Biomarkers screening and related functions identification for spinal cord injury based on guilt-by-association.

Chun Shan Zheng¹, Hui Qin Sun², Xiao Mei Dong³, Ping Li¹, Fuqiang Wang⁴*

¹Department of Traumatic Orthopedics, Shouguang People Hospital, No. 1233, Jiankang Street, Shouguang, Weifang, Shandong, PR China
²Drug Dispensing Center, Shouguang People Hospital, No. 1233, Jiankang Street, Shouguang, Weifang, Shandong, PR China
³Department of Obstetrics, Shouguang People Hospital, No. 1233, Jiankang Street, Shouguang, Weifang, Shandong, PR China
⁴Department of Orthopedics, Lanshan Dist. People's Hospital of Linyi City Shandong Province, No.11, Kaopeng Street, Lanshan District, Linyi City, Shandong Province, PR China

Abstract

Objective: The aim was to research Spinal Cord Injury (SCI) related genes and functions for postoperative recovery and prevention.

Methods: E-GEOD-18179 was used to screen Differently Expressed Genes (DEGs) between SCI samples and controls. Adjacency matrix was constructed as gene co-expression network. Sub-network was identified based on spearman correlation coefficient. Assortativity coefficient was calculated to represent the relationship among nodes. Gene Ontology (GO) functional enrichment was processed and guilt-by-association was undertaken.

Results: Total 129 DEGs were obtained and Adjacency matrix with 129 × 129 genes was constructed. Thereinto, TYROBP and MIA were with weight of 0.9695. Total 129 genes were enriched in 127 GO terms, such as molecular function and catalytic activity. K terms and related AUC values were obtained. The guilt-by-association node degrees AUCs were generated. In addition, the optimal list genes were obtained, such as JUN and PTK2B. Finally, optimal GO terms were obtained, such as regulation of signal transduction, positive regulation of macromolecule metabolic process and cellular component organization.

Conclusion: Screened genes including JUN, PTK2B, CCND1 and ADAM8 might be potential genes for SCI postoperative recovery and complications and prevention.

Keywords: Spinal cord injury, Guilt-by-association, Differently expressed genes, Functions, Adjacency matrix.

Introduction

Spine is a group of nerves which is protected by the individual vertebrae of the spine [1]. The main function of spin cord is to send signals from the brain to other regions [2]. Spinal Cord Injury (SCI) can be classified into complete injury and incomplete injury, which was with different symptoms, such as pain, numbness, paralysis to incontinence and even quadriplegia [3]. Moreover, various complications may occur, including muscle atrophy, pressure sores, infections and respiratory problems.

In 2006, Schmitt et al. [4] found that different degree of recovery was showed in different strains of rats, and difference in functional recovery was closely related with gene expression after SCI. In Wistar rats with high-level SCI, α1-adrenergic receptors gene were a contributing factor for hypertension [5]. Moreover, EPH Receptor A4 (Epha4) was confirmed to be related with deprenyl, which was found to protect nervous system by regulating the expression of p75NTR and TrkB after SCI [6].

Though various genes were researched, recovery and complication were not well controlled. The changes of genes after spinal cord injury were complexly and closely interacted. Gene Ontology was an effective method to describe the function of genes. In addition, Gene Ontology (GO) function was introduced to gene expression network, which could functionally classify the network. In order to research SCI related genes and functions with integrality; guilt-by-association was used in this study to increase the statistical...
reliability of GO terms. It might lay a foundation for SCI treatment and complication prevention.

Materials and Methods

Microarray data acquiring and preprocessing
Raw microarray profile of E-GEOD-18179 including 32 spinal cord injury samples and 30 controls were obtained from Array Express database with the platform of A-AFFY-6-Affymetrix GeneChip Murine Genome U74Av2 (MG_U74Av2). Random Multiple Access (RMA) algorithm was processed for raw data preprocessing with 3 steps: background correction, replicated probes removing and standardization [7]. After mapping probes to genes, gene expression profile was obtained.

Differently expressed genes screening
The expression levels of genes in gene expression profile were calculated by limma package, and examined by T-test and Fisher exact test. LmFit function was used for linear fitting, expected Bayesian estimation (eBayes) was processed for statistics and FDR was corrected p value. The threshold of DEGs was $|\text{logFC}|>0$ and $p<0.05$.

Co-expression network of DEGs
Adjacency matrix was constructed as gene co-expression network. Spearman correlation coefficient between two genes was calculated. Based on the strength of relationships, adjacency matrix was constructed. Nodes of sub-network were identified with the threshold of spearman correlation coefficient $>0.8$ and degree $>1$.

Density of node degree
Based on the occurrence of degree, R package was used to show the density of node degree. Assortativity coefficient was a kind of Pearson correlation coefficient based on degree. The results of assortativity coefficient represented the relationship among these nodes. Positive value represented synergetic relationship of nodes with same degrees, while negative value represented a link between nodes with different degrees. Normally, the value of assortativity coefficient ranged from -1 to 1.1 represented assortativity pattern, 0 represented non-assortativity pattern and -1 represented negative relationship.

GO functional enrichment analysis
A total of 18421 GO terms with 18402 genes were downloaded from GO consortium. Screened DEGs were enriched in GO terms, and enriched GO terms with more than 20 genes were screened.

Inference analysis of network correlation
Gene i in gene co-expression network was selected and all genes except i were enriched in K terms. MF values of each gene i were calculated by following formula:

$$MF_i = \sum_{1|\text{sense}} \text{co}_K \frac{1}{\text{Num}_{\text{ink}} \times \text{Num}_{\text{outk}}}$$

$\text{Num}_{\text{ink}}$ represented numbers of neighbor genes in K terms and $\text{Num}_{\text{outk}}$ represented numbers of neighbor genes which did not exist in K terms.

Three times folding method was used to calculate AUC value of K terms, then average AUC was calculated and evaluated for each GO function.

Based on emergent frequency of genes, cool read was processed to evaluate these predicted functions. If a fine AUC was obtained, the GO term was a good candidate with differently expressed genes. Finally, the most excellent gene list was generated. Then, the most excellent gene list was used to predict GO terms and AUCs were obtained. GO terms with AUC more than 0.6 were screened.

Results

Acquisition of gene expression profile and DEGs screening
After mapping probes to genes, gene expression profile with 8713 genes was obtained. Based on the threshold of DEGs, a total of 129 DEGs were obtained, including Small Proline Rich Protein 1A (SPRR1A), Small Proline Rich Protein 1B (SPRR1B), Activating Transcription Factor 3 (ATF3), Desmoplakin (DSP) and Adenylate Cyclase Activating Polypeptide 1 (ADCYAP1).

Co-expression network of DEGs
Adjacency matrix with 129 × 129 genes was obtained. Based on the threshold of spearman correlation coefficient $>0.8$ and degree $>1$, a total of 84 correlations were obtained. There into, TYRO Protein Tyrosine Kinase Binding Protein (TYROBP) and Melanoma Inhibitory Activity (MIA) were with weight of 0.9695. Small Proline Rich Protein 1B (SPRR1B) and PDZ and LIM Domain 1 (PDLIM1) were with weight of 0.9760.

Density of node degree
Based on the occurrence of degree, density of node degree was shown in Figure 1. The degree of node ranged from 54 to 72. The median of these degrees was 64.7674. In addition, the assortativity coefficient of this density of node degree was 0.3422, which was confirmed that the subnetwork was with assortativity pattern.

GO functional enrichment analysis
A total of 129 genes were enriched in 127 GO terms (Table 1). Various genes were enriched in molecular function, such as Activating Transcription Factor 3 (ATF3), Desmoplakin (DSP), Adenylate Cyclase Activating Polypeptide 1 (ADCYAP1) and Jun Proto-Oncogene, AP-1 Transcription Factor Subunit (JUN).
Similarly, different genes in catalytic activity included Tubulin Beta 6 Class V (TUBB6), ATP Binding Cassette Subfamily A Member 1 (ABCA1) and CCR4-NOT Transcription Complex Subunit 4 (CNOT4).

Figure 1. Density of node degree. Note: The figure represents the frequency of appearance of each degree in network. The red line was the median of all node degrees.

Figure 2. Guilt-by-association node degree AUCs. Note: The figure was used to assess the average value of AUC. The red line was the median. The range between two grey lines was the inter-quartile range.

Relevant inference and cool reading inference

Based on relevant inference analysis, K terms and related AUC values were obtained. As shown in Table 2, top 3 K terms were regulation of cell differentiation (GO: 0045595, degree=66.697, AUC=0.5767), positive regulation of macromolecule metabolic process (GO: 0010604, degree=66.6754, AUC=0.6539) and regulation of macromolecule metabolic process (GO: 0060255, degree=66.5429, AUC=0.6694). The guilt-by-association node degree AUCs were generated (Figure 2). The average of AUC was 0.52198. After cool read of K functions, cool read node degree AUCs were obtained (Figure 3). The average of cool read node degree AUCs was 0.5488. In addition, the optimal list genes were obtained, such as JUN (MF score=0.0338), PTK2B (Protein Tyrosine Kinase 2 Beta, MF score=0.0322), CCND1 (Cyclin D1, MF score=0.0323) and ADAM8 (ADAM Metallopeptidase Domain 8, MF score=0.0319). Finally, optimal GO terms were obtained. Therein, top 3 optimal GO terms with higher AUC were regulation of signal transduction, positive regulation of macromolecule metabolic process and cellular component organization (Table 3).

Figure 3. Cool read node degree AUCs. Note: The figure represent cool read node degree AUCs. The red line was the median. The range between two grey lines was the inter-quartile range.

Table 1. Top 10 GO terms enriched by 129 genes.

<table>
<thead>
<tr>
<th>GO term</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO: 0048513</td>
<td>0.680476</td>
</tr>
<tr>
<td>GO: 0031325</td>
<td>0.674828</td>
</tr>
<tr>
<td>GO: 0065008</td>
<td>0.663036</td>
</tr>
<tr>
<td>GO: 0009893</td>
<td>0.662879</td>
</tr>
<tr>
<td>GO: 0048583</td>
<td>0.662069</td>
</tr>
<tr>
<td>GO: 0023051</td>
<td>0.658678</td>
</tr>
<tr>
<td>GO: 0051239</td>
<td>0.657308</td>
</tr>
<tr>
<td>GO: 0065009</td>
<td>0.65633</td>
</tr>
<tr>
<td>GO: 2000026</td>
<td>0.654762</td>
</tr>
<tr>
<td>GO: 0032879</td>
<td>0.646021</td>
</tr>
</tbody>
</table>

Table 2. Top 10 K terms by relevant inference analysis.

<table>
<thead>
<tr>
<th>GO term</th>
<th>Node degree</th>
<th>AUC</th>
<th>Node degree. AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO: 0045595</td>
<td>66.69697</td>
<td>0.57672</td>
<td>0.64418</td>
</tr>
<tr>
<td>GO: 0010604</td>
<td>66.6754</td>
<td>0.653968</td>
<td>0.655159</td>
</tr>
<tr>
<td>GO: 0060255</td>
<td>66.5429</td>
<td>0.669384</td>
<td>0.663925</td>
</tr>
<tr>
<td>GO: 0042221</td>
<td>66.51661</td>
<td>0.546755</td>
<td>0.645332</td>
</tr>
<tr>
<td>GO: 0010468</td>
<td>66.39146</td>
<td>0.634354</td>
<td>0.642857</td>
</tr>
<tr>
<td>GO: 0010556</td>
<td>66.3754</td>
<td>0.585337</td>
<td>0.628077</td>
</tr>
<tr>
<td>GO: 0005615</td>
<td>66.35419</td>
<td>0.74127</td>
<td>0.62619</td>
</tr>
<tr>
<td>GO: 2000112</td>
<td>66.31894</td>
<td>0.497619</td>
<td>0.623413</td>
</tr>
<tr>
<td>GO: 0006355</td>
<td>66.30374</td>
<td>0.561728</td>
<td>0.619489</td>
</tr>
</tbody>
</table>
Discussion

In most previous studies, microarray studies failed to research sufficiently the biological variability. Thereby, the validation rate was used within a system in this study to identify accurate genes and their interactional functions. In 2014, Ego network was firstly reported [8], which could screen potential connection [10,11]. In this study, several genes including biomarkers and lay the deeper foundation for breast cancer. Importantly, this method increased the prediction accuracy of this method.

Table 3. Top 5 optimal GO terms.

<table>
<thead>
<tr>
<th>GO ID</th>
<th>Term</th>
<th>Domain</th>
<th>Guilt_by_association_go</th>
<th>Cool_read_go</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0009966</td>
<td>Regulation of signal transduction</td>
<td>Biological_process</td>
<td>0.653968</td>
<td>0.680476</td>
<td>Ptk2b, Edn1, Adm, Adcyap1, St7p1, Mit, A464131, Sox11, Trnfsf19, Gadd45a, Mapk42, Adam8, Alp6ap1, C1if1, Fas, Cnd, Lgals3, Sgc, Rhut11, Igbtp2, Pspap, Jun, Abca1</td>
</tr>
<tr>
<td>GO:0010604</td>
<td>Positive regulation of macromolecule metabolic process</td>
<td>Biological_process</td>
<td>0.691506</td>
<td>0.663036</td>
<td>Edn1, Cct2, Ptk2b, Cst3, Jun, Sox11, Hoxa7, Aft6, Adcyap1, Aft3, Cct2, Crem, Foxo4, Sfp1, Ank3, Star, Cdkn1a, Gadd45a, Mapk42, Adam8, Cnd, C1if1, Fas, E330304, G19Rk</td>
</tr>
<tr>
<td>GO:0016043</td>
<td>Cellular component organization</td>
<td>Biological_process</td>
<td>0.603535</td>
<td>0.662069</td>
<td>Asf1a, Porn121, Jun, Mps25, Gadd45a, Tc17, Apc5, Ptk2b, Rhou, St7p1, Dap, Cct2, Maa, Abca1, Stbhp1, Topbp1, Ank3, Cst3, Tub6, Alp6, ap1, Fas, Mnp9, Gmn, Adm, Phn3a, Adcyap1, Edn1, Adam8, Acan, Cst3, Lgals3, Mta</td>
</tr>
<tr>
<td>GO:0050789</td>
<td>Regulation of biological process</td>
<td>Biological_process</td>
<td>0.600944</td>
<td>0.610714</td>
<td>Hoxa7, Cdkn1a, Jun, Edn1, Cct2, Sgc, Cnd, Mmp12, Sox11, St7p1, Igfbp2, Adcyap1, Adm, Aft3, C1if1, Cst3, Ptk2b, Mnp9, Foxo4, Lgals3, Neu1</td>
</tr>
</tbody>
</table>

blocks were effective on SCI treatment but with complex effects. Interestingly, the inhibitor of acid-sensing ion channels could significantly decrease neurological deficits following SCI [20]. In larval lamprey, SCI could induce changes of ion channels of reticulospinal neurons [8]. Thereby, JUN and PTK2B might be potential targets for SCI postoperative recovery and complications and prevention.

Two other genes, **CCND1** and **ADAM8** were screened as optimal list genes in this study. **CCND1** encoded a protein which belongs to the highly conserved cyclin family, whose members were characterized by a dramatic periodicity in protein abundance throughout the cell cycle [21]. In previous study, **CCND1** was found to be targeted by mir-195, which suppressed metastasis of osteosarcoma cells. In addition, et al. found that mir-19a was related with various genes of chondrogenic differentiation, such as **CCND1** [22]. In addition, **CCND1** genomic amplification was occurred during malignant transformation of a cell tumor of bone [23]. Similarly, ADAM8 encoded a protein which might be involved in cell adhesion during neurodegeneration [24]. In adult SCI mice models, **ADAM8** was upregulated in endothelial cells, and also related with angiogenesis [25]. Thereby, both **CCND1** and **ADAM8** might be important genes for SCI postoperative recovery and complications and prevention.

In this study, these screened genes enriched in optimal GO terms, including regulation of signal transduction, positive regulation of macromolecule metabolic process and cellular component organization. Quan et al. [26] found that acupuncture could activate signal transduction and further promote the repair of SCI. After SCI, signal transduction of erythropoietin receptor was also confirmed to inhibit apoptosis of neuron cells [27]. Besides, Zhao et al. [28] also found that multigene signal transduction was contribution to the development of neural stem cells after SCI. Though various macromolecules can’t be used with the limitation of across blood spinal cord barrier, the functions of these macromolecules for SCI recovery couldn’t be ignored [29]. In
rat models, convective delivery of macromolecules was used as an effective treatment method for injury or spinal cord disease [30]. It was worth noting that Fehlings et al. [31] have obtained remarkable advances in recruiting cell therapy for SCI. Besides, cellular transplantation could replace the lost elements, and further regulate cell functions after SCI [32].

**Conclusion**

In this study, relevant inference and cool reading inference using make the functional enrichment prediction was with statistical confidence. Screened genes including JUN, PTK2B, CCND1 and ADAM8 might be potential targets for SCI postoperative recovery and complications and prevention. Various GO terms, such as regulation of signal transduction, positive regulation of macromolecule metabolic process and cellular component organization, were important in these processes.

**Acknowledgments**

None

**References**


*Correspondence to
Funqiang Wang
Department of Orthopedics
Lanshan Dist. People's Hospital of Linyi City Shandong Province
PR China