Identification of hub genes and key pathways in spinal cord injury via bioinformatics analysis

Hao Cheng*, Yanben Wang, Jun Xuan

Department of Spine Surgery, Jinhua Municipal Central Hospital, Jinhua, Zhejiang 321000, China

Article Info

Abstract

This study aimed to explore the hub genes and related key pathways in Spinal Cord Injury (SCI) based on the bioinformatics analysis. Two microarray datasets (GSE45006, GSE45550) were obtained from the GEO database and were merged and batch-corrected. The differentially expressed genes (DEGs) in SCI were explored with the Limma, and the weighted gene co-expression network analysis (WGCNA) was conducted to explore the module genes. Functional enrichment analysis and Gene set variation analysis (GSA) were used to investigate the biological functions and key pathways of the key genes related to SCI. Then the protein-protein interaction (PPI) network was generated using the STING online tool, and the hub genes in SCI were identified. Receiver operating characteristic (ROC) curves were applied to assess the diagnostic value of the selected hub genes. We identified 554 DEGs in SCI, and 1236 key genes in SCI were selected via WGCNA. Totally 111 key genes related to SCI were discovered. Furthermore, the functional enrichment analysis showed that these key mRNAs were primarily enriched in the extracellular matrix (ECM)-related pathways and processes associated with wound healing and cell growth. The PPI network further filtered six hub genes (Cd44, Timp1, Loxl1, Col6a1, Col3a1, Col5a1) ranked by the degree, and the diagnostic value of the six hub genes was confirmed by the ROC curves. Six hub genes including Cd44, Timp1, Loxl1, Col6a1, Col3a1, and Col5a1 were identified in SCI, with differential expression and excellent diagnostic value, which might provide insight into the targeted therapy of SCI.

Keywords: Hub genes, Bioinformatics analysis, Biomarkers, Therapeutic targets, Spinal cord injury.

1. Introduction

Spinal cord injury (SCI) is a devastating neurological disorder characterized by high disability and mortality rates, affecting the life quality of patients and bringing huge burden to society (1). It is reported with approximately 250,000-500,000 new cases annually worldwide (2). SCI is divided into primary and secondary injury, and primary injury is the mechanical damage to the spinal cord and the consequent secondary injury is the following impaired neuronal homeostasis and tissue destruction induced by the local microenvironment alteration, and the pathophysiological mechanisms such as autophagy, inflammatory response, radical accumulation, neuronal death and blood-brain barrier (BBB) disruption are involved in this process (3-5). Currently, the effective management of SCI lacks consensus (6). Early surgical decompression, blood pressure augmentation, corticosteroids and invasive spinal cord pressure monitoring are suggested for SCI patients, with limited effects for the complete restoration of the spinal cord function (7). Therefore, it is imperative to deepen the understanding of the underlying mechanism in the process of SCI and identify promising targets for SCI therapy.

The exploration and identification of biomarkers for SCI have attracted increasing attention in recent years. By targeting the key factors involved in the pathological changes in the acute phase of SCI, various neuroprotective strategies have been investigated to restrain progressive damage to the spinal cord. Studies revealed that growth-associated factors such as PDGF, VGF, BDNF, FGF and BMPs are upregulated in the process following SCI of the neural connection reestablishment (8). Neurosteroid progesterone is revealed to improve recovery after SCI in clinical trials by elevating BDNF expression (9). Inflammatory factors such as TNF-α, IL-1β and IL-6 are reported to reach the peak at 6-12 h following SCI and induce the function loss of endothelial cells and the astrocytes (10). Methylprednisolone is used for SCI treatment and is indicated to promote neurological recovery by reducing the inflammatory factors including TNF-α, IL-1 and IL-6 (11). Macrophages play critical roles in the process of SCI, and the M2 subtype is at low level after SCI. It is reported that deficiency of CD36, the most enriched lipid transporter in the M2 subtype is at low level after SCI. Despite the increasing therapeutic strategies, the translation into clinical practice remains further exploration.

Bioinformatics analysis is widely used to evaluate the expression pattern and biological functions of genes as

* Corresponding author.
E-mail address: Drchenghao1025@163.com (H. Cheng).
# These authors contributed equally
Doi: http://dx.doi.org/10.14715/cmb/2024.70.3.9
well as the potential regulatory mechanisms involved in diverse diseases, SCI included (14). As an effective approach to exploring the pathogenesis in clinical diagnosis and pharmacological study, bioinformatics analysis has provided numerous targets and biomarkers for SCI diagnosis and therapy (15).

In our work, bioinformatics analysis was conducted to analyze the RNA-seq data and explore the hub genes and related key pathways in SCI. DEGs in SCI were explored and functional analysis was applied for further evaluation. Hub genes were selected using the Protein-Protein Interaction (PPI) network. The results of this study might provide promising therapeutic targets and deepen our understanding of the molecular mechanism in SCI.

2. Materials and Methods

2.1. Microarray Data and Screening DEGs

Two microarray datasets (GSE45006, GSE45550) based on the GPL1355 platform ([Rat230_2] Affymetrix Rat Genome 230 2.0 Array) were obtained from the NCBI GEO database (https://www.ncbi.nlm.nih.gov/geo/). The above-mentioned two datasets were merged and batch-corrected by the “sva” package in R software. The runUMAP function in R software was used for the visualization of the batch-removing effects. The DEGs were screened with the R“limma” package with |(log2FC)| >1 and p value less than 0.05 as the threshold value.

2.2. Construction of Weighted gene co-expression network analysis (WGCNA)

WGCNA was applied to investigate the correlation between genes and identify modules closely related to the phenotype of samples in the two datasets using the “WGCNA” R package. The soft thresholding power (β) was calculated by network topology analysis, and the adjacency was computed followed by conversion into the topological overlap matrix. Genes with similar patterns were assigned into modules using the average linkage hierarchical clustering with TOM-based dissimilarity measure. The correlation of gene modules and phenotypes (control, SCI) were calculated with the Pearson correlation coefficients. The module eigengene dissimilarity was then calculated, and the eigengene network was finally visualized. Key genes were defined with large Module Membership and Gene Significance in the modules (16). The key genes related to SCI were intersecting genes of DEGs selected by the limma package and key genes selected by the WGCNA using the Venn diagram web tool.

2.3. Functional Enrichment analysis

The biological functions of selected key genes in SCI were explored with the Gene Ontology (GO) analysis for biological process (BP), cellular component (CC), as well as molecular function (MF) using clusterProfiler R package. KEGG enrichment analysis was also performed with the same package. The results were visualized using the org. Hs.eg.db and GO plot R. Metascape online tool was also used for the enrichment analysis of the selected key genes associated with SCI.

2.4. Identification of hallmark pathways via gene set variation analysis (GSVA)

GSVA analysis was conducted to explore the biological functions of the selected key genes in SCI with the “GSVA” and “Msigdb” packages. The cut-off criteria are set as P. adjust < 0.05, gene size greater than or equal to 50 and enrichment score (ES) >2. The top ten most significantly enriched pathways were selected and visualized in the results.

2.5. Construction of PPI (protein-protein interaction) network and identification of hub genes

STRING (https://string-db.org/) database was applied to predict the direct and indirect relation between proteins. The Cytoscape v 3.8.2 was applied for the visualization of the PPI network. The top 12 genes ranked by degree were selected as hub genes in SCI using the CytoHubba plug-in.

2.6. Statistical analysis and prognostic value of hub genes

The ROC curves were established to calculate the AUC and 95%CI of the six selected hub genes in SCI using SPSS V 26.0 software (IBM Corporation, Armonk, NY, USA).

3. Results

3.1. The data characteristics in the two datasets of spinal cord injury

GSE45006 and GSE45550 datasets were downloaded from the GEO database, merged and batch corrected, and the inter-batch differences were eliminated. As shown in the plot of density, the expression value of the two datasets concentrated on zero after normalization, indicating the general consistency between the samples in the two datasets for further analysis (Figure 1A-B). Consistently, the box plot displayed that the distribution of data in the two datasets became even after data normalization (Figure 1C-D). Consistently, Uniform Manifold Approximation and Projection (UMAP) displayed that data of samples...
in these two datasets were intermingled (Figure 1E-F). Overall, these results suggested that the data source of our study was reliable.

3.2. Identification of DEGs and construction of WGCNA

The differentially expressed genes (DEGs) in spinal cord injury were then explored in the two datasets. The results showed that totally 551 DEGs were screened out under \(|\log2FC| > 1\) and \(p < 0.05\), and the expression pattern of up-regulated and down-regulated genes in spinal cord injury was shown in Figure 2A and B. WGCNA was applied for the cluster analysis of the DEGs in SCI. The network topology analysis was conducted for soft-holding power ranging from 1-30 and confirmed the relatively balanced scale independence and mean connectivity of the WGCNA. We set the power value (β) of 24 for further analysis because the scale-free topology fit index reached 0.89 with relatively high connectivity (Figure 2C-D).

3.3. Identification of key genes in SCI

The gene network was constructed and the modules were identified. Genes were clustered by a TOM-based dissimilarity measure. The soft thresholding power (β) was set at 24, and the genes were totally divided into 14 modules with similar modules merged, each with a unique color (Figure 3A). The connectivity of eigengenes was then analyzed, which provides information of the relation between the pairwise gene coexpression modules. The 14 modules were mainly clustered into 2 clusters, and the adjacencies between modules were shown in the heatmap (Figure 3B). The correlation of different modules with the phenotype is shown in Figure 3C. We found that the white (Cor=-0.39, \(p=6.7e^{-3}\)), cyan (Cor=-0.62, \(p=2.4e^{-6}\)) and grey (Cor=-0.51, \(p=2.4e^{-4}\)) modules were negatively correlated with SCI, while red (Cor=0.41, \(p=3.7e^{-3}\)), darkorange (Cor=0.4, \(p=4.4e^{-3}\)), darkred (0.47, 6.9e-4) and lightcyan (Cor=0.37, \(p=8.9e^{-3}\)) modules were positively correlated with SCI. As shown in the Venn diagram, the 551 DEGs intersected with the 1236 genes in the co-expression modules of WGCNA, and totally 111 key genes in SCI were identified in the intersection for the following analysis (Figure 3D).

3.4. Identification of key biological pathways

GO analysis and KEGG analysis of the selected key genes were performed and the results indicated that the selected key genes were primarily enriched in the wound healing, collagen fibril organization and regulation of phosphatidylcholine metabolic process in terms of BP, collagen-containing extracellular matrix, extracellular matrix and external encapsulating structure in terms of CC, platelet-derived growth factor binding, extracellular matrix structural constituent and mRNA methyltransferase activity in terms of MF. Moreover, KEGG analysis indicated that these key genes associated with SCI were primarily enriched in the pathways such as PPAR signaling, protein digestion and absorption and ECM-receptor interaction (Figure 4A-B). The network of specific genes with the terms and pathways of enrichment analysis was presented in Figure 4C. Then Gene set enrichment analysis (GSEA) revealed that for the selected key genes, the TYROBP Causal Network In Microglia, Cell Cycle, Resolution of Sister Chromatid Cohesion, Separation of Sister Chroma-

Fig. 2. Identification of DEGs and construction of WGCNA. (A) Volcano plot and (B) Heatmap of the differentially expressed genes in SCI. (C) Scale independence and (D) Mean connectivity for the WGCNA network.

Fig. 3. Identification of key genes in SCI. (A) Clustering dendrograms of genes, with dissimilarity according to topological overlap, together with assigned module colors. (B) Heatmap displaying the eigengene adjacency. (C) Correlation between the modules and the phenotypes. (D) Venn diagram of the key genes in SCI selected based on the limma package and WGCNA.

Fig. 4. Identification of key biological pathways enriched by key genes. (A) The bubble chart and (B) bar graph of the top three terms enriched by hub genes in the BP, CC, MF and KEGG. (C) The network of hub genes and terms and pathways based on the GO and KEGG analysis. (D-G) Representative pathways enriched in the selected hub genes were identified by GSEA.
3.5. Construction of the PPI network and the selection of hub genes

Based on the analysis using Metascape, we found that the hub genes were mainly enriched in terms such as extracellular matrix organization, wound healing, regulation of phosphatidylcholine metabolic process focal adhesion and regulation of neutrophil apoptotic process (Figure 5A-C). Protein-protein interaction network was generated as shown in Figure 5D. The top 12 mRNAs ranked by degrees of connectivity were screened and identified as hub genes in SCI (Figure 5E).

3.6. Diagnostic value of the hub genes

Based on the ROC curves, we evaluate the stability of the selected disease targets. We found that the Cd44 has the highest diagnostic value (AUC= 0.913, CI: 0.829,-0.997), followed by Timp1 (AUC=0.897, CI: 0.809,-0.986), Loxl1 (AUC=0.858, CI: 0.751,-0.965), Col6a1 (AUC=0.839, CI: 0.685,-0.993), Col3a1 (AUC= 0.803, CI: 0.640,-0.965), Col5a1 (AUC=0.784, CI: 0.641,-0.927) in spinal cord injury (Figure 6A-F).

4. Discussion

Spinal cord injury is a devastating neurological condition leading to severe neurological dysfunction and disability, affecting the physical and mental health, life quality as well as social participation of patients (17). The enhancement of the neuroplasticity and tissue repair following SCI are key to the functional recovery of SCI patients (18). However, the effects of current treatment for SCI are still limited in neural regeneration, and understanding the underlying mechanisms in the process of SCI provides opportunities for the targeted therapy of SCI (19, 20). In this study, we identified 6 hub genes (Cd44, Timp1, Loxl1, Col6a1, Col3a1, Col5a1) and key biological mechanisms involved in the process of SCI based on bioinformatics analysis. Further analysis confirmed the prognostic value of the selected hub genes for SCI, which may provide promising biomarkers and therapeutic targets for SCI patients.

Based on the GSE45006 and GSE45550 datasets, we discovered 551 differentially expressed genes in SCI. Functional enrichment analysis indicated that the DEGs in SCI were significantly enriched in the pathways such as extracellular matrix (ECM), collagen-containing ECM, ECM structural constituent, collagen fibril organization, wound healing, protein digestion and absorption, and PPAR signaling pathway. Spinal cord ECM contains structural and communication proteins implicated in the repair and regeneration following SCI (21). ECM proteins such as laminin, collagen, and fibronectin are suggested to be associated with fibrotic scar formation and affect axon growth after SCI (22). Wound healing is a critical process in the central nervous system (CNS) injury, during which glial cells are moved to generate a protective barrier to seal the wound and reduce further tissue damage (23). The PPAR signaling pathway is reported to be increased and after SCI, which may cause enhanced bone resorption and related to the bone loss following SCI (24). Moreover, the platelet-derived growth factor has been demonstrated to improve the recovery after SCI in rodent models by promoting cell survival and repair and improving locomotor function (25, 26). The results of GSEA showed that the key genes related to SCI were primarily enriched in the pathways such as TYROBP Causal Network In Microglia, Cell Cycle, Resolution of Sister Chromatid Cohesion, and signaling by interleukins. Microglia TYROBP plays a key role in brain homeostasis, and it is indicated that the TYROBP may be associated with the loss of markers of synaptic integrity in Alzheimer’s disease (27). Several bioinformatics analyses have also confirmed that Tyrobp is highly expressed after SCI and is a candidate key gene for SCI (28, 29). Cell cycle is activation is reported to be involved in the pathophysiologic process of neurodegenerative disorders such as Parkinson’s disease and Alzheimer’s disease (30, 31), and it is also suggested to regulate the neuronal apoptosis and glial proliferation/activation after central nervous system (CNS) injury, SCI included (32).

Totally 6 hub genes (Cd44, Timp1, Loxl1, Col6a1, Col3a1, Col5a1) in SCI were identified. Cd44 is known as an ECM phosphoglycoprotein implicated in cell adhesion, chemotaxis as well as immunomodulation, and serves as a receptor for osteopontin OPN (33). Cd44 has been re-
reported to be upregulated in rat spinal cords after clip compression injury, and it is suggested that Cd44 facilitates cell adhesion and glial cell attraction at the early stage following SCI to improve the recovery of injured spinal cords (34). Tissue inhibitors of metalloproteinases (TIMPs) is revealed to regulate the activities of MMPs, which break down inhibitory ECM molecules and significantly affect the axonal growth (35). Timp1 is suggested to be involved in tissue remodeling after dorsal root injury in a rat model and is a possible candidate for CNS axonal regeneration (36). As a growth factor, TIMP-1 is expressed to promote myelination or promote myelin repair in CNS (37). Loxl1 as a member of the LOX family and an important enzyme in elastic fiber synthesis and homeostasis, is related to the regulation of tensile strength and structural integrity of various tissues (38). Studies have uncovered that Loxl1 can catalyze the covalent cross-linking of ECM proteins collagen and elastin, which concedes to the ECM stiffness and mechanical properties (39-41). Col6a1 is upregulated in chronic spinal cord injury and is targeted by miR-330-3p and co-expressed with IncRNA0632, which may be a candidate target for chronic SCI therapy (42). Col3a1 encodes type I and III collagen in connective tissues and is differentially expressed between rostral and caudal regions in SCI rates and enriched in terms such as blood vessel development, response to mechanical stimulus and wound healing based on GO analysis (43). Col5a1 plays a role in ECM organization and is suggested as a biomarker to distinguish different SCI subtypes (44). The results of our study also indicate that these key genes associated with SCI are closely related to ECM organization, wound healing and elastic fibre formation, etc., and the ROC curves confirmed the diagnostic value of the six selected hub genes in SCI.

In conclusion, based on the two GEO datasets, the results of bioinformatics analysis found 551 differentially expressed genes in SCI and 111 key genes associated with SCI. Further analysis indicated that these key genes in SCI were enriched in the pathways related to ECM, wound healing and cell cycle, etc. Further analysis identified 6 hub genes in SCI, and the diagnostic value of these hub genes was verified using ROC curves. The results of our work may deepen the understanding of the potential mechanism in the process of SCI and contribute to the development of novel therapeutic strategies for SCI.

References


