Metabolic Pathways-based Two-step Selective Cluster Analysis: a New Method for Identification of Different Phenotypes of Breast Cancer

Zhao-qi Wang[1], Ping Chen[1], Hong Xia[2]* and Ping Zhou[2]*

[1]*School of Basic Medical Science, Capital Medical University, Beijing 100069, China*
[2]*School of Biomedical Engineering, Capital Medical University, Beijing 100069, China*
wangburgess@sohu.com, aappalg@sina.com

Abstract

OBJECTIVE: To verify the feasibility of identification of different phenotypes of breast cancer by metabolic pathways-based two-step selective cluster analysis. METHODS: Gene expression data of breast cancer (series number GSE10810) were available through Gene Expression Omnibus (GEO) database of National Center for Biotechnology Information (NCBI). Breast cancer samples served as study materials. Samples were analyzed by two-step selective clustering method based on several pathways. After the first clustering analysis, samples with fair results were retained; the other samples were chosen to do the second clustering analysis and got the final results. Cluster3.0 software was applied to clustering analysis. RESULTS AND CONCLUSION: Finally, breast cancer samples were well identified by analyzing of 3 date sets of KEGG pathways. The results confirm the feasibility of discernment of different phenotypes by this method.

Keywords: *metabolic pathway; cluster analysis; breast cancer; phenotype*

1. Introduction

Breast cancer is the most common female malignant tumor. One in every three cancer patients suffers from breast cancer. To the prevention and control of breast cancer, it is necessary to provide a high-quality screening and diagnosis methods [1]. Different phenotypes of breast cancer indicate different prognosis, so identification of phenotypes have important clinical significance. Methods based on morphology become incomplete in clinical practices. Because of high heterogeneity in gene level, same breast cancer patients in morphology have different prognosis and response to clinical treatment. Under the development of molecular biology, methods based on combination of gene chip and pathological analysis provides a better way to reflect biological processes of tumors and estimate prognosis.

Oxidative phosphorylation pathway is a metabolic pathway, using the energy from oxidation of nutrition to synthesize ATP. VEGF signaling pathway regulates angiogenesis in embryo and contributes to formation of new blood vessels. Meanwhile, VEGF receptors induce numerous biological processes, such as cell migration and proliferation. Chemokine signaling pathway participates in inflammation response, such as leukocyte migration. At the same time, chemokine can regulate leukocyte growth, differentiation and leukocyte activation.

* To whom correspondence should be addressed. Tel. 010-83911805. E-mail: 6357649@qq.com.
* To whom correspondence should be addressed. Tel. 010-83911805. E-mail: 6357649@qq.com.

It is important to understand mechanism of diseases by high-throughput gene expression profiles and data mining technology. Gene expression is not isolated, function-related genes display high relevance [2]. Gene expression in same metabolic pathway tends to be high correlated [3]. KEGG pathway can effectively predict protein interaction network and cell function [4-5]. Tumor is a gene-based disease, whose causes are to be found in disruptions of basic biologic processes. Analysis based on pathways can better integrate phenotypes and biological processes [6]. In order to enhance identification different phenotypes further discuss the biological meaning, particular KEGG pathways were obtained, whose gene expression profiling date was cluster analyzed.

2. Materials and methods

2.1. Date obtaining

Gene expression profile data of breast cancer (series number GSE10810) were obtained from GEO database of NCBI. Removing date of control group, 31 samples served as study materials, including phenotype 1 (lymph node negative, ER-positive, 8 samples), phenotype 2(lymph node negative, ER-negative, 10 samples), phenotype 3(lymph node positive, ER-positive, 13 samples). Meanwhile, 22 samples (remove clinical date missing samples) were divided into 3 groups by degree of differentiation, including high differentiation group (grade 1, 2 samples), medium differentiation group (grade 2, 10 samples), low differentiation group (grade 3, 10 samples).Gene sets of related pathway were downloaded from PATHWAY database (http://www.genome.jp/kegg/pathway.html). Figure 1 showed the gene relationship of the three pathways. Genes can be assayed by one or more probes, results were all retained.

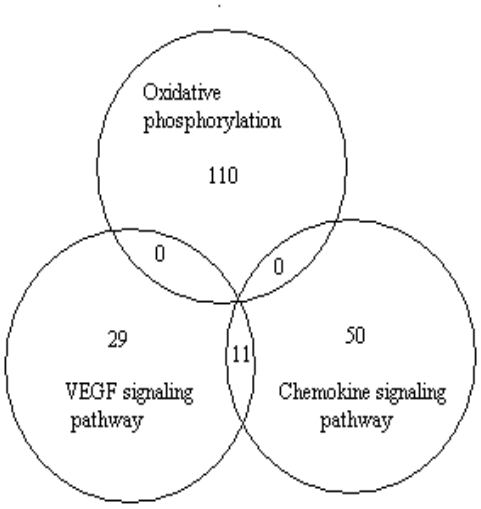


Figure 1. Relationship of genes of the three pathways

2.2. Centroid-linkage clustering analysis

Distance of two clusters is distance of the two centroids c_i and c_j of the two clusters C_i and C_j. The formula as follows:

$$d(C_i, C_j) = d(c_i, c_j) \quad (1)$$

$$c_i = \frac{1}{|C_i|}\sum_{x \in c_i} x \qquad (2)$$

$$c_j = \frac{1}{|C_j|}\sum_{x \in c_j} x \qquad (3)$$

2.3. Clustering analysis process

2.3.1. Clustering analysis on lymph node and ER: Cluster3.0 software (American Axon Company) was applied to clustering analysis [7-9]. First, Centroid-linkage clustering analysis on 31 samples' gene expression profiling of Chemokine signaling pathway, treeview software was applied to observing results. Samples with fair results (validity rate>60%) were retained, residual samples were selected to do the second clustering analysis on VEGF signaling pathway and Oxidative phosphorylation. And Treeview software was applied to observing results.

2.3.2. Clustering analysis on degree of differentiation: Centroid-linkage clustering analysis on 31 samples' gene expression profiling of Chemokine signaling pathway. And Treeview software was applied to observing results.

Figure 2 was showed process of clustering analysis.

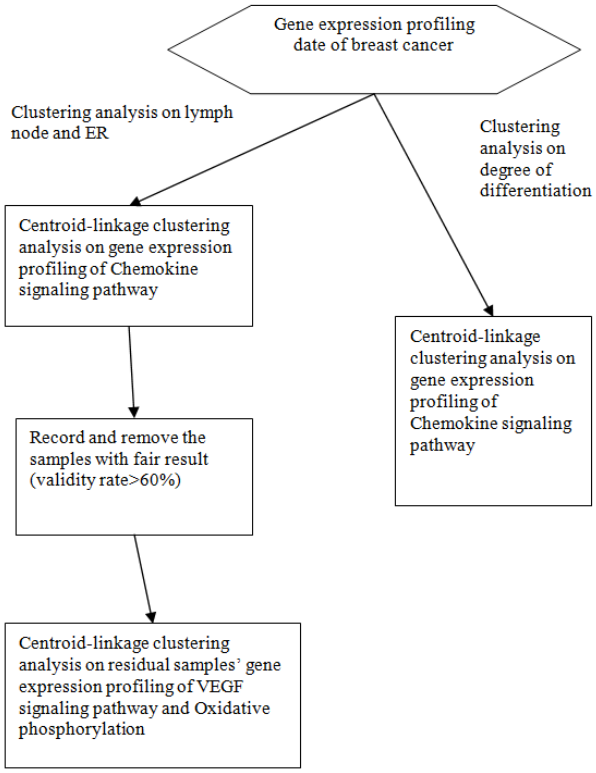
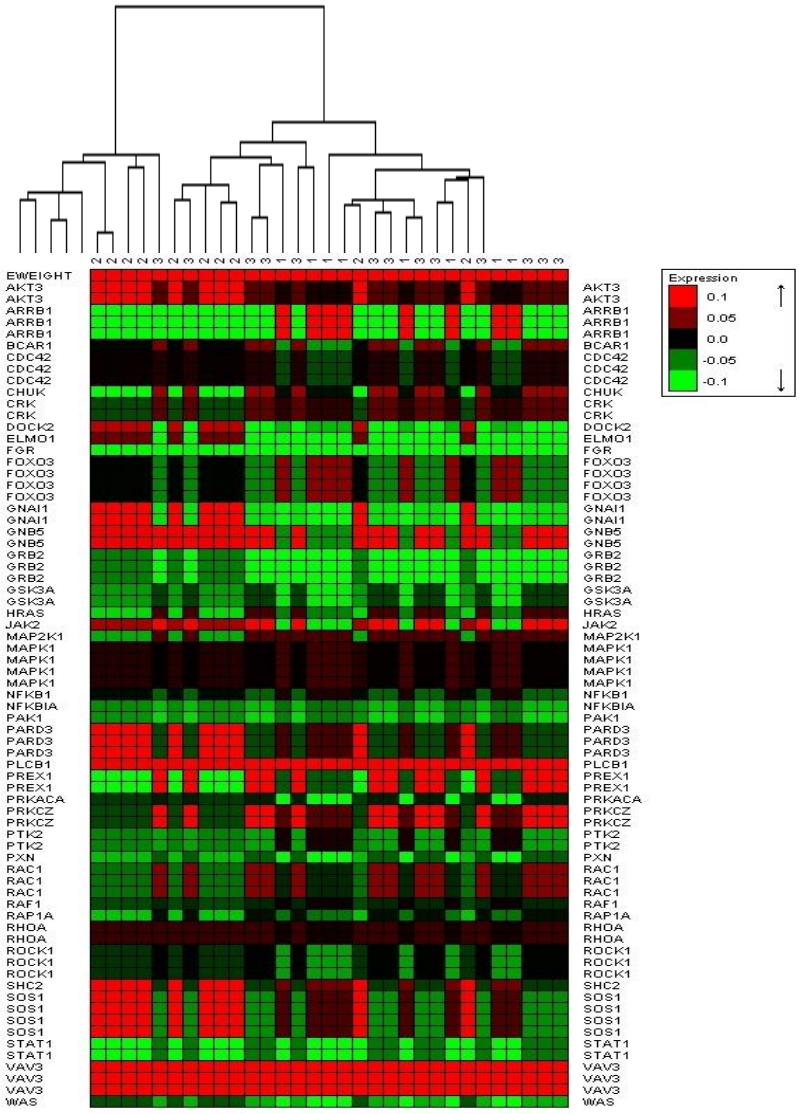


Figure 2. Process of clustering analysis

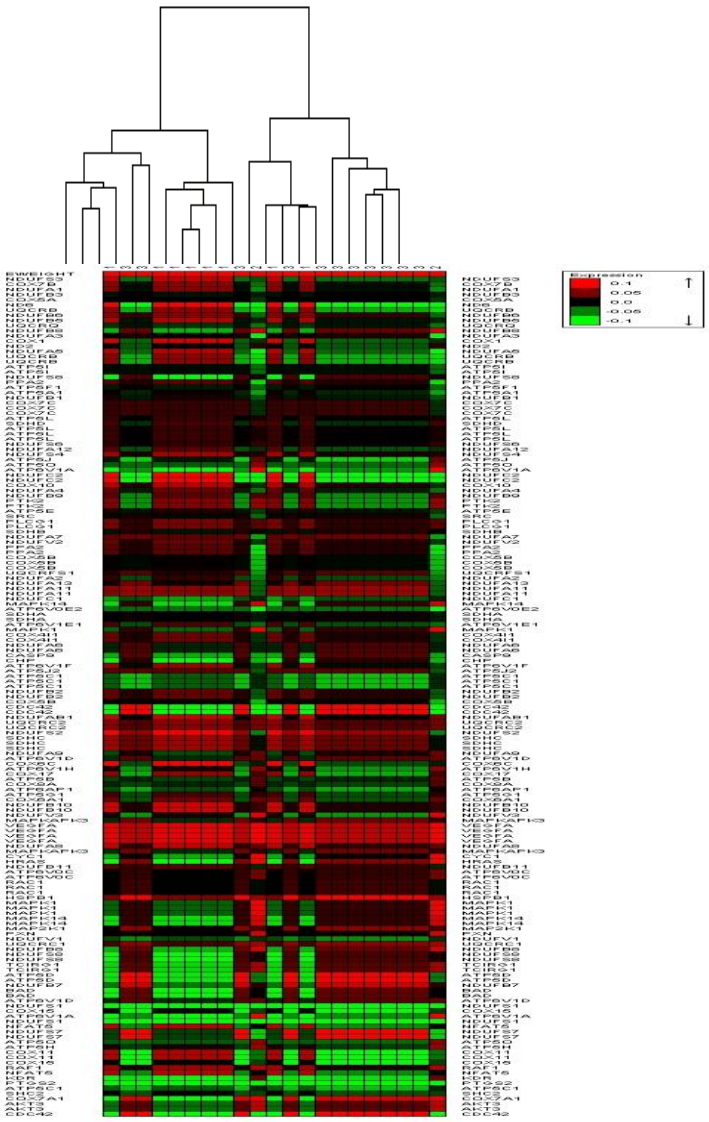
3. Result

3.1. Results of clustering analysis of lymph node and ER

Centroid-linkage clustering analysis on 31 samples' gene expression profiling of Chemokine signaling pathway, Figure 3(A) shows the results. The dendrogram shows that samples are divided into two groups. Remove samples in the left group (group of phenotype 2), residual samples (indistinct result) are selected to do the second clustering analysis, Results are showed in Figure 3(B). Table 1 shows the results after the two steps.



3（A）



3(B)

Figure 3. Clustering analysis on 31 breast caner samples using KEGG pathway

Table 1. Cluster analysis on 31 samples of breast cancer

Group	Sample	Phenotype			P-value	FDR (%)
		1	2	3		
Group of phenotype 1	11	7	1	3	<0.001	36.4
Group of Phenotype 2	10	0	8	2		20.0
Group of phenotype 3	10	1	1	8		20.0

3.2. Clustering analysis result of degree of differentiation

Centroid-linkage clustering analysis on 22 samples' gene expression profiling of Chemokine signaling pathway, results are showed in Figure 4 and Table 2. The dendrogram shows that samples are divided into two groups, the group of low differentiation (left) and the group of high and medium differentiation (right).

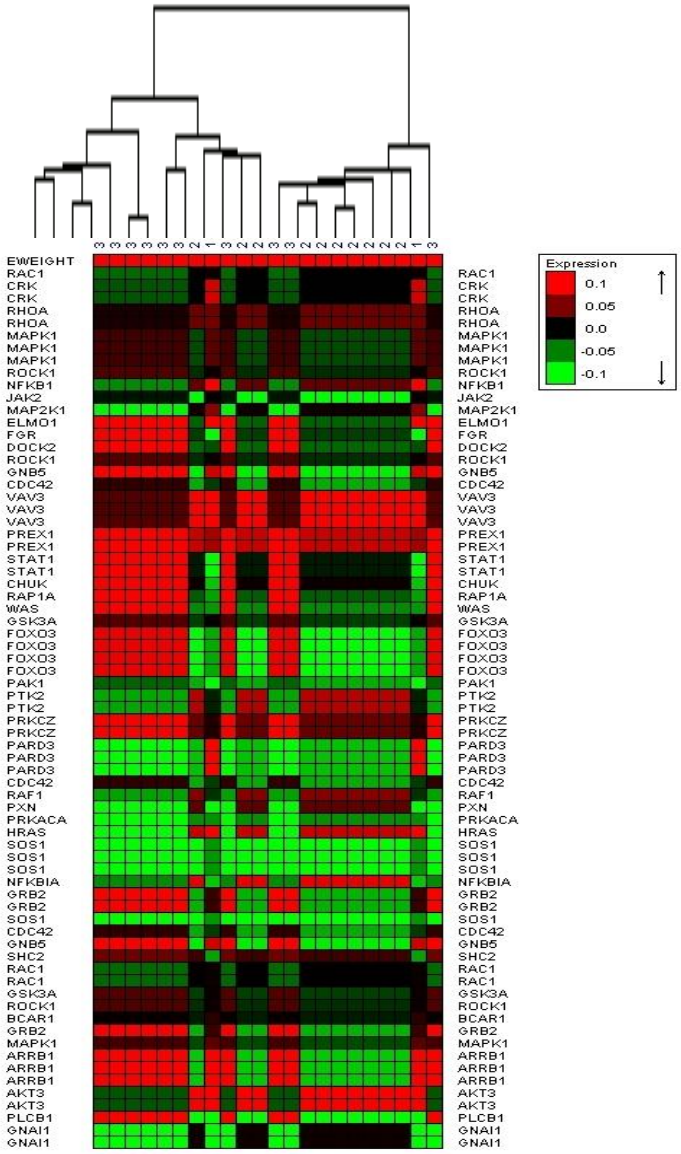


Figure 4. Clustering analysis on 22 breast caner samples using KEGG pathway

Table 2. Cluster analysis on 22 samples of breast cancer

Group	Sample	Cell differentiation		P-value	FDR (%)
		Grade1 and Grade2	Grade3		
Group of Grade1 and Grade2	9	8	1	0.002	11.1
Group of Grade3	13	4	9		30.8

4. Discussion

It is verified that high-throughput gene expression profiles has capacity in dissecting complexity of tumor phenotype and the mechanism [10]. The available analysis methods of gene expression profiles mostly focus on differential expression gene among samples [11]. However, it neglects the genes without obvious multiple change but active, losing lots of information of date [12-15]. Biological phenomenon is the consequence of interaction of genes and their products [6]. So, study based on gene expression profiling of pathways has more biological meaning than differential expression gene.

3 sets of KEGG pathways are obtained, whose gene expression profiling date can identify different grades of tumor differentiation and phenotypes of breast cancer. Samples are well divided into 3 groups ($P<0.001$) by phenotypes: group of phenotype 1 (FDR= 0.364), group of phenotype 2 (FDR=0.200) and group of phenotype 3 (FDR= 0.200). Meanwhile, samples are divided into 2 groups ($P=0.002$) by degree of cell differentiation: group of high and medium differentiation (FDR=0.111) and group of low differentiation (FDR= 0.308). The results confirm the feasibility of identification of different phenotypes of breast cancer by this method.

At the same time, identification of breast cancer phenotypes based on metabolic pathways contributes to discussing biological meaning underlying phenotypes, benefiting further investment on clinical diagnosis and treatment of breast cancer. About 75% breast cancer is ER-positive [16]. Lymph node metastasis always represents tumor metastasis, as a prognosis indicator of diseases progression [17]. In the article, ER-positive breast cancer divides into phenotype 1 and phenotype 2 by condition of lymph node metastasis. Analysis based on VEGF signaling pathway and oxidative phosphorylation can distinguish effectively this two phenotypes, indicating that these pathways make contribution to lymph node metastasis of breast cancer. Studies show that VEGF family participates in lymphoangiogenesis and other biological processes, and is related to tumor lymph node metastasis [18-20]. Bevacizumab is a monoclonal antibody that can specifically blocks receptor binding site of VEGF. Taking Bevacizumab in chemical treatment can improve progression free survival (PFS) and objective remission rate (ORR) of metastatic breast cancer patients [21].

This study shows the feasibility of identification of different phenotypes of breast cancer by metabolic pathways-based two-step selective cluster Analysis. Breast cancer is a heterogeneous disease, different phenotypes of breast cancer affect the outcome and treatment

of patients [22-23]. Our study focuses on gene expression profiling of metabolic pathways, benefiting identification of phenotypes and providing a new idea on clinical diagnosis and treatment.

Acknowledgements

This work was supported by the Science and Technology Development Plan Project of Beijing Municipal Education Commission (Grant No. SQKM201210025008), the Clinical and Basic Cooperation Foundation of Capital Medical University (Grant No. 12JL47).

References

[1] C. DeSantis, R. Siegel, P. Bandi and A. Jemal, CA, CANCER, J., CLIN, vol. 61, no. 409**, (2011)**.
[2] L. Hua, D. G. Li, H. Lin, L. Li, X. Li and Z. C. Liu, J., THEOR. BIOL., vol. 266, **(2010)**, pp. 242.
[3] L. Hua, W. Y. Zheng, H. Liu, H. Lin and L. Gao, Chin. J., Biotech, vol. 24, **(2008)**, pp. 164.
[4] M. Kanehisa, S. Goto, S. Kawashima and A. Nakaya, NUCLEIC. ACIDS. RES., vol. 30, **(2002)**, pp. 42.
[5] A. H. M. Salleh and M. S. Mohamad, International Journal of Bio-Science and Bio-Technology, vol. 4, **(2012)**, pp. 1.
[6] K. H. Buetow, S. Efroni and C. F. Schaefer, PLoS. ONE., vol. 2, **(2007)**, pp. e425.
[7] M. B. Eisen, P. T. Spellman, P. O. Brown and D. Botstein, Proc. Natl. Acad. Sci. USA., vol. 95, pp. 1486, **(1998)**.
[8] N. Tajunisha and V. Saravanan, International Journal of Advanced Science and Technology, vol. 27, **(2011)**, pp. 85.
[9] S. S. Yeo, G. C. Park, S. S. Kim, T. H. Kim and S. K. Kim, International Journal of Multimedia and Ubiquitous Engineering, vol. 2, **(2007)**, pp. 15.
[10] J. R. Nevins and A. Potti, Nat. Rev. Genet., vol. 8, **(2007)**, pp. 601.
[11] Y. H. You and Z. J. Zhang, China Biotechnology, vol. 29, **(2009)**, pp. 87.
[12] E. Segal, N. Friedman, N. Kaminski, A. Regev and D. Koller, Nat. Genet., vol. 37, **(2005)**, pp. 38.
[13] R. Radha, International Journal of Advanced Science and Technology, vol. 33**, (2011)**, pp. 23.
[14] C. K. Verma and N. Srivastava, International Journal of Advanced Science and Technology, vol. 42, pp. 19, **(2012)**.
[15] S. H. Jang, T. Y. Kim, J. K. Kim and J. S. Lee, International Journal of Control and Automation, vol. 5, **(2012)**, pp. 157.
[16] S. C. Baumgarten and J. Frasor, Mol, Endocrinol, vol. 26**, (2012)**, pp. 360.
[17] C. Wissmann and M. Detmar, Clin. Cancer. Res., vol. 12**, (2006)**, pp. 6865.
[18] K. Alitalo, T. Tammela and T. V. Petrova, Nature, vol. 438, **(2005)**, pp. 946**.**
[19] M. Kowanetz and N. Ferrara, Clin. Cancer. Res., vol. 12, **(2006)**, pp. 5018.
[20] P. Saharinen, L. Eklund, K. Pulkki, P. Bono and K. Alitalo, Trends. Mol. Med., vol. 17, **(2011)**, pp. 347.
[21] A. Valachis, N. P. Polyzos, N. A. Patsopoulos, V. Georgoulias, D. Mavroudis and D. Mauri, Breast. Cancer. Res. Treat., vol. 122**, (2010)**, pp. 1.
[22] K. Polyak, J. Clin. Invest., vol. 121, **(2011)**, pp. 3786.
[23] G. V. Minckwitz, M. Untch, E. Nüesch, S. Loibl, M. Kaufmann, S. Kümmel, P. A. Fasching, W. Eiermann, J. U. Blohmer, S. D. Costa, K. Mehta, J. Hilfrich, C. Jackisch, B. Gerber, A. du Bois, J. Huober, C. Hanusch, G. Konecny, W. Fett, E. Stickeler, N. Harbeck, V. Müller, P. Jüni, Breast. Cancer. Res. Treat., vol. 125, **(2011)**, pp. 145.

Authors

Zhao-qi Wang, undergraduate, majoring in Clinical Medicine, School of Basic Medical Science, Capital Medical University.

Ping Chen, undergraduate, majoring in Clinical Medicine, School of Basic Medical Science, Capital Medical University.

Hong Xia, Ph. D., School of Basic Medical Science, Capital Medical University, major research field: Biomedical Informatics.

Ping Zhou, Ph. D., School of Basic Medical Science, Capital Medical University, major research field: Bioinformatics.