Distinguishing Tumors from Normal by Cell Surface Maker with Spherical SOM

Yuh Sugii
Ling Chen
Heizo Tokutaka
Masaharu Seno

dns20307@okayama-u.ac.jp  chenlingdoctor@126.com  tokuhema@hal.ne.jp  mseno@cc.okayama-u.ac.jp

1 Graduate School of Natural Science and Technology, Okayama University, Okayama 700-8530, Japan
2 SOM Japan, Inc. Tottori 680-8550, Japan

Keywords: cell surface marker, oligonucleotide microarray, brain tumor, spherical Self-Organizing Map, gene expression profiling

1 Introduction

Since tumor cells and tissues have gene expression profiles different from those of their counterparts of normal tissues, tumor specific cell surface proteins should be attractive markers for tumor targeting. To detect differentially expressed cell surface markers in brain tumor, we analyzed gene expression profiles of nine human brain tumor cell lines and normal brain tissues by utilizing an oligonucleotide microarray, which we originally were designed the probe oligonucleotides are restricted only to cell surface proteins [1]. Oligonucleotide microarray is a powerful tool to analyze gene expression profiles. However, huge amount of information obtained from one experiment is not always easy to make the character of target cells or tissues comprehensive. There still seem to be needs to improve the procedures of clustering the information of signals from microarray. To evaluate and analyze microarray data, we have been proposing the application of spherical self-organizing map (sSOM) [2, 3].

In this paper we are describing the procedure of clustering two groups, tumor and normal tissues, with signal data filtered with the average values and standard deviations. As a result, we could successfully identified several genes whose expression is significantly consistent in all the cancer cell lines used and higher than those in normal brain tissue.

2 Method and Results

2.1 Expression Analysis

Total RNA of human brain tumor cell lines was isolated from cultured cells with RNeasy Mini kit (Qiagen, #74104). Total RNA of human brain was bought from Stratagen (Cat. #540005 #540157). Twenty micrograms of total RNA was used to synthesize cDNA. Cy3-labeled cDNAs were prepared by indirect labeling method adapted from the Brown web site (http://cmgm.stanford.edu/pbrown/protocols) and hybridized to cell surface marker DNA microarray in 5xSSC/0.5%SDS solution at 55°C for 15h. Arrays were designed to contain 1809 probes (1795 for human genes of cell surface proteins, 9 for housekeeping genes and 5 for exogeneous controls), which are spotted at least in duplicate. After washing, arrays were scanned on a FLA8000 scanner (Fuji Film, Japan). Intensity for each spot of the array was captured by using GenePix® Pro5.1 image analysis software (Axon Instrument) and a single raw expression level for each gene was derived from the 2 spots representing each gene by mean algorithm. Then, the raw fluorescent intensity of each gene referred to as “relative fluorescent intensity (RFI)”, which represented the expression level of each gene. Gene expression levels were compared to one another by RFI value to identify differentially expressed genes.

2.2 Data Filtering and sSOM Clustering

In order to eliminate genes that did not change significantly between tumor cell lines and normal tissue, each gene is given a score by a formula $|N-G| - V_N - V_G$, where “N”, “G”, “V_N” and “V_G” denote the average
expression level of the gene in adult and fetal normal brain, the average expression level of the gene in the nine tumor cell lines, the standard deviation of the gene expression level in adult and fetal normal brain, the standard deviation of the gene expression level in the nine tumor cell lines, respectively. Genes were eliminated if they did not show \( |N - G| - V_N - V_G > 0 \), since only the genes with a score greater than a threshold (i.e., zero) are deemed potentially significant (Fig. 1). The expression levels were then normalized first by the maximal value of each gene among the ten samples, and secondly by the maximal difference value between maximum and minimum of each gene within all the remained genes. The normalized data were clustered and displayed using sSOM.

2.3 Result
As a result, 267 genes with a potentially significant difference were analyzed by sSOM. As shown in Fig. 2, similar expression patterns in all the nine tumor derived cell lines were recognized in each case, whereas red, blue and white denoted the expression levels for a single gene is above, below and same as the median of that gene across all samples. Based on the sSOM patterns, several genes that were highly expressed in the tumor cell lines were selected as candidates of tumor-specific cell surface markers. The candidate genes were validated quantitative PCR and immunostaining. The results were consistent with those from microarray analysis.

![Figure 1: Selection of genes that fulfill the equation \( |N - G| - V_N - V_G > 0 \). Spots correspond to each gene. The specificity for tumor increases along with the direction of arrow.](image1)

![Figure 2: Gene expression profiles of brain tumor cell lines and normal brain tissues clustered by spherical SOM.](image2)

3 Discussions
Our cell surface marker oligonucleotide microarray should be useful for screening the candidates of cell type-specific surface markers, which might serve as molecular targets for diagnosis and treatment of disease. The sSOM clustering is a powerful tool for data mining, knowledge discovery and visualization of multidimensional data.

References