

# A Comprehensive Data Mining Method of Constructing the Interaction Network of the MicroRNA Differential Expression

Rong Wang, Xiaoi Chen, Yafen Chen, Yiwei Wang, Ping Zhou<sup>a\*</sup>, Guoming Yuan<sup>b\*</sup>

*Institute of Biomedical Engineering, Capital Medical University, Beijing 100069, China*

*E-mail: <sup>a</sup>15652523353@163.com, <sup>b</sup>2454358988@qq.com*

**Keywords:** data mining method; interaction network; microRNA; differential expression

**Abstract:** Objective To analyze the expression of microRNA in Chinese and Italy women by a series of data mining methods to rendering microRNA network. Methods Using T test, Pearson correlation method to deal with the data, Using Cytoscape3.2.1 to draw the interaction network, and do the analysis. Results After the methods, we have found the different microRNAs in expression of breast cancer, and it was found that there were differences in the pathogenesis of breast cancer between the two countries in Italy and China. Conclusion The data mining method designed by this research can be successfully found in the gene expression profile of microRNA in the two countries, the mechanism of breast cancer between different races has been studied.

## Introduction

Breast cancer is the most common cancer and the leading cause of death among women both in China and Italy [1, 2]. The mechanisms of breast cancer are different in different ethnic groups. To know more about the differences between ethnics, researchers need to find out the interaction between microRNAs. MicroRNAs plays an important role in cell growth, differentiation, proliferation and apoptosis in various organisms, which indicates their functionality in carcinogenesis as tumour suppressor genes or oncogenes [3, 4]. Studying the interaction between microRNAs in breast cancer patients in both two countries helps the clinical development of breast cancer in different races [5, 6].

We have made a microRNAs data analysis about breast cancer patients, and compared the differences in microRNAs interaction networks between China and Italy.

## Materials and methods

### Materials

#### SeriesGSE59594

Experiment design: Gene and microRNA profiles are collected from Shanghai (China) and Milan (Italy) in breast cancer patients.

**ChIP type: Affymetrix Human Genome U133AArray.**

Software: Access 2010, SPSS 17.0 and Cytoscape 3.2.1

### Flow chart

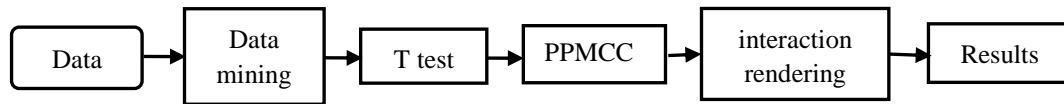


Fig. 1 The flow chart of statistical analysis

### Paired sample t text

The paired sample t test testing whether the difference between the average number of the two samples and the total of their respective representatives is significant. It is divided into two cases. One is the significance test of the difference of the average number of relevant samples. The two sets of data obtained from the test match or the data obtained by the same group are tested in different conditions. Another is the significant test of the average number of independent samples, which no correlation between two groups.

We use the paired sample t test to find the significant difference of breast cancer patients between China and Italy, and separate data out when the value of P is less than 0.05.

T test formula for the relevant sample is:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_{x_1}^2 + \sigma_{x_2}^2 - 2\gamma\sigma_{x_1}\sigma_{x_2}}{n-1}}} \quad (1)$$

$\bar{X}_1$  and  $\bar{X}_2$  are averages of two samples;

$\sigma_{x_1}$  and  $\sigma_{x_2}$  are variance of two samples;

$\gamma$  is the correlation coefficient of two samples.

### Pearson product-moment correlation coefficient

Pearson product-moment correlation coefficient (PPMCC) is used to measure the correlation between the two variables X and Y (linear correlation), its value is between -1 and 1. We can find the correlation between microRNAs and breast cancer by calculating PPMCC, so as to further compare the differences between the two countries [7].

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X\sigma_Y} \quad (2)$$

The above equation defines the overall correlation coefficient, usually expressed as the Greek letter  $\rho$ . Based on the sample covariance and standard deviation estimates, the sample correlation coefficient can be obtained, which is generally expressed as r:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (3)$$

An equivalent expression is expressed as a mean of a standard. Based on (Xi, Yi) of the sample score, the sample Pearson coefficient is

$$r = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{\sigma_X} \right) \left( \frac{Y_i - \bar{Y}}{\sigma_Y} \right) \quad (4)$$

$\frac{X_i - \bar{X}}{\sigma_X}$ ,  $\bar{X}$ , and  $\sigma_X$  are standard score, average and standard deviation, respectively.

We consider the values of r in ranges  $-1 \leq r \leq -0.5$  and  $0.5 \leq r \leq 1$  of the data taken into account.

## Results

After the treatment above, we can find the difference in the effects of different microRNAs on Chinese people and the Italians in the breast cancer [8,9].

### Interaction of microRNAs

The interactions of microRNAs are different between China and Italy (Fig. 2,3).

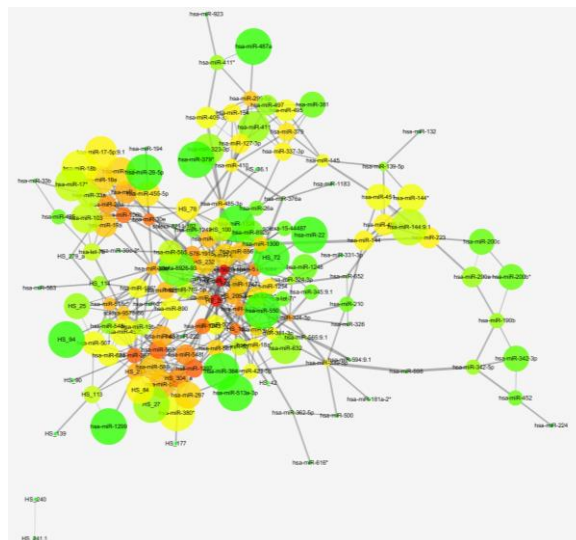


Fig. 2 interaction of Chinese people of breast cancer

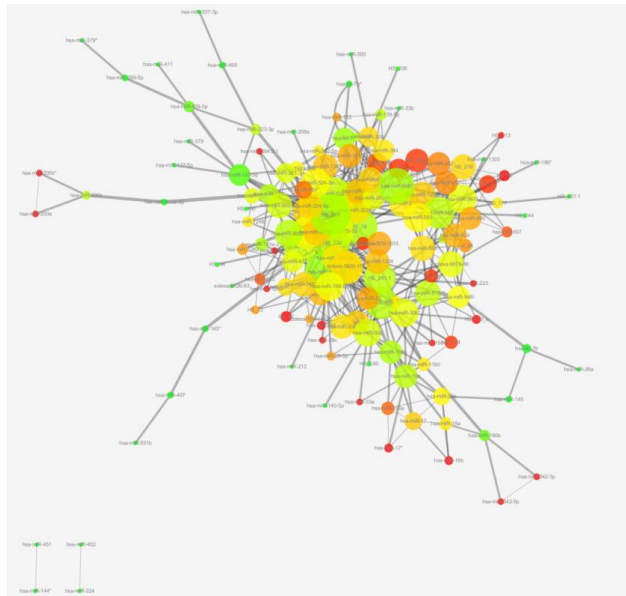
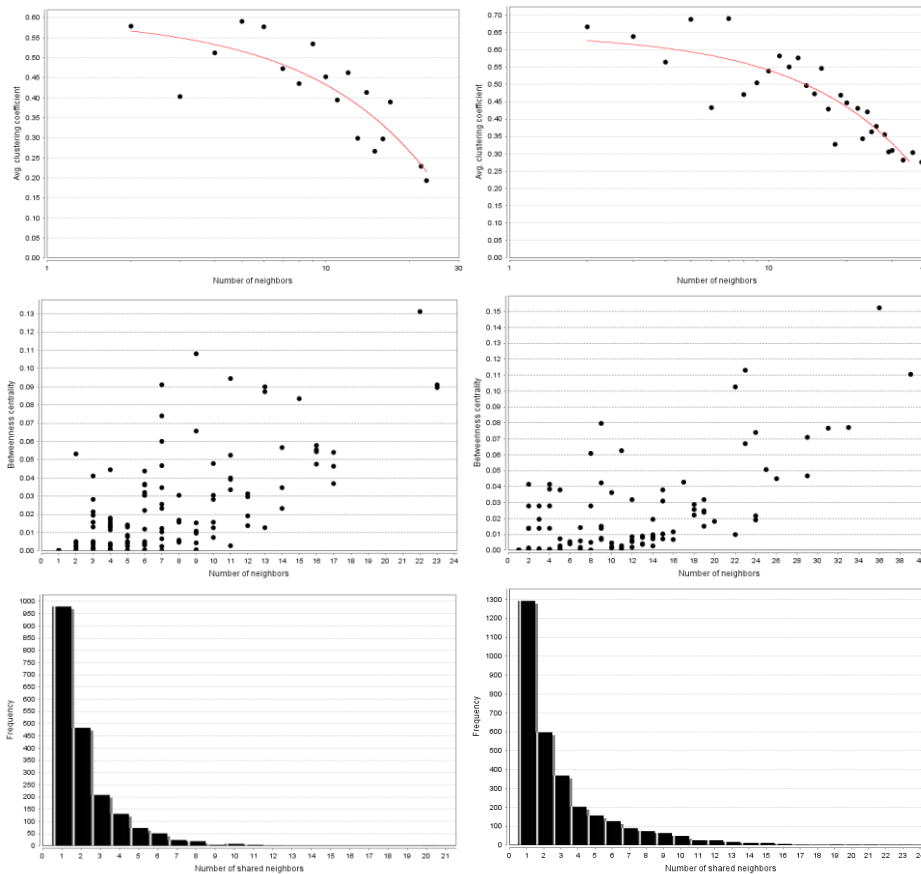
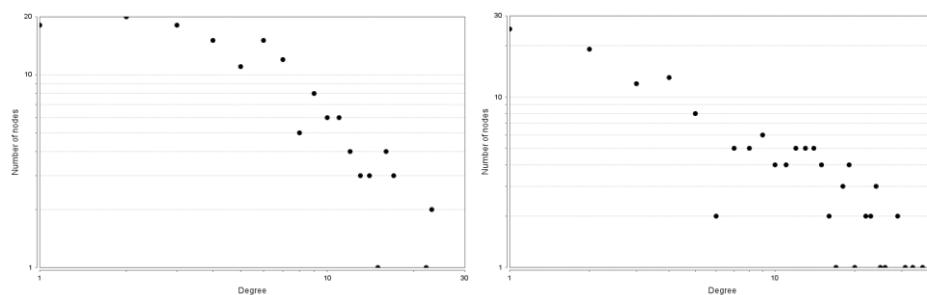


Fig. 3 interaction of Italians of breast cancer

### Characteristic analysis

Several different characteristics among two countries are as follows.





## Summary

The research on the mechanism of gene expression in different ethnic groups is the basis for the consideration of racial differences in the treatment of disease. Micro-RNA plays an important role in cancer and other diseases[10]. The expression of microRNAs in breast cancer can provide a new direction for the diagnosis, prognosis and treatment[11,12].

In this paper, we propose a comprehensive method based on ACCESS data mining, construct the miRNA differential expression network, and this method can construct the miRNA differential expression between Chinese and Italy female successfully.

## Acknowledgments

This research was financially supported by the Scientific Research Common Program of Beijing Municipal Commission of Education (km201510025009).

## References

- [1] YANG, Ling, LI, Lian-di, CHEN, Yu-de, D, M, Parkin. Time trends, estimates and projects for breast cancer incidence and mortality in China[J]. Chinese Journal of Oncology,, 2006, 28(6): 438-440.
- [2] HUANG, Zhe-zhou, CHEN, Wan-qing, WU, Chun-xiao, et, al. ncidence and mortality of female breast cancer in China — a report from 32 Chinese cancer registries, 2003-2007 [J]. Tumor, 2012, 32(6): 435-439.
- [3] ZHANG, Li;FU, Yan-ping. Prospects of microRNA-21 for cancer and other diseases[J]. Biomedical Engineering and Clinical Medicine,, 2015, 19(1): 79-82.
- [4] Jose Antonio Rodriguez-Montes, Pablo Menendez Sanchez. Role of Micro-RNA in Colorectal Cancer Screening[J]. Cirugia Espanola, 2014, 92(10): 654-658.
- [5] Bao, na, Wang. Evaluation of the Correlation Between miR-182, miR-378, miR-34a/c, Ki-67 and Breast Cancer[D]. Beijing Union Medical College:Tumor Hospital Cancer Institute, 2014. 6-35.
- [6] Andrew JG, Sykes PJ. Duplicate thumbs: A survey of results in twenty patients. *J Hand Surg* 1988; 13B: 50–53.
- [7] SHI, Rong, ZHAO, Zhen, GAO, Yang, et, al. Analysis and verification of the interaction of differentially expressed genes in invasive bladder cancer[J]. Journal of Southern Medical University, , 2010, 30(8): 1771-1774.
- [8] Li, Hongbin;He, Guangzhong;Guo, Qiuting. Similarity Retrieval Method of Organic Mass Spectrometry Based on the Pearson Correlation Coefficient[J]. Chemical Analysis and Meterage,, 2015, 24(3): 33-37.
- [9] CHEN, Li-Na, WANG, Qian, SHANG, Yu-Kui, ZHANG, Liang-Cai, et, al. Human protein structure interaction network: the effect of structural domain on the network topology and protein function[J]. Progress in Biochemistry and Biophysics,, 2010, 37(5): 517-519.

- [10] Riccardo, Valdagnia b c, Hendrik, Van, Poppeld, Michael, Aitchisone, et, al. Prostate Cancer Unit Initiative in Europe: A position paper by the European School of Oncology[J]. Elsevier, 2015, (5): 1-11.
- [11] Dobyys JS, Lipscomb PR, Cooney WP. Management of thumb duplication. *Clin Orthop* 1985; 195: 26–44.
- [12] Min, Zhao, Lei, Kong, &, Hong, Qu. A systems biology approach to identify intelligence quotient score-related genomic regions, and pathways relevant to potential therapeutic treatments[J]. *Scientific Reports*, 2014, (4): 1-7.