A Data Mining Method to Find Differentially Expressed miRNAs Using Access Database Language

Ya-Fen CHEN[a], Xiao-Ai CHEN[b] ,Rong WANG[c] , Yi-Wei WANG[d],*Ping ZHOU,

*Ke WANG

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

E-mail: [a]cyfhsps@icloud.com, [b]422750281@qq.com

*Corresponding Author: Contact information: 010-83911544 Email: eduhelp@163.com

Keywords: Database language, data mining, breast cancer, miRNA

Abstract: **Objective** We use database language for data mining in order to find differentially expressed miRNAs. **Methods** We first construct the E-R model, then the data were converted into the appropriate format, and the converted data were imported into the database accurately. In the end, we extracted the data and carried out statistical analysis. **Results** MicroRNA's t test data show that there are differences of miRNA expressions between breast cancer patients of different nationalities. **Conclusion** Access database language can effectively assist data mining, and facilitates data for further mathematical analysis.

INTRODUCTION

MicroRNAs (miRNAs) are small single-stranded RNA molecules, which function as key negative regulators of post-transcriptional modulation in almost all biological processes [1]. Breast cancer is the most common cancer and the leading cause of death among women worldwide [2, 3]. According to the World Health Organization (IARC), the global cancer research center (GLOBOCAN2002), the global annual new female breast cancer cases reached 23%, accounting for 1150000 of women malignant tumor cases, 410000 deaths, accounting for 14% of female malignant tumor [3]. MicroRNAs play 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. Therefore, in order to better carry out data mining and prepare for data analysis, computer technology intervention is essential. ACCESS is a very important tool in data mining and establishing data internal linkages.

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 and EXCEL 2010.

Flow diagram

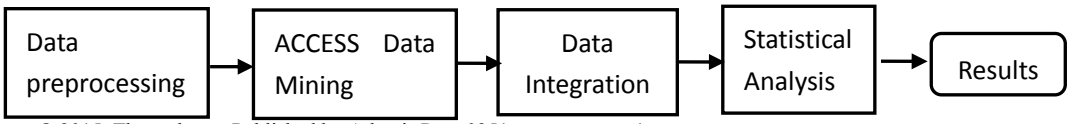

Fig. 1 The flow diagram of statistical analysis

Data preprocessing Module, in the NCBI GEO database, storage format of "series matrix" data is of two parts: "explanation “+”data". After Downloading data, through program, we remove data description. Then the data is processed into a format that can be imported into ACCESS.

ACCESS Data Mining Module, we have the data associated with the probe annotation, extracting miRNA expression profiling data. Association rule shown in Figure 2.

Data Integration Module, according to the corresponding information, we group the corresponding expression profiles of mircoRNAs by nationalities or types of cancer.

Statistical Analysis Module, data imported into statistical analysis software are subjective of t-test.

Results Module, the results of statistical analysis rise.

E-R model

For miRNAs, each platform can obtain only one corresponding data set, and the attribute of data set includes a plurality of samples. So it is possible to use E-R model of the conceptual data model to represent the real world, providing a more intuitive understanding of the relationship between the data set.

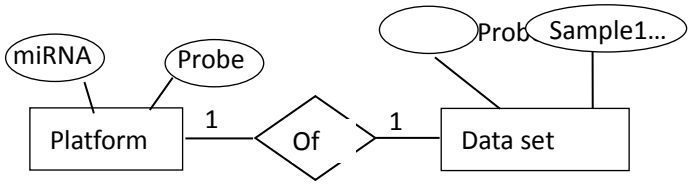

Fig. 2 The E-R model of database

STATISTICAL ANALYSIS

Data contain 1146 microRNAs of each patient. And in the totally 158 cases of patients, there are 66 cases are Italy and 92 cases are Chinese. We make the data subjective of t-test and set parameter: two-tailed, equal variance, unpaired t-test.

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.

The table of significant value

After t-test on the two groups of patients with breast cancer (respectively Chinese and Italian group), microRNAs are ranked in ascending order according to significance value, of which the first 36 set specific value are as below.

Table 1 Significance value of microRNAs

Name	Sig.	Name	Sig.
hsa-miR-663b	8.36927E-21	HS_100	0.000908507
hsa-miR-1226	4.71439E-09	hsa-miR-143*	0.000917441
hsa-miR-1248	1.01349E-06	hsa-miR-378*	0.000975585
hsa-miR-594:9.1	1.73625E-06	hsa-miR-20a	0.000983796
hsa-miR-17-5p:9.1	5.47619E-06	solexa-9578-86	0.0009968
hsa-miR-451	1.30721E-05	hsa-miR-639	0.001102978

hsa-miR-144	1.76754E-05	hsa-miR-18b	0.001136438
hsa-miR-144:9.1	2.12256E-05	hsa-miR-19b-2*	0.001228519
hsa-miR-18a	3.12174E-05	hsa-miR-615-5p	0.001494847
hsa-miR-1299	3.29632E-05	hsa-miR-410	0.00155201
hsa-miR-923	3.63597E-05	HS_42	0.002003915
hsa-miR-598	4.72075E-05	hsa-miR-127-5p	0.002085796
hsa-miR-144*	6.38252E-05	hsa-miR-299-5p	0.002184197
hsa-miR-486-5p	0.000122041	hsa-miR-381	0.002351559
hsa-miR-106b	0.000155256	hsa-miR-181a-2*	0.002409747
hsa-miR-380*	0.000165922	HS_170	0.002437378
hsa-miR-17	0.000166612	hsa-miR-624*	0.002549829
hsa-miR-563	0.000183344	hsa-miR-425	0.002644098

Histogram

List all microRNAs whose significant value smaller than 0.1. The distribution of them is shown below. Fig.3 shows that ,in the total number of 1146 microRNA, the number of microRNAs whose significant value is less than 0.01 is 66, the number less than 0.05 is 181, while the total number less than 0.1 reached 278, accounting for 24.3% of the whole number of microRNAs .

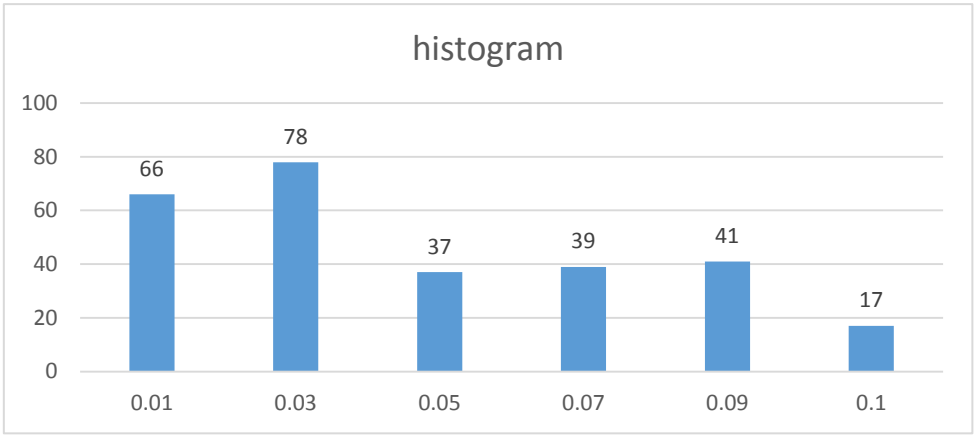

Fig. 3. 0.01 represents the significance below 0.01;0.03 no less than 0.01 and greater than 0.03;0.05 no less than 0.03 and greater than 0.05; 0.07 no less than 0.05 and greater than 0.07; 0.09no less than 0.07 and greater than 0.09; 0.1 no less than 0.09 and greater than 0.1. Ordinate represents the number of microRNAs.

SUMMARY

Micro-RNA plays an important role in various cancers and other diseases. Actually，breast cancer is the best studied cancer in terms of pathobiology, subtypes and treatment, and thus, the role of miRNAs in breast cancer is well characterized. [5].In this paper, we propose a method based on ACCESS data mining. This method greatly facilitated our data collation, and makes subsequent statistical analysis more intuitive and effective. At the same time, by t-test analysis, we got the results of differential expression of microRNA in breast cancer patients of different races, which also prompted us to stand more comprehensive perspective of disease prevention and treatment of breast cancer.

Acknowledgments

This research was financially supported by the Scientific Research Common Program of

Beijing Municipal Commission of Education (km201510025009).

References

[1] Wang W, Luo Y P. MicroRNAs in breast cancer: oncogene and tumor suppressors with clinical potential. Journal of Zhejiang University.science.b, 2015, 16(1):18-31.
[2] Francies F Z, Herd O, Muller X, et al. Chromosomal radiosensitivity of lymphocytes of South African breast cancer patients. South African Medical Journal, 2015.Jose Antonio Rodriguez-Montes, Pablo Menendez Sanchez. Role of Micro-RNA in Colorectal Cancer Screening. Cirugia Espanola, 2014, 92(10): 654-658.
[3] YANG, Ling, LI, Lian-di, CHEN, Yu-de, D, M, Parkin. Time trends, estimates and projects for breast cancer incidence and mortality in China. Chinese Journal of Oncology, 2006, 28(6): 438-440.
[4] Shen Kunwei, Mei Zhangyi, Han Baosan. Research progress on racial differences in breast cancer patients with surgery theory and practice, 2011, 01:82-84.
[5] Mo Y Y. Role of microRNAs in breast cancer. Cancer Biology & Therapy, 2013, 14(3):201-212.