

# Circulating Tumor Cell and Metabolites as Novel Biomarkers for Early-stage Lung Cancer Diagnosis

**Lingling Wan**

Cancer Institute, the Fourth Hospital of Hebei Medical University

**Yutong He**

Cancer Institute, the Fourth Hospital of Hebei University

**Qingyi Liu**

Thoracic Surgery, the Fourth Hospital of Hebei Medical University

**Di Liang**

Cancer Institute, the Fourth Hospital of Hebei Medical University

**Yongdong Guo**

Cancer Institute, the Fourth Hospital of Hebei Medical University

**Guangjie Liu**

Thoracic Surgery, the Fourth Hospital of Hebei Medical University

**Jinxia Ren**

Cancer Institute, the Fourth Hospital of Hebei Medical University

**Baoen Shan** (✉ [baoenshan62@sina.com](mailto:baoenshan62@sina.com))

Hebei Medical University Fourth Affiliated Hospital and Hebei Provincial Tumor Hospital

---

## Research

**Keywords:** Early stage lung cancer, LC-MS untargeted metabolomics, circulating tumor cell, next-generation sequencing

**Posted Date:** October 30th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-97860/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.  
[Read Full License](#)

---

## Abstract

**Background:** Lung cancer is a malignant tumor that has the highest morbidity and mortality rate among all cancers. Early diagnosis of lung cancer is a key factor in reducing mortality and improving prognosis.

**Methods:** In this study, we performed CTC next-generation sequencing (NGS) in early-stage lung cancer patients to identify lung cancer-related gene mutations. Meanwhile, a serum liquid chromatography-tandem mass spectrometry (LC-MS) untargeted metabolomics analysis was performed in the CTC-positive patients, and the early diagnostic value of these assays in lung cancer was analyzed.

**Results:** 62.5% (30/48) of lung cancer patients had  $\geq 1$  CTC. By CTC NGS, we found that > 50% of patients had 4 commonly mutated genes, namely, *NOTCH1*, *IGF2*, *EGFR*, and *PTCH1*. 47.37% (9/19) patients had *ARIDH1* mutations. Additionally, 30 CTC-positive patients and 30 healthy volunteers were subjected to LC-MS untargeted metabolomics analysis. We found 100 different metabolites, and 10 different metabolites were identified through analysis, which may have potential clinical application value in the diagnosis of CTC-positive early-stage lung cancer (AUC > 0.9).

**Conclusions:** Our results indicate that NGS of CTC and metabolomics may provide new tumor markers for the early diagnosis of lung cancer. This possibility requires more in-depth large-sample research for verification.

## Background

According to statistics provided by the International Agency for Research on Cancer (IARC) of the World Health Organization, there were approximately 2.09 million new cases of lung cancer and approximately 1.76 million deaths from lung cancer worldwide in 2018[1]. Both the morbidity and mortality rank first among cancers, and the 5-year survival rate of lung cancer at stage I is only 6%[2]. The main reason for this high mortality is that early-stage lung cancer lacks typical clinical symptoms, and when it is diagnosed, it is often already in the advanced stage, or even has metastases. Therefore, early diagnosis of lung cancer is a key factor in reducing cancer-related death and improving prognosis. At present, the commonly used diagnostic strategies for lung cancer include imaging, such as computed tomography, sputum exfoliation cytology, serum tumor markers and bronchoscopy, which can easily result in missed diagnoses and misdiagnosis. Therefore, it is urgently important to identify new biomarkers for the early diagnosis of lung cancer.

Circulating tumor cell (CTC) refers to tumor cells derived from tumor tissue that enter the peripheral blood circulation. Even in the early stage of tumor, tumor cells may flow into the circulatory system. Therefore, CTC is an important marker for liquid biopsy and can be used as a non-invasive and real-time monitoring tool for tumors to detect micrometastases[3, 4]. A French study showed that CTC could be employed to detect early-stage lung cancer 4–5 years earlier than LDCT in chronic obstructive pulmonary disease (COPD) patients[5]. In the past 10 years, the clinical application of CTC has primarily focused on the number of CTCs[6, 7]. In recent years, single-cell sequencing of CTC was performed by next-generation

sequencing (NGS) technology, which provided a novel and precise tool for CTC research[8, 9]. Our previous research[10] showed that LDCT combined with CTC may be a more effective method for early-stage lung cancer screening. Through NGS analysis of CTC, we identified three cancer-related genes *KIT*, *SMARCB1* and *TP53* in five CTC positive patients.

During the development of tumors, metabolites in the body change. More recently, researchers have turned to metabolomics to analyse specific metabolic markers for the early diagnosis of lung cancer[11]. Untargeted metabolomics can qualitatively identify differential metabolites between different groups, thereby screening potential disease markers. A clinical study on the early diagnosis of lung cancer, including 65 non-smoking female non-small cell lung cancer (NSCLC) patients, 6 benign lung tumors patients, and 65 healthy controls, determine that cysteine, serine and 1-monooleoylglycerol, as a biomarker panel, can be used to diagnose non-smoking female NSCLC[12]. Musharraf et al analysed the plasma of lung cancer patients, COPD patients, healthy smokers and healthy non-smokers using GC-MS technology and observed that the fatty acid and glucose levels of lung cancer patients were higher than those of patients in other groups[13]. Research by Ding et al. suggests that glucose metabolism disorders may be closely related to lung cancer, as indicated by the presence of such metabolites as glycerol phosphate, lactic acid, acetyl-CoA, and 3-phosphoglycerate[14]. However, the accuracy and reliability of the identification of metabolites are low, and no efficient diagnostic markers for lung cancer have been found.

Recent studies have verified that changes in cancer cell metabolites regulate the tumor microenvironment[15, 16]. In addition, CTC can survive in the peripheral blood through the immune escape mechanism and as latent tumor-initiating seeds that eventually break out to replace the host tissue[17]. In this study, we performed CTC NGS and LC-MS untargeted metabolomics analyses in early-stage lung cancer patients, to identify lung cancer-related gene mutations and metabolites in CTC-positive patients, and analyse the value of early diagnosis of lung cancer.

## Patients And Methods

### Patients

A total of 48 pathologically diagnosed as lung cancer patients were enrolled in this study from the Fourth Hospital of Hebei Medical University from Dec 2018 to Jan 2019. All these patients and healthy controls were volunteers from LDCT lung cancer screening from HeBei Province[10, 18]. None of the patients received preoperative radiotherapy or chemotherapy. Among the diagnosed lung cancer patients, 21 were male, and 27 were female. The age range was 38-75 years with an average age of 59.1 years. Postoperative pathology showed 3 squamous cell carcinomas and 45 adenocarcinomas. According to the eighth edition of the TNM staging criteria, 41 cases were stage I, and 7 cases were stage IIa. All patients underwent CTC *in vivo* before surgery, and peripheral blood was taken for metabolomic detection. Meanwhile, 30 healthy controls were enrolled in the group, and their personal characteristics, such as gender, age, and smoking history, matched those of the lung cancer group, there was no

significance. Approval was obtained from the Forth Hospital of Hebei Medical University ethics committee (Shijiazhuang, Hebei, China), and written informed consent was obtained from all the patients.

### **CTC capture and CTC PD-L1 identification**

The CellCollector is a medical stainless-steel wire with a 2 cm functional area coated with EpCAM antibodies and hydrogel coatings. CellCollector was punctured into the peripheral blood of the cubital vein through a 20G indwelling needle and was held in the body for 30 minutes to capture tumor cells. After the collection of CTC was completed, the CellCollector with captured CTC is stained and identified according to the instructions of the staining kit, and a negative control (NK92 cells) and positive control (SK-BR-3 cells) were also provided. CD45 (Exbio, Clone Mem-28-Alexa647) antibody, cytokeratin 7/19/pan-CK antibody (Exbio Praha, Clone A53-B/A2-Alexa488) and PD-L1 (Clone PD-L1, Abcam) antibody staining analysis was performed, and nuclear staining was subsequently performed by Hoechst 33342 (Sigma) to identify tumor cells and analyse the expression of PD-L1 in CTC.

### **Whole genomic amplification and next generation sequencing of CTC**

After identifying CTCs by immunofluorescence staining, the CTC area under the microscope was located and sheared. A small portion of the CTCs contained in the sampling needle was collected into a PCR tube, and the MALBAC method was employed to perform whole-genome amplification. Qubit 3.0 and Nanodrop 2000 (Thermo Fisher) were used for quantitative analysis. The eligibility criteria were as follows: Qubit 3.0>10 ng/ $\mu$ l, ND 2000>40 ng/ $\mu$ l. Quantitative PCR (ABI7500) was employed to detect the coverage of some tumor driver gene fragments. Each amplified sample was tested for coverage of 8 different segments, and 5 or more coverages were qualified. If both the output and the coverage were observed to meet the requirements, the amplification product was determined to pass the inspection.

In total, 50 ng of genomic DNA (Nanodrop concentration as the standard) was used to construct sequencing libraries using the Ion Ampliseq Library Kit 2.0 (Thermo Fisher) and Ion Ampliseq Cancer Hotspot Panel v2 in keeping with the manufacturer's instructions. Then, quality inspection and next-generation sequencing (NGS) were performed. NGS was performed with HiSeq X Ten (Illumina) following the manufacturer's protocols using a paired-end 150-bp (PE150) sequencing strategy with a 127-gene panel.

### **Metabolomics analysis**

#### **Metabolic substance extraction**

Thirty CTC-positive lung cancer patients were selected for fasting blood sampling in 5 mL tubes without anticoagulant, and serum was separated at 4°C (serum  $\geq$  200  $\mu$ L). Meanwhile, 30 healthy volunteers were selected for fasting blood sampling, and serum was separated at 4°C as the control group. LC-MS was used to detect metabolic differences.

A 100- $\mu$ L serum sample was collected, and the metabolite was extracted with 400  $\mu$ L methanol:acetonitrile (1:1, v/v) solution. The mixture was vortexed for 30 seconds and sonicated on ice for 10 min, and this step was repeated 3 times. The sample was placed at -20°C for 30 min. After centrifugation at 13000 g at 4°C for 15 min, the supernatant was carefully transferred to a sample bottle for LC-MS/MS analysis.

### **Metabolite detection**

As part of the system adjustment and quality control process, a combined quality control sample (QC) was prepared by mixing all samples of equal volume. QC samples were treated and tested in the same way as analytical samples. It was preferable to represent the entire sample set, and samples were injected at regular intervals (every 9 samples) to monitor the stability of the analysis.

Metabolites were analysed using the UPLC-Triple-TOF-MS-based platform. The chromatographic separation of metabolites was performed using an ExionLCTMAD system (AB Sciex, USA) equipped with ACQUITY UPLC BEH C18 column (100 mm × 2.1 mm i.d., 1.7  $\mu$ m; Waters, Milford, USA). Mobile phase A is water (containing 0.1% formic acid), mobile phase B is acetonitrile/isopropanol (1/1) (containing 0.1% formic acid); the flow rate is 0.40 mL/min, the injection volume is 20  $\mu$ L, and the column temperature is 40 °C.

### **Metabolite identification**

After UPLC-TOF/MS analysis, the raw data were imported into Progenesis QI 2.3 (Waters Corporation, Milford, USA) for peak detection and comparison. The preprocessing result generated a data matrix consisting of retention time (RT), mass-to-charge ratio ( $m/z$ ) values and peak intensity. At least 50% of the metabolic characteristics detected in all samples were retained. After filtering, half of the lowest metabolite value of a specific metabolite was estimated. In these specific samples, the metabolite level fell below the lower limit of quantification, and each metabolite characteristic was normalized by the sum. The internal standard was used for data quality control (reproducibility), metabolic characteristics were discarded, and the relative standard deviation (RSD) of quality control was > 30%.

By matching with the database (<http://www.hmdb.ca/>, <https://metlin.scripps.edu/> and other public databases and self-built databases), the metabolite list and data matrix were finally obtained. PCA and (O)PLS-DA were used for multivariate statistical analysis. In OPLS-DA analysis, variables with variable importance in projection (VIP) > 1 were considered difference variables. T-test combined with multivariate analysis OPLS-DA method was used to screen out the differential metabolites between groups (while meeting VIP > 1, p value < 0.05). Further advanced analyses, such as pathway analysis, association analysis, and cluster analysis, were used to analyse the biological information of differential metabolism.

### **Statistical analysis**

The data were analysed by SPSS 22.0 software. The measurement data followed the normal distribution using the mean ± standard deviation, those not following the normal distribution used the median

(quartile), and the counting data used frequency or rate. A t-test was used for the comparison of measurement data in a normal distribution, and a rank sum test was used for the comparison of measurement data in a non-normal distribution. The chi-square test was used to compare counting data. ROC curves were used to analyse the diagnostic effect of different indexes on lung cancer. P-values < 0.05 were considered to be significant.

## Results

### CellCollector *in vivo* CTC detection

A total of 30 of 48 patients had  $\geq 1$  CTC detected with the CellCollector *in vivo* strategy, and the detection rate was 62.5% (range, 0–17, median, 1). Figure 1A is a representative diagram of CTC capture. No CTC was detected in healthy controls (Fig. 1B). The detection rates of CTC were 64% (16/25) and 60.86% (14/23) in the < 60-year-old patients and  $\geq 60$ -year-old patients, respectively. The detection rates among male and female patients were 66.67% (14/21) and 59.26% (16/27), respectively. Regardless of whether the patients smoked, the detection rate was equivalent, 63.33% and 61.11%, respectively. In addition, we also found that the detection rate of stage I–II lung cancer patients was 65.85% (27/41), and the detection rate of stage III–IV patients was 42.86% (3/7). There was no correlation between CTC and clinical characteristics (Table 1). Meanwhile, PD-L1 protein was detected and 52.08% (25/48) of patients had PD-L1 expression on CTC (Fig. 1C).

Table 1  
Clinical characteristics of patients

Patients characteristic	CTC		p-value	PD-L1 + CTC		p-value
	negative	positive		negative	positive	
<b>Age</b>						
< 60 years	9(36.00%)	16(64.00%)	0.823	13(52.00%)	12(48.00%)	0.555
≥ 60 years	9(39.13%)	14(60.87%)		10(43.48%)	13(56.52%)	
<b>Gender</b>						
Male	7(33.33%)	14(66.67%)	0.599	11(52.38%)	10(47.62%)	0.585
Female	11(40.74%)	16(59.26%)		12(44.44%)	15(55.56%)	
<b>Smoking</b>						
No	11(36.67%)	19(63.33%)	0.878	13(43.33%)	17(56.67%)	0.412
Yes	7(38.89%)	11(61.11%)		10(55.56%)	8(44.44%)	
<b>Stage</b>						
I	14(34.15%)	27(65.85%)	0.460	20(48.78%)	21(51.22%)	1.000
II	4(57.14%)	3(42.86%)		3(42.86%)	4(57.14%)	
<b>Tumor size</b>						
< 1.5 cm	11(45.83%)	13(54.17%)	0.233	12(50.00%)	12(50.00%)	0.773
≥ 1.5 cm	7(29.17%)	17(70.83%)		11(45.83%)	13(54.17%)	
CTC, circulating tumor cell						

## Next generation sequencing of CTC

Preamplification and quality inspection were performed on 30 patients with CTC. 19 randomly selected samples were qualified for quality inspection. NGS was performed to analyse 127 cancer-related mutated genes (Figure S1). The gene mutation rate of 19 patients was 48.2–74.8%, among which 4 patients had a gene mutation rate of > 70%, 12 patients had a gene mutation rate between 60% and 70%, 2 patients had a gene mutation rate between 50% and 60%, and 1 patient had a gene mutation rate below 50% (Fig. 2A). To explore whether the lung cancer-related mutant genes could be screened early by CTC gene detection, we conducted a common mutation analysis on 19 CTC sequencing samples. According to the proportion of patients with common genetic mutations, we are divided into 3 groups, which are > 10%, > 30% and > 50% (Fig. 2B and Table 2). Among these group, > 50% of patients had 4 common mutation genes, namely,

*NOTCH1*, *IGF2*, *EGFR* and *PTCH1*. 47.37% (9/19) of patients had mutated *ARID1A*. Based on analysis of the COSMIC database, *NOTCH1* and *ARID1A* are reported to be associated with lung cancer.

Table 2  
Mutation gene information in > 10%, > 30% and > 50% patients

Gene	Mutation site	Mutation frequency (%)	Gene annotation
NOTCH1	NM_017617:exon34:c.7244_7246del:p. 2415_2416del	68.42	It belongs to the NOTCH family, and abnormal signals are associated with the occurrence of tumors.
IGF2	NM_001127598:exon5:c.686delC:p.P229fs	57.89	Insulin-like growth factor 2
PTCH1	NM_000264:exon22:c.3606delC:p.P1202fs	52.63	The negative regulatory gene of hedgehog signaling pathway. Hedgehog signaling pathway is associated with tumorigenesis and development
EGFR	NM_005228:exon9:c.1078delA:p.K360fs	57.89	A member of the epidermal growth factor receptor family, associated with tumor cell proliferation, angiogenesis, tumor invasion, and metastasis.
ARID1A	NM_006015:exon20:c.5542delG	47.37	Potential tumor suppressor gene
TP53	NM_000546:exon2:c.58delT:p.S20fs	36.84	Tumor suppressor gene
FGFR3	NM_001163213:exon18:c.2336delC:p.T779fs	31.58	Recombinant human fibroblast growth factor receptor-3, somatic FGFR3 mutations have been reported to be more common in superficial papillary bladder tumors.
PDGFRB	NM_002609:exon19:c.2594delT:p.L865fs	42.11	Human platelets derive growth factor receptors

Gene	Mutation site	Mutation frequency (%)	Gene annotation
ERBB2	NM_001005862:exon20:c.1911delG:p.L637fs	31.58	A member of the epidermal growth factor receptor family, associated with tumor metastasis and prognosis
SETD2	NM_014159:exon3:c.164delT:p.L55fs	47.37	Potential tumor suppressor gene, related to poor prognosis of tumor
SETD2	NM_014159:exon3:c.101delA:p.N34fs	31.58	Potential tumor suppressor gene, related to poor prognosis of tumor
NRG1	NM_001160004:exon10:c.934delA:p.K312fs	36.84	Epidermal growth factor receptor (EGFR) is a member of the epidermal growth factor receptor family
AXIN1	NM_003502:exon7:c.1922delA:p.K641fs	21.05	It is associated with ontogenesis, cell proliferation and carcinogenesis
TP53	NM_000546:exon4:c.98C > T:p.S33F	10.53	Tumor suppressor gene
ALK	NM_004304:exon7:c.1435delT:p.Y479fs	26.32	Oncogenic driver gene
SETD2	NM_014159:exon3:c.4319delC:p.P1440fs	10.53	Potential tumor suppressor gene, related to poor prognosis of tumor
BAP1	NM_004656:exon13:c.1464delC:p.P488fs	21.05	Closely related to tumor development

Gene	Mutation site	Mutation frequency (%)	Gene annotation
FGFR4	NM_022963:exon4:c.453delC:p.H151fs	26.32	Fibroblast growth factor, related to angiogenesis, involved in tumor relapse resistance
FGFR4	NM_022963:exon6:c.734delC:p.S245fs	10.53	Fibroblast growth factor, related to angiogenesis, involved in tumor relapse resistance
PTCH1	NM_001083604:exon23:c.3734G > A:p.G1245E	10.53	The negative regulatory gene of hedgehog signaling pathway. Hedgehog signaling pathway is associated with tumorigenesis and development

### Multivariate analysis for discriminating metabolites between lung cancer patients and control individuals

To determine whether there are specific differential metabolites in the early stage of lung cancer, this study used LC-MS to perform metabolomics analysis on serum samples of lung cancer patients and control individuals. The comparison of metabolite profiles between the lung cancer group and the control group and the overall difference between the groups showed that the early lung cancer group and the healthy control group could be significantly separated (Figure S2).

To further indentify the differential metabolic characteristics of the early-stage lung cancer group, the differential metabolites between groups were screened by t-test combined with multivariate OPLS-DA analysis and fold change analysis (FC) (Fig. 3A and 3B). The quality of PLS-DA was evaluated by  $R^2Y$  and  $Q^2$  values, which respectively represent the interpretive and predictive power of the model (Fig. 3C and 3D), combined with  $VIP > 1$ ,  $P < 0.05$  and  $FC > 1$  or  $FC < 1$ . It was observed that there were 100 metabolites with significant differences between the two groups (Fig. 3E and 3F). The most significant upregulation of differential metabolites in the early lung cancer group involved the lipid material, especially sphingolipids (such as trihexylceramide) and glycerophospholipids (such as cardiolipin), which are components of the cell membrane, and the differential metabolites are most significantly downregulated. Cyclic guanosine phosphate and guanosine 1-phosphate play important roles in purine nucleotide metabolism. In short, it is believed that these 100 differential metabolites are closely related to early lung cancer (Fig. 3G).

## Metabolic pathway analysis of the differentially regulated metabolites

100 differential metabolites screened were classified into compounds by the Human Metabolome Database (HMDB), and the KEGG (Kyoto Encyclopedia of Genes and Genomes) database was used for pathway annotation and enrichment analysis. There were 6 types of differential metabolites that we screened (Fig. 4A), most of which were lipids and their analogues, accounting for 78.87%, followed by organic acids and their derivatives, accounting for 8.45%, and phenylpropyl esters, accounting for 5.63%. The KEGG pathway analysis was divided into six categories (Fig. 4B). The three most significant pathways for enrichment through KEGG are choline metabolism in cancer, glycerophospholipid metabolism, and the retrograde endocannabinoid signalling pathway. Enrichment analysis of the KEGG pathway revealed that the highest rate of differential metabolite enrichment was the cytosolic DNA-sensing pathway and the pathogenic *Escherichia coli* infection pathway (Fig. 4C).

Previous studies have found that the occurrence and development of tumors are closely related to glucose and lipid metabolism[19, 20]. In this study, according to KEGG pathway analysis, there were more differential metabolites enriched in the glucose and lipid metabolism pathways and tumor metabolism pathways between the two groups (Fig. 4B). Among these metabolites, there were 21 differential metabolites involved in lipid metabolism pathways (Table S1), 5 differential metabolites involved in polysaccharide synthesis and metabolic pathways (Table S2), and 15 differential metabolites involved in tumor metabolic pathways (Table S3).

## Evaluation of the metabolic index in the diagnosis of lung cancer

To effectively screen potential biomarkers in the early-stage lung cancer group, we further compared the differential metabolites in lipid metabolic pathways, polysaccharide synthesis and metabolism pathways, and tumor-related pathways between the early-stage lung cancer group and the healthy control group using receiver operating characteristic (ROC) curve and area under the curve (AUC) (Figure S3). It was found that AUC > 0.9 contained 9 differential metabolites related to lipid pathways, 3 differential metabolites related to polysaccharide synthesis and metabolism, and 5 differential metabolites related to tumor pathways. These metabolites were primarily glycerophospholipids, which are integrated with 3 types of upregulated differences and 6 types of downregulated difference, namely, PE (14:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z)), PE (16:0/22:5(7Z,10Z,13Z,16Z, 19Z)), PE (14:0/20:4(5Z,8Z,11Z,14Z)), PC (18:0/20:4(8Z,11Z,14Z,17Z)), PC (16:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z)), PC (16:0/20:4(5Z,8Z,11Z,14Z)), LysoPC (16:1(9Z)/0:0), L-isoleucine, LysoPC (18:0), L-palmitoylcarnitine (Fig. 5 and Table 3).

Table 3  
The diagnostic efficacy of various metabolites in lung cancer

Metabolite	AUC	95%CI	p-value	Sensitivity	Specificity
PC(16:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	0.968	0.930 ~ 1.000	< 0.001	90.00%	96.67%
PC(18:0/20:4(8Z,11Z,14Z,17Z))	0.979	0.950 ~ 1.000	< 0.001	93.33%	93.33%
LysoPC(18:0)	0.933	0.876 ~ 0.991	< 0.001	93.33%	80.00%
PC(16:0/20:4(5Z,8Z,11Z,14Z))	1.000	1.000 ~ 1.000	< 0.001	100.00%	100.00%
LysoPC(16:1(9Z))	0.951	0.904 ~ 0.998	< 0.001	96.67%	83.88%
PE(14:0/20:4(5Z,8Z,11Z,14Z))	1.000	1.000 ~ 1.000	< 0.001	100.00%	100.00%
PE(14:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	1.000	1.000 ~ 1.000	< 0.001	100.00%	100.00%
PE(16:0/22:5(7Z,10Z,13Z,16Z,19Z))	1.000	1.000 ~ 1.000	< 0.001	100.00%	100.00%
L-Isoleucine	0.967	0.914 ~ 1.000	< 0.001	93.33%	96.67%
L-Palmitoylcarnitine	0.993	0.9817 ~ 1	< 0.001	93.33%	100.0%

## Discussion

In this study, CTC counting, CTC next-generation sequencing, and LC-MS untargeted metabolomics were combined to characterize the potential gene mutation and energy metabolism disturbance characteristics of lung cancer, to provide a better detection method for the early screening and diagnosis of lung cancer patients. We used CellCollector *in vivo* CTC capture technology to detect CTC in early lung cancer patients. The CTC detection rate was 62.5%. Compared with *in vitro* technology, it has a higher detection rate, which is consistent with previous research results[18, 21]. The high CTC detection rate provides convenience for CTC molecular typing and CTC next-generation sequencing.

Through CTC NGS, this study found that more than 50% of early lung cancer patients have 4 common mutated genes, namely *NOTCH1*, *IGF2*, *EGFR* and *PTCH1*. Also, 47.37% of patients have *ARID1A* mutations. *EGFR* has the highest mutation rate in *NSCLC*[22]. *NOTCH1*, a member of the PCG gene family, was first discovered in mouse B-cell lymphoma and is regarded as a co-oncogene of C-MYC, closely related to cell proliferation, differentiation and apoptosis[23]. Studies have shown that the stimulation of

the Notch signalling pathway by high *NOTCH1* expression can induce BM-1 to mediate the production of related intracellular signals to accelerate the transmission efficiency of lung cancer, thereby inducing the occurrence, development, metastasis and invasion of lung cancer[24]. In our study, 68.42% of patients had a mutation in *NOTCH1*. Huang et al.[25] found that homozygous *ARID1A* was deleted at the 5' end of the lung adenocarcinoma cell line, strongly suggesting that *ARID1A* is a tumor suppressor gene. Imielinska et al. reported exons and genome sequences of 183 cases of lung adenocarcinoma, and the results showed that mutations of the *ARID1A* gene existed in lung adenocarcinoma cells[26]. This study is the first to find *IGF2* and *RTCH1* mutations in peripheral blood CTC NGS of early lung cancer. Whether these mutations can be combined with *NOTCH1*, *EGFR* and *ARID1A* mutations as tumor markers in the diagnosis of early lung cancer merits further investigation.

Studies have found that CTC already exists in the early stages of cancer[27], and disturbances of metabolism are produced in the body[28], including disorders of glucose and lipid metabolism[13, 19, 29, 30], and the homeostasis of the microenvironment of the body is disrupted. Through metabolomics analysis, we found 100 different metabolites, which mainly occurred in lipid metabolism, polysaccharide synthesis and metabolism, amino acid metabolism and other pathways and were dominated by lipid metabolism, being especially enriched in choline metabolism and glycerophospholipid metabolic pathways. Chen et al. found that abnormal sphingolipid metabolism is the most important metabolic change in lung cancer patients[31]. A study on lung adenocarcinoma in female non-smokers found that abnormal lipid metabolism may play role in the development of lung cancer[12]. High-lipid molecules, including phospholipids (e.g., glycerophospholipids and sphingomyelin), and cholesterol are the main component of cell membranes and participate in cell signalling and cell proliferation. Lipid metabolism changes cause abnormal cell signals and lead to tumor formation[32, 33].

Tumor growth requires the uptake of a large amount of energy in the blood. The body ensures the normal energy metabolism of other organs by increasing the "raw materials" in the aerobic oxidation pathway, resulting in increased glucose metabolism and decreased fat metabolism[34]. The purpose of this study was to discover a combination of serum metabolite biomarkers for the early detection of non-small cell lung cancer. In our study, the most obvious differences in the screened metabolites can be divided into four categories, namely phosphatidylethanolamine (PE), phosphatidylcholine (PC), lysophosphatidylcholine (LyoPC), L-isoleucine and L-palmitoylcarnitine. A large number of metabolomic studies have been undertaken to identify robust biomarkers for lung cancer diagnosis using plasma, serum, or urine. However, we found remarkably few metabolomic studies that specifically attempted to detect early-stage lung cancer.

In our study, the concentration of LyoPC was reduced in stage I/II NSCLC, which is similar to previous research[35]. Another targeted metabolomics study found and verified that  $\beta$ -hydroxybutyric acid, LyoPC 20:3, PC ae C40:6, citric acid, and fumaric acid differed significantly between healthy controls and stage I/II NSCLC. Robust predictive models with  $AUC > 0.9$  were developed and validated using these metabolites and other, easily measured clinical data for detecting different stages of NSCLC[36]. It has been observed and reported that in mouse and human models, the plasma concentration of total LyoPC

is usually inversely related to the risk of various types of cancer[37–39]. In our study, we found that another member of the phosphatidylcholine family, PC(16:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z)), PC(18:0/20:4(8Z,11Z,14Z,17Z)) and PC(16:0/20:4(5Z,8Z,11Z,14Z)), also appears to play a role in both stage I and stage II NSCLC. A study reported that PC levels were dysregulated in early-stage NSCLC patients[40]. Decreased lipid membrane unsaturation levels were observed to protect tumor cells from free radicals or chemotherapeutics and promote invasion and infiltration[41]. Clearly, more detailed lipidomic studies need to be conducted to investigate the biological significance of these PC alterations.

Lysophosphatidylethanolamine (LPE) is a group of lipids that has been recently shown to be related to breast cancer[42]. In addition, PE (16:0/18:1) is associated with the stage and prognosis of pancreatic cancer and may be a potential diagnostic marker[43]. Yang et al. found 25 different lipid metabolites, including PE, between malignant pleural effusion (MPE) and benign pleural effusion (NPE), indicating that lipid metabolites may be used to partition MBE and BPE[44]. In our study, PE (14:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z)), PE (16:0/22:5(7Z,10Z,13Z,16Z, 19Z)), and PE (14:0/20:4(5Z,8Z,11Z,14Z)) were upregulated and effectively distinguished the control group, with specificity and sensitivity close to 100% being observed. In addition, we found that the level of amino acids (L-isoleucine) was significantly increased in the lung cancer model group compared with that of the control group, indicating disorder in amino acid metabolism in the cancer model group. Then, cell lesions occurred, causing diseases. Maeda et al. reported 6 significantly different amino acid metabolites, with AUCs of 0.817 and 0.801 (on their validation sets), for diagnosing stage I and stage II lung cancer[45]. One Study showed that L-palmitoylcarnitine is significantly reduced in advanced lung cancer patients[46]. In another study, the level of palmitoylcarnitine was lower in the hepatocellular carcinoma group than in the cancer-free control group, and blood acylcarnitine levels may be influenced by hepatic fatty acid metabolism, in other words, decreased acylcarnitine levels may reflect the decreased production of acyl groups in the liver or other tissues. Indeed, palmitoylcarnitine and palmitic acid are associated with fatty acid metabolism, and this group displayed an impact factor of 0.030 based on metabolic pathway analysis[47]. The decrease in L-palmitoylcarnitine in our study may also be related to a disorder of lipid metabolism in patients with lung cancer.

This study combines genomics and metabolomics to conduct a comprehensive assessment of early lung cancer in the search for biomarkers to enable early lung cancer screening and diagnosis and to improve the diagnosis rate of early lung cancer. However, there are still some shortcomings to this research. First, due to the small sample size in this study, a large sample study is needed to further verify the reliability of our research results. Second, the differential metabolites screened by untargeted metabolomics were not further verified by targeted metabolomics. In the end, we only performed CTC gene mutation detection but did not sequence the tissue. Whether there is a relationship between the mutations carried by CTCs and the changes in metabolic substances, which affect the microenvironment of patients, and whether there is a connection with the occurrence and development of lung cancer warrants further study.

## Conclusions

In our study, there was a higher CTC detection rate, with 62.5% in Ⅲ/Ⅳ NSCLC, and 4 high frequency mutation genes, namely, NOTCH1, IGF2, EGFR and PTCH1. Meanwhile, we found that 10 different metabolites may have potential clinical application value for the diagnosis of CTC-positive early-stage lung cancer (AUC > 0.9). Later, a larger cohort of patients will be required for verification, which may help determine whether these markers can be used for the early diagnosis of lung cancer.

## List Of Abbreviations

CTC, circulating tumor cell. NGS, next-generation sequencing. LC-MS, liquid chromatography-tandem mass spectrometry. IARC, International Agency for Research on Cancer. COPD, chronic obstructive pulmonary disease. NSCLC, non-small cell lung cancer. QC, quality control. RT, retention time. RSD, relative standard deviation. FC, fold change. HMDB, Human Metabolome Database. KEGG, Kyoto Encyclopedia of Genes and Genomes. ROC, receiver operating characteristic. AUC, area under the curve. PE, Phosphatidylethanolamine. PC, Phosphatidylcholine. LysoPC, lysophosphatidylcholine.

## Declarations

### Ethics approval and consent to participate

Approval was obtained from the Forth Hospital of Hebei Medical University ethics committee (Shijiazhuang, Hebei, China), and written informed consent was obtained from all the patients.

### Availability of data and materials

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

### Conflicts of Interest

All authors state that they have no conflicts of interest and declare that they have neither financial nor non-financial competing interests.

### Funding

This work was supported by the Financial Department of Hebei Province [No. [2016] 361006 and No. 2017043367-2].

### Author Contributions

WLL performed data analysis and interpretation, and drafted the manuscript. LQY and LGJ recruited the patients and collected clinical data. LD were responsible for the CTC data collection. GYD and RJJ participated in the LC-MS untargeted metabolomics analysis. HYT participated in the design of the study and performed the statistical analysis and helped to draft. SBE conceived of the study, and participated in

its design and coordination and helped to draft and revise the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We acknowledge Hongxia Yang from Viroad Biotechnology Co., Ltd. for technical support and CTC data collection.

## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA Cancer J Clin* 2018, **68**(6):394-424.
2. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE *et al*: **The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer.** *J Thorac Oncol* 2016, **11**(1):39-51.
3. Pantel K, Brakenhoff RH: **Dissecting the metastatic cascade.** #N/A 2004, **4**(6):448-456.
4. Wicha MS, Hayes DF: **Circulating tumor cells: not all detected cells are bad and not all bad cells are detected.** *J Clin Oncol* 2011, **29**(12):1508-1511.
5. Ilie M, Hofman V, Long-Mira E, Selva E, Vignaud JM, Padovani B *et al*: **"Sentinel" circulating tumor cells allow early diagnosis of lung cancer in patients with chronic obstructive pulmonary disease.** *PLoS One* 2014, **9**(10):e111597.
6. Pailler E, Adam J, Barthelemy A, Oulhen M, Auger N, Valent A *et al*: **Detection of circulating tumor cells harboring a unique ALK rearrangement in ALK-positive non-small-cell lung cancer.** *J Clin Oncol* 2013, **31**(18):2273-2281.
7. Krebs MG, Sloane R, Priest L, Lancashire L, Hou JM, Greystoke A *et al*: **Evaluation and prognostic significance of circulating tumor cells in patients with non-small-cell lung cancer.** *J Clin Oncol* 2011, **29**(12):1556-1563.
8. Heitzer E, Auer M, Gasch C, Pichler M, Ulz P, Hoffmann EM *et al*: **Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing.** *Cancer Res* 2013, **73**(10):2965-2975.
9. Bendall SC, Nolan GP: **From single cells to deep phenotypes in cancer.** *Nat Biotechnol* 2012, **30**(7):639-647.
10. He Y, Shi J, Shi G, Xu X, Liu Q, Liu C *et al*: **Using the New CellCollector to Capture Circulating Tumor Cells from Blood in Different Groups of Pulmonary Disease: A Cohort Study.** *Scientific Reports* 2017, **7**(1):9542.
11. Duarte IF, Rocha CM, Gil AM: **Metabolic profiling of biofluids: potential in lung cancer screening and diagnosis.** *Expert Rev Mol Diagn* 2013, **13**(7):737-748.

12. Mu Y, Zhou Y, Wang Y, Li W, Zhou L, Lu X *et al*: Serum Metabolomics Study of Nonsmoking Female Patients with Non-Small Cell Lung Cancer Using Gas Chromatography-Mass Spectrometry. *J Proteome Res* 2019, **18**(5):2175-2184.
13. Musharraf SG, Mazhar S, Choudhary MI, Rizi N, Atta ur R: Plasma metabolite profiling and chemometric analyses of lung cancer along with three controls through gas chromatography-mass spectrometry. *Sci Rep* 2015, **5**:8607.
14. Ding M, Li F, Wang B, Chi G, Liu H: A comprehensive analysis of WGCNA and serum metabolomics manifests the lung cancer-associated disordered glucose metabolism. *J Cell Biochem* 2019, **120**(6):10855-10863.
15. Renner K, Singer K, Koehl GE, Geissler EK, Peter K, Siska PJ *et al*: Metabolic Hallmarks of Tumor and Immune Cells in the Tumor Microenvironment. *Front Immunol* 2017, **8**:248.
16. Sekine H, Yamamoto M, Motohashi H: Tumors sweeten macrophages with acids. *Nat Immunol* 2018, **19**(12):1281-1283.
17. Massague J, Obenauf AC: Metastatic colonization by circulating tumour cells. *Nature* 2016, **529**(7586):298-306.
18. He Y, Shi J, Schmidt B, Liu Q, Shi G, Xu X *et al*: Circulating Tumor Cells as a Biomarker to Assist Molecular Diagnosis for Early Stage Non-Small Cell Lung Cancer. *Cancer Manag Res* 2020, **12**:841-854.
19. Yu S, Li J, Gao W, Wu Y, Qin X, Li Z: Uncovering the anticancer mechanism of petroleum extracts of Farfarae Flos against Lewis lung cancer by metabolomics and network pharmacology analysis. *Biomed Chromatogr* 2020:e4878.
20. Farre PL, Scalise GD, Duca RB, Dalton GN, Massillo C, Porretti J *et al*: CTBP1 and metabolic syndrome induce an mRNA and miRNA expression profile critical for breast cancer progression and metastasis. #N/A 2018, **9**(17):13848-13858.
21. Gorges TM, Penkalla N, Schalk T, Joosse SA, Riethdorf S, Tucholski J *et al*: Enumeration and Molecular Characterization of Tumor Cells in Lung Cancer Patients Using a Novel In Vivo Device for Capturing Circulating Tumor Cells. *Clin Cancer Res* 2016, **22**(9):2197-2206.
22. da Cunha Santos G, Shepherd FA, Tsao MS: EGFR mutations and lung cancer. *Annu Rev Pathol* 2011, **6**:49-69.
23. Pesch B, Casjens S, Stricker I, Westerwick D, Taeger D, Rabstein S *et al*: NOTCH1, HIF1A and other cancer-related proteins in lung tissue from uranium miners—variation by occupational exposure and subtype of lung cancer. *PLoS One* 2012, **7**(9):e45305.
24. Cai Hua-rong, Yue-quan J: Expression and significance of Notch1 and Bmi-1 protein in lung cancer tissue. *Chinese Journal of Tissue Engineering Research* 2016, **20**:683-687.
25. Huang J, Zhao YL, Li Y, Fletcher JA, Xiao S: Genomic and functional evidence for an ARID1A tumor suppressor role. *Genes Chromosomes Cancer* 2007, **46**(8):745-750.
26. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E *et al*: Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 2012, **150**(6):1107-1120.

27. Hosseini H, Obradovic MMS, Hoffmann M, Harper KL, Sosa MS, Werner-Klein M *et al*: **Early dissemination seeds metastasis in breast cancer.** *Nature* 2016, **540**(7634):552-558.
28. Ward PS, Thompson CB: **Metabolic reprogramming: a cancer hallmark even warburg did not anticipate.** *Cancer Cell* 2012, **21**(3):297-308.
29. 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 Analyt Technol Biomed Life Sci* 2014, **966**:147-153.
30. Perrotti F, Rosa C, Cicalini I, Sacchetta P, Del Boccio P, Genovesi D *et al*: **Advances in Lipidomics for Cancer Biomarkers Discovery.** *Int J Mol Sci* 2016, **17**(12).
31. Chen Y, Ma Z, Min L, Li H, Wang B, Zhong J *et al*: **Biomarker identification and pathway analysis by serum metabolomics of lung cancer.** *Biomed Res Int* 2015, **2015**:183624.
32. Ogretmen B, Hannun YA: **Biologically active sphingolipids in cancer pathogenesis and treatment.** #N/A 2004, **4**(8):604-616.
33. Lee GK, Lee HS, Park YS, Lee JH, Lee SC, Lee JH *et al*: **Lipid MALDI profile classifies non-small cell lung cancers according to the histologic type.** *Lung Cancer* 2012, **76**(2):197-203.
34. Luan Y, Li Y, Yue X, Cao Y, Xiang F, Mao D *et al*: **Metabonomics of mice intestine in Codonopsis foetens induced apoptosis of intestine cancer cells.** *Saudi J Biol Sci* 2019, **26**(5):1003-1010.
35. Guo S, Wang Y, Zhou D, Li Z: **Significantly increased monounsaturated lipids relative to polyunsaturated lipids in six types of cancer microenvironment are observed by mass spectrometry imaging.** *Sci Rep* 2014, **4**:5959.
36. Zhang L, Zheng J, Ahmed R, Huang G, Reid J, Mandal R *et al*: **A High-Performing Plasma Metabolite Panel for Early-Stage Lung Cancer Detection.** *Cancers (Basel)* 2020, **12**(3).
37. Raynor A, Jantscheff P, Ross T, Schlesinger M, Wilde M, Haasis S *et al*: **Saturated and mono-unsaturated lysophosphatidylcholine metabolism in tumour cells: a potential therapeutic target for preventing metastases.** *Lipids in health and disease* 2015, **14**:69.
38. Taylor LA, Arends J, Hodina AK, Unger C, Massing U: **Plasma lyso-phosphatidylcholine concentration is decreased in cancer patients with weight loss and activated inflammatory status.** *Lipids in health and disease* 2007, **6**:17.
39. Kuhn T, Floegel A, Sookthai D, Johnson T, Rolle-Kampczyk U, Otto W *et al*: **Higher plasma levels of lysophosphatidylcholine 18:0 are related to a lower risk of common cancers in a prospective metabolomics study.** *BMC medicine* 2016, **14**:13.
40. 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**(1):7137.
41. Rysman E, Brusselmans K, Scheys K, Timmermans L, Derua R, Munck S *et al*: **De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation.** *Cancer Res* 2010, **70**(20):8117-8126.

42. Cui M, Wang Q, Chen G: **Serum metabolomics analysis reveals changes in signaling lipids in breast cancer patients.** *Biomed Chromatogr* 2016, **30**(1):42-47.
43. Tao L, Zhou J, Yuan C, Zhang L, Li D, Si D *et al*: **Metabolomics identifies serum and exosomes metabolite markers of pancreatic cancer.** *Metabolomics* 2019, **15**(6):86.
44. Yang Z, Song Z, Chen Z, Guo Z, Jin H, Ding C *et al*: **Metabolic and lipidomic characterization of malignant pleural effusion in human lung cancer.** *J Pharm Biomed Anal* 2020, **180**:113069.
45. Maeda J, Higashiyama M, Imaizumi A, Nakayama T, Yamamoto H, Daimon T *et al*: **Possibility of multivariate function composed of plasma amino acid profiles as a novel screening index for non-small cell lung cancer: a case control study.** *BMC Cancer* 2010, **10**:690.
46. Tian Y, Wang Z, Liu X, Duan J, Feng G, Yin Y *et al*: **Prediction of Chemotherapeutic Efficacy in Non-Small Cell Lung Cancer by Serum Metabolomic Profiling.** *Clin Cancer Res* 2018, **24**(9):2100-2109.
47. Jee SH, Kim M, Kim M, Yoo HJ, Kim H, Jung KJ *et al*: **Metabolomics Profiles of Hepatocellular Carcinoma in a Korean Prospective Cohort: The Korean Cancer Prevention Study-II.** *Cancer Prev Res (Phila)* 2018, **11**(5):303-312.

## Figures

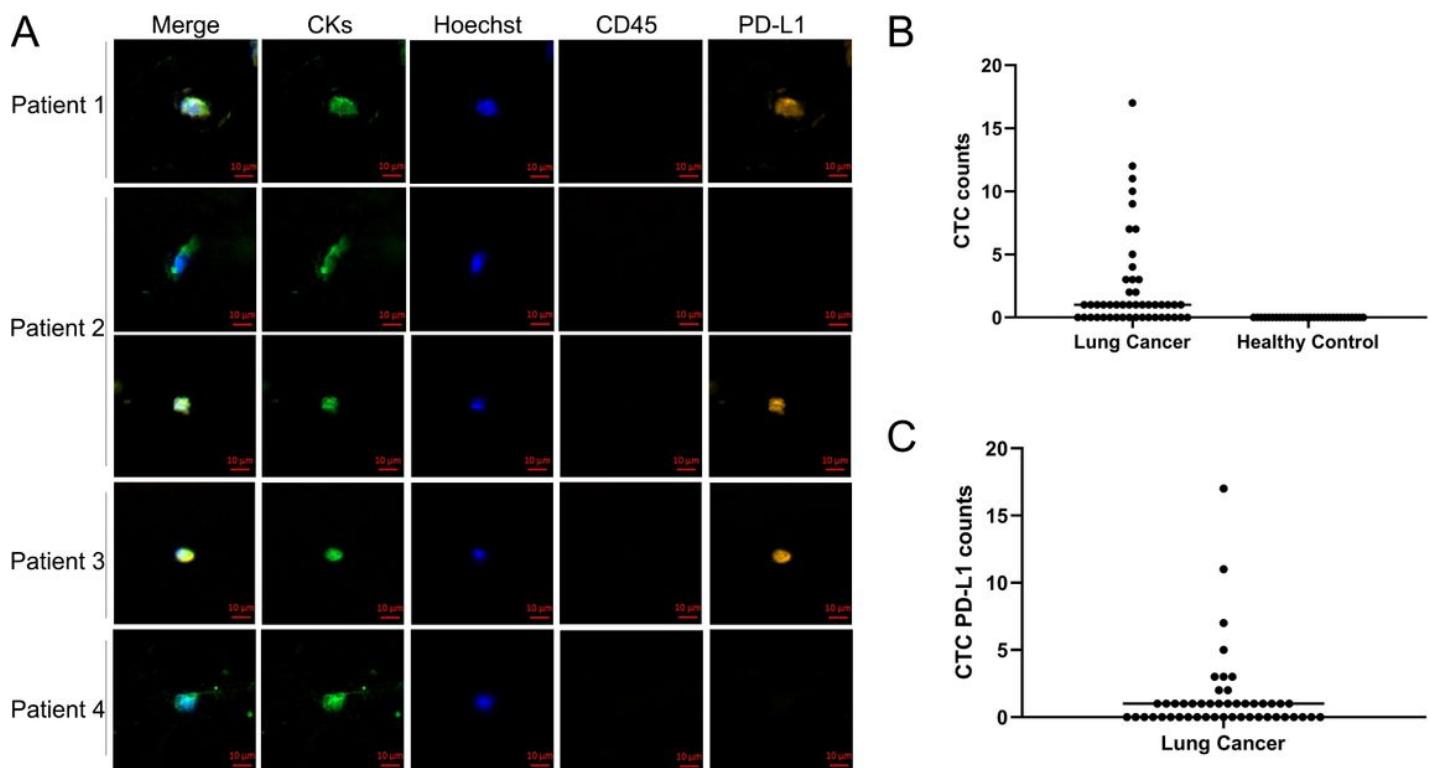
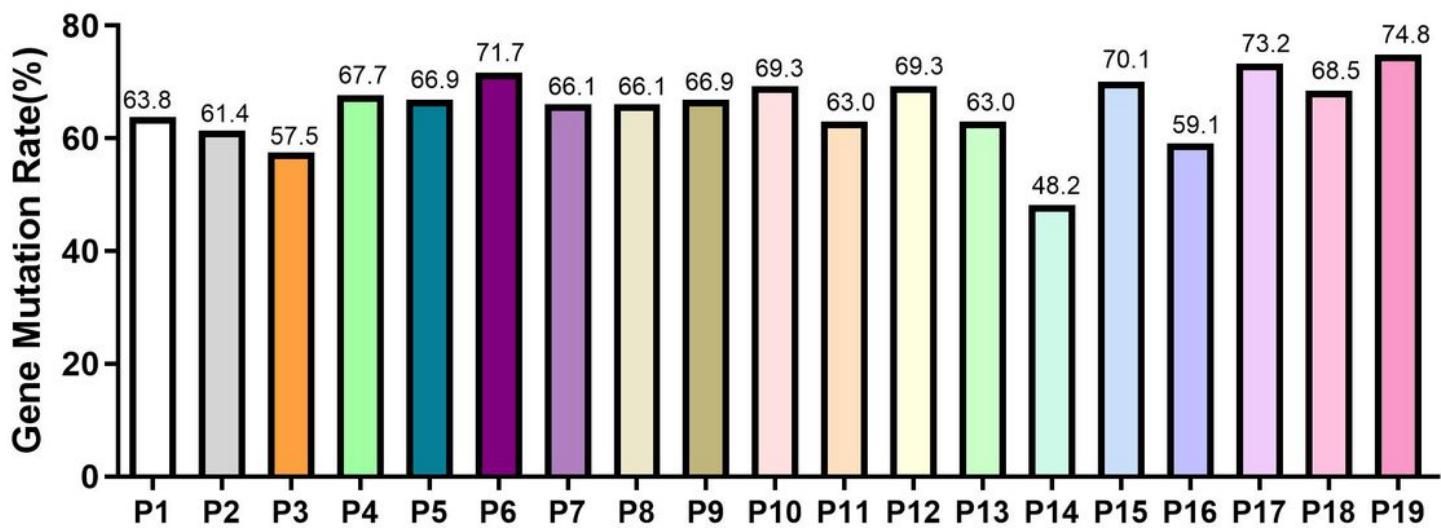


Figure 1

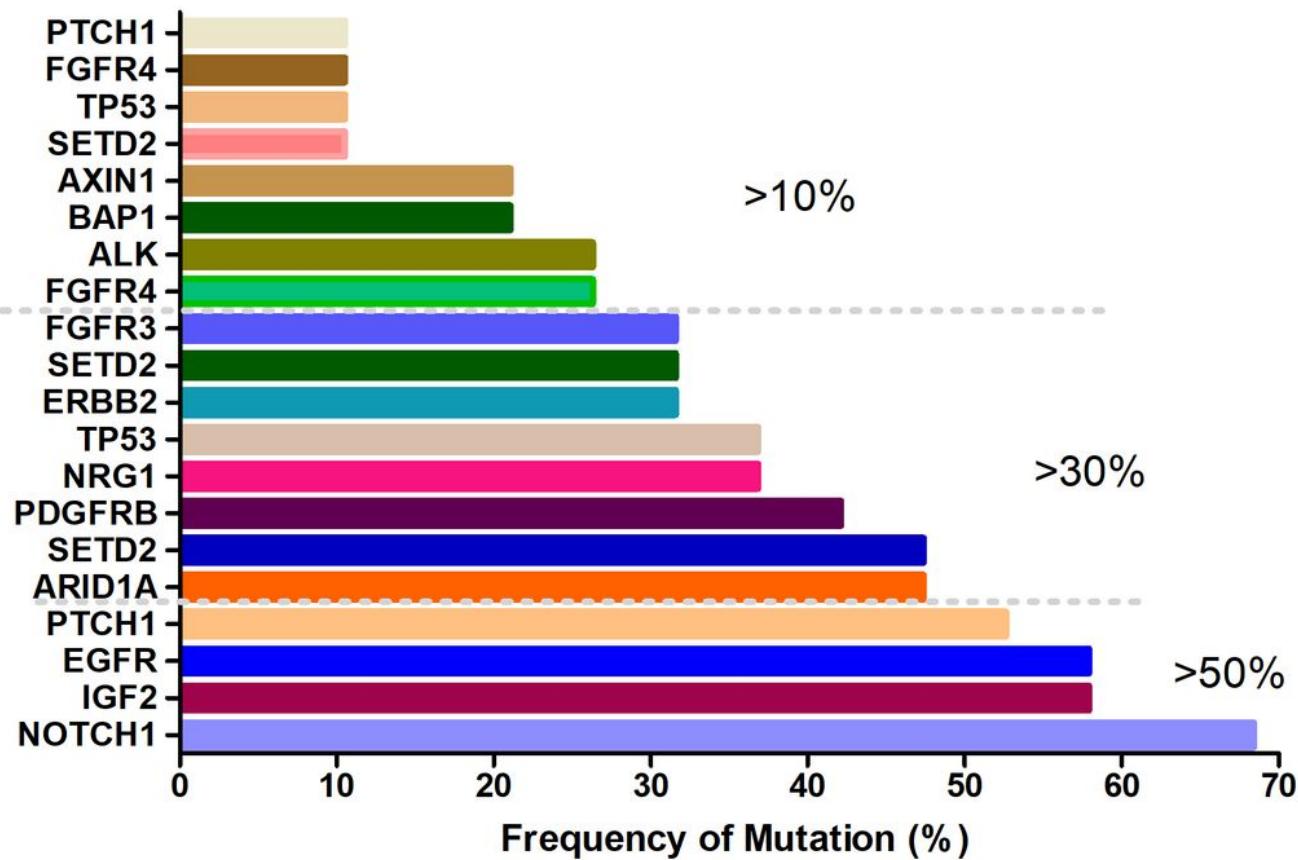
CellCollector captures representative CTC graphs and CTC and CTC PD-L1 detection. (A) CellCollector captures CTC identification charts. Patient 1 captures 7 CTCs, patient 2 captures 3 CTCs, patients 3 and 4

capture 1 CTC. (B) Detection rate and number of CTC in lung cancer patients and healthy controls. (C) Detection rate of PD-L1+ CTC in patients with lung cancer.

**A**

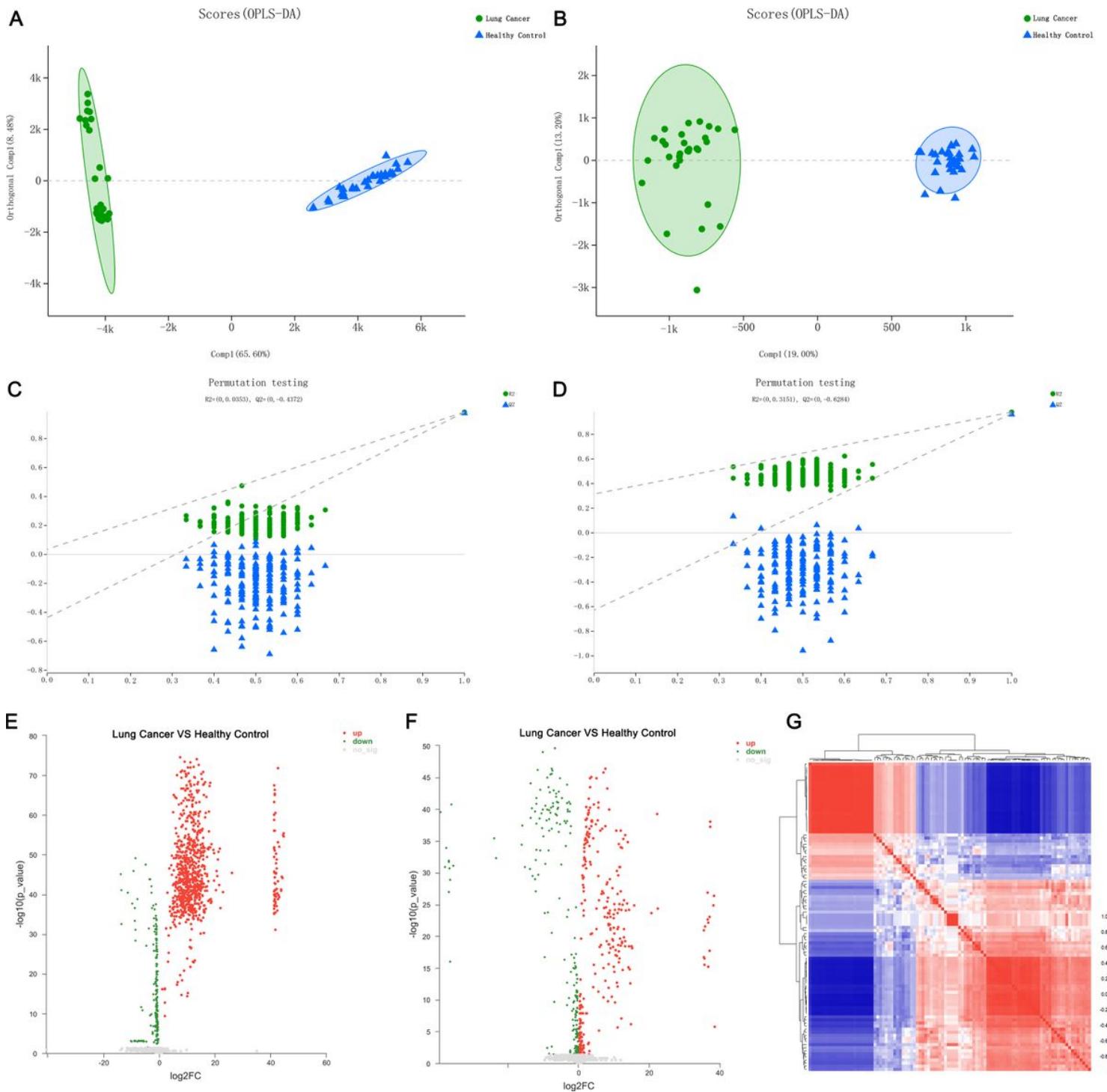


**B**



**Figure 2**

CTC NGS analysis of cancer-related gene mutations in lung cancer patients. (A) Proportion of mutated genes in 19 lung cancer patients. (B) > 10%, > 30%, > 50% of patients with mutated genes.

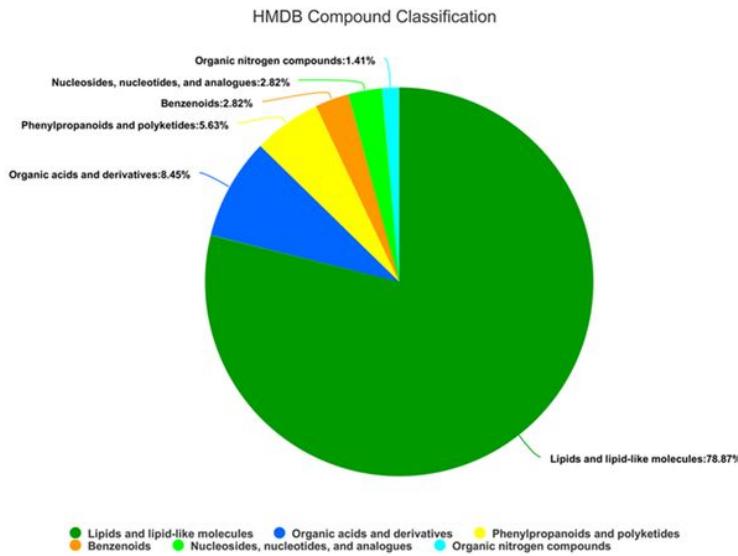


**Figure 3**

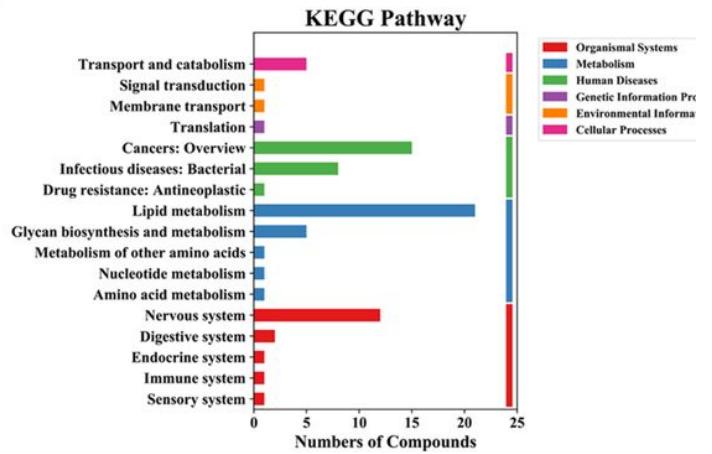
LC-MS untargeted metabolomics analysis of serum from the lung cancer group and the healthy control group. (A) OPLS-DA score of metabolite serum anions between the lung cancer group and healthy control group. (B) OPLS-DA score of metabolite serum cations in the lung cancer group and the healthy control group. (C) OPLS-DA differential metabolite serum anion permutation testing.  $R^2 = (0, 0.0353)$ ,  $Q^2 = (0, -0.4372)$ . (D) OPLS-DA differential metabolite serum cation permutation testing.  $R^2 = (0, 0.3151)$ ,  $Q^2 = (0, -0.6284)$ . (E) Volcano plot of serum anion metabolites in the control group and lung cancer group. (F)

Volcano plot of differential metabolites of serum cations in the healthy controls and lung cancer group.  
(G) Metabolite heat map of serum differences between healthy controls and lung cancer groups.

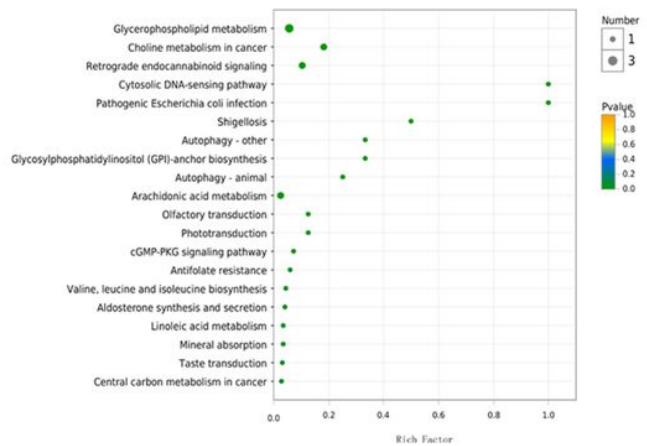
**A**



**B**

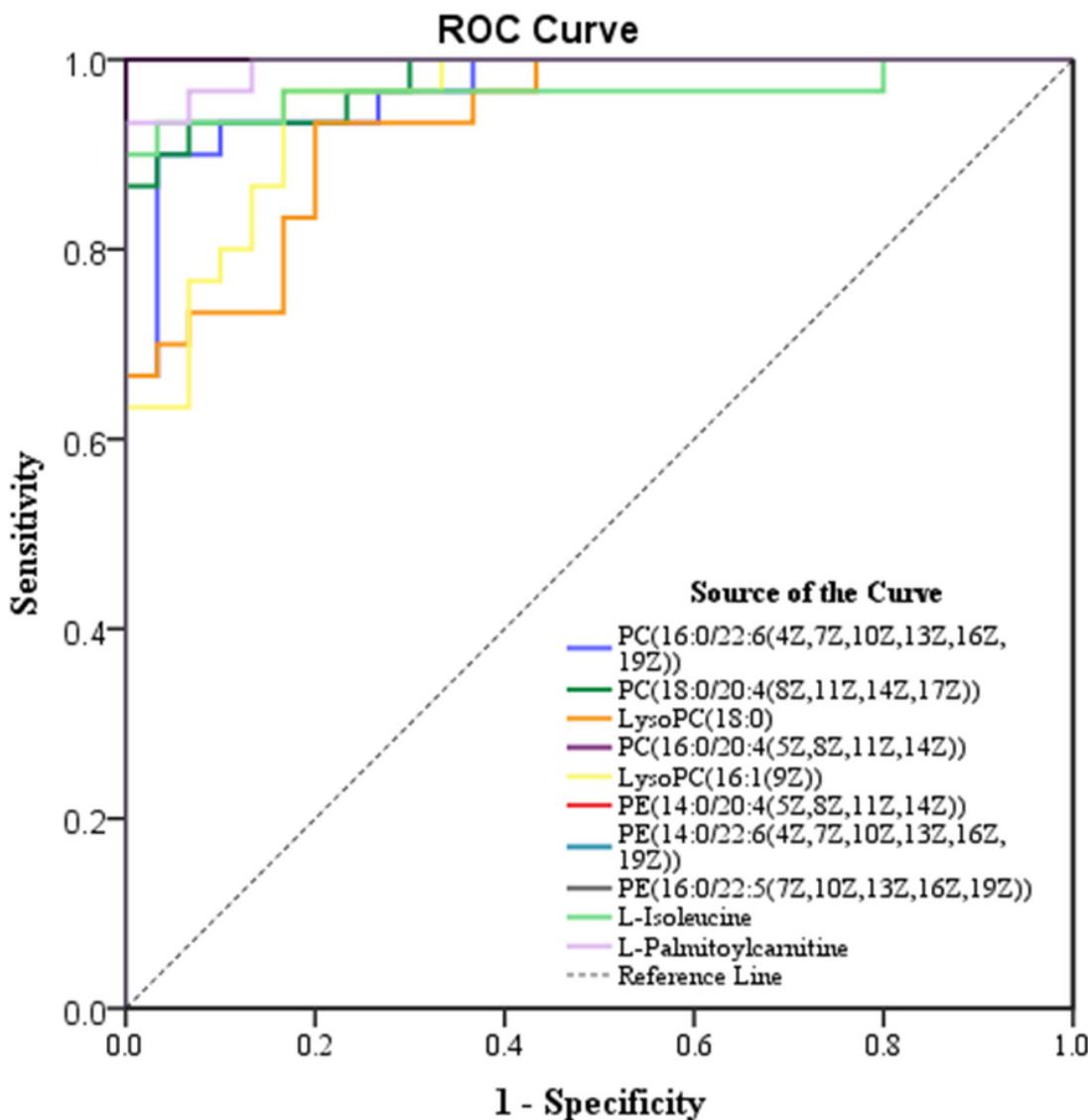


**C**



**Figure 4**

HMDB compound classification and KEGG pathway analysis of serum metabolites in the lung cancer group and healthy control group. (A) HMDB compound classification of serum metabolites between the lung cancer group and the healthy control group. (B) Serum KEGG pathways between the lung cancer group and the healthy control group. (C) KEGG enrichment pathways for serum metabolites in the lung cancer group and the healthy control group.



**Figure 5**

ROC curve of nine metabolites in the diagnosis of lung cancer or-related pathways.

## Supplementary Files

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

- [SupplementaryFile.doc](#)