Molecular weight statistical model of blood sugar in pancreatic cancer

Fan Shihao

Abstract —**The molecular weight of sugar in blood is an important index to recognize pancreatic cancer. The molecular weight data of sugar in the blood of pancreatic cancer were collected, the statistical mathematical model was established, the mean variance cancer recognition method was constructed, and the mean and mean variance of molecular weight were calculated. The calculation results are given in chart form. The analysis of the results showed that the variance of patient and health data of mean and mean variance was completely separated, and the recognition rate of pancreatic cancer was as high as 86%.**

Index Terms—**Pancreatic cancer, blood, sugar, statistics.**

I. INTRODUCTION

Cancer leads to death, cancer patients often have abnormal glycosylation, one of which is abnormal N-glycan. Studying the abnormal expression of N-glycan is an important way to identify cancer. In the serum of cancer patients, N-glycan has significant abnormal expression [1]. Therefore, in recent years, N-glycan in serum is a hot spot in cancer research. Rudd et al. Pointed out that N-glycan is an important cancer molecular marker, especially suitable for early diagnosis, stage recognition and prognosis monitoring of cancer [2]. A large number of studies have shown that in the serum of cancer patients, the expression of N-glycan is significantly abnormal [3]. For example, in the study of breast cancer and lung cancer, abnormal sialylation and fucosylation were found [4] [5]; in the study of liver cancer, the change of glycosyltransferase can cause the change of N-glycan content [6]. Similar phenomena have been found in prostate cancer, colorectal cancer and other cancers [7] [8]. However, while the abnormal expression of N-glycan caused by cancer has been widely concerned, the relationship between N-glycan and cancer stage is rarely reported.

The survival rate and poor treatment rate of pancreatic cancer are the most challenging problems in biomedical and clinical fields. How to improve the long-term survival rate and prognosis of patients with pancreatic cancer has become an urgent problem[9].

In this paper, we will study the molecular weight of sugar in the blood of pancreatic cancer, and construct the mean variance pancreatic cancer recognition method with statistical theory.

II. AVERAGE STATISTICS

Let u be the molecular weight of the sugar in the blood of pancreatic cancer. The value is u_j, j = 1,2,..., J. the molecular

Fan Shihao,Shanghai Xuhui high school, Shanghai 200030, China.

weight measurement data of the k th pancreatic cancer patient is u_{jk}, k = 1,2,... K. The sum of the measurement data of each patient is

$$S_k = \sum_{j=1}^{J} u_{jk} \qquad k = 1, 2, ..., K$$

Divide the measured data by this number, and then express it in percentage

$$x_{kj} = \frac{u_{jk}}{S_k} * 100 \qquad k = 1, 2, ..., K, j = 1, 2, ..J$$

The average of the j-th value is

$$Exj = \frac{1}{K} \sum_{k=1}^{K} x_{jk} \qquad j = 1, 2, ..., J \tag{1}$$

If the patient measurement data is set as one row and the k-th patient measurement data is set as the k-th row, then the k-th patient data will form the k-th row j-column table. The above formula is the average value of the data in column J of the table.

In the experiment, the molecular weight of glucose in the blood of 27 patients with pancreatic cancer was measured. The total molecular weight was 47 and 1269 data were obtained. Figure 1 shows part of the calculated data of the mean molecular weight of patients with pancreatic cancer. According to the calculation data, when the molecular weight is 2792
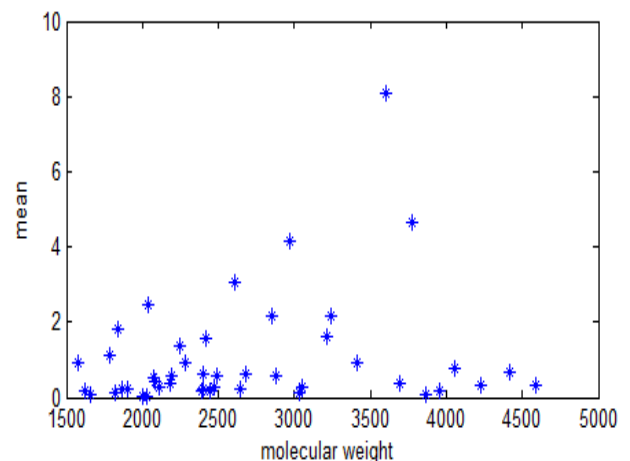

Fig.1.Mean varied with molecular weight

III. VARIANCE STATISTICS

The calculation formula of mean square deviation is:

$$Dxj = \sqrt{\frac{1}{K} \sum_{k=1}^{K} (x_{jk} - Exj)^2} \qquad j = 1, 2, ..., J$$

(2)

Figure 2 shows the calculation data of molecular weight mean square deviation of pancreatic cancer patients. It can be seen from the figure that When the molecular weight is 1998.9, the minimum corresponding mean square deviation is 0.033291.
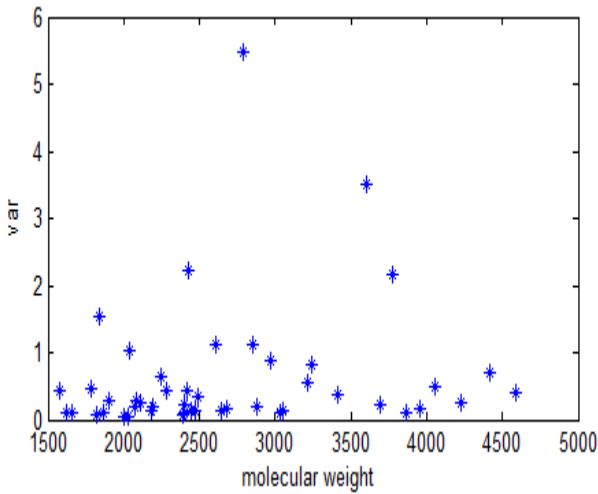

Fig.2, Mean square deviation

IV. VARIANCE OF MEAN

Let x(j,k)= X_{jk} Is the j-th molecular weight measurement data of the k-th individual, Let transformation:

$$y(j,k) = \frac{x(j,k)}{Ex_j} - 1 \qquad k = 1,2,..,K; j = 1,2,...J$$

here

$$Exj = \frac{1}{K} \sum_{k=1}^{K} x_{jk} \qquad j = 1,2,...,J$$

The mean of y(j,k) is

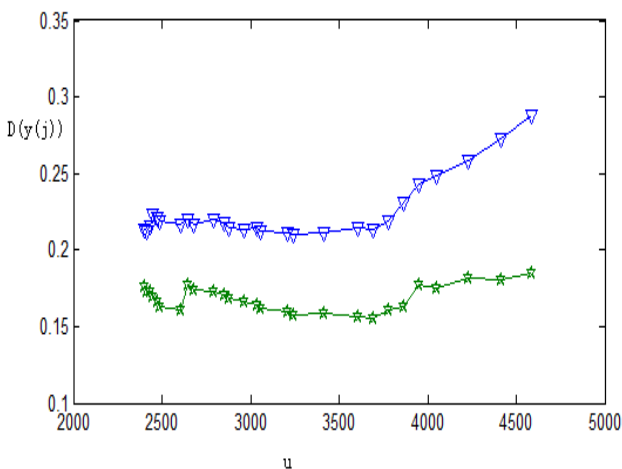$$z(j) = Ey(j) = \frac{1}{K} \sum_{k=1}^{K} y(k,j) \qquad j = 1,2,...,J$$

(3)

Fig.3, Mean square deviation of patients and healthy people, patient triangle. Pentagram of healthy people

Then calculate the mean square deviation of {Z (1), Z (2),..., Z (P)} of the first P data, and record it as Dy (P), P = 2,3,..., J

According to the measurement data, according to the measurement data of patients and healthy people, the variance is calculated, and the results are shown in Figure 3. The patient triangle in the figure. Healthy person pentagram. It can be seen from the figure that the mean square deviation of patients and healthy people is completely separated.

V. PANCREATIC CANCER RECOGNITION

The above steps of calculating mean square deviation are：

1) To collect the molecular weight data of glucose in the blood of pancreatic cancer ujk, j=1,2,...,Jk=1,2,..K.

2) Calculate the sum of each person's measurements

$$S_k = \sum_{k=1}^{K} u_{jk} \qquad k = 1,2,...,K$$

3）Measured data in percentage

$$x_{kj} = \frac{u_{jk}}{S_k} * 100 \qquad k = 1,2,...,K, j = 1,2,..J$$

4）Calculate the j-th molecular weight average：

$$Exj = \frac{1}{K} \sum_{k=1}^{K} x_{jk} \qquad j = 1,2,...,J$$

5）Let x(j,k)= xjk is the measurement data of the j-th molecular weight of the k-th person, Let transformed

$$y(j,k) = \frac{x(j,k)}{Ex_j} - 1 \qquad k = 1,2,..,K; j = 1,2,...J$$

6) The mean of y(j,k) are

$$z(j) = Ey(j) = \frac{1}{K} \sum_{k=1}^{K} y(k,j) \qquad j = 1,2,...,J$$

7) Then, the mean square deviation of {Z (1), Z (2),..., Z (P)} of the first P data is calculated, which is recorded as Dy (P).

Using the above mean square deviation calculation results, the following gives the pancreatic cancer recognition method. Set up the molecular weight data set of sugar in the blood of known pancreatic cancer patients U={ujk, j=1,2,...,J, k=1,2,..K}, Molecular weight data set of sugar in blood of healthy people V={vjk, j=1,2,...,n, k=1,2,..m}, Molecular weight data set of sugar in the blood of cancer patients to be examined is A={vjk, j=1,2,...,n, k=1}. The steps of recognition of pancreatic cancer are:

1) The mean square deviation of known pancreatic cancer patients was calculated by using the data in set U, which was recorded as Dy (P,b). The mean square deviation of known healthy people is calculated by the data in set V, which is recorded as Dy (P,a);

2) Consider to be check data in set A, Add the data of A to U and V, and calculate the mean square deviation with the data in the set (U + A), which is recorded as Dy (P,b '). Calculate the mean square deviation with the data in the set (V + A), and record it as Dy (P,a ');

3) Calculate the difference value|dy (P,b) - Dy (P,b') |, and the difference value|dy (P,a) - Dy (P,a') |, where|x|is the

absolute value of x.

4) If $\sum |Dy(P,b) - Dy(P,b')| < \sum |Dy(P,a) - Dy(P,a')|$, the subject is a cancer patient, otherwise it is a healthy person. Where $\sum$ is the sum of all data.

VI. CALCULATION RESULTS AND DISCUSSION

The above mean variance method is used to calculate the statistical value of 51 people. The results are listed in table 1. According to the data in the table, the correct probability of identifying cancer patients is (26-4) / 26 = 84.6%, and the correct probability of identifying healthy people is (25-3) / 25 = 88%. The overall correct rate is (51-7) / 51 = 86%. The highest correct probability of traditional classification method is 43% [10].

Table 1,Calculation results

class	Accuracy rate	Healthy sample	Pancreatic cancer samples
Healthy sample	88%	22	3
Pancreatic cancer samples	84.60%	4	22
total	86.30%	26	25

VII. CONCLUSION

This paper studies the recognition of pancreatic cancer. After analyzing the measurement data, we found that the measurement data of pancreatic cancer patients fluctuated greatly, so we constructed the mean variance recognition method. This method first calculates the mean value of the measure data, and then calculates the mean square deviation of the mean value. According to the mean variance method, the calculation shows that the mean variance of pancreatic cancer patients is 100% separated from the mean variance of healthy people. The results show that the recognition rate of pancreatic cancer is as high as 86%.

ACKNOWLEDGMENT

REFERENCES

[1] Adamczyk B, Tharmalingam T, Rudd P M. Glycans as cancer biomarkers[J]. Biochim Biophys Acta. 2012, 1820(9): 1347-1353.

[2] Imre T, Kremmer T, Heberger K, et al. Mass spectrometric and linear discriminant analysis of N- glycans of human serum alpha-1-acid glycoprotein in cancer patients and healthy individuals[J]. J Proteomics. 2008, 71(2): 186-197.

[3] Kudelka M R, Ju T, Heimburg-Molinaro J, et al. Simple sugars to complex disease--mucin-type O- glycans in cancer[J]. Adv Cancer Res. 2015, 126: 53-135.

[4] Kyselova Z, Mechref Y, Kang P, et al. Breast cancer diagnosis and prognosis through quantitative measurements of serum glycan profiles[J]. Clin Chem. 2008, 54(7): 1166-1175.

[5] Nakano M, Nakagawa T, Ito T, et al. Site-specific analysis of N-glycans on haptoglobin in sera of patients with pancreatic cancer: a novel approach for the development of tumor markers[J]. Int J Cancer. 2008, 122(10): 2301-2309.

[6] Tsai T H, Wang M, Di Poto C, et al. LC-MS profiling of N-Glycans derived from human serum samples for biomarker discovery in hepatocellular carcinoma[J]. J Proteome Res. 2014, 13(11): 4859- 4868.

[7] Basu P S, Majhi R, Batabyal S K. Lectin and serum-PSA interaction as a screening test for prostate cancer[J]. CLINICAL BIOCHEMISTRY. 2003, 36(5): 373-376.

[8] Liu S, Cheng L, Fu Y, et al. Characterization of IgG N-glycome profile in colorectal cancer progression by MALDI-TOF-MS[J]. J Proteomics. 2018, 181: 225-237.

[9] Wang Junpu, Zhang Hui, Yan Jun, Zhou WENCE, research progress in the treatment of pancreatic cancer [J], progress in clinical medicine, 2019(05),638-644.

[10] Miyamoto S, Taylor S L, Barupal D K, et al. Systemic Metabolomic Changes in Blood Samples of Lung Cancer Patients Identified by Gas Chromatography Time-of-Flight Mass Spectrometry[J]. Metabolites. 2015, 5(2): 192-210.