

# Identification of Different Phenotypes of Breast Cancer Based on Two-Step Selective Clustering Analysis of Gene Expression Profiling of Several Signal Transduction, Immune and Metabolic Pathways

Zhao-qi Wang<sup>1,a</sup>, Ping Chen<sup>1,b</sup>, Hong Xia<sup>2,\*</sup>, Ping Zhou<sup>2,\*</sup>

<sup>1</sup>School of Basic Medical Science, Capital Medical University, Beijing 100069, China

<sup>2</sup>School of Biomedical Engineering, Capital Medical University, Beijing 100069, China a [wangburgess@sohu.com](mailto:wangburgess@sohu.com), b [aappalg@sina.com](mailto:aappalg@sina.com)

**Abstract.** OBJECTIVE: In order to enhance identification of different phenotypes of breast cancer, a two-step selective clustering method was proposed. METHODS: Gene expression profile data of breast cancer (series number GSE10810) were obtained from Gene Expression Omnibus (GEO) database of National Center for Biotechnology Information (NCBI). Gene expression profiling of several metabolic pathways were analyzed by two-step selective clustering method. After the first clustering analysis, samples with fair results were retained; residual samples were selected to do the second clustering analysis for the final results. Cluster3.0 software was applied to clustering analysis. RESULTS AND CONCLUSION: 3 sets of KEGG pathways were obtained that can identify different phenotypes of breast cancer. The results confirm the feasibility of identification of different phenotypes by two-step selective clustering method.

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

## 1 Introduction

Breast cancer is the most common female malignant tumor. Identification of phenotypes have important clinical significance. Under the development of molecular biology, methods based on gene chip provides a better way to reflect biological processes of tumors and estimate prognosis.

Oxidative phosphorylation pathway is a metabolic pathway, using the energy from nutrition to synthesize ATP. VEGF signaling pathway regulates angiogenesis and formation of new blood vessels. Chemokine signaling pathway participates in inflammation response.

Gene expression is not isolated, function-related genes display high relevance<sup>[1]</sup>. Gene expression in same metabolic pathway tends to be high correlated. KEGG

\* To whom correspondence should be addressed. Tel. 010-83911805. [E-mail: 6357649@qq.com](mailto:6357649@qq.com).

\* To whom correspondence should be addressed. Tel. 010-83911805. [E-mail: wjzpwyz@163.com](mailto:wjzpwyz@163.com).

pathway can predict protein interaction network and cell function. Analysis based on pathways can better integrate phenotypes and biological processes<sup>[2]</sup>.

## 2 Materials and Methods

### 2.1 Data Obtaining:

Gene expression profile data of breast cancer (series number GSE10810) were obtained from GEO database of NCBI. Removing data 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). Gene sets of related pathway were downloaded from PATHWAY database (<http://www.genome.jp/kegg/pathway.html>).

### 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:

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

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

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

### 2.3 Clustering Analysis Process:

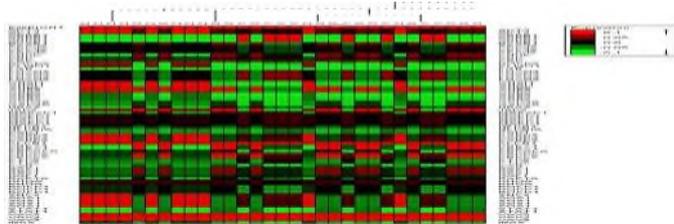
Cluster3.0 software (American Axon Company) was applied to clustering analysis. First, Centroid-linkage clustering analysis on 31 samples' gene expression profiling of Chemokine signaling pathway, treeview software was applied to observing results. Samples with fair results (validity rate > 60%) were 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.

## 3 Result

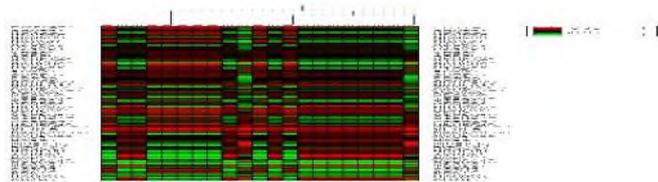
Centroid-linkage clustering analysis on 31 samples' gene expression profiling of Chemokine signaling pathway, figure 3(A) shows the results. Samples of the right

Identification of Different Phenotypes of Breast Cancer Based on Two-Step Selective Clustering Analysis of Gene Expression Profiling of Several Signal Transduction, Immune and Metabolic Pathways

group(indistinct result) are selected to do the second analysis, Results are showed in figure 3(B). Table 1 shows the results after the two steps.



1 ( A )



1 ( B )

**Fig.1** Clustering analysis on 31 breast cancer samples using KEGG pathway

Table1 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

#### 4 Discussion

It is verified that high-throughput gene expression profiles has capacity in dissecting complexity of tumor phenotype and the mechanism. The available analysis methods of gene expression profiles mostly focus on differential expression gene among samples [3]. However, it neglects the genes without obvious multiple change but active, losing lots of information of data [4]. Biological phenomenon is the consequence of interaction of genes and their products [2]. So, study based on gene expression profiling of pathways has more biological meaning than differential expression gene.

Three sets of KEGG pathways are obtained, whose gene expression profiling data can identify different phenotypes of breast cancer. After two steps clustering analysis, samples divide into 3 groups ( $P < 0.001$ ), with 74.2% (23/31) coincidence rate between results and phenotypes. The results confirm the feasibility of

identification of different phenotypes of breast cancer by two-step selective clustering analysis.

At the same time, identification of breast cancer phenotypes based on metabolic pathways contributes to discussing biological meaning underlying phenotypes. About 75% breast cancer is ER-positive<sup>[5]</sup>. Lymph node metastasis always represents tumor metastasis, as a prognosis indicator of diseases progression. 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<sup>[6]</sup>. 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<sup>[7]</sup>.

Breast cancer is a heterogeneous disease, different phenotypes of breast cancer affect the outcome of patients. 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).

## Reference

1. Hua L, Li D G, Lin H, Li L, Li X and Liu Z C: The correlation of gene expression and co-regulated gene patterns in characteristic KEGG pathways. *J THEOR BIOL.* 266, 242–249 (2010)
2. Buetow K H, Efroni S, Schaefer C F: Identification of Key Processes Underlying Cancer Phenotypes Using Biologic Pathway Analysis. *PLoS ONE.* 2, e425 (2007)
3. You Y H, Zhang Z J: Data Mining from Microarray Gene Expression Profile. *China Biotechnology.* 29, 87–91 (2009)
4. Segal E, Friedman N, Kaminski N, Regev A and Koller D: From signatures to models: understanding cancer using microarrays. *Nat Genet.* 37, 38–45 (2005)
5. Baumgarten S C, Frasar J: Minireview: Inflammation: An Instigator of More Aggressive Estrogen Receptor (ER) Positive Breast Cancers. *Mol Endocrinol.* 26, 360–371 (2012)
6. Saharinen P, Eklund L, Pulkki K, Bono P and Alitalo K: VEGF and angiopoietin signaling in tumor angiogenesis and metastasis. *Trends Mol Med.* 17, 347–362 (2011)
7. Valachis A, Polyzos N P, Patsopoulos N A, Georgoulas V, Mavroudis D and Mauri D: Bevacizumab in metastatic breast cancer: a meta-analysis of randomized controlled trials. *Breast Cancer Res Treat.* 122, 1–7 (2010)