Transcriptome network-based method to identify genes associated with unruptured intracranial aneurysms


1Department of Neurosurgery, East Hospital, Tongji University School of Medicine, Shanghai, China
2Department of Neurosurgery, The Second Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China

*These authors contributed equally to this study.
Corresponding authors: Z.Y. Sun / W. Xu
E-mail: weiliangdf@163.com

Received January 23, 2013
Accepted March 12, 2013
Published September 3, 2013
DOI http://dx.doi.org/10.4238/2013.September.3.2

ABSTRACT. Intracranial aneurysm is a balloon or sac-like dilatation of blood vessels inside the brain. Despite their importance, the biological mechanisms of intracranial aneurysms are not totally understood. We used public genome-wide gene expression profile data to identify potential genes that are involved in intracranial aneurysm in order to construct a regulation network. Some of the transcription factors and target genes that we identified in this network had been identified as related to intracranial aneurysm in previous studies. We found additional transcription factors and target genes that are apparently related to intracranial aneurysm with this method. The confirmation of previously identified genes and transcription factors supports the usefulness of this transcriptome network analysis for the identification of candidate genes involved in intracranial aneurysm.

Key words: Unruptured intracranial aneurysms; Transcription factors; Transcriptome network
INTRODUCTION

An intracranial aneurysm (IA) is a balloon or sac-like dilatation of blood vessels inside the brain. These vascular anomalies pose potential risk of rupture usually resulting in bleeding into the meninges or the brain itself, leading to a subarachnoid hemorrhage (SAH) or intracranial hematoma (ICH), either of which constitutes a stroke (Zhang and Claterbuck, 2008). Rebleeding, hydrocephalus, vasospasm, or multiple aneurysms may also occur. A common location of cerebral aneurysms is in the arteries at the base of the brain. Aneurysms may result from congenital defects, preexisting conditions such as high blood pressure and atherosclerosis, or head trauma (Li et al., 2012).

Many genes have been identified associated with IA on the basis of many reports: the extracellular matrix (ECM) turnover factors, such as the tissue inhibitors of metalloproteinase-3, osteoblast-specific factor-2, connective tissue growth factor; the ECM components including fibronectin, several collagens (COL3A1, COL1A1, COL1A2, COL6A1, COL6A2), and elastin; cell adhesion and antiadhesion factors including osteonectin and hevin; the cytokine cdc-rel2a/PNUTL2; and the cell migration factor tetraspanin-5 (Skirgaudas et al., 1996; Peters et al., 2001; Zhang and Claterbuck, 2008).

Efforts of intensive research to increase our understanding of the molecular basis of IA have been made, and there is growing evidence showing that transcription factors (TFs) play a critical role in disease by regulating the transcription and expression of many genes. An understanding of these new developments in TFs and pathways may pave the way for innovative combinatorial approaches for treatment of IA. However, most previous studies on transcription are primarily on proteins and mRNA expression, and high-throughput functional analysis of multiple TFs in IA is rare. In this research, we aimed to identify potential candidate genes associated with IA by transcriptional network analysis and to explore new predictive biomarkers and target genes.

MATERIAL AND METHODS

Data source

**Affymetrix microarray data**

Transcription profiles of IA GSE26969 were obtained from the public functional genomics data repository Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/), which are based on the Affymetrix platform data (Wachi et al., 2005).

**Pathway data**

Kyoto Encyclopedia of Genes and Genomes (KEGG) is a collection of online databases dealing with genomes, enzymatic pathways, and biological chemicals (Kanehisa, 2002). PATHWAY database records networks of molecular interactions in cells, and variants of those specific to particular organisms (http://www.genome.jp/kegg/). A total of 130 pathways, involving 2287 genes, were collected from KEGG.
Regulation data

There are approximately 2600 proteins in the human genome that contain DNA-binding domains, and most of these are presumed to function as TFs (Wachi et al., 2005). The combinatorial use of a subset of approximately 2000 human TFs easily accounts for the unique regulation of each gene in the human genome during development (Brivanlou and Darnell, 2002).

These TFs are grouped into 5 superclass families, on the basis of the presence of conserved DNA-binding domains. TRANSFAC database contains data on TFs, their experimentally proven binding sites, and regulated genes (Wingender, 2008).

Transcriptional Regulatory Element Database (TRED) has been built in response to increasing needs of an integrated repository for both cis- and trans-regulatory elements in mammals (Jiang et al., 2007). TRED has done the curation for transcriptional regulation information, including transcription factor-binding motifs and experimental evidence. Curation is currently focusing on target genes of 36 cancer-related TF families.

A total of 774 pairs of regulatory relationships between 219 TFs and 265 target genes were collected from TRANSFAC (http://www.gene-regulation.com/pub/databases.html). A total of 5722 pairs of regulatory relationships between 102 TFs and 2920 target genes were collected from TRED (http://rulai.cshl.edu/TRED/). Combined, the 2 regulation datasets, totaling 6328 regulatory relationships between 276 TFs and 3002 target genes, were collected (Table 1).

<table>
<thead>
<tr>
<th>Source</th>
<th>Regulation</th>
<th>TFs</th>
<th>Targets</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRED</td>
<td>5722</td>
<td>102</td>
<td>2920</td>
<td><a href="http://rulai.cshl.edu/TRED/">http://rulai.cshl.edu/TRED/</a></td>
</tr>
<tr>
<td>Total</td>
<td>6328</td>
<td>276</td>
<td>3002</td>
<td></td>
</tr>
</tbody>
</table>

TFs = transcription factors; TRED = transcriptional regulatory element database.

Differentially expressed gene (DEG) analysis

For the GSE26969 dataset, the Limma method (Smyth, 2004) was used to identify DEGs. The original expression datasets from all conditions were processed into expression estimates using the RMA method with the default settings implemented in Bioconductor, and the linear model was then constructed. The DEGs with a fold-change greater than 2 and P value less than 0.05 were selected.

Co-expression analysis

For demonstrating the potential regulatory relationship, the Pearson correlation coefficient (PCC) was calculated for all pairwise comparisons of gene expression values between TFs and the DEGs. The regulatory relationships with absolute Pearson correlation coefficient greater than 0.6 were considered to be significant.

Gene ontology (GO) analysis

The BiNGO analysis (Maere et al., 2005) was used to identify overrepresented GO categories in biological processes.
Regulation network construction

Using the regulation data collected from TRANSFAC and TRED databases, we matched the relationships between differentially expressed TFs and their differentially expressed target genes.

On the basis of the above regulation datasets and the pathway relationships of the target genes, we built the regulation networks using Cytoscape (Shannon et al., 2003). According to the significant relationships (PCC > 0.6 or PCC < -0.6) between TFs and their target genes, 16 putative regulatory relationships were built. The result of the regulation network is shown in Figure 1. To visualize the relationship between KEGG pathways and TFs, KEGG pathways were enriched among the DEG genes in the regulatory network.

Figure 1. Regulation network construction in intracranial aneurysm.

Significance analysis of pathway

We adopted an impact analysis that included the statistical significance of the set of pathway genes and that also considered other crucial factors such as the magnitude of each gene’s expression change, the topology of the signaling pathway, their interactions, etc. (Draghici et al., 2007). In this model, the impact factor (IF) of a pathway $P_i$ is calculated as the sum of two terms:

$$IF(P_i) = \log \left( \frac{1}{p_i} \right) + \frac{\sum |PF(g)|}{\Delta E \cdot N_{a}(P_i)}$$
The first term is a probabilistic term that captures the significance of the given pathway $P_i$ from the perspective of the set of genes contained in it. It is obtained using the hypergeometric model in which $p_i$ is the probability of obtaining at least the observed number of differentially expressed genes ($N_{de}$) just by chance (Tavazoie et al., 1999; Draghici et al., 2003).

The second term is a functional term that depends on the identity of the specific genes that are differentially expressed as well as on the interactions described by the pathway (i.e., its topology). The second term sums up the absolute values of the perturbation factors (PFs) for all genes $g$ on the given pathway $P_i$.

The PF of a gene $g$ is calculated as follows:

$$PF(g) = \Delta E(g) + \sum_{u \in U_g} \beta_{ug} \cdot \frac{PF(u)}{N_d(u)}$$

In this equation, the first term $\Delta E(g)$ captures the quantitative information measured in the gene expression experiment. The factor $\Delta E(g)$ represents the normalized measured expression change of the gene $g$. The first term $\Delta E(g)$ in the above equation is a sum of all PFs of the genes $u$ directly upstream of the target gene $g$, normalized by the number of downstream genes of each such gene $N_{ds}(u)$, and weighted by a factor $\beta_{ug}$, which reflects the type of interaction: $\beta_{ug} = 1$ for induction, $\beta_{ug} = -1$ for repression (KEGG supplies this information about the type of interaction of 2 genes in the description of the pathway topology). $U_g$ is the set of all such genes upstream of $g$. We needed to normalize with respect to the size of the pathway by dividing the total perturbation by the number of differentially expressed genes in the given pathway, $N_{de}(P_i)$. To make the IFs as independent as possible from the technology, and also comparable between problems, we also divided the second term in Equation 1 by the mean absolute fold-change $\Delta E$, calculated across all differentially expressed genes. The result of the significance analysis of the pathway is shown in Table 2.

<table>
<thead>
<tr>
<th>Pathway name</th>
<th>Impact factor</th>
<th>% Pathway genes in input</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adherens junction</td>
<td>31.077</td>
<td>28.205</td>
<td>8.17E-04</td>
</tr>
<tr>
<td>Phosphatidylinositol signaling system</td>
<td>29.861</td>
<td>17.105</td>
<td>0.298667</td>
</tr>
<tr>
<td>Ribosome</td>
<td>27.101</td>
<td>58.416</td>
<td>5.69E-12</td>
</tr>
<tr>
<td>Circadian rhythm</td>
<td>20.59</td>
<td>15.385</td>
<td>0.534339</td>
</tr>
<tr>
<td>Parkinson’s disease</td>
<td>17.006</td>
<td>27.737</td>
<td>2.84E-07</td>
</tr>
<tr>
<td>Huntington’s disease</td>
<td>12.773</td>
<td>24.339</td>
<td>1.45E-05</td>
</tr>
<tr>
<td>Focal adhesion</td>
<td>12.754</td>
<td>25.123</td>
<td>2.19E-05</td>
</tr>
<tr>
<td>Ubiquitin mediated proteolysis</td>
<td>12.462</td>
<td>27.536</td>
<td>1.02E-05</td>
</tr>
<tr>
<td>Alzheimer’s disease</td>
<td>11.519</td>
<td>23.596</td>
<td>6.19E-05</td>
</tr>
<tr>
<td>Regulation of actin cytoskeleton</td>
<td>10.034</td>
<td>22.581</td>
<td>2.58E-04</td>
</tr>
</tbody>
</table>

### RESULTS

#### Regulation network construction in IA

To get pathway-related DEGs of IA, we obtained the publicly available microarray
dataset GSE26969 from GEO. After microarray analysis, the differentially expressed genes with a fold-change greater than 2 of GSE26969 and P value less than 0.05 were selected. A total of 3661 genes were selected as DEGs from GSE26969. To get the regulatory relationships, the co-expressed value (PCC ≥ 0.6) was chosen as the threshold. Finally, we obtained 16 regulatory relationships between 7 differentially expressed TFs and their 15 differentially expressed target genes. By integrating the regulatory relationships above, a regulation network of IA was built between TFs and their target genes (Figure 1). In this network, FLI1, ETS2, and STAT1 showed higher degrees to form a local network, which suggested that these genes may play an important role in IA.

GO analysis of the regulation network in IA

Several GO categories were enriched among these genes in the regulatory network, including response to chemical stimulus, response to external stimulus, response to stimulus, response to extracellular stimulus, positive regulation of biological process, and others (Figure 2, list of top 10).

![Gene ontology biological process analysis.](image)

Figure 2. Gene ontology biological process analysis.

Significant pathway in IA

Several significant pathways were enriched among these genes in the regulatory network, including adherens junction, phosphatidylinositol signaling system, ribosome, circadian rhythm, Parkinson’s disease, and others (Table 2).
Regulation network between TFs and pathways in IA

To further investigate the regulatory relationships between TFs and pathways, we mapped DEGs to pathways and obtained a regulation network between TFs and pathways (Figure 3). In the network, XBP1, STAT1, and SMAD2 were shown as hub nodes linked to IA-related pathways.

**Figure 3.** Regulation network between transcription factors and pathways in intracranial aneurysm.

**DISCUSSION**

From the results of regulation network construction in IA, we found that many TFs and pathways closely related to IA were linked by our methods. The genes JUN, Fli-1, ETS2, XBP1, and others appearing in our transcriptome network have been proved to be closely related to IA in previous study.

JUN is the putative transforming gene of avian sarcoma virus 17. It encodes a protein that is highly similar to the viral protein, and that interacts directly with specific target DNA sequences to regulate gene expression. Phosphorylated c-Jun is increased in vascular walls of cerebral aneurysms (Takagi et al., 2002).

Fli-1, a member of the ETS family of DNA-binding TFs, is involved in cell proliferation and tumorigenesis. Fli1 knock-out mice die during embryogenesis with a loss of vascular integrity leading to cerebral hemorrhage, suggesting that Fli1 is involved in the regulation of genes critical for vascular remodeling (Asano and Trojanowska, 2009).
STAT1 is a member of the STAT protein family. In response to cytokines and growth factors, STAT family members are phosphorylated by receptor-associated kinases, and then form homo- or heterodimers that translocate to the cell nucleus where they act as transcription activators. STAT1 is found to be phosphorylated in rat basilar artery after SAH, suggesting that STAT1 signaling may contribute to morphological arterial wall changes in cerebral vasospasm (Osuka et al., 2010).

ETS2 regulates numerous genes and is involved in stem cell development, cell senescence and death, and tumorigenesis. The conserved ETS domain within these proteins is a winged helix-turn-helix DNA-binding domain that recognizes the core consensus DNA sequence GGAA/T of target genes. Some reports suggest that apoptosis may be a therapeutic target for acute brain injury after SAH. Apoptosis has been observed in human patients with SAH, and ETS2 has been induced in endothelial cells to promote the apoptosis process (Lunardi et al., 2007).

XBP1 encodes a transcription factor that regulates MHC class II genes by binding to a promoter element referred to as an X box. After transient cerebral ischemia, unfolded proteins are abundantly accumulated in the endoplasmic reticulum (ER). The genes coding for ER stress proteins are induced to be expressed. The genetic response is based on the endonuclease inositol-requiring enzyme 1-induced processing of xbp1 messenger RNA (mRNA), resulting in the synthesis of a new XBP1proc protein that functions as a potent transcription factor for ER stress genes (Paschen et al., 2003). It has been found that the mRNA of this gene is processed to an active form by an unconventional splicing mechanism that is mediated by the endonuclease inositol-requiring enzyme 1. The XBP1 (-116C/G) gene polymorphism, especially in the homozygous state, is suggested to play an active role in ER stress-related ischemia/reperfusion injury (Yilmaz et al., 2010).

Adducins are a family of cytoskeleton proteins encoded by 3 genes (alpha, beta, and gamma). ADD1 belongs to alpha adducin. ADD1 G/W460 has been significantly associated with ICH, showing higher expression in an intracerebral hemorrhage sample than in the control (Dou et al., 2004). ADD1 is found in primary spontaneous intracerebral hemorrhage, especially nonlobar hematoma. Of the ADD1 variant genotypes, the GW genotype (1.9%) is found to be more associated with hemorrhage than WW (42.7%) in patients (Kalita et al., 2011).

Tenascin-C (TNC), a large ECM glycoprotein with a molecular weight of 180-250 kDa, is present in several normal adult tissues. TNC is upregulated in brain tumors (Leins et al., 2003). Overexpressed TNC has been found in trileaflet aortic valve aneurysms, suggesting that it is strongly associated with this disease (Majumdar et al., 2007).

Type I collagen found in ECM of the intimal, medial, and adventitial layers of the aortic wall, is responsible for the tensile strength of the vessel. COL1A1 has been shown to be overexpressed in IA patients, suggesting that IA undergoes significant tissue remodeling (Peters et al., 2001). Col1a1-mutated mice exhibit thinner cerebral artery wall (Marjamaa et al., 2006).

The BARD1 protein can interact with the N-terminal region of BRCA1. On the other hand, it shares homology with the 2 most conserved regions of BRCA1: the N-terminal RING motif and the C-terminal BRCT domain. The RING motif is a cysteine-rich sequence found in a variety of proteins that regulate cell growth, including the products of tumor suppressor genes and dominant protooncogenes. BARD1 has been found to be downregulated in IA (Pera et al., 2010).

HSPB1 (heat shock protein 27) plays a major role in mediating survival responses to a range of central nervous system insults, functioning as a protein chaperone and an antioxidant,
and acting through inhibition of cell death pathways. The infarct size is detected to be reduced in overexpressing HSP27tg mice compared with wild-type littermates in cerebral ischemia condition. This result indicates that HSP27 has powerful neuroprotective effects in cerebral ischemia (van der Weerd et al., 2010). HSPB1 is also identified to be expressed as a candidate gene for IA by genetic mapping analysis (Weinsheimer et al., 2007).

Von Willebrand factor (vWF) plays an important role in the coagulation system. It affects platelet aggregation at the site of vessel endothelium damage. vWF is elevated in the plasma and cerebrospinal fluid of patients after SAH. Also, vWF elevation in cerebrospinal fluid is correlated with the clinical condition of patients and also the presence of cerebral vasospasm and cerebral infarct, suggesting that vWF may play an important role in the pathogenesis of human cerebral vasospasm and ischemic complications (Beuth et al., 2001; McGirt et al., 2002).

Adherens junction pathway is the most significant biological pathway for IA. Adherens junctions mainly maintain the integrity of the vessel wall and are also involved in endothelium remodeling in various physiological and pathological processes. The analysis of brain tissue sections in IA has shown that large intracranial vessels and microvessels can express the molecular components of adherens-type junctions, including catenins, plakoglobin, vinculin, and alpha-actinin (Navratil et al., 1997). Adherens junction proteins can transfer intracellular signals to catenin, which has the ability to translocate to the nucleus and associate with TFs to regulate the expression of genes involved in endothelial cell growth, differentiation, and apoptosis (Weinsheimer et al., 2007).

Phosphatidylinositol 3-kinases (PI3-K) belong to a group of lipid kinases that catalyze phosphorylation in the D-3 position of the inositol ring of phosphatidylinositol. It is well established that PI3-K play a pivotal role in cell signaling and in the regulation of cellular functions including proliferation, differentiation, cytoskeletal organization, membrane trafficking, and apoptosis (Kimura et al., 2002). Of them, cell signaling is activated by the tyrosine kinase receptor, such as serine-threonine kinase, Akt. Growth factors activate Akt and promote cell survival through the PI3-K pathway. DNA fragmentation at 24 h is significantly increased in the PI3-K inhibitor, LY294002-treated animals compared with the vehicle-treated animals, suggesting that inhibition of Akt/glycogen synthase kinase-3 signaling through the PI3-K pathway increases acute brain injury after SAH (Endo et al., 2006).

The percentage of polyribosome-forming ribosomes with respect to total ribosomes (aggregation rate) is a morphological index of the activity of protein synthesis. The aggregation rate is found to be significantly decreased in smooth muscle cells of canine basilar arteries after SAH. The contraction-inhibiting proteins are also markedly reduced after the initial SAH to cause severe vasospasm. These abnormalities may cooperate to cause cerebral arterial narrowing accompanied by degeneration of smooth muscle cells after SAH (Oka et al., 1996).

Circadian rhythm in the onset of SAH is significant. The risk of aneurysmal SAH is low during the night (from 12:00 pm to 6:00 am) and at noon and high during working hours. This circadian fluctuation parallels that of blood pressure and suggests that an increase in blood pressure is a risk factor for aneurysmal rupture (Vermeer et al., 1997; Inagawa et al., 2000).

A patient with a right temporal hematoma and an overlying subdural hematoma following rupture of a right middle cerebral artery aneurysm is found to develop a levodopa-responsive parkinsonian syndrome, finally involving the right limbs (Turjanski et al., 1997). An 18F-dopa positron-emission tomography shows a marked reduction of uptake in the left
putamen of this patient, raising the possibility that the intracranial hemorrhage unmasks latent Parkinson’s disease. Parkinsonism resulting from unruptured giant aneurysms is rare. A patient with parkinsonism due to a giant middle cerebral aneurysm, which is abolished after surgical excision of the aneurysm, has been reported, but there is another case in the literature with asymmetric parkinsonism, due to internal carotid artery aneurysm, which does not improve after internal carotid artery ligation (Bostantjopoulou et al., 2006).

From the results of regulation network between TFs and pathways in IA, we found that many pathways closely related to IA were linked by our method. The basic understanding of the mechanisms underlying the functioning of IA genes is important. A deeper understanding of TFs and their regulated genes remains an area of intense research activity. Our regulation network is useful in investigating the complex interacting mechanisms of IA. However, further experiments are still needed to confirm this conclusion.

REFERENCES


Genes associated with unruptured intracranial aneurysms


