Journal of Clinical Neuroscience 21 (2014) 316–319

Contents lists available at ScienceDirect

Journal of Clinical Neuroscience

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

Laboratory studies

Protein–protein interaction network analysis and gene set enrichment analysis in epilepsy patients with brain cancer

Bin Kong [a,c], Tao Yang [a,c], Lin Chen [b], Yong-qin Kuang [a], Jian-wen Gu [a,*], Xun Xia [a], Lin Cheng [a], Jun-hai Zhang [a]

[a] Department of Neurosurgery, Chengdu Military General Hospital, 270 Rong Du Road, Chengdu 610083, Sichuan Province, China
[b] Department of Neurology, Chengdu Military General Hospital, Chengdu, Sichuan Province, China
[c] Third Military Medical University, Chongqing, China

ARTICLE INFO

Article history:
Received 27 January 2013
Accepted 15 June 2013

Keywords:
Differentially expressed genes
Gene set enrichment analysis
Protein-protein interaction network
Tumor-associated epilepsy

ABSTRACT

Many patients with brain cancer experience seizures or epilepsy and tumor-associated epilepsy (TAE) significantly decreases their quality of life. This study aimed to achieve a better understanding of the mechanisms of TAE. The differentially expressed genes (DEG) between epilepsy patients with or without brain tumor were firstly screened using the Linear Models for Microarray Data package using GSE4290 datasets from the USA National Center for Biotechnology Information Gene Expression Omnibus database. Then the protein-protein interaction (PPI) network, using data from the Human Protein Reference Database and the Biological General Repository for Interaction Datasets, was constructed. For further analysis, the PPI network structure and clusters in this PPI network were identified by ClusterOne. Meanwhile, gene set enrichment analysis was performed to illuminate the biological pathways and processes which generally affect patients with TAE. A total of 5113 DEG were identified and a PPI network, which contained 114 DEG and 21 normal genes, was established. Proteins, which mainly belonged to the mini chromosome maintenance and collagen families, were discovered to be enriched in the three identified clusters in the PPI network. Finally, several biological pathways (including cell cycle, DNA replication and transforming growth factor β1 signaling pathways) and processes (such as nucleocytoplasmic transport, nuclear transport and regulation of phosphorylation) were identified. Proteins in these three clusters may become new targets for TAE treatment. Our results provide some potential underlying biomarkers for understanding the pathogenesis of epilepsy in patients with brain tumor.

1. Introduction

A brain tumor is a life-threatening solid neoplasm within the brain or central spinal canal, usually with a poor prognosis [1]. Moreover, tumors can affect any part of the brain and induce many complicating symptoms [2]. Epileptic seizures are one of the most common symptoms of brain tumor. Epilepsy, which was first described in southern Turkey in 1951 [3], is one of the most common neurological disorders in humans and is characterized by seizures [4]. About 30% of patients with brain tumor experience seizures or epilepsy and this tumor-associated epilepsy (TAE) can severely affect quality of life, particularly in patients with slow-growing primary brain tumors [5]. Moreover, TAE is often refractory to antiepileptic drugs, accounting for the poor prognosis in patients with brain tumors [6]. Therefore, there is an urgent need to explore the underlying mechanism of TAE and identify TAE biomarkers.

TAE is a poorly understood manifestation of many gliomas, meningiomas and metastatic brain tumors, and has important clinical and social implications [7]. One of the most important pathogenetic features of TAE is changes in amino acid neurotransmission [5]. Aberrant expression of neurotransmitters and receptors is also important [8]. For example, gamma-amino butyric acid (GABA) is the main inhibitory neurotransmitter in the brain. Altered expression of the GABA receptor contributes to seizures in patients with brain tumor, and increased GABA activity suppresses epileptic activity [9]. Although several reports have been published about the underlying mechanism, studies undertaking screening and further analysis of differentially expressed genes (DEG) in TAE by bioinformatics methods remain rare.

In this study, to explore the underlying molecular mechanism of TAE and identify its biomarkers, DEG were screened and a protein–protein interaction (PPI) network was constructed. Furthermore, gene set enrichment analysis (GSEA) was performed to reveal the functions of the DEG in the network. This study provides a basis for researchers to understand the pathogenesis of TAE and identify novel treatments.

* Corresponding author. Tel.: +86 28 8657 0361.
E-mail address: jianwengujwg249@hotmail.com (J.-w. Gu).

0967-5868/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jocn.2013.06.026

2. Materials and methods

2.1. Data source

The transcription profile of GSE4290 [10] was obtained from the USA National Center for Biotechnology Information Gene Expression Omnibus (GEO) database [11], which was based on the Affymetrix Human Genome U133 Plus 2.0 Array. A total of 180 chips (purchased from the USA National Cancer Institute in Neurological Disorders and Stroke, Bethesda) [12] were used for the analysis.

Samples were collected from 180 epilepsy patients. Among them, 23 samples were from epilepsy patients without tumor, and the remaining 157 samples were collected from epilepsy patients with brain tumor (astrocytoma = 26, oligodendroglioma = 50 and glioblastoma = 81). All gliomas were pathologically standardized according to the World Health Organization classification and all 180 tumor and non-tumor brain tissue samples were obtained from surgical patients and frozen immediately after surgery. The clinical protocol was approved by the National Cancer Institute of China Institutional Review Board and informed consent was obtained from all patients. The frozen samples were disrupted and homogenized with TRIZOL (Invitrogen, Carlsbad, CA, USA) and the total RNA isolated from each sample was further purified using RNeasy Mini Kit (Qiagen, Valencia, CA, USA).

2.2. DEG analysis

The DEG between epilepsy patients with and without tumor were identified using the Linear Models for Microarray Data (LIMMA) package (http://svitsrv25.epfl.ch/R-doc/library/limma/html/00Index.html) [13], and then a linear model was constructed. Only the genes which met our criterion ($|\log fc| > 1$ and p value < 0.05) were selected as DEG in this study.

2.3. PPI network construction

The relationships between DEG were matched according to the PPI data collected from the Human Protein Reference Database [14] and the Biological General Repository for Interaction Datasets [15]. A Pearson correlation coefficient (PCC) [16] was calculated for these interaction relationships. Based on the significant relationships ($|\log PCC| > 0.75$), the PPI network was constructed using Cytoscape [17]. Cytoscape is a free software package for integrating biomolecular interaction networks with high-throughput expression data and other molecular states into a unified conceptual framework.

2.4. Cluster identification

ClusterONE [18] is designed to discover densely connected and possibly overlapping regions within the Cytoscape network. In PPI networks, dense regions usually correspond to protein complexes or fractions of them. ClusterONE works by "growing" dense regions out of small seeds guided by a quality function. The quality of a group is evaluated by the number of internal edges.

In our PPI network, sub-graphs with highly interconnected proteins were identified as protein complexes or functional modules. In addition, sub-graphs with fewer than six nodes or with a density (number of edges within the cluster divided by the number of theoretically possible edges) lower than five were discarded.

2.5. GSEA enrichment analysis

GSEA is a statistical method to determine if predefined sets of genes are differentially expressed in different phenotypes [19]. Predefined gene sets may be genes in a known metabolic pathway, located in the same cytogenetic band, sharing the same Gene Ontology category [20], or any user-defined set. In the present study, GSEA was performed on the identified clusters to reveal their functions in the pathogenesis of TAE.

3. Results

3.1. Microarray data analysis

The publicly available microarray data set GSE4290 was obtained from GEO. Based on the research criterion ($|\log fc| > 1$ and p value < 0.05), a total of 5113 DEG were identified between the samples collected from the epilepsy patients with and without brain tumor using the LIMMA package.

3.2. PPI network

To obtain a PPI network with a minimum size of four, the relationship of proteins encoded by 114 DEG and 21 normal genes were matched, according to the PPI database. By integrating interaction relationships, a PPI network was constructed (Supp. Fig. 1).

3.3. Cluster

ClusterONE was used to identify the clusters in our PPI network and three clusters were identified with a minimum size of six and minimum density of 0.5. Most of the genes in these three clusters belonged to the mini chromosome maintenance (MCM) and collagen (COL) families (Fig. 1).

3.4. GSEA enrichment analysis

GSEA was performed on the DEG in the PPI network. It was discovered that genes differentially expressed in epilepsy patients with brain tumor were significantly enriched in the several biological processes, including nucleocytoplasmic transport, nuclear transport and regulation of phosphorylation with normalize enrichment score (NES) > 1.6, p value < 0.01, and a false discovery rate (FDR) < 0.4 (Table 1). In addition, the significantly enriched biological pathways, including cell cycle (NES = 1.59, p value = 0.039, FDR = 0.256), a pathway which can regulate cell cycle and DNA replication, and transforming growth factor β1 (TGF β1) signaling pathways with NES = 1.61, p value = 0.0012, and FDR = 0.357, were also analyzed by GSEA (Table 2).

4. Discussion

Although not all epileptic symptoms can be explained by tumor-related factors, the complex relationship between tumor and epilepsy has been confirmed by previous studies. In order to provide clues for discovering the underlying mechanism of TAE, DEG between the samples collected from the epilepsy patients with and without brain tumor were identified. In addition, a PPI network was constructed, and then three clusters were identified in this PPI network. Consequent GSEA analysis revealed that genes which were differentially expressed in patients with TAE were enriched in several biological pathways including cell cycle, DNA replication and the TGF β1 signaling pathway.

In Clusters A and C, most proteins belonged to the MCM family (Fig. 1). The MCM proteins are essential for DNA replication initiation and elongation in eukaryotes [21]. In this network, they were predicted to interact not only with each other, but also with Dumbbell former 4 (DBF4). As the regulatory subunit of the DBF4-dependent kinase complex, DBF4 can also regulate the activity of