follow-up time was restricted to 0–48 months ($p = 0.0615$, Supplementary Figs. 4 and 5). Although PLS-DA component analysis resulted in significant Cox proportional hazards regression models (Supplementary Figs. 4D, 5D), they were not as significant as the metabolite-based risk scores (Figs. 3D, 4D). Further, to ensure that missing value imputation was not contributing to the survival analysis, the metabolomic data was one-hot-encoded as detected (0) vs. not-detected (1) and a Cox proportional hazards model was performed (Supplementary Table 3). None of the metabolites were significant, while only one and three different metabolites were respectively marginally significant ($0.05 \leq p \leq 0.1$) for OS and PFS, indicating that the presence or absence of metabolites in patient tissue samples was not affected by the events of disease progression or death.

This study includes a small and heterogeneous patient population, which limits the interpretation of results. Potential confounding factors include type and duration of chemotherapy/ immunotherapy cycles, smoking status, and differences in biopsy collection date relative to date of diagnosis. However, it was previously reported that adjusting for confounding factors such as age and cancer stage at diagnosis, smoking status, hypertension, diabetes mellitus, BMI, HDL/LDL levels, and time between sample collection and diagnosis had minimal significant effect on metabolite hazard ratios associated with prostate cancer-specific mortality (Huang et al., 2019). Our study along with (Huang et al., 2019) found that metabolomic profiles were able to uncover biomarkers relevant to cancer survival prediction without adjusting for these potential confounding factors, emphasizing the predictive capability of the metabolomics data. Additionally, it was previously found that number of chemotherapy cycles was a weak predictor of survival in epithelial ovarian cancer, being overpowered by pathological stage and a metabolite-based risk score (Xie et al., 2017). Further limitations of this study are that the data on association of metabolic score and OS or PFS would have to be validated in a larger and independent cohort. Considerations for future studies include integrating machine learning algorithms into a more comprehensive multivariable predictive modeling framework for high-accuracy predictions of patients in survival groups and simultaneous determination of a reliable set of key metabolites as biomarkers.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11306-022-01891-x.

Acknowledgements Authors acknowledge Melissa Hall, Andrei Smolenkov, and Danyelle Clark for James Graham Brown Cancer Center Biorepository support; Andrea Spencer, Lauren Whelan, Dr. John Hamm, and Dr. Stephen Wyatt for collection of samples at Norton Hospital.

Funding This work was supported by National Institutes of Health/National Cancer Institute Grant R15CA203605 (Frieboes).

Data availability Datasets used are in [Metabolomics Workbench (MetWB), RRID:SCR_013794, www.metabolomicsworkbench.org] (Study ST001527).

Declarations

Conflict of interest The authors declare that they have no competing interests.

Informed consent Informed consent was obtained from all individual participants included in the study.

Research involving human and animals participants This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by Internal Review Board protocols at University of Louisville Hospital (IRB 05.0523) and Norton Hospital (IRB 18.0264).

References

Abusalamah, H., Reel, J. M., & Lupfer, C. R. (2020). Pyruvate affects inflammatory responses of macrophages during influenza A virus infection. *Virus Research, 286*, 198088.

Anantharaju, P. G., Reddy, B. D., Padukudru, M. A., Kumari Chitturi, C. M., Vimalambike, M. G., & Madhunapantula, S. V. (2017). Naturally occurring benzoic acid derivatives retard cancer cell growth by inhibiting histone deacetylases (HDAC). *Cancer Biology & Therapy, 18*, 492–504.

Bamji-Stocke, S., van Berkel, V., Miller, D. M., & Frieboes, H. B. (2018). A review of metabolism-associated biomarkers in lung cancer diagnosis and treatment. *Metabolomics, 14*, 81.

Berker, Y., Vandergrift, L. A., Wagner, I., Su, L., Kurth, J., Schuler, A., Dinges, S. S., Habbel, P., Nowak, J., Mark, E., Aryee, M. J., Christiani, D. C., & Cheng, L. L. (2019). Magnetic resonance spectroscopy-based metabolomic biomarkers for typing, staging, and survival estimation of early-stage human lung cancer. *Science and Reports, 9*, 10319.

Cascino, A., Muscaritoli, M., Cangiano, C., Conversano, L., Laviano, A., Ariemma, S., Meguid, M. M., & Rossi Fanelli, F. (1995). Plasma amino acid imbalance in patients with lung and breast cancer. *Anticancer Research, 15*, 507–510.

Collino, S., Martin, F. P., & Rezzi, S. (2013). Clinical metabolomics paves the way towards future healthcare strategies. *British Journal of Clinical Pharmacology, 75*, 619–629.

Cooke, J. P., & Ghebremariam, Y. T. (2008). Endothelial nicotinic acetylcholine receptors and angiogenesis. *Trends in Cardiovascular Medicine, 18*, 247–253.

Dercle, L., Fronheiser, M., Lu, L., Du, S., Hayes, W., Leung, D. K., Roy, A., Wilkerson, J., Guo, P., Fojo, A. T., Schwartz, L. H., & Zhao, B. (2020). Identification of non-small cell lung cancer sensitive to systemic cancer therapies using radiomics. *Clinical Cancer Research, 26*, 2151–2162.

Fan, T. W., Lane, A. N., Higashi, R. M., Farag, M. A., Gao, H., Bousamra, M., & Miller, D. M. (2009). Altered regulation of metabolic pathways in human lung cancer discerned by (13)C stable isotope-resolved metabolomics (SIRM). *Molecular Cancer, 8*, 41.

Girouard, J., Belgorosky, D., Hamelin-Morrissette, J., Boulanger, V., D'Orio, E., Ramla, D., Perron, R., Charpentier, L., van Themsche, C., Eijan, A. M., Berube, G., & Reyes-Moreno, C. (2020). Molecular therapy with derivatives of amino benzoic acid inhibits tumor growth and metastasis in murine models of bladder cancer through inhibition of TNFalpha/NFKappaB and iNOS/NO pathways. *Biochemical Pharmacology, 176*, 113778.

Goldstraw, P., Crowley, J., Chansky, K., Giroux, D. J., Groome, P. A., Rami-Porta, R., Postmus, P. E., Rusch, V., Sobin, L., International Association for the Study of Lung Cancer International Staging, C. & Participating, I. (2007). The IASLC Lung Cancer Staging Project: Proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM classification of malignant tumours. *Journal of Thoracic Oncology, 2*, 706–714.

Gueta, I., Perach Ovadia, Y., Markovits, N., Schacham, Y. N., Epsztein, A., & Loebstein, R. (2020). Is pyroglutamic acid a prognostic factor among patients with suspected infection? A Prospective Cohort Study. *Scientific Reports, 10*, 10128.

Hori, S., Nishiumi, S., Kobayashi, K., Shinohara, M., Hatakeyama, Y., Kotani, Y., Hatano, N., Maniwa, Y., Nishio, W., Bamba, T., Fukusaki, E., Azuma, T., Takenawa, T., Nishimura, Y., & Yoshida, M. (2011). A metabolomic approach to lung cancer. *Lung Cancer, 74*, 284–292.

Howlader, N., Noone, A., Krapcho, M., Miller, D., Brest, A., Yu, M., Ruhl, J., Tatalovich, Z., Mariotto, A., Lewis, D. R., Chen, H. S., Feuer, E. J., & Cronin, K. A. (2021). *Based on November 2020 SEER data submission, posted to the SEER web site, April 2021. SEER Cancer Statistics Review, 1975–2018*. National Cancer Institute, Bethesda, MD.

Huang, J., Weinstein, S. J., Moore, S. C., Derkach, A., Hua, X., Mondul, A. M., Sampson, J. N., & Albanes, D. (2019). Pre-diagnostic serum metabolomic profiling of prostate cancer survival. *Journals of Gerontology. Series a, Biological Sciences and Medical Sciences, 74*, 853–859.

Ji, X., Qian, J., Rahman, S. M. J., Siska, P. J., Zou, Y., Harris, B. K., Hoeksema, M. D., Trenary, I. A., Heidi, C., Eisenberg, R., Rathmell, J. C., Young, J. D., & Massion, P. P. (2018). xCT (SLC7A11)-mediated metabolic reprogramming promotes non-small cell lung cancer progression. *Oncogene, 37*, 5007–5019.

Jin, X., Yun, S. J., Jeong, P., Kim, I. Y., Kim, W. J., & Park, S. (2014). Diagnosis of bladder cancer and prediction of survival by urinary metabolomics. *Oncotarget, 5*, 1635–1645.

Jung, S. Y., Song, H. S., Park, S. Y., Chung, S. H., & Kim, Y. J. (2011). Pyruvate promotes tumor angiogenesis through HIF-1-dependent PAI-1 expression. *International Journal of Oncology, 38*, 571–576.

Kami, K., Fujimori, T., Sato, H., Sato, M., Yamamoto, H., Ohashi, Y., Sugiyama, N., Ishihama, Y., Onozuka, H., Ochiai, A., Esumi, H., Soga, T., & Tomita, M. (2013). Metabolomic profiling of lung and prostate tumor tissues by capillary electrophoresis time-of-flight mass spectrometry. *Metabolomics, 9*, 444–453.

Klavins, K., Drexler, H., Hann, S., & Koellensperger, G. (2014). Quantitative metabolite profiling utilizing parallel column analysis for simultaneous reversed-phase and hydrophilic interaction liquid chromatography separations combined with tandem mass spectrometry. *Analytical Chemistry, 86*, 4145–4150.

Loo, J. M., Scherl, A., Nguyen, A., Man, F. Y., Weinberg, E., Zeng, Z., Saltz, L., Paty, P. B., & Tavazoie, S. F. (2015). Extracellular metabolic energetics can promote cancer progression. *Cell, 160*, 393–406.

Maas, M. N., Hintzen, J. C., Porzberg, M. R., & Mecinović, J. (2020). Trimethyllysine: From carnitine biosynthesis to epigenetics. *International Journal of Molecular Sciences, 21*, 9451.

McShane, L. M., Altman, D. G., Sauerbrei, W., Taube, S. E., Gion, M., Clark, G. M., Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics. (2005). REporting recommendations for tumor MARKer prognostic studies (REMARK). *Nature Clinical Practice. Urology, 2*, 416–422.

Medina, V. A., & Rivera, E. S. (2010). Histamine receptors and cancer pharmacology. *British Journal of Pharmacology, 161*, 755–767.

Miller, H. A., Yin, X., Smith, S. A., Hu, X., Zhang, X., Yan, J., Miller, D. M., van Berkel, V. H., & Frieboes, H. B. (2021). Evaluation of disease staging and chemotherapeutic response in non-small cell lung cancer from patient tumor-derived metabolomic data. *Lung Cancer, 156*, 20–30.

Peng, F., Liu, Y., He, C., Kong, Y., Ouyang, Q., Xie, X., Liu, T., Liu, Z., & Peng, J. (2018). Prediction of platinum-based chemotherapy efficacy in lung cancer based on LC-MS metabolomics approach. *Journal of Pharmaceutical and Biomedical Analysis, 154*, 95–101.

Puchades-Carrasco, L., Jantus-Lewintre, E., Perez-Rambla, C., Garcia-Garcia, F., Lucas, R., Calabuig, S., Blasco, A., Dopazo, J., Camps, C., & Pineda-Lucena, A. (2016). Serum metabolomic profiling facilitates the non-invasive identification of metabolic biomarkers associated with the onset and progression of non-small cell lung cancer. *Oncotarget, 7*, 12904–12916.

Sellers, K., Fox, M. P., Bousamra, M., Slone, S. P., Higashi, R. M., Miller, D. M., Wang, Y., Yan, J., Yuneva, M. O., Deshpande, R., Lane, A. N., & Fan, T. W. (2015). Pyruvate carboxylase is critical for non-small-cell lung cancer proliferation. *The Journal of Clinical Investigation, 125*, 687–698.

Serpa, J. (2020). Cysteine as a carbon source, a hot spot in cancer cells survival. *Frontiers in Oncology, 10*, 947.

Shen, J., Ye, Y., Chang, D. W., Huang, M., Heymach, J. V., Roth, J. A., Wu, X., & Zhao, H. (2017). Circulating metabolite profiles to predict overall survival in advanced non-small cell lung cancer patients receiving first-line chemotherapy. *Lung Cancer, 114*, 70–78.

Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2021). Cancer statistics, 2021. *CA: A Cancer Journal for Clinicians, 71*, 7–33.

Stout, J. T., & Caskey, C. T. (1985). HPRT: Gene structure, expression, and mutation. *Annual Review of Genetics, 19*, 127–148.

Suzuki, T., & Takahashi, E. (1975). Metabolism of xanthine and hypoxanthine in the tea plant (*Thea sinensis* L.). *The Biochemical Journal, 146*, 79–85.

Tanner, J. J., Fendt, S. M., & Becker, D. F. (2018). The proline cycle as a potential cancer therapy target. *Biochemistry, 57*, 3433–3444.

Tian, Y., Wang, Z., Liu, X., Duan, J., Feng, G., Yin, Y., Gu, J., Chen, Z., Gao, S., Bai, H., Wan, R., Jiang, J., Liu, J., Zhang, C., Wang, D., Han, J., Zhang, X., Cai, L., He, J., & Wang, J. (2018). Prediction of chemotherapeutic efficacy in non-small cell lung cancer by serum metabolomic profiling. *Clinical Cancer Research, 24*, 2100–2109.

Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society, 61*, 611–622.

Townsend, M. H., Anderson, M. D., Weagel, E. G., Velazquez, E. J., Weber, K. S., Robison, R. A., & O'Neill, K. L. (2017). Non-small-cell lung cancer cell lines A549 and NCI-H460 express hypoxanthine guanine phosphoribosyltransferase on the plasma membrane. *Oncotargets and Therapy, 10*, 1921–1932.

van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics, 7*, 142.

Wald, L. L., Nelson, S. J., Day, M. R., Noworolski, S. E., Henry, R. G., Huhn, S. L., Chang, S., Prados, M. D., Sneed, P. K., Larson, D. A., Wara, W. M., McDermott, M., Dillon, W. P., Gutin, P. H., & Vigneron, D. B. (1997). Serial proton magnetic resonance spectroscopy imaging of glioblastoma multiforme after brachytherapy. *Journal of Neurosurgery, 87*, 525–534.

Wikoff, W. R., Grapov, D., Fahrmann, J. F., Defelice, B., Rom, W. N., Pass, H. I., Kim, K., Nguyen, U., Taylor, S. L., Gandara, D. R., Kelly, K., Fiehn, O., & Miyamoto, S. (2015). Metabolomic markers of altered nucleotide metabolism in early stage adenocarcinoma. *Cancer Prevention Research (philadelphia, Pa.), 8*, 410–418.

Wilson, J. M., Tarr, G. E., & Kelley, W. N. (1983). Human hypoxanthine (guanine) phosphoribosyltransferase: An amino acid substitution in a mutant form of the enzyme isolated from a patient with gout. *Proceedings of the National Academy of Sciences USA, 80*, 870–873.

Xie, H., Hou, Y., Cheng, J., Openkova, M. S., Xia, B., Wang, W., Li, A., Yang, K., Li, J., Xu, H., Yang, C., Ma, L., Li, Z., Fan, X., Li, K., & Lou, G. (2017). Metabolic profiling and novel plasma biomarkers for predicting survival in epithelial ovarian cancer. *Oncotarget, 8*, 32134–32146.

Yu, H., Xia, H., Tang, Q., Xu, H., Wei, G., Chen, Y., Dai, X., Gong, Q., & Bi, F. (2017). Acetylcholine acts through M3 muscarinic receptor to activate the EGFR signaling and promotes gastric cancer cell proliferation. *Scientific Reports, 7*, 40802.

Yue, P., Lopez-Tapia, F., Paladino, D., Li, Y., Chen, C. H., Namanja, A. T., Hilliard, T., Chen, Y., Tius, M. A., & Turkson, J. (2016). Hydroxamic acid and benzoic acid-based STAT3 inhibitors suppress human glioma and breast cancer phenotypes in vitro and in vivo. *Cancer Research, 76*, 652–663.

Zhang, X., Wang, Y., Li, X., Wu, J., Zhao, L., Li, W., & Liu, J. (2021). Dynamics-based discovery of novel, potent benzoic acid derivatives as orally bioavailable selective estrogen receptor degraders for ERalpha+ breast cancer. *Journal of Medicinal Chemistry, 64*, 7575–7595.