Gene expression profiling reveals candidate biomarkers and probable molecular mechanism in diabetic peripheral neuropathy

Han Zhou
WenChuan Zhang
Department of Neurosurgery, Ninth People Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200011, People’s Republic of China

Purpose: To investigate the molecular mechanism and search for candidate biomarkers in the gene expression profile of patients with diabetic peripheral neuropathy (DPN).

Methods: Differentially expressed genes (DEGs) of progressive vs non-progressive DPN patients in dataset GSE24290 were screened. Functional enrichment analysis was conducted, and hub genes were extracted from the protein–protein interaction network. The expression level of hub genes in serum samples in another dataset GSE95849 was obtained, followed by the ROC curve analysis.

Results: A total of 352 DEGs were obtained from dataset GSE24290. They were involved in 14 gene ontology terms and 10 Kyoto Encyclopedia of Genes and Genomes pathways, mainly related to lipid metabolism. Eight hub genes (LEP, APOE, ADIPOQ, FABP4, CD36, GPAM, CIDEC, and PNPLA4) were revealed, and their expression level was obtained in dataset GSE95849. The receiver operating characteristic curve analysis indicated that CIDEC (AUC=1), APOE (AUC=0.833), CD36 (AUC=0.803), and PNPLA4 (AUC=0.861) might be candidate serum biomarkers of DPN.

Conclusion: Lipid metabolism of Schwann cells might be inhibited in progressive DPN. CIDEC, APOE, CD36, and PNPLA4 might be potential predictive biomarkers in the early DPN diagnosis of patients with DM.

Keywords: differentially expressed genes, functional enrichment analysis, demyelination, lipid metabolism, diabetic peripheral neuropathy

Introduction
It was estimated that 415 million people aged 20–79 years suffered from diabetes in 2015, and the number was predicted to rise to 642 million by 2040.1 Approximately 50% of those with diabetes may develop a diabetic peripheral neuropathy (DPN). The number will only increase as the diabetes epidemic grows.2 DPN is characterized by pain, paraesthesia, and sensory loss.3 Patients complaint unbearable lancinating, tingling and burning sensation, even depression, anxiety, and sleep deprivation.4 5 Moreover, insensitivity to trauma often results in foot ulcers which can lead to some levels of amputation.6 Several attempts have been made to treat this disease; however, none of the pharmacotherapy have proven to be effective in altering the progressive course to date.7 Various symptomatic medications are nonspecific with serious side effects.8 9 The treatment dilemma reflects our present knowledge for the pathogenesis of DPN is far from to be clear.

Demyelination is known as an early pathological feature in DPN, and it precedes the degeneration of axon.10 Therefore, studying the molecular mechanism...
associated with demyelination may conduce to a better understanding of DPN and finding the predictive biomarkers of this disease. In 2010, Hur et al, analyzed genes that differed between DPN progressors and non-progressors, which is classified by the loss of myelinated fiber density of sural nerves.11 Based on the microarray data, the present study analyzed the differentially expressed genes (DEGs) via function enrichment analysis and topological approaches. We found several hub genes according to the most significant cluster of protein–protein interaction (PPI) network. The expression level of each hub gene was obtained in serum samples of another dataset, followed by receiver operating characteristic (ROC) curve study. The results may reveal the molecular mechanism of demyelination in DPN and provide potential serum biomarkers.

Materials and methods

Data preprocessing and DEGs screening
From the Gene Expression Omnibus database, we collected the gene expression data of GSE24290 deposited by Hur et al.11 There were 35 specimens including progressors and non-progressors sural nerve biopsies for analysis. The raw expression data underwent background correction, normalization, and summarization using the robust multi-array average (RMA) algorithm in oligo.12,13 The limma package in R was applied to identify DEGs between two groups. In this analysis, P-value <0.05 and log|FC| >0.5 were used as the cutoff criteria.

Functional enrichment and PPI network analysis
To investigate the main functional pathways of DPN, we submitted the DEGs to Database for Annotation, Visualization, and Integrated Discovery (DAVID), which was used to perform the Gene Ontology (GO) analysis and KEGG pathways enrichment analysis of DEGs.14–16 Then, pathview was used to describe the most important pathway.17,18 Criteria for this step were set as P-value <0.05 and gene counts ≥3. The functional protein interactions of DEGs and the encoding proteins were predicted using the Search Tool for the Retrieval of Interacting Genes (STRING).19 Cytoscape 3. 6. 1 was used for visualization and to calculate the properties of the PPI network. Subsequently, the Network Analyzer plug-in of Cytoscape was utilized to analyze the topology properties of the network.20 Connectivity degree analysis was performed and the most highly connected cluster was extracted from the PPI network through MCODE analysis.21

ROC curve analysis
The genes that constituted the most highly connected cluster in PPI network were considered to be hub genes. Then, we downloaded the gene expression profiling data of GSE95849, which contains serum samples from six DM patients and six DPN patients.22 We obtained the expression level of hub genes and used the pROC package in R software to draw the ROC curves and calculate the area under curve (AUC).23 Larger AUC value means the gene can well distinguish DPN from the DM patient samples. The diagnosis effect of hub genes was further investigated according to the AUC value.

Results

DEGs identified between progressive and non-progressive DPN patients
The raw gene expression data of GSE24290 were normalized by RMA. The boxplots of the intensity of all samples demonstrated that the expression values of each sample were close to the same after normalization (Figure S1). A total of 352 DEGs between non-progressive and progressive DPN patients were obtained with the criteria of P-value <0.05 and log|FC| >0.5. The 142 up-regulated genes and the 210 down-regulated genes were shown in the heat map and the volcano plot (Figure 1A and B and Figure S2).

Functional annotation and enrichment analysis of the DEGs
To reveal further insights into the biological functions of DEGs, functional enrichment analyses were performed using DAVID. For the GO analysis, we focused on the categories of biological processes (BP) and set the criteria with P-value <0.05 as significantly enriched GO terms. As a result, DAVID identified two terms significantly enriched from 142 up-regulated genes and 12 terms significantly enriched from 210 down-regulated genes. These terms demonstrated that the gene expression differences of nerve samples in progressive DPN were associated with “fatty acid homeostasis”, “glucose homeostasis”, and other BP (Table 1). Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis unveiled ten significantly enriched KEGG pathways with P-value <0.05.
Among these pathways, the PPAR signaling pathway was the most important one according to the P-value, gene counts, and fold enrichment. The DEGs in the PPAR pathway were concentrated in PPAR-γ (Figure S3). There were seven DEGs genes that participated in this pathway (CD36, OLR1, SCD, ACSBG2, FABP4, ADIPOQ, MMP1). We listed other useful information of these KEGG pathways in Table 2.
PPI network clusters analysis and selection of hub genes

PPI network was constructed involving 73 nodes (DEGs) and 132 edges. The nodes represented the proteins expressed by DEGs and the edges between two nodes means the physical interactions. The connectivity degree is an important parameter and the high connectivity degree indicated the protein interacted with more surrounding proteins and play a more important role. The nodes with higher connectivity degree were shown as larger sizes with red or orange color (Figure 2A). Subsequently, we analyzed the PPI network and extracted the most highly connected cluster by MCODE plug-in in Cytoscape (Figure 2B). Genes in this cluster were at the core of the whole network, including LEP, APOE, ADIPOQ, FABP4, CD36, PNPLA4, GPAM, and CIDEC. Hence, we consider the eight genes as the hub surrounding proteins and play a more important role. The nodes with higher connectivity degree were shown as larger sizes with red or orange color (Figure 2A). Subsequently, we analyzed the PPI network and extracted the most highly connected cluster by MCODE plug-in in Cytoscape (Figure 2B). Genes in this cluster were at the core of the whole network, including LEP, APOE, ADIPOQ, FABP4, CD36, PNPLA4, GPAM, and CIDEC. Hence, we consider the eight genes as the hub

Table 1 GO terms enrichment results of DEGs

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>P-value</th>
<th>Enriched genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enriched from up-regulated genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0007154→cell communication</td>
<td>3</td>
<td>0.0169</td>
<td>GJB6, GJB2, GJA3</td>
</tr>
<tr>
<td>GO:0050896→response to stimulus</td>
<td>3</td>
<td>0.0233</td>
<td>CLRN1, KERA, RGRIP1</td>
</tr>
<tr>
<td>Enriched from down-regulated genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0055089→fatty acid homeostasis</td>
<td>3</td>
<td>0.0311</td>
<td>APOE, MLXIPL, GPAM</td>
</tr>
<tr>
<td>GO:0042593→glucose homeostasis</td>
<td>5</td>
<td>0.0042</td>
<td>LEP, CAV3, GPR21, MLXIPL, ADIPOQ</td>
</tr>
<tr>
<td>GO:0006936→muscle contraction</td>
<td>5</td>
<td>0.0052</td>
<td>CAV3, MYH1, ACTA1, MYH13, MYH7</td>
</tr>
<tr>
<td>GO:0042362→cholesterol homeostasis</td>
<td>4</td>
<td>0.0083</td>
<td>CAV3, MALRD1, APOE, FABP4</td>
</tr>
<tr>
<td>GO:0019433→triglyceride catabolic process</td>
<td>3</td>
<td>0.0114</td>
<td>APOE, FABP4, PNPLA3</td>
</tr>
<tr>
<td>GO:0051897→positive regulation of protein kinase B signaling</td>
<td>4</td>
<td>0.0173</td>
<td>LEP, TNFAP8L3, STOX1, IL26</td>
</tr>
<tr>
<td>GO:0002027→regulation of heart rate</td>
<td>3</td>
<td>0.0193</td>
<td>CAV3, RYR2, MYH7</td>
</tr>
<tr>
<td>GO:0043407→negative regulation of MAP kinase activity</td>
<td>3</td>
<td>0.0223</td>
<td>CAV3, APOE, ADIPOQ</td>
</tr>
<tr>
<td>GO:0006810→transport</td>
<td>7</td>
<td>0.0259</td>
<td>CLCA2, AQP8, GRIK4, FABP4, CLVS1, CNGA3, TRPM1</td>
</tr>
<tr>
<td>GO:0035338→long-chain fatty-acyl-CoA biosynthetic process</td>
<td>3</td>
<td>0.0304</td>
<td>SCD, AC5B2G, ACOT6</td>
</tr>
<tr>
<td>GO:0006635→fatty acid beta-oxidation</td>
<td>3</td>
<td>0.0331</td>
<td>LEP, ABCD2, ADIPOQ</td>
</tr>
<tr>
<td>GO:0034220→ion transmembrane transport</td>
<td>5</td>
<td>0.0480</td>
<td>CLCA2, AQP8, GRIK4, ANO3, RYR2</td>
</tr>
</tbody>
</table>

Notes: Significantly enriched GO terms with P-value <0.05 count ≥3 and were screened out. Five terms were directly associated with lipid metabolism.

Abbreviations: GO, Gene Ontology; DEGs, differentially expressed genes.

Table 2 Information of KEGG analysis of dataset GSE24290

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>P-value</th>
<th>Enriched genes</th>
<th>Fold enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa03320:PPAR signaling pathway</td>
<td>7</td>
<td>&lt;0.0001</td>
<td>CD36, OLRL1, SCD, ACSB2G, FABP4, ADIPOQ, MMP1</td>
<td>11.0569</td>
</tr>
<tr>
<td>hsa04923:Regulation of lipolysis in adipocytes</td>
<td>5</td>
<td>0.0002</td>
<td>PTGER3, NPY, PDE3B, FABP4, PRKG2</td>
<td>9.4492</td>
</tr>
<tr>
<td>hsa00982:Drug metabolism – cytochrome P450</td>
<td>5</td>
<td>0.0004</td>
<td>CYP2A6, ADH1A, GSTM5, UGT2B28, MGST1</td>
<td>7.7817</td>
</tr>
<tr>
<td>hsa04920:Adipocytokine signaling pathway</td>
<td>5</td>
<td>0.0004</td>
<td>LEP, CD36, NPY, ACSB2G, ADIPOQ</td>
<td>7.5593</td>
</tr>
<tr>
<td>hsa00980:Metabolism of xenobiotics by cytochrome P450</td>
<td>5</td>
<td>0.0005</td>
<td>CYP2A6, ADH1A, GSTM5, UGT2B28, MGST1</td>
<td>7.1507</td>
</tr>
<tr>
<td>hsa05204:Chemical carcinogenesis</td>
<td>5</td>
<td>0.0006</td>
<td>CYP2A6, ADH1A, GSTM5, UGT2B28, MGST1</td>
<td>6.6144</td>
</tr>
<tr>
<td>hsa04060:Cytokine-cytokine receptor interaction</td>
<td>8</td>
<td>0.0007</td>
<td>LEP, IL20RB, INHBE, GDF5, IL26, EDAR, CCL27, CCL26</td>
<td>3.4841</td>
</tr>
<tr>
<td>hsa00830:Retinol metabolism</td>
<td>4</td>
<td>0.0213</td>
<td>CYP2A6, SDR16C5, ADH1A, UGT2B28</td>
<td>6.6144</td>
</tr>
<tr>
<td>hsa01520:AMPK signaling pathway</td>
<td>5</td>
<td>0.0271</td>
<td>LEP, CD36, SCD, GYS2, ADIPOQ</td>
<td>4.3021</td>
</tr>
<tr>
<td>hsa04024:cAMP signaling pathway</td>
<td>6</td>
<td>0.0364</td>
<td>PTGER3, NPY, PDE3B, RYR2, CNGA3, TRPM1</td>
<td>3.2070</td>
</tr>
</tbody>
</table>

Notes: KEGG biological pathway enrichment analysis found that the PPAR signaling pathway (P-value <0.0001, count =7, and fold enrichment =11.0569) was the most important one among the ten significantly enriched pathways.

Abbreviation: KEGG, Kyoto Encyclopedia of Genes and Genomes.
genes for further analysis, and they were all down-regulated as demyelination.

**ROC curves and candidate biomarkers**

The ROC curve analysis was performed by the pROC package in R (Figure 3). The AUCs of each gene were more than 0.500. Among these hub genes, CIDEC was the outstanding one with the AUC = 1.000, which represented it might have great value for the diagnosis of DPN in patients with DM. Besides, APOE (AUC = 0.833), CD36 (AUC = 0.803), and PNPLA4 (AUC = 0.861) could also be predictive biomarkers. The AUCs of other hub genes were less than 0.8000.

**Discussion**

The mechanism producing DPN is multifactorial and extremely complex. To further understand the molecular mechanism and search for novel serum biomarkers for DPN, we analyzed two different datasets of expression profile by bioinformatic approaches in the current study. We first compared the microarray data of two groups of sural nerve samples from DPN progressors and non-progressors. They were divided by the level of demyelination. Samples in the progressor group lost ≥500 fibers/mm², while samples in the non-progressor group lost ≤100 fibers/mm² over 52 weeks.11

Totally 352 DEGs were identified for following functional enrichment analysis and PPI network analysis. The results demonstrated the function of DEGs was closely related to lipid metabolism. In GO analysis, most of the terms were derived from down-regulated DEGs. We only focused on the most prominent parts of these terms. It was not difficult to find that five terms were directly associated with lipid metabolism in total 14 GO terms of BP. They were “fatty acid homeostasis”, “cholesterol homeostasis”, “Adipocytokine signaling pathway”, “triglyceride catalytic process”, “long-chain fatty-acyl-CoA biosynthetic process”, and “fatty acid beta-oxidation” (Table 1). The changes were likely to occur in Schwann cells. In the peripheral nerve system, myelin results from the circumferential wrapping of the Schwann cell plasma membrane, which is enriched in lipid-like glycosphingolipids, saturated long-chain fatty acids and, particularly, cholesterol.24

After being damaged by hyperglycemia, the balance of Schwann cells de-differentiation and re-differentiation
can be destroyed. As the fatty acid homeostasis, cholesterol homeostasis was changed and the triglyceride catabolic process, long-chain fatty-acyl-CoA biosynthetic process was inhibited, re-differentiated mature Schwann cells might be difficult to generate new myelin, which might be an important cause of DPN progression.

The KEGG analysis further revealed changes in the signaling pathways in progressive DPN. Three important signaling pathways were associated with lipid metabolism. They were “PPAR signaling pathway”, “regulation of lipolysis in adipocytes”, and “adipocytokine signaling pathway”. The “PPAR signaling pathway” is highlighted with smallest P-value and largest fold enrichment among them. Almost all DEGs involved in the PPAR signaling pathway were down-regulated, which was consistent with the study of Kim et al. They demonstrated chronic high glucose inhibited the function of PPAR-γ due to the reduction of PPAR-γ binding to target genes in Schwann cells. Montani et al, described the endogenous fatty acid synthesis, which was potentially critical process of myelination, could trigger activation of the PPAR-γ transcriptional program in Schwann cells and the PPAR-γ agonist could partially rescue Schwann cell myelination in the setting of deficient endogenous fatty acid synthesis. Our study suggested the inhibition of PPAR-γ signaling pathway caused by hyperglycemia might be crucial in the progression of DPN.

The PPI network contained multiple clusters. We applied the MCODE in Cytoscape to extract the most significant cluster of PPI network. All the genes that make up this cluster play important roles in lipid metabolism. For example, CD36 facilitates cell membrane free fatty acid transport in adipocytes and ADIPOQ and LEP were secreted by adipocytes. Accordingly, we assumed that Schwann cells could secrete some adipokines like adiponectin and take in lipid and free fatty acid via CD36. Uptake of palmitic acids, which is the most abundant plasma free fatty acid involved in inducing insulin resistance, proven to cause Schwann cells dysfunction and death. Downregulation of CD36 in progressive DPN might reduce lipid uptake and influence myelination, but might play a protective role.

Furthermore, we suggested that CIDEC, PNPLA4, APOE, and CD36 might be used as potential molecules for liquid biopsy of DPN based on the results of ROC curve analysis. In particular, CIDEC might be an important molecule that has not been thoroughly studied but is of great value to the diagnosis of DPN. CIDEC protein is a member of the cell death-inducing DNA fragmentation factor-like effector family, which are crucial for multiple lipid metabolism-related processes.
metabolic pathways and lipid homeostasis. PPAR-γ agonist may lead to upregulation of CIDE© thereby increasing lipid accumulation. Although the expression of CIDE© proven to positively correlate with the development of insulin sensitivity in obese people, it has not been studied within the context of DPN by now. The only investigation of CIDE© associated with diabetic complications is that silence of CIDE© may partially reverse diabetic pulmonary vascular in type 2 diabetes.

The present study re-analyzed the two publicly available microarray gene expression profiling via different methods and arrived at different conclusions. We constructed the PPI network of DEGs of the dataset GSE24290 and extracted the most important cluster via MCODE. This topological analysis differs from gene co-citation analysis in primary publication. Combining with our results from analysis of the dataset GSE95849, we proposed our new point that lipid metabolism of Schwann cells might be inhibited in progressive DPN and the inhibition of PPAR-γ signaling pathway might be crucial in the pathogenesis of the disease. In addition, in the primary publication of the dataset GSE95849, the research found the downregulation of the neurotrophin-MAPK signaling pathway may be crucial for DPN pathogenesis, while our analysis only focused on the expression level of several selected genes in serum samples of the dataset.

Conclusion
The present study aimed to investigate the molecular mechanism in gene expression profiling in the pathogenesis of DPN. Totally 352 DEGs and eight hub genes were screened via this bioinformatic approaches of two microarray datasets (GSE24290 and GSE95849). Our salient findings were that lipid metabolism of Schwann cells might be inhibited in progressive DPN and CIDE©, APOE, CD36, PNPLA4 were identified as candidate predictive biomarkers in the early DPN diagnosis of patients with DM. However, there were some limitations in the present study such as small sample size and lack of verification test. Further basic experiments with large sample size are needed to validate our results.

Acknowledgments
This study was supported by a grant (No. 81771320) from the National Natural Science Foundation of China.

Disclosure
The authors report no conflicts of interest in this work.

References


Figure S1 The boxplots of sample data before and after normalization. 
Notes: The lateral axis represented 35 samples from 18 progressive and 17 non-progressive DPN patients. The longitudinal axis represents expression levels. The horizontal line in the middle of post represents the expression levels of each sample. Each sample was close to the same following normalization.
Abbreviation: DPN, diabetic peripheral neuropathy.
Figure S2. The heat map of all 352 DEGs.
Abbreviation: DEGs, differentially expressed genes.
Figure S3 The PPAR signaling pathway and DEGs.
Abbreviation: DEGs, differentially expressed genes.