European Review for Medical and Pharmacological Sciences 2014; 18: 2306-2310

Partial least squares based gene expression analysis in posttraumatic stress disorder

K. DONG, F. ZHANG, W. ZHU, Z. WANG, G. WANG

Department of Psychology, Wuxi Mental Health Center of Nanjing Medical University, Wuxi, China

Abstract. – OBJECTIVE: Posttraumatic stress disorder (PTSD) is an adverse psychological response to traumatic events. Microarray technology for large-scale gene expression analysis facilitates the identification of signatures that underlie the pathogenesis of PTSD. Previous studies mostly used variance/regression analysis without considering array specific factors. We aim to investigate the underlying mechanism of PTSD through partial least squares (PLS) based analysis.

MATERIALS AND METHODS: With a gene expression profile data set for 17 chronic PTSD patients and 16 controls recovered from psychological trauma from the Gene Expression Omnibus (GEO) database, we performed Partial Least Square (PLS) based analysis.

RESULTS: We acquired 230 down-regulated genes and 335 up-regulated genes. Significantly increased representations of dysregulated genes in immune, endocrine and nervous pathways were identified. Among the top 5 hub genes in the network, PRKCA has been reported to be related with PTSD before. Three other genes, TP53, EP300 and CALM1 might also contribute to the pathogenesis of PTSD since they are all related with other neuronal disorders.

CONCLUSIONS: Our findings shed light on expression signatures of PTSD with the hope to give further theoretical supports for future therapeutic study.

Key Words:

Posttraumatic stress disorder, Gene expression, Partial least squares, Pathway, Network.

Introduction

Posttraumatic stress disorder (PTSD) is an adverse psychological response to traumatic events, such as sexual assault, serious injury or the threat of death[1]. As a common mental disorder, the lifetime prevalence of PTSD is various across cultures[2]. Typically, the female-to-male ratio is 2:1[3], and the lifetime prevalence of PTSD in general population was about 6.8%[4]. Since 1980, when PTSD was firstly proposed by the American Psychiatric Association, a great deal of efforts has been devoted to explore the molecular mechanism of PTSD. Biological alterations, which may be associated with differential gene transcription, have been suggested to be related with the pathogenesis of PTSD[5,6]. Advances in microarray technology for large-scale gene expression analysis facilitate the identification of signatures that underlie the pathogenesis of PTSD. Several studies have investigated gene expression from PTSD patients using microarray analysis[7-9]. These studies implemented the standard variance/regression analysis, which cannot remove unaccounted array specific factors. For example, it may happen that some genes are identified to be induced or depressed as result of specific demographic profile. Previous study[10] proposed that partial least squares (PLS) based gene expression analysis was robust in selecting disease specific genes. Compared with variance/regression analysis, PLS-base analysis yielded higher sensitivity, reasonably high specificity and impressively small false discovery rate and false non-discovery rate. Characterize the gene expression signatures of PTSD using PLS based method may conduce to further understanding of pathogenesis and the development of novel preventing or therapeutic procedure. In this study, with an expression profile downloaded from the gene expression omnibus (GEO) database, we investigated the pathological mechanism in PTSD using PLS-based analysis. Corresponding pathways with significantly increased dysregulated genes were also acquired. Moreover, protein-protein interaction (PPI) network analysis was carried out to identify key genes among the differentially expressed genes.

Materials and Methods

Microarray Data

The gene expression array data (GSE860) was downloaded from the GEO (http://www.ncbi.nlm.nih.gov/geo/) database.

Corresponding Author: Guoqiang Wang, MD; e-mail: kunlundong2013@gmail.com

This series represents transcription profile of 17 chronic PTSD patients and 16 controls recovered from psychological trauma. All samples were taken from peripheral mononuclear blood cells of survivors of psychological trauma. The data set was based on platform GPL91: [HG_U95A] Affymetrix Human Genome U95A Array.

Identification of Differentially Expressed Genes (DEGs)

Raw data in CEL format which were generated from satisfactory image files were obtained from the GEO database. Raw intensity values normalization was performed by Robust Multi-array Analysis (RMA)[11]. The resulting log2-transformed expression value for each probe was used in subsequent analysis. Briefly, PLS was used to estimate the effects of each probe in PTSD patients and healthy controls. Firstly, PLS latent variables derived from expression profile were calculated by using the non-linear iterative partial least squares (NIPALS) algorithm[12]; Secondly, the variable importance in the projection (VIP)[13] was calculated to evaluate the importance of the expressed genes on the status of the subjects. Finally, the empirical distribution of PLS-based VIP was obtained with a permutation procedure (n=10000 times). False discovered rate (FDR) of each probe was then calculated based on the empirical distribution. In this study, probes with FDR less than 0.05 were considered as significantly differentially expressed.

Pathway Analysis

Data in simple omnibus format in text (SOFT) format were also downloaded from the GEO database. All probes were annotated according to the SOFT format file. All genes were annotated based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database[14]. Enrichment analysis was then carried out by using hypergeometric distribution test to identify pathways enriched with dysregulated genes.

Network Analysis

We used Cytoscape (V 2.8.3, available on http://www.cytoscape.org/)[15] and the National Center for Biotechnology Information (NCBI) database (http://ftp.ncbi.nlm.nih.gov/gene/GeneRIF/, 2013-2-25) to construct the interaction network for differentially expressed genes.

Results

Compared with normal controls, a total of 565 genes were differentially expressed, including 230 down-regulated genes and 335 up-regulated genes. For all well-characterized genes in the array, 4359 genes can be mapped to various pathways while 286 differentially expressed genes were mapped to KEGG pathways. The top ten pathways enriched with dysregulated genes are listed in Table I. We identified pathways involved in immune response, endocrine system, and signal transduction. The Glioma pathway, which is a type of cancer in the nervous system, was also identified to be enriched with differentially expressed genes. Significantly increased representation of genes involved in the dopaminergic synapse pathway was also found.

Most proteins function with other proteins, thus, protein-protein interaction (PPI) is critical for all biological processes[16], which should be considered as complex PPI networks constituted with interconnected proteins. The known interactions between proteins encoded by differentially expressed genes were visualized using Cytoscape

Table I. Pathways enriched with differentially expressed genes.

KEGG id	Pathway description	Pathway class	*p*-value
05164	Influenza A	Infectious diseases	2.83E-05
04940	Type I diabetes mellitus	Endocrine and metabolic diseases	1.63E-04
04672	Intestinal immune network for IgA production	Immune system	1.83E-04
05214	Glioma	Cancers	3.87E-04
04728	Dopaminergic synapse	Nervous system	8.85E-04
04961	Endocrine and other factor-regulated calcium reabsorption	Excretory system	1.08E-03
04912	GnRH signaling pathway	Endocrine system	1.08E-03
04151	PI3K-Akt signaling pathway	Signal transduction	1.47E-03
04914	Progesterone-mediated oocyte maturation	Endocrine system	1.76E-03
5169	Epstein-Barr virus infection	Infectious diseases	1.85E-03

(V 2.8.3)[15]. Degree is a measure that is used to identify critical molecules called 'hubs'. The degree of one node is defined as its number of interactions. We defined the top 5 nodes with the highest degrees as hubs (Table II). All the hub genes except *EWSR1* were known to be associated with neuronal disorders.

Discussion

Pathophysiology of PTSD is highly complex. Gene expression profiling has offered great ease for investigating the underlying pathophysiological cascades in PTSD. However, using a suitable model to handle large number of genes and relatively small sample size is challenging. Previous gene expression studies for PTSD mostly used variance/regression analysis, which cannot remove hidden biological effects. Here, we used a PLS based model to identify differentially expressed genes in PTSD patients and healthy controls. Subsequent pathway and network analysis was also used to explore underlying mechanism.

The pathway analysis results were generally consistent with previous findings. Over-representation of differentially expressed genes was found in the immune system. Dysregulation of immune genes in PTSD has been reported in previous studies[7], and alterations in the immune system were suggested to be the characteristic of the disorder[17]. Identification of increased presentation of the endocrine system is also expected since over-reactive adrenaline response is a common clinical manifestation of PTSD. In addition, the dopaminergic synapse pathway was found with significantly increased dysregulated genes. Dopamine functions as a neurotransmitter and it has been considered as a potential target for the treatment of PTSD[18].

Network analysis revealed that *TP53* was a hub gene with the highest degree (Figure 1, Table II). No previous report has proposed the association of *TP53* and PTSD. However, TP53 protein expression was increased in Alzheimer's disease patients[15]. It was suggested that *TP53* may be implicated in the pathological mechanism of Alzheimer's disease by repair of oxidative DNA damage and/or apoptosis[19]. Implication of TP53 in neuronal disorders suggested that TP53 may be involved in PTSD through similar biological process.

PRKCA was also identified as a hub gene with the second highest degree (Figure 1, Table II). Previous study[20] has indicated a genetic link between this gene and PTSD since genetic variability of the gene was associated with memory capacity. Our investigation here confirmed the relationship between *PRKCA* and PTSD.

EP300 encodes a transcriptional co-activator protein, which stimulates CREB-dependent gene expression by integrating various signal-responsive transcription factors. The promoted signaling mechanism of EP300 is important in neuronal survival[21] and it has also been related with familial Alzheimer's disease[21]. Therefore, the association of EP300 and PTSD warrants further investigation. Another hub gene might be related with PTSD is *CALM1*. Protein encoded by this gene is a member of the calcium-binding protein family. No previous report of the relationship between PTSD and *CALM1* has been proposed. However, according to KEGG database, it is involved in several nervous pathways, including the Neurotrophin signaling pathway, Dopaminergic synapse, Alzheimer's disease, Glioma and Amphetamine addiction.

Conclusions

With gene expression profile data from the GEO database, we carried out a PLS based analysis to identify differentially expressed genes which may contribute to the pathology of PTSD patients. Further analysis was also implemented to identify pathways and hub genes related with the disease. Our results facilitate the disclosure of the molecular mechanism underlying PTSD.

Table I. The top 5 hub molecules in the network constructed by differentially expressed genes.

Gene	Implications in PTSD or other neuronal disorders (PubMed ID)	Degree
TP53	Alzheimer's disease (22503900)	20
PRKCA	PTSD (22586106)	13
EP300	Alzheimer's disease (17197080); neuronal survival (10350644)	12
EWSR1	No known association with neuronal disorders.	10
CALM1	Drug addiction (17205118)	8

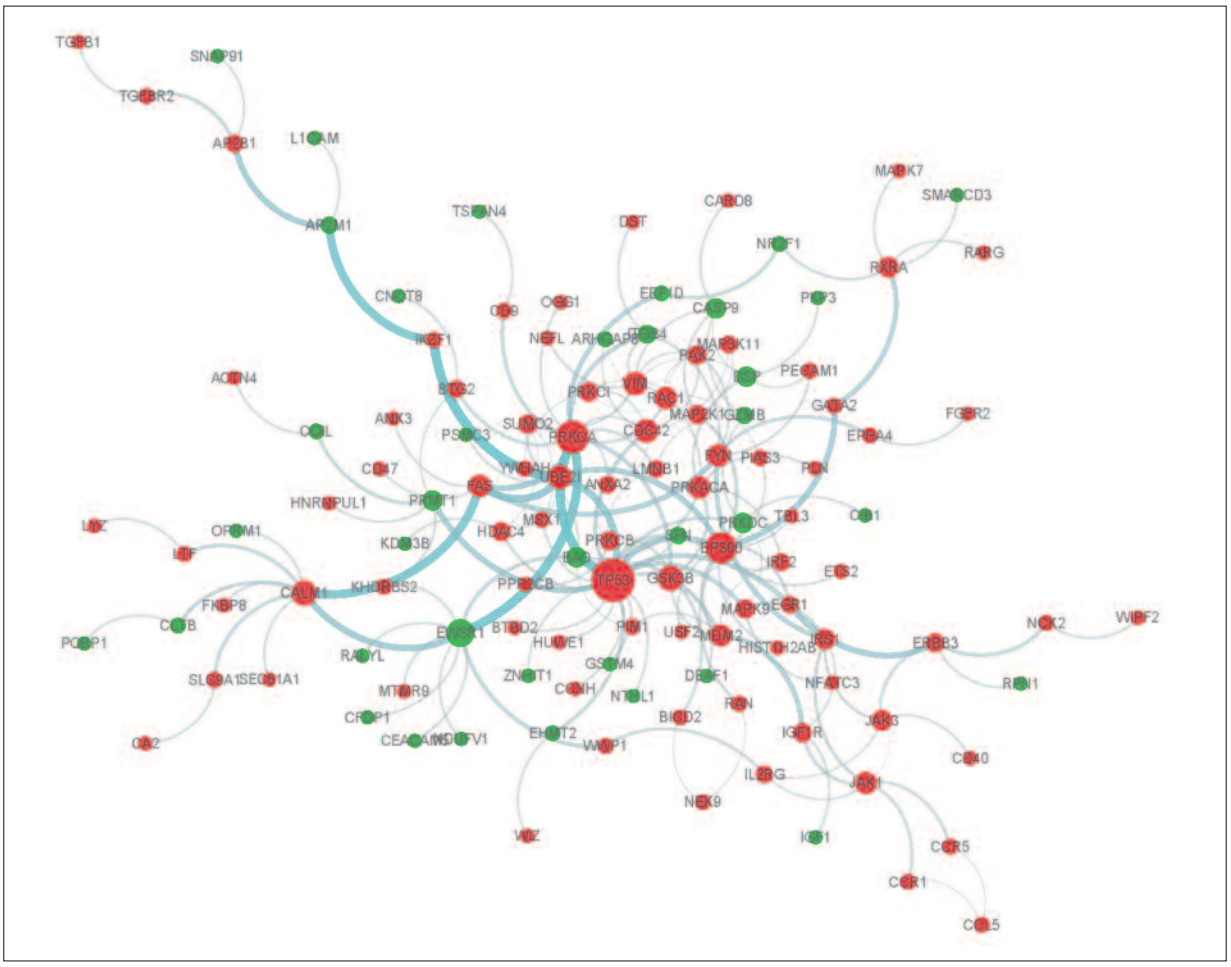

Figure 1. Interaction network constructed by differentially expressed genes. Only genes with more than two direct or indirect links were shown. Genes or with higher degree (*more links*) are shown in bigger size. Genes shown in red are overexpressed while genes in green are down regulated.

Conflict of Interest

The Authors declare that there are no conflicts of interest.

References

1) Association AP. Diagnostic and Statistical Manual of Mental Disorders (Fifth ed.). American Psychiatric Association, Arlington, 2013.
2) Keane TM, Marshall AD, Taft CT. Posttraumatic stress disorder: etiology, epidemiology, and treatment outcome. Annu Rev Clin Psychol 2006; 2: 161-197.
3) Breslau N. The epidemiology of posttraumatic stress disorder: what is the extent of the problem? J Clin Psychiatry 2001; 62 Suppl 17: 16-22.
4) Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005; 62: 593-602.
5) Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB. Posttraumatic stress disorder in the National Comorbidity Survey. Arch Gen Psychiatry 1995; 52: 1048-1060.
6) Yehuda R. Post-traumatic stress disorder. N Engl J Med 2002; 346: 108-114.
7) Segman RH, Shefi N, Goltser-Dubner T, Friedman N, Kaminski N, Shalev AY. Peripheral blood mononuclear cell gene expression profiles identify emergent post-traumatic stress disorder among trauma survivors. Mol Psychiatry 2005; 10: 500-513, 425.
8) Zieker J, Zieker D, Jatzko A, Dietzsch J, Nieselt K, Schmitt A, Bertsch T, Fassbender K, Spanagel R, Northoff H, Gebicke-Haerter PJ. Differential gene expression in peripheral blood of patients suffering from post-traumatic stress disorder. Mol Psychiatry 2007; 12: 116-118.
9) Yehuda R, Cai G, Golier JA, Sarapas C, Galea S, Ising M, Rein T, Schmeidler J, Muller-Myhsok B, Holsboer F, Buxbaum JD. Gene expression pat-

terns associated with posttraumatic stress disorder following exposure to the World Trade Center attacks. Biol Psychiatry 2009; 66: 708-711.

10) Chakraborty S, Datta S, Datta S. Surrogate variable analysis using partial least squares (SVA-PLS) in gene expression studies. Bioinformatics 2012; 28: 799-806.

11) Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003; 4: 249-264.

12) Martins JPA, Teofilo RF, Ferreira MMC. Computational performance and cross-validation error precision of five PLS algorithms using designed and real data sets. J Chemometrics 2010; 24: 320-332.

13) Gosselin R, Rodrigue D, Duchesne C. A Bootstrap-VIP approach for selecting wavelength intervals in spectral imaging applications. Chemometrics and Intelligent Laboratory Systems 2010; 100: 12-21.

14) Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 2012; 40: D109-114.

15) Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498-2504.

16) Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE. A human protein-protein interaction network: a resource for annotating the proteome. Cell 2005; 122: 957-968.

17) Uddin M, Aiello AE, Wildman DE, Koenen KC, Pawelec G, de Los Santos R, Goldmann E, Galea S. Epigenetic and immune function profiles associated with posttraumatic stress disorder. Proc Natl Acad Sci U S A 2010; 107: 9470-9475.

18) Dunlop BW, Mansson E, Gerardi M. Pharmacological innovations for posttraumatic stress disorder and medication-enhanced psychotherapy. Curr Pharm Des 2012; 18: 5645-5658.

19) Dezor M, Dorszewska J, Florczak J, Kempisty B, Jaroszewska-Kolecka J, Rozycka A, Polrolniczak A, Bugaj R, Jagodzinski PP, Kozubski W. Expression of 8-oxoguanine DNA glycosylase 1 (OGG1) and the level of p53 and TNF-alphalpha proteins in peripheral lymphocytes of patients with Alzheimer's disease. Folia Neuropathol 2011; 49: 123-131.

20) de Quervain DJ, Kolassa IT, Ackermann S, Aerni A, Boesiger P, Demougin P, Elbert T, Ertl V, Gschwind L, Hadziselimovic N, Hanser E, Heck A, Hieber P, Huynh KD, Klarhofer M, Luechinger R, Rasch B, Scheffler K, Spalek K, Stippich C, Vogler C, Vukojevic V, Stetak A, Papassotiropoulos A. PKCalpha is genetically linked to memory capacity in healthy subjects and to risk for posttraumatic stress disorder in genocide survivors. Proc Natl Acad Sci U S A 2012; 109: 8746-8751.

21) Yukawa K, Tanaka T, Tsuji S, Akira S. 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. Brain Res Mol Brain Res 1999; 69: 124-134.