Indian Journal of Biochemistry & Biophysics
Vol. 56, February 2019, pp. 53-56

Applications of serum amino acid levels in identification of cancer

Yankun Li[1, 2]*, Xinpeng Ma[1, 2], Kenan Huang[3] & Zhichao Bai[3]

[1]MOE Key Laboratory of Resources and Environmental Systems Optimization, Beijing -102 206, China
[2]School of Environmental Science and Engineering, North China Electric Power University, Baoding -071 003, China
[3]Oncology, 252 Hospital of Chinese People's Liberation Army, Baoding -071 000, China

Received 06 February 2016; *revised 06 November 2018*

Chemometric method of uncorrelated linear discriminant analysis (ULDA) was applied to the data of amino acid levels in serum of lung cancer patients and healthy people, eventually successfully classifies the samples of cancer patients and healthy people. Simultaneously, several potential amino acid biomarkers were possibly chosen. So the method of amino acid levels combined with ULDA algorithm could be applied to the identification of cancer and exploration of tumor biomarker, which has certain practical value and application prospects.

Keywords: Amino acid, Cancer identification, Uncorrelated linear discriminant analysis (ULDA)

Human serum contains rich information that can provide important clues for cancer diagnosis and treatment. Compared with the tissue and cell, a serum sample is simple, fast and easy to be extracted and detected, and it brings small hurt to patients. But at the present, clinically early cancer screening method based on traditional protein tumor marker (TM) of serum[1-5] including α-fetoprotein (AFP), carcinoembryonic antigen (CEA), carbohydrate antigen (CA) and neuron-specific enolase (NSE) *etc*., has many problems such as low sensitivity, low accuracy and poor specificity.

To solve the above problems, new highly specific tumor biomarkers of serum are always looked for[6]. Amino acid metabolism is the basis of life activities, and amino acid level changes in the human body are closely associated with many diseases. The bodies of malignant tumor patients mostly lie at a high metabolic state, and protein synthesis and catabolism in the body are synchronously increased. As the raw material of protein synthesis and product of protein catabolism, composition, and level of amino acid can reflect the metabolic state of the patients. Tumor tissues of rapid growth and unlimited proliferation cells need to absorb and consume large amounts of amino acids, which cause amino acid metabolism defects. The high demand of amino acids by malignant tumor is known as the host of "nitrogen trap"[7,8].

*Correspondence:
Phone: +86-312-7525528
E-mail: 309267061@qq.com

Through literature researches, it has been found that few papers of amino acids levels in the human body combined with chemometric methods are applied in cancer screening. The related research work is expected to carry out systematically. In this paper, uncorrelated linear discriminant analysis (ULDA) was used to find the best classification of subspace and characteristic variables and applied to 17 kinds of amino acid content data in the serum of cancer patients and healthy people.

Feature extraction and dimension reduction are very important in data processing. Many methods have been proposed for feature reduction[9-13], such as principal component analysis (PCA), partial least squares-discriminate analysis (PLS-DA), *k*-means cluster analysis, and linear discriminant analysis (LDA) *etc*. Uncorrelated linear discriminant analysis[14-19] was first put forward by Jin and others in the field of face recognition. Now it has been successfully applied in the data of metabolomics, proteomics and gene expression profile. ULDA algorithm based on linear discriminant analysis aims to find discriminant directions by maximizing the distance between two target classes and simultaneously minimizing the distance between the Universum and the mean of the target classes.

The results of this article showed that the cancer patients and healthy people were identified completely by ULDA modeling to 17 kinds of amino acid content data. And several characteristic amino acid biomarkers were chosen by compared with a significant difference

through the mean equation of *t*-test. It indicates that ULDA is an effective feature extraction and dimension reduction method, and applied to the analysis of amino acids levels in serum, possibly provides a new cancer identification method and possibly gives new potential tumor biomarkers.

Materials and Methods

Materials

Chinese People's Liberation Army 252th Hospital provided serum of 31 healthy people and 50 lung cancer patients (age range from 22 to 83). The 17 kinds of free amino acids in serum were detected by high performance liquid chromatography (HPLC) instrument including aspartic acid (ASP), glutamate (GLU), histidine (HIS), serine (SER), arginine (ARG), glycine (GLY), threonine (THR), taurine (TAU), proline (PRO), alanine (ALA), valine (VAL), methionine (MET), cystine (CYS), isoleucine (ISO), leucine (LEU), phenylalanine (PHE), lysine (LYS).

Agilent 1200 high performance liquid chromatography (American Agilent company); Kromat Universal C18 column (The United States); Ultrasonic broken instrument (Branson S-450D, America); The thermostatic water bath pot (Beijing changfeng instrument co., Ltd); pH meter (Mettle Toledo instrument co., Ltd., Shanghai; LE438 pH electrode); 17 kinds of amino acid standard substances (Ltd. Wako, Japan).

Amino acids were determined by HPLC with 2,4-dinitrochlorobenzene derivatization. Gradient elution was adopted with mobile phase A of acetonitrile and mobile phase B of acetic acid-sodium acetate buffer (pH 5.25). The flow rate was 1 mL/min and the injection volume was 10 μL. The column temperature was 40°C and the detection wavelength was 360 nm.

Chemometric method

ULDA aims to find an optimal transformation matrix G. The data X in high dimensional space is projected into low dimensional space, as well as maximize the Fisher discriminant equation with constraint condition. ULDA algorithm considers uncorrelated between column vectors on the transformation matrix based on LDA, so it can reduce the data redundancy after dimension reduction. The uncorrelated discriminant vector (UDV) in dimension reduction space is a linear combination of the variables in an original high dimensional space, and the coefficient of combination depends on the transformation matrix G. Then the new low dimensional matrix Z can be calculated by $Z = XG$[20].

Statistical analysis

The method of *t*-test (also called Student's *t*-test in statistics) is also used in this study. A *t*-test is any statistical hypothesis test in which the test statistic follows a Student's *t*-distribution if the null hypothesis is supported. It is used to determine the probability (*p*) that two populations are the same in respect to the variable that you are testing, that is if two sets of data are significantly different from each other[21,22].

Matlab 7.0 and SPSS 19.0 were used as the calculation software.

Results and Discussion

The results of the ULDA method

Firstly, the sample data were divided into two parts. 25 samples of cancer patients and 15 samples of healthy people were arbitrarily chosen to modeling, and the other 25 samples of cancer patients and 16 samples of healthy people were used to prediction. Because a number of categories is two, only one uncorrelated discriminant vector was obtained, and transformation matrixes G was also one line. So data visualization in two-dimensional spatial was achieved, consequently, the complexity of the model was decreased and explain the ability of the model was enhanced. The uncorrelated discriminant vectors of samples obtained are shown in (Fig. 1).

In the data of amino acids levels, the 30th sample of healthy people was a bad sample, which had become red and cloudy obviously due to improper storage. Also, amino acids levels of this sample are mostly

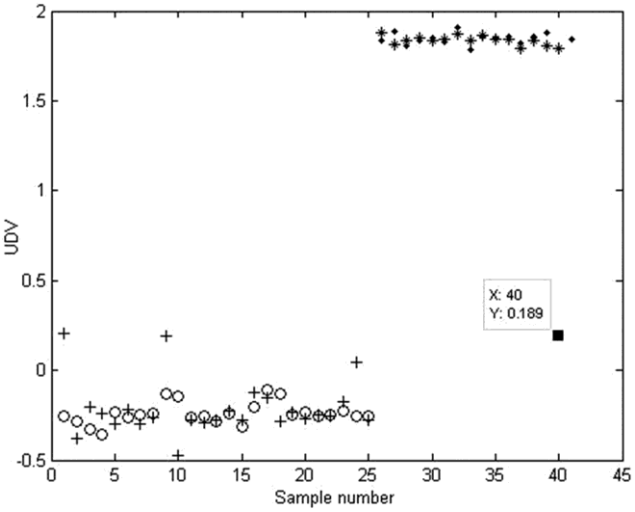

Fig. 1 — Uncorrelated discriminant vectors of ULDA (" + ": modeling samples of cancer patients; " * ": modeling samples of healthy people; " o ": prediction samples of cancer patients; " · ": prediction samples of healthy people)

much lower than those of other normal samples. In (Fig. 1), a sample deviation is clearly found, which is just the 30th sample of healthy people. So it is seems that ULDA can be used as a good method of judging for the singular sample.

In the rest 80 samples data after the bad sample was deleted, in the same way, 25 samples of cancer patients and 15 samples of healthy people were arbitrarily chosen to modeling. The other 25 samples of cancer patients and 16 samples of healthy people were used to prediction. The analysis results are shown in (Fig. 2).

It can be clearly seen from (Fig. 2), the UDVs got by ULDA can completely distinguish cancer patients and healthy people. The classification accuracy rate reaches 100%, which means that the method of ULDA combined with amino acids levels could be applied in cancer detection.

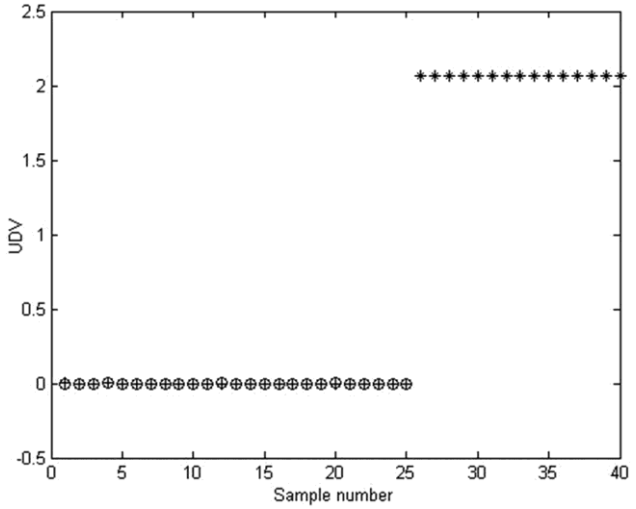

Fig. 2 — Uncorrelated discriminant vectors of ULDA without the bad sample (“ + ”: modeling samples of cancer patients; “ * ”: modeling samples of healthy people; “ o ”: prediction samples of cancer patients; “ · ”: prediction samples of healthy people)

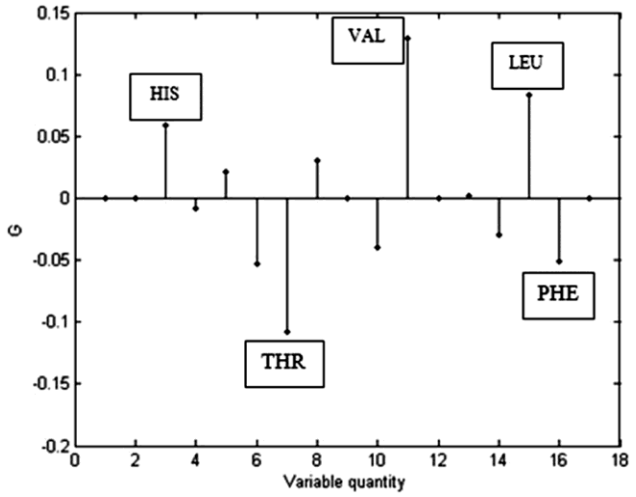

Fig. 3 — Transformation matrix of ULDA

In (Fig. 3), transformation matrixes G of 17 variables were presented, in which biomarkers could be clearly and precisely selected. Several variables with high absolute values were chosen and marked in the figure, which is successively valine, threonine, leucine, phenylalanine, and histidine.

The results of statistical processing

Then, statistical processing was adopted to analyze the data. The determination results of amino acids were expressed by “mean ± standard deviation”, and *t*-test was used to find amino acid with a significant difference between the data of cancer patients and healthy people. The corresponding analysis results are given in (Table 1).

According to (Table 1), firstly, it can be seen that the mean levels of 5 kinds of free amino acids in serum including GLU, SER, ARG, MET and PHE in the cancer group are higher than those of normal group. Levels of other amino acids in cancer group are mostly lower than those of normal group. The results proved "nitrogen trap" phenomenon of the high demand of amino acids by a malignant tumor.

Secondly, levels of SER, THR, TAU, ALA and CYS in the cancer group compared with those of normal group, there was no statistically significant

Table 1—Serum amino acid levels of cancer patients compared with normal group (mg/L, $\bar{X} \pm s$)

Amino acid	Normal group (n=30)	Cancer group (n=50)	*t*
ASP	14.59±0.21	12.58±0.42	4.256 (*P* = 0.001)
GLU	13.46±0.28	15.39±0.43	-3.727 (*P* =0 .002)
HIS	10.55±0.15	7.18±0.21	12.794 (*P* = 0.000)
SER	19.10±0.46	21.10±0.90	-1.974 (*P* = 0.052)
ARG	8.23±0.25	11.86±0.39	-7.827 (*P* = 0.000)
GLY	19.81±0.35	18.32±0.52	2.370 (*P* = 0.020)
THR	11.06±0.27	11.54±0.42	-0.979 (*P* = 0.331)
TAU	10.13±0.22	9.78±0.38	0.794 (*P* = 0.430)
PRO	13.87±0.24	11.81±0.37	4.657 (*P* = 0.001)
ALA	27.86±0.52	27.60±0.56	0.347 (*P* = 0.729)
VAL	26.74±0.31	15.64±0.42	21.061 (*P* = 0.000)
MET	0.48±0.09	1.17±0.23	-2.804 (*P* =0.007)
CYS	0.01±0.002	0.01±0.003	-0.044 (*P* = 0.965)
ISO	5.09±0.11	4.22±0.22	3.493 (*P* = 0.001)
LEU	16.91±0.24	11.46±0.28	14.672(*P* = 0.000)
PHE	4.58±0.12	6.18±0.15	-8.258(*P* = 0.000)
LYS	0.84±0.06	3.92±0.73	-4.207(*P* = 0.002)

difference (P >0.05). Other amino acids levels all show significant differences between cancer patients and normal group, especially amino acids of HIS, ARG, VAL, LEU and PHE (P = 0.000). Comparatively, VAL, THR, LEU, PHE and HIS were chosen by ULDA. So the results of the *t*-test method verify that ULDA could be used to select potential tumor biomarkers distinguishing cancer patients and healthy people.

Conclusion

Analysis of free amino acid level in serum combined with the chemometric method of ULDA applied to the classification of cancer patients and healthy people, and as well as an exploration of potential tumor biomarker. The method has certain practical significance and application prospects.

Acknowledgment

This study is supported by the National Natural Science Foundation of China (No. 21305043), Beijing Natural Science Foundation (No. 7142102).

References

1 Nicholson JK, Holmes E, Kinross JM, Darzi AW, Takats Z & Lindon JC, Metabolic phenotyping in clinical and surgical environments. *Nature*, 491 (2012) 384.

2 Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J, Liu Q, Iorio F, Surdez D, Chen L, Milano RJ, Bignell GR, Tam AT, Davies H, Stevenson JA, Barthorpe S, Lutz SR, Kogera F, Lawrence K, McLaren-Douglas A, Mitropoulos X, Mironenko T, Thi H, Richardson L, Zhou W, Jewitt F, Zhang T, O'Brien P, Boisvert JL, Price S, Hur W, Yang W, Deng X, Butler A, Choi HG, Chang JW, Baselga J, Stamenkovic I, Engelman JA, Sharma SV, Delattre O, Saez-Rodriguez J, Gray NS, Settleman J, Futreal PA, Haber DA, Stratton MR, Ramaswamy S, McDermott U & Benes CH, Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, 483 (2012) 570.

3 Chaubard JL, Krishnamurthy C, Yi W, Smith DF & Hsieh-Wilson LC, Development of orally active inhibitors of protein and cellular fucosylation. *J Am Chem Soc*, 134 (2012) 4489.

4 Krishnan S, Mani V, Wasalathanthri D, Kumar CV & Rusling JF, Attomolar Detection of a Cancer Biomarker Protein in Serum by Surface Plasmon Resonance Using Superparamagnetic Particle Labels. *Angew Chem Int Edit*, 50 (2011) 1175.

5 Taguchi A & Hanash MS, Unleashing the power of proteomics to develop blood-based cancer markers. *Clin Chem*, 59 (2013) 119.

6 Indumathi B, Katkam SK & Kutala VK, A review on application of biomarkers in heart failure. *Indian J Biochem Biophys*, 55 (2018) 303.

7 Kuroda H, Ushio A, Miyamoto Y, Sawara K, Oikawa K, Kasai K, Endo R, Takikawa Y, Kato A & Suzuki K, Effects of branched-chain amino acid-enriched nutrient for patients with hepatocellular carcinoma following radiofrequency ablation: a one-year prospective trial. *J Gastroenterol Hepatol*, 25 (2010) 1550.

8 Wie GA, Cho YA, Kim SY, Kim SM, Bae JM & Joung H, Prevalence and risk factors of malnutrition among cancer patients according to tumor location and stage in the National Cancer Center in Korea. *Nutrition*, 26 (2010) 263.

9 Yang QX, Zhang LX, Wang LX & Xiao HB, MultiDA: Chemometric software for multivariate data analysis based on Matlab. *Chemom Intell Lab Syst*, 116 (2012) 1.

10 Chen XH, Chen SC & Xue H, Universum linear discriminant analysis. *Electron Lett*, 48 (2012) 1407.

11 Duraipandian S, Zheng W, Ng J, Low Jeffrey JH, IIancheran A & Huang ZW, Elucidation of Molecular Changes in Cervical Epithelial Cells by Raman Microspectroscopy. *Analyst*, 136 (2011) 4328.

12 Li SX, Zeng QY, Li LF, Zhang YJ, Wan MM, Liu ZM, Xiong HL, Guo ZY & Liu SH, Study of support vector machine and serum surface-enhanced Raman spectroscopy for noninvasive esophageal cancer detection. *J Biomed Opt*, 18 (2013) 027008.

13 Barman I, Dingari NC, Singh GP, Kumar R, Lang S & Nabi G, Selective sampling using confocal Raman spectroscopy provides enhanced specificity for rinary bladder cancer diagnosis. *Ana Bioana Chem*, 404 (2012) 3091.

14 Jin Z, Yang JY, Hu ZS & Lou Z, Face recognition based on the uncorrelated discriminant transformation. *Pattern Recogn*, 34 (2001) 1405.

15 Liang YZ & Xu QS, *Instrumental analysis of complex systems: white, gray and black analytical systems and their multivariate methods.* (Chemical industry Press, Beijing) 2012, 525.

16 Chu DL, Goh ST & Hung YS, Characterization of all solutions for under sampled uncorrelated linear discriminant analysis problems. *SIAM J Matrix Anal Appl*, 32 (2011) 820.

17 Zhang MJ, Tong PJ, Wang WM, Geng JP & Du YP, Proteomic profile analysis and biomarker discovery from mass spectra using independent component analysis combined with uncorrelated linear discriminant analysis. *Chemom Intell Lab Syst*, 105 (2011) 207.

18 Xu WJ, Zhang LX, Huang YH, Yang QX, Xiao HB & Zhang DQ, Discrimination of type 2 diabetes mellitus corresponding to different traditional Chinese medicine syndromes based on plasma fatty acid profiles and chemometric methods. *J Ethnopharmacol*, 143 (2012) 463.

19 Ye JP, Characterization of a family of algorithms for generalized discriminant analysis on under sampled problems. *J Mach Learn Res*, 6 (2005) 483.

20 Li YK & Zeng XC, Serum SELDI-TOF MS analysis model applied to benign and malignant ovarian tumor identification. *Anal Methods*, 8 (2016) 183.

21 Sawilowsky SS, Misconceptions leading to choosing the *t*-test over the Wilcoxon Mann-Whitney Test for Shift in Location Parameter. *J Mod Appl Statis Methods*, 4 (2005) 598.

22 Richard M, *The Story of Mathematics*, (Princeton University Press, Princeton) 2004, 158