Identification of potential biomarkers and small-molecule compounds related to intracerebral hemorrhage with bioinformatics analysis

Ziqi Yang, Ruonan Wang, Xingyu Chen, Dexi Zhao*

Changchun University of Chinese Medicine, Jilin, China,
* Email: zdx02@163.com

This study aimed to further explore the underlying molecular mechanism of intracerebral hemorrhage (ICH), gene expression profile GSE24265, containing perihematomal tissues, contralateral grey matter, and contralateral white matter was retrieved and analyzed. The data was hierarchically clustered and the differentially expressed genes (DEGs) were screened. Functional analysis and protein interaction analysis of DEG hubs were performed, and the miRNA-transcription factor (TF)-target network was built. In addition, the candidate small-molecule compounds that might reverse the expression of an ICH-linked gene were identified by CMap. This method revealed a total of 408 DEGs. Five modules including chemokine-related, antigen immune-related, pathogen infection, cell reaction, and positive regulation of tyrosine phosphorylation and MAPK cascade were identified. The expression levels of CCL5, CXCL8, ICAM1, IL-1B, IL-6, VCAM1, and VEGFA were correlated with ICH among the top 10 hub genes obtained in the protein protein interaction (PPI) network. A total of 237 miRNA-TF-target regulatory relationships were obtained, including 6 TFs, 11 miRNAs and 105 target genes. Finally, the CMap database identified Prestwick-1083, xamoterol, ifosfamide, methylpate, nifurtimox, propranolol, and methoxamine as potential therapeutic agents for ICH while doxorubicin, menadione and azacitidine may increase its pathogenicity. Furthermore, CCL5, CXCL8 and VEGFA may be novel candidate susceptibility genes for ICH. Some small-molecule drugs, including xamoterol may be used for the treatment of ICH.

Key words: spontaneous intracerebral hemorrhage, gene expression profile, small molecules, candidate genes

INTRODUCTION

Spontaneous intracerebral hemorrhage (ICH), also known as cerebral bleed, usually presents with bleeding in both the brain tissue and ventricles. In the United States, approximately 37,000–52,400 people are affected by ICH each year, which accounts for 20%–30% of all cases of cerebrovascular diseases (Caceres and Goldstein, 2012; Zhong et al., 2016). The risk factors for ICH included hypertension, amyloidosis, alcoholism, and cocaine use (Caceres and Goldstein, 2012), and the primary brain injury caused by ICH is mainly due to compression and destruction of brain tissue contained within the skull (Zheng et al., 2016). Identifying the pathological mechanisms for ICH is essential. Genome-wide gene expression profiles can simultaneous compare transcripts of thousands of differentially expressed genes (DEG) between diseased and normal tissue. Currently, expression profiling has been used to identify prognostic and therapeutic markers involved in the pathomechanism of many diseases, such as ICH. Stamova et al. (2018) analyzed the differentially expressed transcripts of ICH from patients with ischemic stroke (IS) and matched controls (CTRL). Sang et al. (2017) demonstrated a complex mechanism of peripheral blood mononuclear cells (PBMCs) in response to ICH, which could help understand its pathogenesis. The expression level of golgin A8 family member

Received 24 July 2021, accepted 23 November 2021
A (GOLGA8A) was increased in the blood of patients with ICH (Merino-Zamorano et al., 2016). Interesting experiments using mRNA microarrays have identified some dysregulated mRNAs, and biological processes enriched by differentially expressed genes (DEGs) may help explain the pathophysiological changes in ICH (Hanjin et al., 2018). Furthermore, similar work has demonstrated sex differences in ICH patients (Xie et al., 2018). However, the underlying molecular mechanism of ICH remains largely unknown.

Rosell et al. (2011) conducted a microarray study of brain tissues of deceased patients with ICH and found that the overexpressed genes were related to cytokines, chemokines, and coagulation factors. Furthermore, Yang et al. (2014) applied the same dataset of brain tissues after ICH and found that there was significant dysregulation in the calcium signaling pathway of DEGs in ICH. Our investigations built on this previous work by examining DEGs and functional gene analysis, together with PPI and miRNA–TF–target network analysis. In addition, we identified some of the ICH-related small-molecule drugs that targeted the DEGs. These results may help illustrate the molecular mechanism and explore the key genes involved in ICH.

METHODS

Data preprocessing

The gene expression profile GSE24265 (Rosell et al., 2012) was obtained from the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) database (Barrett et al., 2005). A total of eleven brain samples were obtained, from 4 deceased patients with ICH comprising perihematomal area (PH), contralateral white matter (CW), and contralateral gray matter (CG). Using GPL570 (HG-U133_Plus_2) Affymetrix Human Genome U133 Plus 2.0 Array as the detection platform, the raw data (CEL file) was then preprocessed by the Affy package (Gautier, Cope, Bolstad, and Irizarry, 2004) in R (version 1.50.0; http://www.bioconductor.org/packages/release/bioc/html/affy.html). These data were normalized using the Robust Multi-Array Average (RMA) algorithm. The RMA algorithm is a comprehensive means for extracting signals and standardization, comprising background correction, normalization, and expression calculation. This was used to establish an expression matrix based on the Affymetrix data. Once the probe is mapped to the same gene, the average expression value of the probe is used as the gene expression value, after which the annotation file was downloaded and probes that did not fit the gene symbol were deleted.

Principal component analysis (PCA) and cluster analysis

We then used PCA and hierarchical clustering was used to investigate the similarity of samples. PCA was performed by applying the R prcomp function (https://stat.ethz.ch/R-manual/R-devel/library/stats/html/prcomp.html) with standard parameters, which reduces the dimensionality of the data, as well as identify the latent variables (principal components). Then, the scatterplot3d package (version 0.3-41; https://mirrors.tuna.tsinghua.edu.cn/CRAN/bin/windows/contrib/3.4/scatterplot3d_0.3-41.zip) was performed to build the PCA map. To observe whether it is consistent with the pre-grouping, hierarchical cluster analysis was used to investigate the hclust function (https://stat.ethz.ch/R-manual/R-devel/library/stats/html/hclust.html) in R.

Identification of DEGs

To identify DEGs within the data, the Limma package (Smyth, 2011) (version 3.10.3; http://www.bioconductor.org/packages/2.9/bioc/html/limma.html) was performed on CG vs. PH, CW vs. PH, and CG vs. CW. Adjusted P value < 0.05 and |log 2 FC (fold change)| > 1 were set as the cutoff values for DEG identification. To observe the expression level of the DEGs between the contralateral hemisphere of cerebral hemorrhage and PH, and to exclude any heterogeneity caused by different contralateral areas, the overlapping DEGs of the CG vs. PH and CW vs. PH groups were obtained, then genes with the same expression trend in the CG vs. PH and CW vs. PH groups were screened. We then removed genes from the CG vs. CW group and any remaining DEGs of interest were used for subsequent analysis.

Functional analysis

The Gene Ontology (GO) (Ashburner et al., 2000) biological process (BP) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) were analyzed to examine DEG pathways by applying DAVID (version 6.8; https://david-d.ncifcrf.gov/), (Huang et al, 2008) with P value < 0.05 as the cutoff criterion and an enrichment count of 5 being considered significant.

The results of the enrichment analyses were grouped using the functional annotation clustering feature of DAVID by integrating the kappa statistics algorithm and the fuzzy heuristic clustering algo-
rithm. The parameter was set to Classification Stringency: Medium, and Enrichment Thresholds: 0.05. Finally, the R FGNet package (http://bioconductor.org/packages/release/bioc/html/FGNet.html) (Aibar et al., 2015) was performed to visualize the functional modules.

**PPI network and module construction**

The PPIs among the DEGs were analyzed by applying the STRING (Szklarczyk et al., 2014) (version 10.0; http://www.string-db.org/) database, with a confidence of >0.7 being the required threshold. Subsequently, the PPI network was built and visualized by applying the Cytoscape software (version 3.4.0; http://chianti.ucsd.edu/cytoscape-3.4.0/) (Shannon et al., 2003). We then applied the CytoNCA (Yu and Tang, 2014) plugin (version 2.1.6; http://apps.cytoscape.org/apps/cytonca) to obtain the hub nodes and to further analyze topology properties (degree centrality, betweenness centrality, and closeness centrality) of the node network. Then, the major protein of the PPI network that participated in protein interaction was obtained based on the ranking of the topological property score of each node.

**MiRNA prediction and MicroRNA–target network construction**

Based on the miRWalk, miRanda, miRDB, miRMap, miRNAMap, RNA22, and Targetscan databases, the miRNAs targeting DEGs were identified by applying the miRWalk2.0 (http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/) tool (Dweep and Gretz, 2015). We assume that the predicted miRNAs can be considered to have high reliability when they appeared in more than 6 of the aforementioned databases. Then, the miRNA–mRNA relationship pair was acquired, and miRNAs regulated 15 or more target genes were selected. Finally, the miRNA–target network was constructed by applying the Cytoscape software.

**TF prediction and TF–target network construction**

The TRRUST online tool (Han et al., 2017) (version 2; http://www.grnpedia.org/trrust/) was used to predict TFs, with the threshold value of q < 0.05. Then, the intersection of the predicted TF with DEGs was initiated, and the TF related to disease occurrence was obtained. Finally, we applied the Cytoscape software to build the TF–target network.

**MiRNA–TF–target network construction**

The miRNA–TF–target network was built by integrating the miRNA–target and TF–target interaction pairs acquired above using the Cytoscape software.

**Cerebral hemorrhage-related small-molecule drugs**

The drug-specific gene expression profiles in the CMap reference database (http://www.broadinstitute.org/cmap/) can find the connections between small-molecule drugs and diseases, when submitting the genes that may be related to a specific disease. Any identified DEGs between the contralateral tissue of cerebral hemorrhage and the normal tissue surrounding the hematoma were divided into the upregulated and downregulated groups and were transformed into a probe set on the HG-U133A platform. Then, DEGs were mapped to the CMap database, and the enrichment values representing the degree of similarity were obtained. The enrichment value ranged between −1 and 1, and a value closer to 1 demonstrated that the small-molecule compound can simulate the state of normal tissue, while those closer to −1 indicate that the small-molecule compound can simulate the state of the cerebral hemorrhage. A P value < 0.01 and |enrichment value| > 0.8 were chosen as the thresholds.

**RESULTS**

**Data preprocessing, PCA, and cluster analysis**

The distribution of the expression values is displayed before and after preprocessing (Fig. 1A, B). Finally, a total of 20,514 genes were annotated and the PCA (Fig. 1C) and hierarchical clustering (Fig. 1D) results show that all samples were more clearly grouped together according to the pregrouping, except for the GSM596850 (PH) sample, which was therefore excluded from further analysis.

**Identification of DEGs**

In total, there are 2,203 DEGs obtained in the CG vs. PH group, including 1,388 upregulated and 815 downregulated. Moreover, there are 1,095 DEGs obtained in the CW vs. PH group, containing 400 upregulated and 695 downregulated, with a P value < 0.05 and log₂FC > 1 set as the cutoff value. The co-upregulated or co-down-
regulated DEGs were screened via intersected DEGs of the CG vs. PH and CW vs. PH group for further analysis. Then, in total, there are 433 common DEGs (including 359 uniformly downregulated and 74 uniformly upregulated) were obtained (Fig. 2). After excluding 2,383 DEGs in the CG vs. CW group, 408 DEGs were obtained, and are related to disease occurrence as they are not affected in the contralateral area (Fig. 2A). Bidirectional hierarchical clustering of DEGs is shown in Fig. 2B. The PH and contralateral samples were clearly separated based on the gene expression profile, while the CG and CW samples were not distinguished, suggesting that these DEGs were significantly related to the occurrence of cerebral hemorrhage after removing the influence of the contralateral area.

### Functional enrichment analysis of DEGs

The GO-BP and KEGG pathway enrichment analysis of DEGs identified 143 GO-BP terms and 42 KEGG pathways. The top 10 GO-BP terms and top 10 KEGG terms are displayed in Fig. 3. According to the functional annotation clustering feature of DAVID, five functional mod-
ules were obtained (cl 1: chemokine-related, cl 2: antigen immune-related, cl 3: pathogen infection, cl 4: cell reaction, and cl 5: positive regulation of tyrosine phosphorylation and MAPK cascade) (Fig. 4A). In addition, 29 versatile genes associated with two or more modules are shown in Fig. 4B, such as IL-6 involved in modules 2, 3, 4, and 5. A distance heat map of all modules was produced to analyze the similarity of the modules (Fig. 4C). In a distance heat map, the closer the two modules were, the higher the similarity was, and the more the number of genes shared. As shown (Fig. 4C), modules 1 and 4 and modules 4 and 5 exhibited higher similarity.

PPI network and module construction

In total, there are 755 protein interaction pairs were obtained, and the PPI network containing 228 nodes (9 upregulated and 219 downregulated) was constructed. After the topological property analysis of the PPI network, the top 10 genes with supreme DC, BC, or CC scores, such as IL-6 and VEGF-A, are shown in Table 1. The nodes with the highest DC, BC, or CC scores in the PPI network may play crucial roles in ICH.

The miRNA–TF–target network

There are 3,466 miRNA–target relationship pairs were calculated totally, among them, 11 miRNAs were able to regulate 15 or added target genes, and 189 miRNA–target relationship pairs were found between 11 miRNAs and 84 targets (23 upregulated and 61 downregulated). In addition, 48 TF–target relationship pairs were calculated, contains 6 TFs (5 downregulated and 1 upregulated) and 30 downregulated target genes. Integrating the 189 miRNA–target pairs and 48 TF–target pairs mentioned previously, 237 miRNA–TF–target regulatory relationships were obtained, including six TFs (1 upregulated and 5 downregulated), 11 miRNAs, and 105 target genes (22 upregulated and 83 downregulated) (Fig. 5).

Cerebral hemorrhage-related small-molecule drugs

A total of 23 small-molecule drugs with highly significant correlations with cerebral hemorrhage were obtained in the study (Table 2). The enrichment values of Prestwick-1083, xamoterol, ifosfamide, methylidopate,
nifurtimox, propranolol, and methoxamine were greater than 0.8, indicating that these 7 small-molecule drugs were potential therapeutic drugs for the management of cerebral hemorrhage. In addition, small-molecule drugs with enrichment values of less than −0.9, such as doxorubicin, menadione, and azacitidine, exhibited highly significant negative scores, indicating potential pathogenic substances for ICH.

**DISCUSSION**

ICH is the second most common type of hemorrhagic stroke, with a 30-day mortality estimated to be 44%–52% (Dekker et al., 2018). The research presented herein identified 408 overlapped DEGs in the CG vs. PH, CW vs. PH, and CG vs. CW groups. Prior to our study, 35 common DEGs have been identified in the PH vs. CG and PH vs. CW groups, although this was with tighter statistical parameters (Yang et al., 2014). We set our cutoff values at \( p \) value < 0.05 and \(|\log\text{FC}| > 1\), which enlarged the DEG search scope. Furthermore, we identified the overlapped genes in the CG vs. CW group, which ensured the accuracy of the DEG analysis. DEGs also act as hub genes and play an important role in the PPI network and the miRNA–TF–target network. These crucial genes may participate in the underlying mechanism of ICH, and therefore our study could be used to further illustrate the etiology and pathogenesis of ICH.

Functional enrichment analysis revealed five modules, including chemokine-related process, antigen immune-related process, pathogen infection, cell reaction, and positive regulation of tyrosine phosphorylation and MAPK cascade. Similar to our research results, Carmichael et al. (2008) have demonstrated that immune responses and signal transduction were associated with ICH. The key genes of CXCL8 and CCL5 have previously been shown to be associated with chemokine signaling in the accumulation of inflammatory cells to the ICH site via their high-affinity receptors (Yao and Tsirka, 2012). Studies indicated that the serum concentration of CXC chemokine ligand-12 significantly increased after ICH (Shen et al., 2017). Serum levels of CCL2 and CXCL10 were also elevated post-ICH. Interestingly, brain tissue injuries in patients with ICH occur after an inflammatory reaction and these pro-inflammatory chemokines are associated with worse functional outcomes after ICH (Landreneau et al., 2018). ICAM1, IL-6, IL-1B, and VACM1 were found to be related to the antigen immune-related process. ICAM1 encodes a cell surface glycoprotein that was found to be highly expressed in rat brains after ICH (Zhi et al., 2007), together with IL-1B (Wagner et al., 2006). Furthermore,
Fig. 4. The network and heat map of five functional modules and genes. (A) An intersection network of the five modules. (B) The multifunctional genes in the modules. (C) Distance heat map for the five modules.

Table 1. Top 10 genes with highest degrees in protein–protein interaction network.

<table>
<thead>
<tr>
<th>Gene Names</th>
<th>DC</th>
<th>Gene Names</th>
<th>BC</th>
<th>Gene Names</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL-6</td>
<td>43</td>
<td>VEGFA</td>
<td>7411.906</td>
<td>IL-6</td>
<td>0.038267028</td>
</tr>
<tr>
<td>CXCL8</td>
<td>39</td>
<td>IL6</td>
<td>7123.1206</td>
<td>CXCL8</td>
<td>0.03822192</td>
</tr>
<tr>
<td>VEGFA</td>
<td>36</td>
<td>VCAM1</td>
<td>5109.4194</td>
<td>VEGFA</td>
<td>0.038100034</td>
</tr>
<tr>
<td>IL-1B</td>
<td>32</td>
<td>CXCL8</td>
<td>4742.5576</td>
<td>CCL5</td>
<td>0.038061704</td>
</tr>
<tr>
<td>CCL5</td>
<td>31</td>
<td>IL-1B</td>
<td>4301.455</td>
<td>CCL5</td>
<td>0.038055323</td>
</tr>
<tr>
<td>CXCL12</td>
<td>28</td>
<td>TYROBP</td>
<td>3997.5818</td>
<td>ICAM1</td>
<td>0.038017083</td>
</tr>
<tr>
<td>VCAM1</td>
<td>27</td>
<td>ITGAM</td>
<td>3485.4338</td>
<td>VCAM1</td>
<td>0.0380170716</td>
</tr>
<tr>
<td>ICAM1</td>
<td>27</td>
<td>ITGAS5</td>
<td>3059.5947</td>
<td>CXCL12</td>
<td>0.038010716</td>
</tr>
<tr>
<td>CXC4</td>
<td>27</td>
<td>ICAM1</td>
<td>2807.8218</td>
<td>CCL2</td>
<td>0.037978917</td>
</tr>
<tr>
<td>PPBP</td>
<td>26</td>
<td>CCL5</td>
<td>2403.2915</td>
<td>PTPRC</td>
<td>0.037896495</td>
</tr>
</tbody>
</table>

*Note: DC, Degree Centrality; BC, Betweenness Centrality; CC, Closeness Centrality.*
VCAM1 may be involved in neuronal apoptosis and the pathophysiology of ICH (Dongmei et al., 2015). The inflammatory cytokine IL-6 is involved in neuronal survival and axonal regeneration following ICH (Xiong and Yang, 2015), although its upregulation is associated with the clinical conditions of ICH patients (Antunes et al., 2010). In addition, IL-6 and VEGF-A were associated with positive regulation of tyrosine phosphorylation and MAPK cascade. Overexpressed VEGF may play a pathophysiological role in metastatic brain tumor-associated ICH (Shin et al., 2006). These genes were also hub genes in the PPI network. Moreover, in the miRNA–TF–target network, these key DEGs, except VCAM1, were all regulated by one or more TFs and miRNAs. The results demonstrate the significant roles of these genes in the progression of ICH.

Additionally, the CMap database was applied to analyze small-molecule drugs that may reverse the expression of ICH genes. Seven small-molecule compounds, namely Prestwick-1083 (enrichment value 0.965), xamoterol (enrichment value 0.876), ifosfamide (enrichment value 0.867), methylidopate (enrichment value 0.842), nifurtimox (enrichment value 0.833), propranolol (enrichment value 0.824), and methoxamine (enrichment value 0.816) were identified to be candidate anti-ICH drugs. This discovery could contribute to a more targeted study of potential reversing agents for ICH.

Adoption, doxorubicin (enrichment value -0.993), menadione (enrichment value -0.990), and azacitidine (enrichment value -0.976) may be involved in the progression of ICH. Ardestani et al. (2017) have demonstrated that xamoterol can reduce neuroinflammation in the brain of mice. Phelan et al. (2015) have found that β-adrenergic antagonists have a certain neuroprotective effect on ischemic stroke. Wajngarten and Silva (2019) have summarized that labetalol, nicardipine or sodium nitroprusside can be used for the treatment of hypertensive encephalopathy, and drugs that affect the central nervous system should be avoided, such as nifurtimox. Doxorubicin is a chemotherapy medication for cancers.
and may induce congestive heart failure (Von Hoff et al., 1979). Moreover, neonatal brain damage is one of the adverse outcomes of menadione, and Azacitidine is another potential chemotherapeutic agent for cancers. At present, there are no reports on the relationship between these molecules and ICH. Additionally, the targets of the molecules participating in ICH need to be further investigated.

However, this study is primarily based on a bioinformatics analysis. Therefore, inevitable limitations exist. First, this study was based on only one dataset, as there were no other qualified datasets available in the GEO database. Second, the key DEGs and small-molecule compounds identified in this study were not validated using laboratory experiments because of limited experimental conditions. In future studies, the authors will provide up-to-date literature searches and verify the results of this study.

CONCLUSION

In conclusion, key DEGs including CCL5, CXCL8, ICAM1, IL-1β, IL-6, VCAM1, and VEGF-A play essential roles in ICH. Of these, CCL5, CXCL8, and VEGF-A have not been identified previously as candidate susceptibility genes for ICH. Some small-molecule drugs may be useful for the treatment of ICH.

REFERENCES


Antunes AA, Sotomaior VS, Sakamoto KS, de Camargo Neto CP, Mar‑

aconditions, including the regulation of cell adhesion and migration. These results suggest that the identified molecular targets may serve as potential biomarkers and drug targets for ICH.

However, this study is primarily based on a bioinformatics analysis. Therefore, inevitable limitations exist. First, this study was based on only one dataset, as there were no other qualified datasets available in the GEO database. Second, the key DEGs and small-molecule compounds identified in this study were not validated using laboratory experiments because of limited experimental conditions. In future studies, the authors will provide up-to-date literature searches and verify the results of this study.

CONCLUSION

In conclusion, key DEGs including CCL5, CXCL8, ICAM1, IL-1β, IL-6, VCAM1, and VEGF-A play essential roles in ICH. Of these, CCL5, CXCL8, and VEGF-A have not been identified previously as candidate susceptibility genes for ICH. Some small-molecule drugs may be useful for the treatment of ICH.

REFERENCES


