*Oncologie*
DOI: 10.32604/oncologie.2020.012376
***Article***

# Diagnosis of Lung Cancer Based on Plasma Metabolomics Combined with Serum Markers

**Dong Zhang[1,#], Lingxia Tong[2,#], Qingjun Wang[1], Yunfeng Cao[3], Yu Gao[1], Donghua Yang[1], Tianhao Bao[4,5] and Zhitu Zhu[1,*]**

[1]Department of Oncology, The First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China
[2]Ultrasound Department, Jilin Cancer Hospital, Changchun, China
[3]Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China
[4]Mental Health Centre, University of Alberta, Edmonton, Canada
[5]Kunming Medical University, Kunming, China
[#]Dong Zhang and Lingxia Tong contributed equally to this article
[*]Corresponding Author: Zhitu Zhu. Email: 18742497386@163.com

**Abstract:** Predicting the onset and metastasis of early tumor is the primary means of improving lung cancer prognosis. The purpose of this study was to identify the ability of plasma metabolomics combined with blood markers to establish benign lung disease versus lung cancer regression models. Blood samples were collected from 174 lung cancer patients, 350 benign lung disease patients and 100 healthy volunteers and the metabolites were analyzed by mass spectrometry. The target metabolites consisted of 23 amino acids, 26 acylcarnitines and 45 conventional blood markers. A regression analysis model was established based on 12 metabolites and five blood markers selected by elastic network analysis. Two-thirds of the data were used in a training set for modeling and signature construction, and the remaining one-third were used in a validation set to test the model. This model was identified to have good specificity and sensitivity in distinguishing between lung cancer and lung disease. The performance of the model was evaluated using the area under the receiver operating curve, which was 0.915 in training set and 0.875 in validation set. In conclusion, this study demonstrates that regression model established by plasma metabolomics in combination with conventional serum markers has potential for the diagnosis of lung cancer.

**Keywords:** Lung cancer; metabolomics; blood; biomarker

## 1 Introduction

Lung cancer is one of the most common cancers in humans, accounting for nearly 15% of all new cancers [1]. Lung cancer is also one of the most deadly cancers due to early stage detection difficulties, with a 5-year survival rate of less than 15% [2]. Most early stage lung cancer patients have no obvious signs and symptoms, which reduces the likelihood of early diagnosis and treatment. Currently, the main measures for the diagnosis of lung cancer include protein biomarker quantification, radiography, sputum cytology and endoscopy. These traditional methods are used to determine the stage, location and metastasis of lung cancer. However, they have some limitations, such as early diagnosis difficulties, high cost of examination, and lack of application for population screening. Therefore, new diagnostic methods for lung cancer are urgently needed [3,4]. Notably, the biomarkers of lung cancer also show markedly high levels in benign lung diseases. Due to its low sensitivity, tumor markers are generally not recommended as a tool for early detection of lung cancer [5,6]. To date, no serum biomarkers are

available for lung cancer diagnosis.

Metabolomics is an emerging field in which the levels of small molecule metabolites in biological fluids or tissues are monitored at specific times under specific environmental conditions to study global metabolic differences in biological systems [7]. Cancer metabolism is an essential aspect of tumorigenesis [8]. Disordered metabolism is a key event in cancer development and progression [9]. Plasma has been used to identify cancer metabolic biomarkers by using targeted and non-targeted methods [10]. NMR, GC-MS and LC-MS are the main techniques used in metabolomics analysis, and LC-MS is the latest technological advancement in metabolomics that further improves the detection sensitivity and data reliability of cancer metabolomics [11]. In recent years, plasma or serum metabolomics has been widely applied to the differentiation of malignant tumors and benign diseases.

The purpose of this study was to identify the ability of plasma metabolomics combined with blood markers to establish benign lung disease versus lung cancer regression models.

## 2 Materials and Methods

### *2.1 Subjects*

Flow chart for patient selection was shown in Fig. 1. A total of 174 lung cancer (LC) patients, 350 benign lung disease (LD) patients and 100 healthy volunteers (CONT) from the First Affiliated Hospital of Jinzhou Medical University were collected. LC patients with non-small cell carcinoma (NSCLC) included adenocarcinoma (n = 93), squamous cell carcinoma (n = 42), and adenosquamous carcinoma (n = 3). There were 27 small cell lung cancer (SCLC) patients and 13 lung cancer with undefined molecular subtypes. The mean ages of patients with LC and LD were 60.3 (33–84) years and 62.8 (31–91) years, respectively. The study was approved by the Ethics Committee of the First Affiliated Hospital of Jinzhou Medical University, Liaoning. Of LC patients collected, 30 patients were in stage I, and remaining were in Stage II (16 patients), Stage III (43 patients), and Stage IV (71 patients) according to AJCC (Tab. 1). All the patients signed informed consents. The diagnosis of patients was confirmed by histology; all the patients received no medical treatment and surgical operation before this study.

**Table 1:** Clinical characteristics of the patients included in this study

| Number of samples, n | LD | LC | CONT |
|---|---|---|---|
| Age (years) (range) | 62.8 (31–92) | 60.3 (33–84） | 57.8 (29–81) |
| Gender | | | |
| Male | 79 | 72 | 69 |
| Female | 53 | 62 | 31 |
| Clinical stage according to TNM classification | | | |
| I, n | / | 30 | / |
| II, n | / | 17 | / |
| III, n | / | 43 | / |
| IV, n | / | 71 | / |
| Undefined stage | / | 13 | / |

LD: Benign Lung Disease; LC: Lung Cancer; TNM: Tumor, Node, and Metastasis

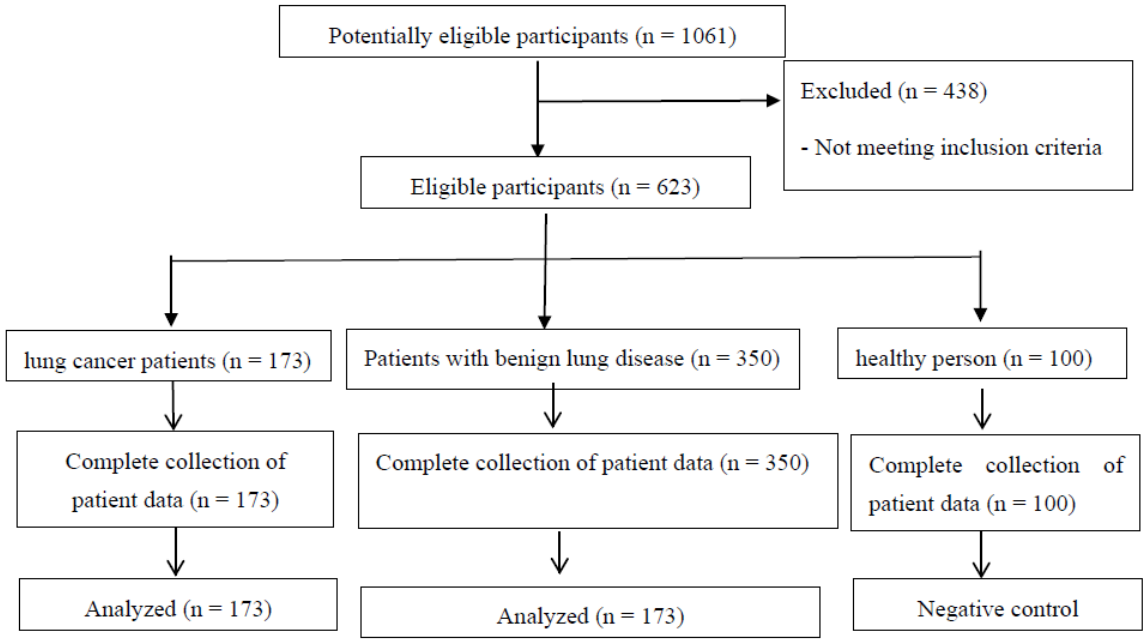


**Figure 1:** Patient flow diagram

### *2.2 Chemicals*

High-purity water and HPLC grade acetonitril (ACN) were provided by Thermo Fisher Scientific (Waltham, MA, USA). 1-Butanol and Acetyl chloride were from Sigma-Aldrich (St Louis, MO, USA), internal standards for the 12 isotopically labeled amino acids (catalog number: NSK-A) and 8 acylcarnitines (catalog number: NSK-B) were from Cambridge Isotope Laboratories (Andover, MA, USA). For quality control of the sample, the kit containing the mixed standard amino acids and acylcarnitines was obtained from Chromsystems (Grafelfng, Germany).

### *2.3 Sample Preparation*

Dried Blood Spot (DBS) paper was drilled into a blood spot disc with a diameter of 3 mm, and transferred to Millipore MultiScreen HV 96-well plate (Merck KGaA, Darmstadt, Germany) to extract amino acids, and 100 μL of working solution was added to each well. The 96-well plate was gently shaken at room temperature for 20 min. Two high levels and two low levels of QC solution were randomly added to the blank wells to ensure the stability of the analysis. The plate containing the disc was centrifuged at 1,500 x g for 2 min, the filtrate was collected and dried with pure nitrogen (50°C). The dried sample was added to 60 μL of acetyl chloride and 1-butanol mixture at 65°C for 20 min for derivatization, and then 100 μL of the mobile phase solution was added to dissolve the derivative. The samples from each patient were subjected to mass spectrometry. Briefly, samples were separated by Ekspert UltraLC coupled with Triple TOF (AB Sciex, Framingham, MA, USA). UPLC HSS T3 column (2.1 mm × 100 mm, 1.8 μm, Waters, USA) was used. The mobile phase A was 1:9 acetonitrile/water (v/v) solution and the mobile phase B was 9:1 acetonitrile/water (v/v) solution. The flow rate was 0.3 mL/min.

### *2.4 Statistical Analysis*

All data were normalized and discretized into categorical variables by dichotomy of the phenotype using the upper or lower limits of the normal range as cutoff values. Two-thirds of each comparison subgroup were used as a training set for modeling and signature construction, and the remaining third were

used as a validation set. The elastic network algorithm was used to model the equal penalty of the training set and validation set and build a composite biomarker signature. Youden Index was used to choose the best cutoff value. Receiver Operating Curve (ROC) was drawn, and the performance of the model was evaluated using the area under ROC (AUR). The area between ROC curve and the no-discrimination line was multiplied by two to calculate coefficient (COEF). All statistical analyses were performed using Python 3.6 and IBM SPSS Statistics software package v.17.0 (IBM SPSS Inc., Chicago, IL, USA).

## 3 Results

### *3.1 Differences between the two Groups in the Diagnostic Model*

Total 49 metabolites (Ala , Arg, Asn, Asp, Cit, Cys, Gln,Glu, Gly, Hcy, His, Leu, Lys, Met, Orn, Phe, Pip, Pro, Ser, Thr, Trp, Tyr, Val, C0, C2, C3, C4, C4-OH, C4DC, C5, C5-OH, C5DC, C5:1, C6, C8, C10, C12, C14, C14-OH, C14DC, C14:1, C16, C16-OH, C16:1-OH, C18, C20, C22, C24, C26) were added to each sample of LD and LC, and additional 45 routine blood markers (Leukocyte, Monocytes/Leukocytes, Monocyte, Hematocrit, Red blood cell, Red blood cell distribution width, Lymphocytes/white blood cells, Lymphocytes, Average red blood cell hemoglobin, Mean red blood cell hemoglobin concentration, Mean red blood cell volume, Mean platelet volume, Basophils/Leukocytes, Basophils, Eosinophils/White Blood Cells, Eosinophils, Hemoglobin, Platelet hematocrit, Platelets, Platelet distribution width, Nucleated red blood cells, Nucleated red blood cells/white blood cells, Neutrophils/White Blood Cells, Neutrophils, 5'-nucleotidase, gamma-glutaminase, α1-microglobulin, α-L-Fucosidase, Albumin, total protein, globulin , Cholinesterase, bile acid, Alanine aminotransferase, Aspartate aminotransferase/Alanine aminotransferase, Aspartate aminotransferase, Cystatin C, Creatinine, Alkaline phosphatase, Leucine aminopeptidase, Urea Nitrogen, Adenosine Deaminase, Conjugated bilirubin, Bilirubin) were added for measurement and calculation. All of these indicators were used in the models to distinguish LC and LD. The accuracy of the training set and the validation set was stable about 80%. Therefore, the model composed of blood index and metabolite parameters selected by the elastic network could distinguish LC and LD groups.

### *3.2 Differential Parameter Selection*

In the elastic network, all indexes of LC and LD groups compared with healthy group were screened. 17 related markers were selected from 137 detection indicators (Ala , Arg, Asn, Asp, Cit, Cys, Gln, Glu, Gly, Hcy, His, Leu, Lys, Met, Orn, Phe, Pip, Pro, Ser, Thr, Trp, Tyr, Val, C0, C2, C3, C4, C4-OH, C4DC, C5, C5-OH, C5DC, C5:1, C6, C8, C10, C12, C14, C14-OH, C14DC, C14:1, C16, C16-OH, C16:1-OH, C18, C20, C22, C24, C26,Arg/Orn, Cit/Arg, Gly/Ala, Met/Leu, Met/Phe, Orn/Cit, Phe/Tyr, Tyr/Cit, Val/Phe, C2/C0, C3/C0, C3/C2, C3/C16, C4/C2, C4/C3, C4/C8, C5/C0, C5/C2, C5/C3, C5-OH/C8, C5-OH/C0, C5DC/C5-OH, C5DC/C16, C8/C2, C8/C10, C16-OH/C16, C14:1/C16, C3DC, C3DC/C10, C18:1, C18-OH, C18:1-OH, C10:1, C10:2, C14:2, C18:2, C10:2/C10, C6DC, C5DC/C8, (0 + 2 + 3 + 16 + 18:1)/Cit, (C16 + C18)/C0, C0/(C16 + C18), C3/Met, Leukocyte, Monocytes/Leukocytes, Monocyte, Hematocrit, Red blood cell, Red blood cell distribution width, Lymphocytes/white blood cells, Lymphocytes, Average red blood cell hemoglobin, Mean red blood cell hemoglobin concentration, Mean red blood cell volume, Mean platelet volume, Basophils/Leukocytes, Basophils, Eosinophils/White Blood Cells, Eosinophils, Hemoglobin, Platelet hematocrit, Platelets, Platelet distribution width, Nucleated red blood cells, Nucleated red blood cells/white blood cells, Neutrophils/White Blood Cells, Neutrophils, 5'-nucleotidase, gamma-glutaminase, α1-microglobulin,α-L-Fucosidase, Albumin, total protein, globulin, Cholinesterase, bile acid, Alanine aminotransferase, Aspartate aminotransferase/Alanine aminotransferase, Aspartate aminotransferase, Cystatin C, Creatinine, Alkaline phosphatase, Leucine aminopeptidase, Urea Nitrogen, Adenosine Deaminase, Conjugated bilirubin, Bilirubin) to build the model (Tab. 2). These included 16 endogenous metabolites and 1 calculated ratio. Compared with CONT group, Gly/Ala, Ala, Glu, Gln, Pip, Lys, Leu, Pro, Thr, alkaline phosphatase, and uric acid levels increased in LC group, while Phe, Gly, Cit decreased in LD group, platelets, γ-glutamine transferase, and alanine aminotransferase levels

decreased in LC group, but increased in LD group. These results suggest that LD causes reversible metabolic changes compared to LC.

**Table 2:** Regression model coefficient

| Marker | COEF |
|---|---|
| Gly/Ala | 0.000285811 |
| Phe | –0.004698505 |
| Ala | 0.000413463 |
| Gly | $-7.82341 \times 10^{-5}$ |
| Glu | 0.003392312 |
| Gln | 0.000299832 |
| Cit | –0.001598781 |
| Pip | 0.000279602 |
| Lys | 0.000101959 |
| Leu | 0.000288229 |
| Pro | 0.000181724 |
| Thr | 0.000215315 |
| Platelets | $-9.61985 \times 10^{-5}$ |
| γ-glutamine | –0.000129896 |
| | –0.000802072 |
| | 0.000387214 |
| | 0.000547153 |

Gly: glycine; Ala: alanine; Phe: phenylalanine; Gln: glutamine; Glu: glutamate; Pro: Proline; Thr: Threonine; Pip: piperonylamide; Cit: Citrulline; Lys: Lysine

### *3.3 Diagnostic Regression Equation*

In the statistical model, the AUC of the training set was 0.915, the specificity was 0.702, the sensitivity was 0.965, and the data screening accuracy was 0.789. In the validation set, the AUC was 0.875, the specificity was 0.756, the sensitivity was 0.948, and the data screening accuracy was 0.791 (Figs. 2A and 2B). The AUC of our model was better than that of cancer markers CEA, CY211 and NSE (AUC of 0.717, 0.690 and 0.698, respectively, Fig. 3), indicating that our diagnostic model has advantage over conventional tumor markers.

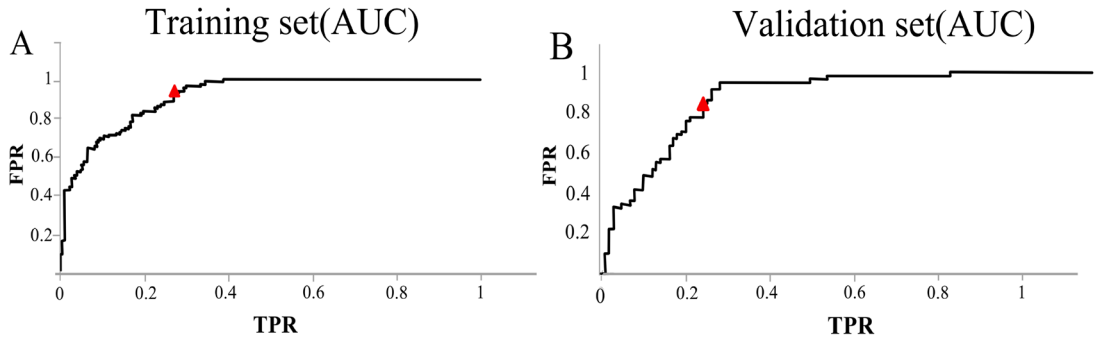


**Figure 2:** The performance of the elastic network based on 16 endogenous metabolites and 1 calculated ratio. A. AUC results of training set obtained by two-thirds patients in the model. B. AUC results of validation set obtained by two-thirds patients in the model

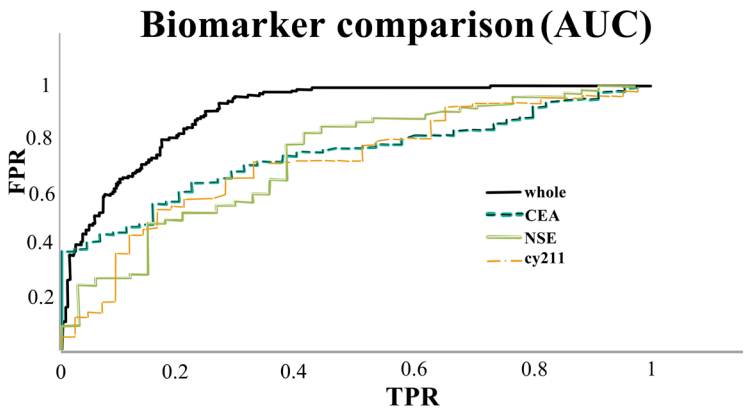


**Figure 3:** Comparison of AUC of our model with that of tumor markers. The AUC of the overall model (whole) was significantly better than that of CEA, NSE, CY211

## 4 Discussion

At present, the most difference among lung cancer and benign lung diseases is histological pathology, but it is not easy to identify benign and malignant lung diseases in the early stage. Low-dose CT is a common method of diagnosis, but the false positive rate may exceed 90% [3]. Tumor markers such as CY211, NSE and CEA are used for detecting lung cancer, but they may be detected in benign diseases, resulting in low positive rates [12–14]. The combined assessment of multiple serum tumor markers was more accurate for LC detection than the use of single serum tumor marker [15]. On the other hand, it has been proposed that metabolites can distinguish LC and LD as novel biomarkers. For example, Callejon-Leblic et al. [16] identified 11 metabolites that could be as biomarkers of LC to distinguish LC from other lung diseases. Therefore, in this study we validated the potential of metabolites in combination with multiple blood biomarkers to establish a multiple regression model for the early diagnosis of LC.

Amino acids are valuable clues in the metabolomics of cancer [17]. Amino acid metabolism is reversible in benign lung disease but is irreversible in malignant lung disease. Previous studies showed that amino acid metabolomics has obvious advantages in predicting disease in gastric cancer, breast cancer, and colorectal cancer [18–20]. The concentrations of tryptophan, glycine, citrulline, ornithine and proline reduced in LC, while the concentration of phenylalanine increased [21]. Therefore, the metabolism of amino acids is different in LD and LC, and provides more information for differential diagnosis of diseases.

Previous studies have demonstrated that changes in amino acid metabolism in cancer cells are associated with gluconeogenesis, especially in patients with advanced cancer. Most cancer cells are highly dependent on glucose [22]. Known organs associated with gluconeogenesis include tissues such as the liver, kidneys, and muscles. Moreover, these organs are closely related to specific amino acid metabolism [23,24]. In addition, platelets are key factors in cancer progression, metastasis, and cancer-related thrombosis [25]. Platelets interact with immune cells to stimulate tumor response, allowing platelets to aggregate or activate in cancer patients. The number of platelets varies between cancer patients and healthy individuals [26,27].

In this study, the AUC of the regression model was significantly better than the AUC of CEA, CY211, and NSE in serum, indicating that the quantification of differential metabolites identified by metabolomics can help distinguish between benign lung disease and malignant lung cancer. Based on metabolomics techniques combined with routine blood marker analysis, a reasonable regression model was constructed to accurately distinguish between LC and LD. However, our study has several limitations. First, our sample size is relatively small. Further studies with larger sample size are needed to confirm our conclusion. Second, metabolomics, unlike other omics techniques, could be easily affected by

environmental factors, especially for LC patients who often smoke. Therefore, better controls are needed.

**Ethics Approval and Consent to Participate:** The present study was approved by the Ethics Committees of the First Affiliated Hospital of Jinzhou Medical University. Written informed consent was obtained from all patients.

**Acknowledgement:** The present study was supported by the Liaoning Province Science and Technology Project (JYTFUDF201752).

**Funding Statement:** This work was supported by a grant from the Foundation of Liaoning Educational Committee (No. XLYC1902026).

**Conflict of Interest:** The authors declare that they have no competing interests.

## References

1. Chen, D. W., Cheng, L., Huang, F., Cheng, L., Zhu, Y. et al. (2014). Ge11-modified liposomes for non-small cell lung cancer targeting: preparation, *ex vitro* and *in vivo* evaluation. *International Journal of Nanomedicine, 9,* 921–935.
2. Simińska, E., Koba, M. (2016). Amino acid profiling as a method of discovering biomarkers for early diagnosis of cancer. *Amino Acids, 48(6),* 1339–1345.
3. Huang, T., Cao, Y., Zeng, J., Dong, J., Sun, X. et al. (2016). Tandem mass spectrometry-based newborn screening strategy could be used to facilitate rapid and sensitive lung cancer diagnosis. *Onco Targets and Therapy, 9,* 2479–2487.
4. Tang, Y., Li, Z., Lazar, L., Fang, Z., Zhao, J. (2019). Metabolomics workflow for lung cancer: discovery of biomarkers. *Clinica Chimica Acta, 495,* 436–445.
5. Colella, S. (2014). Endoscopic ultrasound in the diagnosis and staging of lung cancer. *Endoscopic Ultrasound, 3(4),* 205–212.
6. Spiro, S. G., Porter, J. C. (2002). Lung cancer—where are we today? Current advances in staging and nonsurgical treatment. *American Journal of Respiratory and Critical Care Medicine, 166(9),* 1166.
7. Kosmides, A. K., Kamisoglu, K., Calvano, S. E., Corbett, S. A., Androulakis, I. P. (2013). Metabolomic fingerprinting: challenges and opportunities. *Critical Reviews in Biomedical Engineering, 41(3),* 205–221.
8. Neagu, M., Constantin, C., Popescu, I. D., Zipeto, D., Tzanakakis, G. et al. (2019). Inflammation and metabolism in cancer cell-mitochondria key player. *Frontiers in Oncology, 9,* 348.
9. Deberardinis, R., Thompson, C. (2012). Cellular metabolism and disease: what do metabolic outliers teach us? *Cell, 148(6),* 1144.
10. Sreekumar, A., Poisson, L. M., Rajendiran, T. M., Khan, A. P., Chinnaiyan, A. M. (2009). Corrigendum: metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature, 457(7231),* 910–914.
11. Tao, L., Zhou, J., Yuan, C., Zhang, L., Zhong, L. (2019). Metabolomics identifies serum and exosomes metabolite markers of pancreatic cancer. *Metabolomics, 15(6),* 86.
12. Gao, P., Zhou, C., Zhao, L., Zhang, G., Zhang, Y. (2015). Tissue amino acid profile could be used to differentiate advanced adenoma from colorectal cancer. *Journal of Pharmaceutical and Biomedical Analysis, 118,* 349–355.
13. Okamura, K., Takayama, K., Izumi, M., Harada, T., Furuyama, K. et al. (2013). Diagnostic value of CEA and CYFRA 21-1 tumor markers in primary lung cancer. *Lung Cancer, 80(1),* 45–49.
14. Li, Y., Tian, X., Gao, L., Jiang, X., Fu, R. et al. (2019). Clinical significance of circulating tumor cells and tumor markers in the diagnosis of lung cancer. *Cancer Medicine, 8(8),* 3782–3792.
15. Molina, R., Marrades, R. M., Augé, J. M., Escudero, J. M., Viñolas, N. et al. (2016). Assessment of a combined

panel of six serum tumor markers for lung cancer. *American Journal of Respiratory and Critical Care Medicine, 193(4),* 427–437.

16. Callejon-Leblic, B., Pereira-Vega, A., Vazquez-Gandullo, E., Sanchez-Ramos, J. L., Gomez-Ariza, J. L. et al. (2018). Study of the metabolomic relationship between lung cancer and chronic obstructive pulmonary disease based on direct infusion mass spectrometry. *Biochimie, 157,* 111–122.
17. Shingyoji, M., Iizasa, T., Higashiyama, M., Imamura, F., Saruki, N. et al. (2013). The significance and robustness of a Plasma Free Amino Acid (PFAA) profile-based multiplex function for detecting lung cancer. *BMC Cancer, 13,* 77.
18. Jing, Y., Wu, X., Gao, P., Fang, Z., Wu, J. et al. (2017). Rapid differentiating colorectal cancer and colorectal polyp using dried blood spot mass spectrometry metabolomic approach. *IUBMB Life, 69(5),* 347–354.
19. Hu, L., Gao, Y., Cao, Y. F., Zhang, Y. X., Xu, M. H. et al. (2016). Identification of arginine and its "downstream" molecules as potential markers of breast cancer. *IUBMB Life, 68(10),* 817–822.
20. Jing, F., Hu, X., Cao, Y., Xu, M., Wang, Y. et al. (2018). Discriminating gastric cancer and gastric ulcer using human plasma amino acid metabolic profile. *IUBMB Life, 70(6),* 553–562.
21. Klupczynska, A., Derezinski, P., Dyszkiewicz, W., Pawlak, K., Kasprzyk, M. et al. (2016). Evaluation of serum amino acid profiles' utility in non-small cell lung cancer detection in polish population. *Lung Cancer, 100,* 71–76.
22. Yu, J., Zhao, J., Zhang, M., Guo, J., Liu, X. et al. (2019). Metabolomics studies in Gastrointestinal cancer: a systematic review. *Expert Review of Gastroenterology & Hepatology, 14,* 9–25.
23. Karaca, M., Martin-Levilain, J., Grimaldi, M., Li, L., Dizin, E. et al. (2018). Liver glutamate dehydrogenase controls whole-body energy partitioning through amino acid-derived gluconeogenesis and ammonia homeostasis. *Diabetes, 67(10),* 1949–1961.
24. Cahill, G. F., Aoki, T. T. (1975). Renal gluconeogenesis and amino acid metabolism in man. *Medical Clinics of North America, 59(3),* 751–761.
25. Larsen, J. B., Hojbjerg, J. A., Hvas, A. M. (2020) The role of platelets in cancer-related bleeding risk: a systematic review. *Seminars in Thrombosis and Hemostasis, 46(3),* 328–341.
26. Meikle, C. K. S., Kelly, C. A., Priyanka, G., Wuescher, L. M., Ali, R. A. et al. (2017). Cancer and thrombosis: the platelet perspective. *Frontiers in Cell and Developmental Biology, 4,* 147.
27. Honn, K. V., Tang, D. G., Crissman, J. D. (1992). Platelets and cancer metastasis: a causal relationship? *Cancer and Metastasis Reviews, 11(3–4),* 325–351.