Research Article

Gene Expression Profile Analysis in Epilepsy by Using the Partial Least Squares Method

Dong Wang, Xixiao Song, Yan Wang, Xia Li, Shanshan Jia, and Zhijing Wang

Department of Neurology, Xi'an Children's Hospital, Xi'an, Shaanxi 710003, China

Correspondence should be addressed to Dong Wang; wangdong6928@163.com

Purpose. Epilepsy is a common chronic neurological disorder. We aim to investigate the underlying mechanism of epilepsy with partial least squares- (PLS-) based gene expression analysis, which is more sensitive than routine variance/regression analysis. *Methods*. Two microarray data sets were downloaded from the Gene Expression Omnibus (GEO) database. PLS analysis was used to identify differentially expressed genes. Gene ontology and network analysis were also implemented. *Results*. A total of 752 genes were identified to be differentially expressed, including 575 depressed and 177 overexpressed genes in patients. For GO enrichment analysis, except for processes related to the nervous system, we also identified overrepresentation of dysregulated genes in angiogenesis. Network analysis revealed two hub genes, *CUL3* and *EP300*, which may serve as potential targets in further therapeutic studies. *Conclusion*. Our results here may provide new understanding for the underlying mechanisms of epilepsy pathogenesis and will offer potential targets for producing new treatments.

1. Introduction

Epilepsy is a common chronic neurological disorder, which has devastating effects on patients and their families. There are about 50 million epilepsy patients worldwide and the occurrence in developing countries is more than twice that in developed countries [1]. Currently there are more than 20 antiepileptic drugs available for epilepsy patients [2]; however, multidirectional interactions between seizures and the medications are still challenging for treating patients [3]. Exploring the biological alterations of patients may provide insights into the pathology and new targets for treatments.

Large-scale microarray expression strategy has provided greater ease for investigating the underlying mechanisms of epilepsy. Several gene expression profiling studies have been carried out earlier and most of them used the routine variance/regression analysis [4–6]. However, this procedure cannot remove unaccounted array specific factors, such as certain demographic profiles. Compared with the routine analysis, previous studies [7, 8] proposed that partial least squares- (PLS-) based analysis is more robust in proceeding

gene expression profile data with higher sensitivity. Therefore, using PLS analysis may provide new understanding of the pathogenesis of epilepsy.

In the current study, to identify truly differentially expressed genes between epilepsy patients and normal controls, we carried out a PLS analysis with two combined data sets from the Gene Expression Omnibus (GEO) database. Gene ontology (GO) enrichment analysis was also carried out for the selected genes to capture the biological relevant signatures. A network constructed by proteins encoded by dysregulated genes was used to identify key molecules among the differentially expressed genes. Our results here may provide new understanding on the pathogenesis of epilepsy.

2. Materials and Methods

2.1. Microarray Data. Two data sets (GSE4290 and GSE50161) from the GEO (http://www.ncbi.nlm.nih.gov/geo/) database, which include 23 epilepsy patients and 13 healthy controls, were used in this study.

TABLE 1: Characteristics of the expression profile used in this study.

Accession ID	Description
GSM97800	Brain tissue from epilepsy patient
GSM97803	Brain tissue from epilepsy patient
GSM97804	Brain tissue from epilepsy patient
GSM97805	Brain tissue from epilepsy patient
GSM97807	Brain tissue from epilepsy patient
GSM97809	Brain tissue from epilepsy patient
GSM97811	Brain tissue from epilepsy patient
GSM97812	Brain tissue from epilepsy patient
GSM97816	Brain tissue from epilepsy patient
GSM97817	Brain tissue from epilepsy patient
GSM97820	Brain tissue from epilepsy patient
GSM97825	Brain tissue from epilepsy patient
GSM97827	Brain tissue from epilepsy patient
GSM97828	Brain tissue from epilepsy patient
GSM97833	Brain tissue from epilepsy patient
GSM97834	Brain tissue from epilepsy patient
GSM97840	Brain tissue from epilepsy patient
GSM97846	Brain tissue from epilepsy patient
GSM97848	Brain tissue from epilepsy patient
GSM97849	Brain tissue from epilepsy patient
GSM97850	Brain tissue from epilepsy patient
GSM97853	Brain tissue from epilepsy patient
GSM97855	Brain tissue from epilepsy patient
GSM1214936	Brain tissue from normal control
GSM1214937	Brain tissue from normal control
GSM1214938	Brain tissue from normal control
GSM1214939	Brain tissue from normal control
GSM1214940	Brain tissue from normal control
GSM1214941	Brain tissue from normal control
GSM1214942	Brain tissue from normal control
GSM1214943	Brain tissue from normal control
GSM1214944	Brain tissue from normal control
GSM1214945	Brain tissue from normal control
GSM1214946	Brain tissue from normal control
GSM1214947	Brain tissue from normal control
GSM1214948	Brain tissue from normal control

The two data sets were both based on the GPL570platform Affymetrix Human Genome U133 Plus 2.0 Array. Detailed information of the samples is listed in Table 1.

2.2. Detection of Differentially Expressed Genes. Normalization of raw intensity values was carried out with robust multiarray analysis (RMA) [9]. The resulting log2-transformed expression values of all probes were used for further PLS analysis [10, 11], which is a dimension reduction method for modeling without imposing strong assumptions, to estimate the effects for each probe in epilepsy patients. Briefly, NIPALS algorithm [12] was firstly used to obtain PLS latent variables derived from the expression profile; variable importance on the projection (VIP) [13] was then calculated to estimate

the effect of the expressed probes on the disease status of the patients. Finally, the empirical distribution of PLS-based VIP was obtained with a permutation procedure (N = 10000 times) and false discovered rate (FDR) of each probe was calculated based on the empirical distribution. Probes with FDR value less than 0.05 were selected as differentially expressed genes in this study.

2.3. Enrichment Analysis. Identified differentially expressed probes were annotated by using the simple omnibus format in text (SOFT) format files. All genes were then mapped to the Gene Ontology database [14], which provides a controlled vocabulary of terms for describing gene product characteristics. Hyper geometric distribution test was carried out to identify GO items enriched with differentially expressed genes.

2.4. Network Analysis. Most proteins function through interactions with other proteins. Proteins with more interactions with other proteins are supposed to play more important roles in biological processes. To identify key molecules among the differentially expressed genes, we constructed an interaction network with the proteins encoded by selected genes by using the software Cytoscape (V 2.8.3, http://www.cytoscape.org/) [15]. Interaction information of the proteins was obtained from the NCBI database (http://ftp.ncbi.nlm.nih.gov/gene/GeneRIF/). The number of links (interactions) for each protein was defined as its degree. Proteins with degrees more than 10 were selected as hub molecules in this study.

3. Results

After quality control, two samples (GSM1214938 and GSM1214939) were excluded from subsequent analysis due to aberrant RNA degradation. Thus, 23 epilepsy patients and 11 healthy controls were used in PLS analysis. Sample classification according to the three selected latent variables is illustrated in Figure 1. After FDR control, a total of 752 genes were identified to be differentially expressed, including 575 depressed and 177 overexpressed genes in patients. The top ten GO items enriched with differentially expressed genes are listed in Table 2. Most of them (60%) are related to the nervous system, including nervous system development (GO:0007399), central nervous system myelin maintenance (GO:0007399), neuroligin clustering (GO:0007399), synapse assembly (GO:0007416), spinal cord motor neuron differentiation (GO:0021522), and glial cell development (GO:0021782).

Figure 2 represents the interaction network of proteins encoded by selected genes. Two proteins, CUL3 and EP300, were identified to be hub molecules, with degrees of 56 and 21, respectively.

4. Discussion

Pathophysiology of epilepsy is highly complex. Gene expression profiling is useful in investigating the underlying mechanism of epilepsy. For the data analysis, creating a suitable

TABLE 2: The top 10 GO items enriched with differentially expressed genes.

| GO_id	GO_description	GO_class	P value
GO:0007155	Cell adhesion	Process	$3.74E-05$
GO:0072659	Protein localization to plasma membrane	Process	$4.75E-05$
GO:0001525	Angiogenesis	Process	$2.64E-04$
GO:0007399	Nervous system development	Process	$9.26E-04$
GO:0032286	Central nervous system myelin maintenance	Process	$1.43E-03$
GO:0097118	Neuroligin clustering	Process	$1.43E-03$
GO:0007416	Synapse assembly	Process	$1.59E-03$
GO:0021522	Spinal cord motor neuron differentiation	Process	$2.57E-03$
GO:0005509	Calcium ion binding	Function	$4.01E-03$
GO:0021782	Glial cell development	Process	$4.18E-03$
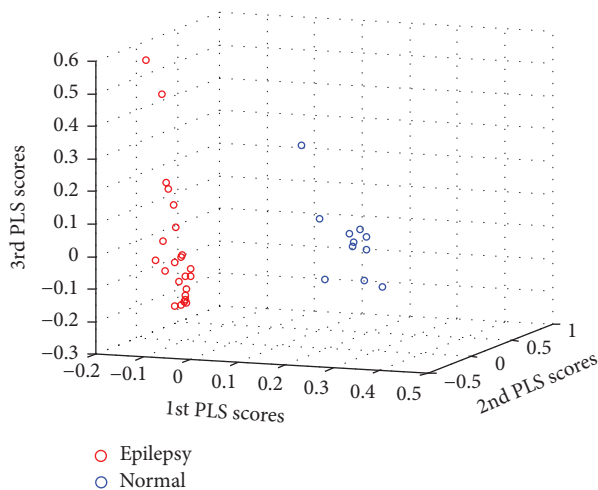
FIGURE 1: Sample classification by using the three selected partial least squares (PLS) latent variables.

model to handle small sample sizes and large number of genes [7] remains challenging. Previous studies [7, 8] have demonstrated better performance of the PLS-based method than common variance/regression analysis, which cannot remove hidden biological effects. Here we used PLS analysis to identify differentially expressed genes between epilepsy patients and healthy controls.

As shown in Figure 1, the selected three latent variables performed well in classification of the samples. GO enrichment analysis of the selected genes revealed the overrepresentation of differentially expressed genes in the nervous system. This is consistent with previous studies. For example, glial cells have been reported to play prominent roles in seizure precipitation and recurrence [16], and the glial cell development was identified to be enriched with dysregulated genes in our study ($P = 4.18 \times 10^{-3}$). In addition, angiogenesis (GO:0001525) was also found to be overrepresented with dysregulated genes. Dysregulation of angiogenesis may be related to the dysfunction of blood-brain barrier, contributing to epileptogenesis [17]. Signs of angiogenesis were also reported to be corresponding with seizure-induced neuronal death in animal models of familial epilepsy [18]. Our results here further confirmed the involvement of angiogenesis process in the pathogenesis of epilepsy.

According to the network analysis, CUL3 was identified to be a hub molecule with the highest degree (Figure 2). CUL3 is a core component and scaffold protein of an E3 ubiquitin ligase complex. Previous expression studies have not reported the differential expression of CUL3 in epilepsy patients. However, E3 ubiquitin ligase may affect the synaptic functions in the central nervous system and the stability of kainate receptors, which form a class of glutamate receptors implicated in epilepsy [19, 20]. Our results suggested that CUL3 may serve as a potential target in therapeutic studies.

EP300 is another hub gene with the degree of 21. No report of this gene and epilepsy has been proposed before. However, protein encoded by this gene is a transcriptional coactivator, which stimulates CREB-dependent gene expression. Seizure disorder was reported to be more frequent in Rubinstein-Taybi syndrome patients with CREBBP mutations [21]. In addition, the promoted signaling mechanism of EP300 is important in the neuronal survival process [22] and this gene was related to other neuronal disorders, such as familial Alzheimer's disease [22]. Thus, the correlation of this gene and epilepsy pathogenesis may involve the activity of CREB-dependent proteins and further investigation is warranted.

In summary, using two data sets from the GEO database, we carried out PLS-based gene expression analysis to investigate the underlying pathology of epilepsy. Except for processes related to the nervous system, we also identified overrepresentation of dysregulated genes in angiogenesis. Network analysis revealed two hub genes, *CUL3* and *EP300*, which may serve as potential targets in further therapeutic studies. Our results here may provide new understanding for the underlying mechanisms of epilepsy pathogenesis and will offer potential targets for producing new treatments.

Conflict of Interests

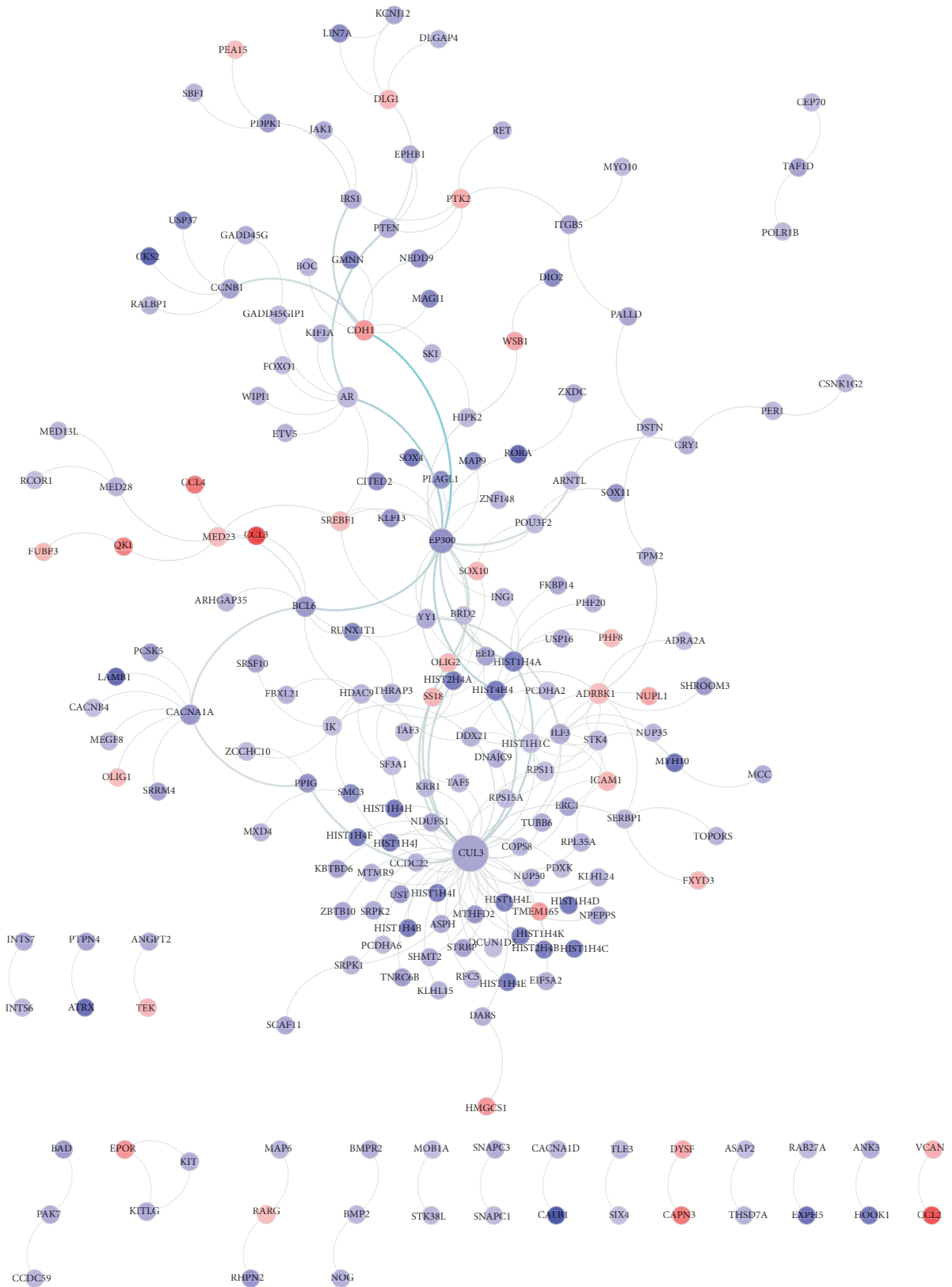The authors have no financial conflict of interests.

Figure 2: Interaction network constructed with proteins encoded by selected genes. Proteins with more links are shown in bigger size. Proteins shown in red are encoded by overexpressed genes in patients while those in blue are encoded by depressed genes in patients.

Acknowledgment

References

[1] C. R. Newton and H. H. Garcia, "Epilepsy in poor regions of the world," *The Lancet*, vol. 380, no. 9848, pp. 1193–1201, 2012.

[2] E. Perucca and T. Tomson, "The pharmacological treatment of epilepsy in adults," *The Lancet Neurology*, vol. 10, no. 5, pp. 446–456, 2011.

[3] C. L. Harden and P. B. Pennell, "Neuroendocrine considerations in the treatment of men and women with epilepsy," *The Lancet Neurology*, vol. 12, no. 1, pp. 72–83, 2013.

[4] J. Lachos, M. Zattoni, H. G. Wieser et al., "Characterization of the gene expression profile of human hippocampus in mesial temporal lobe epilepsy with hippocampal sclerosis," *Epilepsy Research and Treatment*, vol. 2011, Article ID 758407, 11 pages, 2011.

[5] S.-K. Kim, K.-C. Wang, S. J. Hong et al., "Gene expression profile analyses of cortical dysplasia by cDNA arrays," *Epilepsy Research*, vol. 56, no. 2-3, pp. 175–183, 2003.

[6] Y. Tang, T. A. Glauser, D. L. Gilbert et al., "Valproic acid blood genomic expression patterns in children with epilepsy—a pilot study," *Acta Neurologica Scandinavica*, vol. 109, no. 3, pp. 159–168, 2004.

[7] T. R. Golub, D. K. Slonim, P. Tamayo et al., "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, no. 5439, pp. 531–537, 1999.

[8] S. Chakraborty, S. Datta, and S. Datta, "Surrogate variable analysis using partial least squares (SVA-PLS) in gene expression studies," *Bioinformatics*, vol. 28, no. 6, pp. 799–806, 2012.

[9] R. A. Irizarry, B. Hobbs, F. Collin et al., "Exploration, normalization, and summaries of high density oligonucleotide array probe level data," *Biostatistics*, vol. 4, no. 2, pp. 249–264, 2003.

[10] I. S. Helland, "On the structure of partial least squares regression," *Communications in Statistics*, vol. 17, no. 2, pp. 581–607, 1988.

[11] I. S. Helland, "Partial least squares regression and statistical models," *Scandinavian Journal of Statistics*, vol. 17, no. 2, pp. 97–114, 1990.

[12] J. P. A. Martins, R. F. Teófilo, and M. M. C. Ferreira, "Computational performance and cross-validation error precision of five PLS algorithms using designed and real data sets," *Journal of Chemometrics*, vol. 24, no. 6, pp. 320–332, 2010.

[13] R. Gosselin, D. Rodrigue, and C. Duchesne, "A Bootstrap-VIP approach for selecting wavelength intervals in spectral imaging applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 100, no. 1, pp. 12–21, 2010.

[14] M. Ashburner, A. Ball, J. A. Blake et al., "Gene ontology: tool for the unification of biology. The gene ontology consortium," *Nature Genetics*, vol. 25, no. 1, pp. 25–29, 2000.

[15] P. Shannon, A. Markiel, O. Ozier et al., "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.

[16] A. Vezzani, E. Aronica, A. Mazarati et al., "Epilepsy and brain inflammation," *Experimental Neurology*, vol. 244, pp. 11–21, 2013.

[17] M. Morin-Brureau, V. Rigau, and M. Lerner-Natoli, "Why and how to target angiogenesis in focal epilepsies," *Epilepsia*, vol. 53, supplement 6, pp. 64–68, 2012.

[18] M. Sakurai, T. Morita, T. Takeuchi et al., "Relationship of angiogenesis and microglial activation to seizure-induced neuronal death in the cerebral cortex of Shetland Sheepdogs with familial epilepsy," *American Journal of Veterinary Research*, vol. 74, no. 5, pp. 763–770, 2013.

[19] J. Marshall, L. A. C. Blair, and J. D. Singer, "BTB-kelch proteins and ubiquitination of kainate receptors," *Advances in Experimental Medicine and Biology*, vol. 717, pp. 115–125, 2011.

[20] G. D. Salinas, L. A. C. Blair, L. A. Needleman et al., "Actinfilin is a Cul3 substrate adaptor, linking GluR6 kainate receptor subunits to the ubiquitin-proteasome pathway," *The Journal of Biological Chemistry*, vol. 281, no. 52, pp. 40164–40173, 2006.

[21] E. K. Schorry, M. Keddache, N. Lanphear et al., "Genotype-phenotype correlations in Rubinstein-Taybi syndrome," *American Journal of Medical Genetics A*, vol. 146, no. 19, pp. 2512–2519, 2008.

[22] K. Yukawa, T. Tanaka, S. Tsuji, and S. Akira, "Regulation of transcription factor C/ATF by the cAMP signal activation in hippocampal neurons, and molecular interaction of C/ATF with signal integrator CBP/p300," *Molecular Brain Research*, vol. 69, no. 1, pp. 124–134, 1999.