Biochemical and Biophysical Research Communications 376 (2008) 414–418

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Spherical self-organizing map as a helpful tool to identify category-specific cell surface markers

Tuoya [a], Yuh Sugii [a], Hitomi Satoh [b], Dongwei Yu [a], Yasaburo Matsuura [c], Heizo Tokutaka [d], Masaharu Seno [a,*]

[a] Department of Medical and Bioengineering Science, Graduate School of Natural Science and Technology, Okayama University, Rm361, Bldg ENG-6, 3.l.1 Tsushima-naka, Okayama 700-8530, Japan
[b] Department of Bioscience and Biotechnology, Faculty of Engineering, Okayama University, Okayama 700-8530, Japan
[c] Department of Electrical and Electronic Engineering, Faculty of Engineering, Tottori University, Tottori 680-8552, Japan
[d] SOM Japan, Inc., Tottori 680-8550, Japan

ARTICLE INFO

Article history:
Received 28 August 2008
Available online 13 September 2008

Keywords:
Tumor-specific
Cell surface marker
DNA microarray
Spherical self-organizing map

ABSTRACT

We analyzed gene expression profiles of five tumor cell lines (NB2a, NB41A3, C1300N18, BC3H1, and Neuro2a) derived from a category of nervous system using our originally developed cell surface marker DNA microarray in order to search for tumor-specific cell surface markers common to these cells. To visualize the expression patterns and to extract candidate genes of interest based on the expression profiles of several cell lines, we employed the clustering procedure of spherical self-organizing-map. As the result, three candidates of tumor-specific cell surface markers were picked up when the expression profiles were compared with that from normal brain tissue. RT-qPCR showed the expression of these genes was higher in tumor cells than in normal brain. Here we demonstrated the spherical self-organizing-map analysis should be useful to identify the candidates of cell surface markers common and specific to the group of cells or tissues of interest.

Cell surface proteins such as growth factor and cytokine receptors, cell adhesion molecules, extracellular matrices and so on are involved in biological functions. Further, it has been found that there is a difference in their gene expression between tumor and corresponding adjacent normal tissue [1]. This heterogeneity of cell surface protein expression makes these molecules themselves useful as the markers for specific tumor targeting [2–6]. Therefore, it is extremely important to find cell- or tissue-specific surface markers to distinguish the cells or tissues in some diseases from those in a normal state. For this purpose we have developed DNA microarrays carrying the probes for the genes coding cell surface proteins [7]. The probes are mainly designed to include the region essential for membrane binding potential eliminating alternatively spliced by-products released from membrane. We consider this design is highly efficient for the downstream procedure of screening for cell surface markers.

DNA microarray technology allows us large-scale and high-throughput screening of differentially expressed genes, whereas Northern hybridization or RT-PCR procedures are designed only for the characterization of the target genes. Sophisticated data analysis has demonstrated that various tumor types can be distinguished on the basis of their gene expression patterns, and that a large number of genes are over-expressed in tumor compared to their normal tissue counterparts [8]. A method to evaluate and analyze the massive data generated by series of microarray experiments plays an important role to reveal the hidden patterns of gene expression. Computational tools, for example, hierarchical clustering and self-organizing map (SOM) clustering, have widely been used to extract useful information from expression profiles. Compared with hierarchical clustering, SOM has a number of features well suited to cluster genes by their expression patterns [9,10]. Although SOM has good computational properties and is easy to implement, reasonably fast, and to scale large data sets, a conventional plane SOM (2D SOM), whose neighborhood relationship is defined by a 2D rectangular or hexagonal lattice, is not yet familiar to gene clustering procedure. The reasons might be the grid units at the boundary of the SOM have fewer neighbors than the units inside the map, which often leads to the notorious "border effect"—the weight vectors of these units "collapse to the center of the input space" [11]. To solve this problem, Ritter proposed that spherical SOM (sSOM) would be suitable for data with underlying directional structures [12]. sSOM has been shown effective to remove the boundary effect conveying the information of distance and direction with running speed comparable to the conventional 2D SOM [13,14].

In this study, sSOM was applied to analyze the data from cell surface marker DNA microarray and visualize the expression patterns. By comparing the sSOM patterns of tumor cells with that of normal tissues, we demonstrated a simple and easy way of

* Corresponding author. Fax: +81 86 251 8216.
E-mail address: mseno@cc.okayama-u.ac.jp (M. Seno).

0006-291X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2008.09.010

screening for some candidates of common and specific cell surface markers in nervous system-derived tumor cells.

Methods

Cell lines and cell culture. Five cell lines of mouse nervous system-derived tumors were collected. Cell lines Neuro-2a and BC3H1 were from American Type Culture collection (ATCC). They were cultured in Eagle's Minimum essential medium (EMEM) with 10% (v/v) fetal bovine serum (FBS), Dulbecco's modified Eagle's medium (DMEM) with 20% FBS, respectively. Cell lines NB2a, C1300N18, and NB41A3 were kindly provided by Cell Resource Center for Biomedical Research Institute of Development, Aging and Cancer, Tohoku University. They were cultured in RPMI-1640 with 10% FBS, DMEM with 10% FBS, Ham's F12 medium with 15% horse serum, and FBS 2.5%, respectively. All of cell lines were maintained at 37 °C in a 5% CO_2-humidified incubator.

DNA microarray gene expression analysis. Total RNA of mouse normal brain was purchased from Stratagene (Cat. No. 736001). Total RNA preparation from cultured cells and DNA microarray gene expression analysis was performed as described previously [7]. Briefly, 20 μg 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/protcols) and hybridized to cell surface marker DNA microarray originally designed to contain 1402 probes (1391 for mouse genes of cell surface proteins, 10 for housekeeping genes, and 5 for exogenous controls), which are spotted in duplicate. Arrays were scanned on a FLA8000 scanner (Fuji Film, Japan). Fluorescence signal intensities were captured by using GenePix® Pro 5.1 image analysis software (Axon Instruments). The signals were calibrated to a control RNA exogenously supplemented to the total RNA before cDNA synthesis.

Data filtering and sSOM analysis. In order to eliminate genes of which expression did not change significantly between tumor derived cell lines and normal tissue, we evaluated the scores for each gene by a filtering formula $|A-G|-V$, where "A", "G", and "V" denote the expression level of a gene in normal brain, the average expression level of the gene in the five tumor cell lines, and the standard deviation of the gene expression level among the five tumor cell lines, respectively. Genes were eliminated from further process if the score did not show $|A-G|-V>0$, since only the genes with a score greater than the threshold (i.e., zero) are deemed potentially significant. Then the expression levels in each gene were first normalized the maximal value of each gene as 1 among the six samples. The difference between the maximum and minimum values of each gene was calculated. The difference values were secondly normalized the maximal as 1 within all the remained genes. The normalized data set was used for clustering genes by the sSOM software "Blossom" (SOM Japan, Co. Ltd., Tottori, Japan)

Real-time quantitative PCR analysis. Single-strand cDNAs were reverse-transcribed using oligo dT18 primer. Real-time quantitative PCR (RT-qPCR) was performed using Light Cycler DX400 (Roche) with SYBR® Green Realtime PCR Master Mix (Toyobo) in triplicate. The PCR was 40 cycles of 10 s at 95 °C, 10 s at 60 °C and 25 s at 72 °C. The melting curve analysis was 0 s at 95 °C, 15 s at 65 °C, up to 98 °C, and cooled to 40 °C. The primer pairs used were following: 5′-AACGAAGACCTGAAAACGTG-3′ and 5′-CTGGGAGAGACAGATCAGAG-3′ for H-2 class I histocompatibility antigen K-Q alpha chain, 5′-TTACATCGCCCTGAACGAAG-3′ and 5′-AGCTCCAGTGACTATTGCAG-3′ for H-2 class I histocompatibility antigen K-W28 alpha chain, 5′-CCCTTACTGTGCCTGGAATG-3′ and 5′-TTGGCACCCTTTTCTAGCTC-3′ for semaphorin-4C. The results were shown by relative expression level, which was a multiple of each gene expression level in normal brain. The PCR products at 20 or 30 cycles were also subjected to agarose gel electrophoresis and stained with ethidium bromide and confirmed for expression.

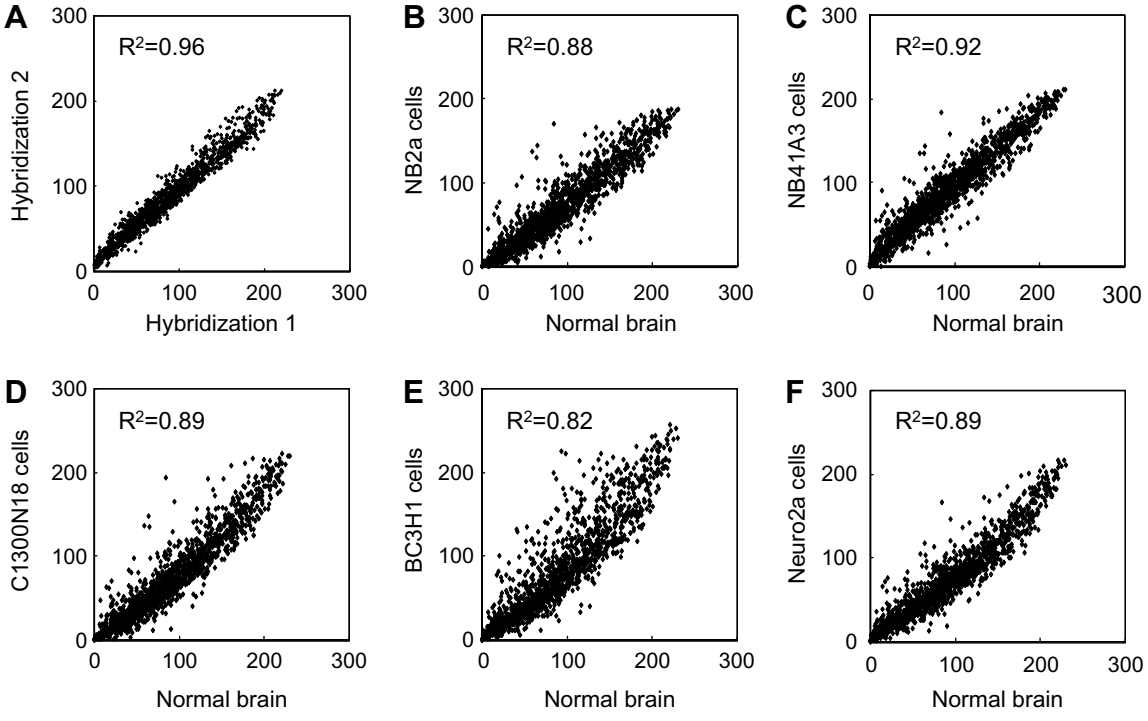

Fig. 1. Scatter plots of gene expression pattern. (A) Two independent hybridizations of normal brain are compared. (B–F) NB2a, NB41A3, C1300N18, BC3H1, and Neuro-2a cells vs. normal brain, respectively. Each R^2 denotes the correlation coefficient.