



Screening of key genes of unruptured intracranial aneurysms by using DNA microarray data analysis techniques

L. Chen, Y. Fan and J. Wan

Department of Neurosurgery, Ren Ji Hospital, Shanghai Jiao Tong University, Shanghai, China

Corresponding author: J. Wan
E-mail: jieqingwan@126.com

Genet. Mol. Res. 13 (1): 758-767 (2014)
Received August 11, 2013
Accepted September 18, 2013
Published January 31, 2014
DOI <http://dx.doi.org/10.4238/2014.January.31.2>

ABSTRACT. This study aimed to identify differentially expressed genes (DEGs) of unruptured intracranial aneurysms (IAs) and provide beneficial information for early diagnosis and treatment of IAs. The gene expression profile GSE26969 from the Gene Expression Omnibus database was downloaded, which included six human IA samples: three intracranial arterial aneurysm samples and three normal superficial temporal artery samples (control). Based on these data, we identified the DEGs between normal and disease samples with packages in the R language. The selected DEGs were further analyzed using bioinformatic methods. First, the STRING software was used to search co-expression relationships among DEGs, and the most important hub gene was found. We then used the plugins of the Cytoscape software, Mcode and Bingo, to conduct a module analysis. Next, pathways of the module genes were annotated using the FuncAssociate program. Compared with the control group, we obtained 169 DEGs in total, and by mining a module with the hub gene *MYH11*, we retrieved the *ACTA2*, *MYH11*, *MYLK*, and *MYL9* genes, which were all in the module and were most significantly related to vascular smooth muscle contraction. We hypothesize that the

genes identified here can be beneficial for early diagnosis and treatment of IAs.

Key words: Co-expression network; Module analysis; Pathway analysis; Expressed genes

INTRODUCTION

Intracranial aneurysms (IAs) are caused by common vascular abnormalities of the brain. The formation of IAs is assumed to be due to diverse exo- and endogenous factors such as inflammatory mechanisms, hypertension, smoking, excessive alcohol intake, and genetic factors (Hellingman et al., 2007). The intracranial vessels affected have a unique histological structure. They consist of a tunica intima with an endothelial layer and smooth muscle cells (SMCs), a tunica media with the internal elastic lamina and SMCs, and a tunica adventitia with loose connective tissue. Additionally, IAs usually develop branching points within the circle of Willis. The continuity of the tunica media shows gaps at these branching points, which are thought to be a predisposing location for the development of aneurysmal pouches (Chalouhi et al., 2012).

Although technical advances have resulted in dramatic progress in the treatment of IAs in recent years, the precise molecular mechanism by which IAs develop and rupture remains unknown. Three mitogen-activated protein kinases, c-Jun N-terminal kinase (JNK), p38, and extracellular signal-regulated kinase, are known to be involved in the IA wall (Szklarczyk et al., 2011). Furthermore, phosphorylation of p54 JNK is associated with IA size, phospho-p38 levels are associated with IA size and rupture, and phospho-Bad levels differ between ruptured and unruptured IAs (Laaksamo et al., 2008). Moreover, phospho-mTOR is down-regulated in ruptured IAs (Laaksamo et al., 2012).

There are approximately 20,000 aneurysm ruptures per year, and a prevalence of 3.2% in the general population of the United States (Juvola, 2011). Approximately 8% of these patients die before reaching the hospital, and half die within the first 30 days after rupture. Despite recent diagnostic and therapeutic advances, an episode of subarachnoid hemorrhage resulting from the rupture of IAs remains lethal in up to 65% of cases (Wiebers et al., 2003), and results in significant disabilities in 50% of those who survive (Nieuwkamp et al., 2009). The life threatening and debilitating sequence of IAs may be prevented if the aneurysm can be occluded before rupture.

Several high-throughput screening approaches for understanding the genetics of unruptured intracranial aneurysms (UIAs) have been conducted in recent years (Li et al., 2009). However, high-throughput screening studies that focus on functional identification of differential gene expression are rare. In this study, we performed a genomic microarray analysis based on UIA expression data, mined a module containing the hub gene *MYH11*, and identified the *ACTA2*, *MYLK*, and *MYL9* genes in this module, which were all significantly related to vascular smooth muscle contraction.

MATERIAL AND METHODS

Affymetrix microarray data

The gene chip GSE26969 was downloaded from the gene expression database Gene

Expression Omnibus (GEO) (Li et al., 2009), which contained data from six intracranial arterial tissue samples formed by three normal samples and three UIAs [GPL570 (HG-U133_Plus_2) platform with Affymetrix HU133 Plus 2.0 microarrays]. Annotation information of the chip probe was performed using Affymetrix, which contained the complete Affymetrix ATH1 (25K) chip probe. The complete original file and platform probe annotation information file were downloaded.

Data pre-processing and differential gene expression analysis

We first transferred the downloaded raw data to a recognizable expression profiling format, and then filled in and standardized the missing data (Troyanskaya et al., 2001; Fujita et al., 2006). Both treated and untreated samples were analyzed for differential gene expression using the Limma package in R language. Differentially expressed genes (DEGs) were selected based on $P < 0.01$ and $|\log_{2}FC| > 1$ thresholds (Smyth, 2005).

Constructing the protein-protein interaction (PPI) network

PPIs have been recognized as the main actors of cellular processes, and therefore great efforts have been made toward understanding their biological functions. Here, we used the online database resource Search Tool for the Retrieval of Interacting Genes (STRING; Szklarczyk et al., 2011) to predict interactions between DEGs based on characteristics of the input sequence, structure, and other information. The PPI network was built based on the scores calculated by the database of known functions.

Hub protein screening

Several studies have shown that the majority of biological networks obey scale-free network attributes (Jeong et al., 2000). This means that a small number of nodes in the network have a large number of connections while most nodes have only a few connections, and these few nodes represent the critical node of the network, or the hub (Yu et al., 2004). In this study, we analyzed the statistics and distribution of the hub based on the PPI network node. Using the scale-free nature of interactions in the protein network, we identified the center of the protein network.

PPI network functional modules

Proteins in PPI networks and in the same module usually have the same biological processes and function by co-expression. In this study, after obtaining network visualized by Cytoscape (Duncan et al., 2010), we used the plugins Mcode (Goldberg et al., 2007) (module parameters: degree cutoff ≥ 2 , K-core ≥ 2) and Bingo (Bernstein et al., 2012) to partition off and annotate functions of the whole network based on the hypergeometric distribution (adjusted $P < 0.01$).

Genes involved in the annotation pathway

Genetic data of each group was included in the construction of the regulatory network using the FuncAssociate software (Berriz et al., 2009). Annotation of DEGs in the pathway

involved in the module was accomplished by enrichment analysis, with the selected threshold value of $P < 0.05$. FuncAssociate is a web-based tool that accepts a list of genes as input and returns the input list of over- or under-represented Gene Ontology (GO) attributes. Only statistically significant over- or under-represented genes will be reported after multiple hypothesis testing.

RESULTS

Screening for DEGs

Differences in normalized expression data were compared after data pre-processing (Figure 1; box plots of normalized expression levels). A total of 169 genes met the difference threshold ($P < 0.01$ and $|\log_{2}FC| > 1$), which included 165 down-regulated and 4 up-regulated genes.

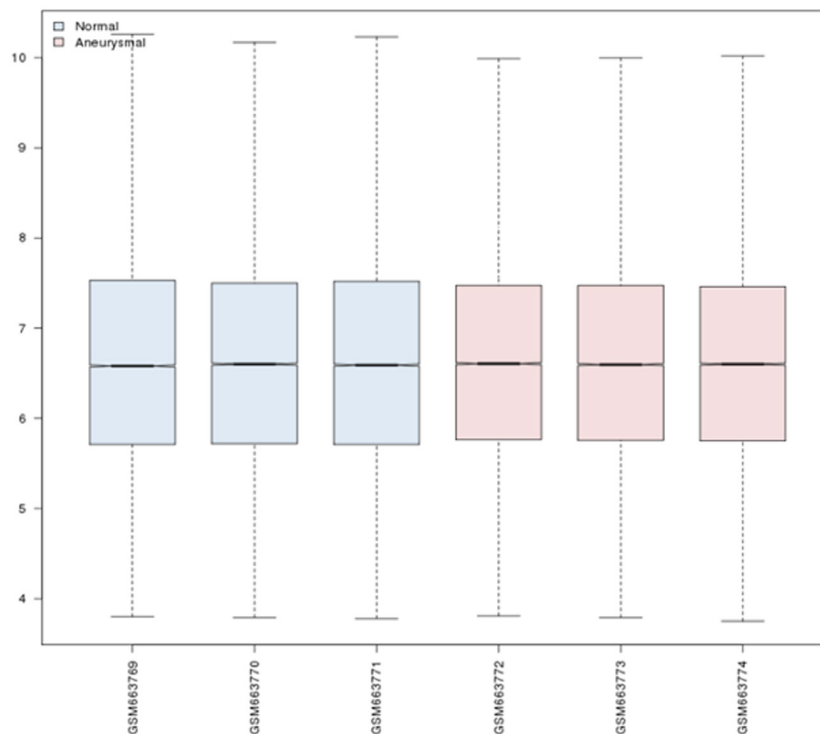


Figure 1. Box plots of normalized expression levels. The blue boxes show data of three normal intracranial samples and the pink boxes show data of patients with intracranial aneurysm samples. The black lines in boxes indicate the median of each set of data, whose distribution determined the degree of standardization of the data; black lines within the same straight line indicate a high degree of standardization.

Constructing the PPI network

Using the STRING software to search for PPIs of DEGs, we obtained 158 paired (106) differences in the gene interaction network (Figure 2; PPI network based on DEGs).

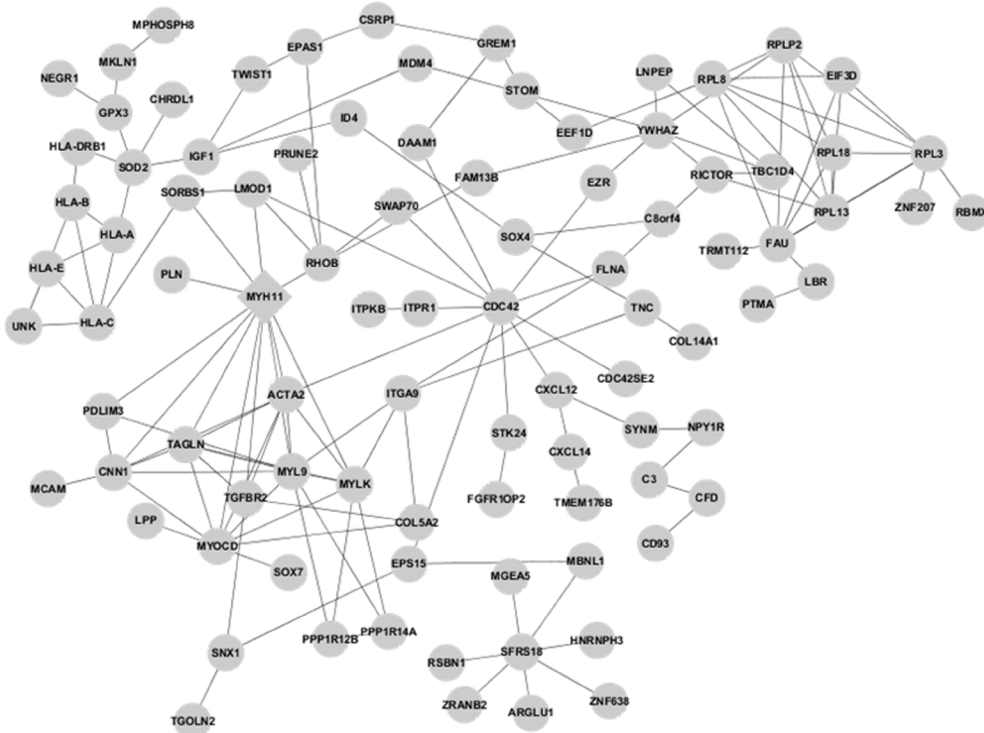


Figure 2. Protein-protein interaction network based on differentially expressed genes (DEGs). Hub genes are shown in the diamond-shaped node, and the circular node shows the other DEGs.

Hub protein screening

We analyzed the network structure of the resulting network. Topology analysis of the network revealed that the PPI network had a scale-free property, which indicated that a small number of nodes in the network had a large number of connections and most nodes had only a few connections (Figure 3; degree of nodes). By using the power law, we obtained the formula: $y = 61.309x^{-1.62}$ (Figure 3; degree of nodes). The number of nodes of the x-axis represents the degree of the node, i.e., the number of direct connections with the node, and the y-axis represents the number of nodes at each degree value. Statistical analysis of the node degree revealed that the highest node was the *MYH11* gene (node 12), suggesting that this gene plays an important role in the network.

Co-expression interaction network of functional modules

In this study, we used the Mcode plugin of Cytoscape to cluster the network module, and the Bingo plugin to divide and annotate the whole network function module based on the hypergeometric distribution (adjusted $P < 0.01$). We obtained a module that contained the hub gene *MYH11*, and retrieved eight notable functions of this module (Figure 4; modules and module functions of genes enriched in the node). These functions are listed in Table 1,

which shows that the *ACTA2*, *MYH11*, *MYLK*, and *MYL9* genes had close relationships with muscle contraction functions. The other significant functions also showed associations with the muscular system. These results indicate that the hub genes and their corresponding function modules mainly play important roles in the vascular muscular system.

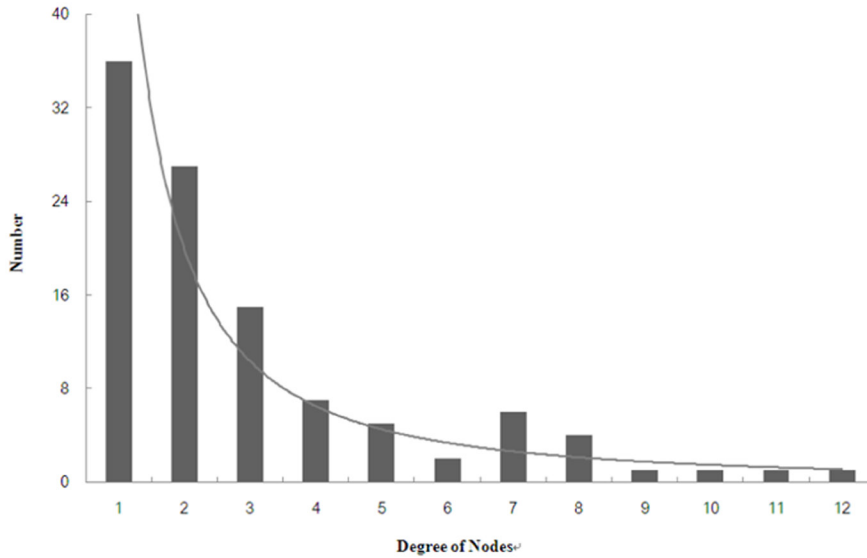


Figure 3. Degree of nodes. The x-axis indicates the degree of the node, which means the number of directly connected nodes with the node. The y-axis indicates the number of nodes at each degree value. As shown in the first column, in 38 cases, the network node degree was one.

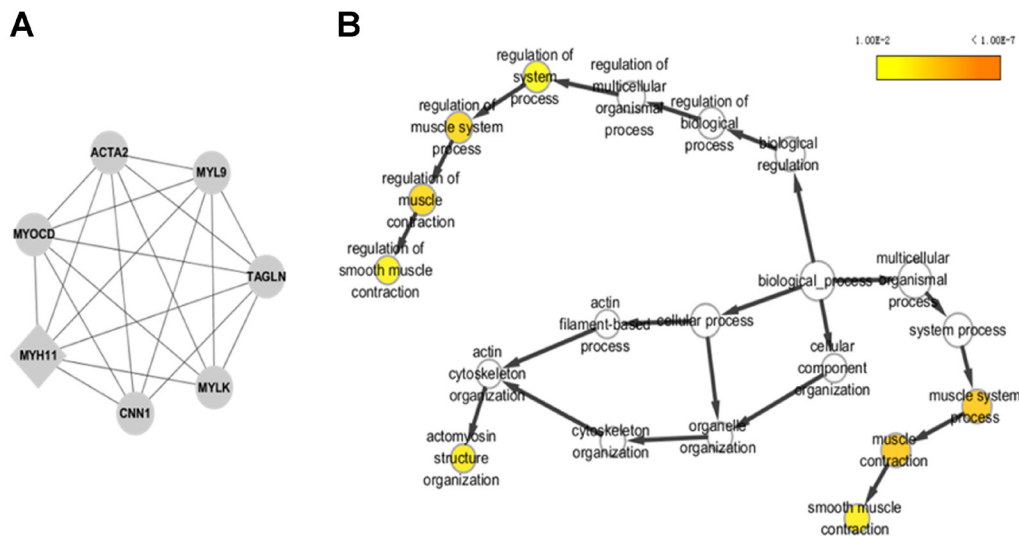


Figure 4. Modules and module functions of genes enriched in the node. **A.** The module of the hub genes was separated from the rest of the network. **B.** Enrichment of the node graph. Color indicates the size of the P value of the hypergeometric distribution. The upper right corner of the color strips indicates $0.01 < P < 10^{-7}$.

Table 1. Gene Ontology (GO) enrichment analysis of differentially co-expressed genes.

GO - ID	P value	Corrected P value	Description	Genes in test set
6936	4.41E-07	4.28E-05	Muscle contraction	ACTA2 MYH11 MYLK MYL9
3012	6.25E-07	4.28E-05	Muscle system process	ACTA2 MYH11 MYLK MYL9
6937	4.58E-06	2.09E-04	Regulation of muscle contraction	MYOCD CNN1 MYL9
90257	6.72E-06	2.30E-04	Regulation of muscle system process	MYOCD CNN1 MYL9
31032	5.17E-05	1.42E-03	Actomyosin structure organization	MYH11 CNN1
6939	7.71E-05	1.76E-03	Smooth muscle contraction	ACTA2 MYH11
6940	1.43E-04	2.80E-03	Regulation of smooth muscle contraction	MYOCD CNN1
44057	3.40E-04	5.82E-03	Regulation of system process	MYOCD CNN1 MYL9

Genes involved in the pathway annotation

Inputting the genetic information of the modules into the FuncAssociate software to analyze the pathway based on the hypergeometric distribution (threshold of $P < 0.05$) revealed that the module genes were involved in the vascular smooth muscle contraction pathway (Figure 5; vascular smooth muscle contraction pathway). The main function of vascular smooth muscle is to reduce the diameter by contraction in order to regulate blood flow and vascular pressure.

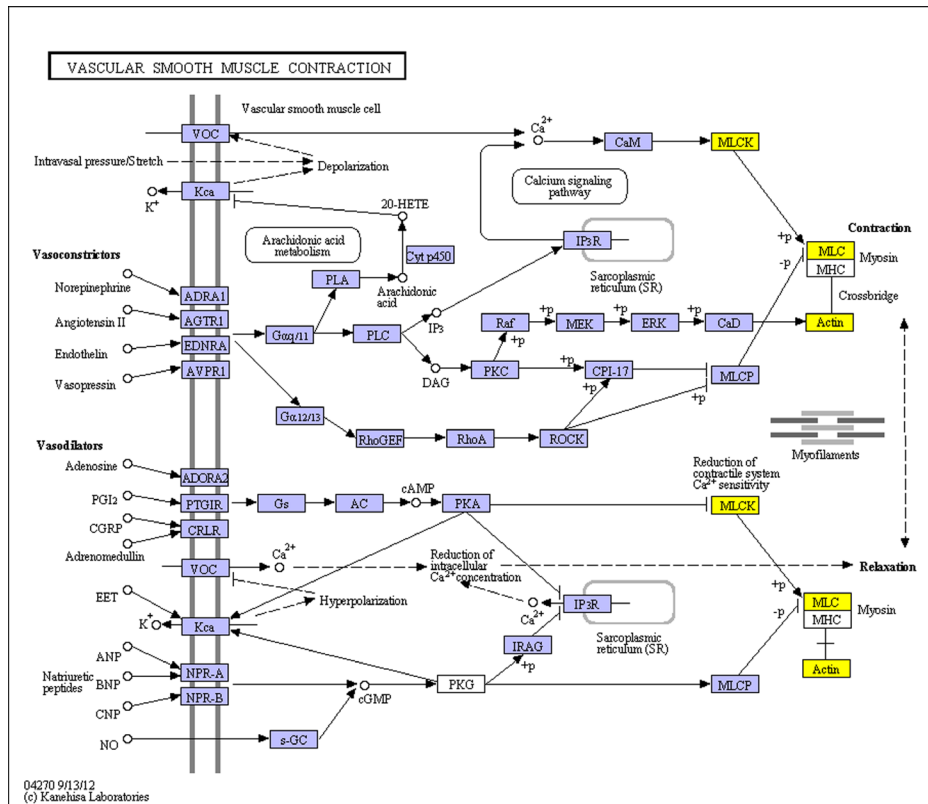


Figure 5. Vascular smooth muscle contraction pathway.

DISCUSSION

Microscopic findings have revealed that factors such as the lack of an outer elastic lamina, disorganization of the extracellular matrix, atherosclerotic and inflammatory changes caused by complement activation, a decrease in reticular fibers, and continuous blood pressure all appear to play important roles in the development of IAs; however, the underlying biological mechanism remains poorly understood. In this study, we identified DEGs of UIA based on DNA microarray data, which provided valuable information for diagnosis and drug therapy interventions at the early stage of IA. By constructing the PPI network using STRING, we obtained the hub protein MYH11 and its co-expression genes, *ACTA2*, *MYLK*, and *MYL9*. All of these genes were found to contribute to vascular smooth muscle contraction.

The myosin heavy chain 11 (*MYH11*) gene encodes the smooth muscle myosin heavy chain beta, which is a major specific contractile protein produced in the structural maintenance of chromosomes. It is identified as one of the defective genes associated with thoracic aortic aneurysms and dissections (TAAD) and patent ductus arteriosus (PDA) (Zhu et al., 2006). Several studies have also found the presence of the *CBFB/MYH11* fusion gene in high frequency in association with acute myeloid leukemia (Rethmeier et al., 2006; Monma et al., 2007; Lazarczyk et al., 2009). Because MYH11 is likely to be specific to the phenotype of non-syndromic TAAD and PDA, it potentially results in a distinct aortic and occlusive vascular pathology, which is driven by insulin-like growth factor 1 and angiotensin II (Pannu et al., 2007). The vascular SMC-specific isoform of α -actin (*ACTA2*) is a contractile protein that is the most abundant protein in vascular SMCs (Guo et al., 2008). Heterozygous *ACTA2* mutations cause familial TAAD, but only half of mutation carriers have aortic disease. Linkage analysis and association studies of individuals with *ACTA2* mutations have also indicated that mutation carriers can have a diversity of vascular diseases, including premature onset of coronary artery disease, premature ischemic strokes (including Moyamoya disease), as well as TAAD (Guo et al., 2009). The myosin light chain kinase (*MYLK*) gene, which localizes with the marker D3S3552, is duplicated on human chromosome 3 (3q13→q21) (Giorgi et al., 2001). *MYLK* has been shown to be a potential cause of inflammatory lung diseases such as asthma (Gao et al., 2007). Furthermore, hyper-methylated *FAM5C* and *MYLK* genes can be used as potential biomarkers for the diagnosis of gastric cancer (Chen et al., 2012), *MYLK* polymorphisms have been suggested to influence increases in blood eosinophil levels among asthmatic patients (Lee et al., 2009), and *MYLK* is also a novel coronary artery disease susceptibility gene (Wang et al., 2007). Myosin light chain 9 (*MYL9*; also known as MLC20, RLC-C, or Mylc2c) is one regulator of the myosin light chain polypeptide (Licht et al., 2010). Previous studies have shown that *MYL9* is critical in vascular function and remodeling processes with age (Shehadeh et al., 2011), and *MYL9* is a major platelet protein based on abundance (Marcus et al., 2000; O'Neill et al., 2002; McRedmond et al., 2004; Sun et al., 2007). At the molecular level, post-translational modifications of *MYL9* are predominantly triggered by the RhoA/Rho-associated kinase signaling pathway (Sanders et al., 1999; Ridley et al., 2003). Expression of *MYL9* is dependent on myocardin-related transcription factors and the serum response factor, and is necessary for cytoskeletal dynamics and experimental metastasis (Medjkane et al., 2009). A recent study also reported that by governing Junb, *MYL9* regulates fiber formation, cellular motility, and contractile capacity in mice (Licht et al., 2010).

Although no studies have investigated the direct interaction between MYH11 and *ACTA2*, *MYLK*, *MYL9*, many have presented convincing evidence that these genes are as-

sociated. The major function of vascular SMCs is contraction to regulate blood pressure and flow, and the SMC contractile force requires cyclic interactions between SMC alpha-actin (encoded by *ACTA2*) and the beta-myosin heavy chain (encoded by *MYH11*) (Guo et al., 2007). Secondly, MYLK regulates SMC contraction and endothelial cell permeability through catalyzing the phosphorylation of the regulatory myosin light chain (Wang et al., 2007), while MYH11 also plays a role in SMC contractility. Last but not least, MYL9 regulates the arterial contraction capacity (Licht et al., 2010). Therefore, due to their roles in vessel contraction and regulation of blood pressure and flow, the expression of the hub protein MYH11 and its co-expression genes *ACTA2*, *MYLK*, and *MYL9* identified in this study, along with their transcripts, may provide beneficial tools for the early diagnosis and treatment of IAs.

REFERENCES

- Bernstein BE, Birney E, Dunham I, Green ED, et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57-74.
- Berriz GF, Beaver JE, Cenik C, Tasan M, et al. (2009). Next generation software for functional trend analysis. *Bioinformatics* 25: 3043-3044.
- Chalouhi N, Ali MS, Jabbour PM, Tjoumakaris SI, et al. (2012). Biology of intracranial aneurysms: role of inflammation. *J. Cereb. Blood Flow Metab.* 32: 1659-1676.
- Chen L, Su L, Li J, Zheng Y, et al. (2012). Hypermethylated FAM5C and MYLK in serum as diagnosis and pre-warning markers for gastric cancer. *Dis. Markers* 32: 195-202.
- Duncan D, Prodduturi N and Zhang B (2010). WebGestalt2: an updated and expanded version of the Web-based Gene Set Analysis Toolkit. *BMC Bioinformatics* 11 (Suppl 4): P10.
- Fujita A, Sato JR, Rodrigues LO, Ferreira CE, et al. (2006). Evaluating different methods of microarray data normalization. *BMC Bioinformatics* 7: 469.
- Gao L, Grant A, Chi P and Gao P (2007). Myosin light chain kinase (MYLK) variants that confer increased risk of sepsis and acute lung injury are associated with asthma and associated phenotypes. *J. Allergy Clin. Immunol.* 119: 146-147.
- Giorgi D, Brand-Arpon V and Rouquier S (2001). The functional Myosin Light Chain Kinase (MYLK) gene localizes with marker D3S3552 on human chromosome 3q21 in a >5-Mb yeast artificial chromosome region and is not linked to olfactory receptor genes. *Cytogenet. Cell Genet.* 92: 85-88.
- Goldberg AD, Allis CD and Bernstein E (2007). Epigenetics: a landscape takes shape. *Cell* 128: 635-638.
- Guo DC, Pannu H, Tran-Fadulu V, Papke CL, et al. (2007). Mutations in smooth muscle alpha-actin (*ACTA2*) lead to thoracic aortic aneurysms and dissections. *Nat. Genet.* 39: 1488-1493.
- Guo DC, Pannu H, Tran-Fadulu V and Papke CL (2008). Mutations in smooth muscle alpha-actin (*ACTA2*) lead to thoracic aortic aneurysms and dissections. *Nat. Genet.* 40: 255.
- Guo DC, Papke CL, Tran-Fadulu V, Regalado ES, et al. (2009). Mutations in smooth muscle alpha-actin (*ACTA2*) cause coronary artery disease, stroke, and Moyamoya disease, along with thoracic aortic disease. *Am. J. Hum. Genet.* 84: 617-627.
- Hellingman CA, van den Bergh WM, Beijer IS, van Dijk GW, et al. (2007). Risk of rebleeding after treatment of acute hydrocephalus in patients with aneurysmal subarachnoid hemorrhage. *Stroke* 38: 96-99.
- Jeong H, Tombor B, Albert R, Oltvai ZN, et al. (2000). The large-scale organization of metabolic networks. *Nature* 407: 651-654.
- Juvela S (2011). Prevalence of and risk factors for intracranial aneurysms. *Lancet Neurol.* 10: 595-597.
- Laaksamo E, Tulamo R, Baumann M, Dashti R, et al. (2008). Involvement of mitogen-activated protein kinase signaling in growth and rupture of human intracranial aneurysms. *Stroke* 39: 886-892.
- Laaksamo E, Ramachandran M, Frosen J, Tulamo R, et al. (2012). Intracellular signaling pathways and size, shape, and rupture history of human intracranial aneurysms. *Neurosurgery* 70: 1565-1572.
- Lazarczyk E, Soszynska K, Mucha B and Skonieczka K (2009). Analysis of gene fusions PML/RARA and CBFβ/MYH11 in acute myeloid leukemia. *Acta Haematol. Pol.* 40: 633-644.
- Lee SO, Cheong HS, Park BL, Bae JS, et al. (2009). MYLK polymorphism associated with blood eosinophil level among asthmatic patients in a Korean population. *Mol. Cells* 27: 175-181.
- Li L, Yang X, Jiang F, Disting GJ, et al. (2009). Transcriptome-wide characterization of gene expression associated with unruptured intracranial aneurysms. *Eur. Neurol.* 62: 330-337.

- Licht AH, Nubel T, Feldner A, Jurisch-Yaksi N, et al. (2010). Junb regulates arterial contraction capacity, cellular contractility, and motility via its target Myl9 in mice. *J. Clin. Invest.* 120: 2307-2318.
- Marcus K, Immler D, Sternberger J and Meyer HE (2000). Identification of platelet proteins separated by two-dimensional gel electrophoresis and analyzed by matrix assisted laser desorption/ionization-time of flight-mass spectrometry and detection of tyrosine-phosphorylated proteins. *Electrophoresis* 21: 2622-2636.
- McRedmond JP, Park SD, Reilly DF, Coppinger JA, et al. (2004). Integration of proteomics and genomics in platelets: a profile of platelet proteins and platelet-specific genes. *Mol. Cell Proteomics* 3: 133-144.
- Medjkane S, Perez-Sanchez C, Gaggioli C, Sahai E, et al. (2009). Myocardin-related transcription factors and SRF are required for cytoskeletal dynamics and experimental metastasis. *Nat. Cell Biol.* 11: 257-268.
- Monma F, Nishii K, Shiga J, Sugahara H, et al. (2007). Detection of the CBFβ/MYH11 fusion gene in de novo acute myeloid leukemia (AML): a single-institution study of 224 Japanese AML patients. *Leuk. Res.* 31: 471-476.
- Nieuwkamp DJ, Setz LE, Algra A, Linn FH, et al. (2009). Changes in case fatality of aneurysmal subarachnoid haemorrhage over time, according to age, sex, and region: a meta-analysis. *Lancet Neurol.* 8: 635-642.
- O'Neill EE, Brock CJ, von Kriegsheim AF, Pearce AC, et al. (2002). Towards complete analysis of the platelet proteome. *Proteomics* 2: 288-305.
- Pannu H, Tran-Fadulu V, Papke CL, Scherer S, et al. (2007). MYH11 mutations result in a distinct vascular pathology driven by insulin-like growth factor 1 and angiotensin II. *Hum. Mol. Genet.* 16: 2453-2462.
- Rethmeier A, Aggerholm A, Olesen LH, Juhl-Christensen C, et al. (2006). Promoter hypermethylation of the retinoic acid receptor beta2 gene is frequent in acute myeloid leukaemia and associated with the presence of CBFβ-MYH11 fusion transcripts. *Br. J. Haematol.* 133: 276-283.
- Ridley AJ, Schwartz MA, Burridge K, Firtel RA, et al. (2003). Cell migration: integrating signals from front to back. *Science* 302: 1704-1709.
- Sanders LC, Matsumura F, Bokoch GM and de Lanerolle P (1999). Inhibition of myosin light chain kinase by p21-activated kinase. *Science* 283: 2083-2085.
- Shehadeh LA, Webster KA, Hare JM and Vazquez-Padron RI (2011). Dynamic regulation of vascular myosin light chain (MYL9) with injury and aging. *PLoS One* 6: e25855.
- Smyth GK (2005). Limma: Linear Models for Microarray Data. In: *Bioinformatics and Computational Biology Solution Using R and Bioconductor* (Gentleman R, Carey V, Dudoit S, Irizarry R, et al., eds.). Springer, New York, 397-420.
- Sun L, Gorospe JR, Hoffman EP and Rao AK (2007). Decreased platelet expression of myosin regulatory light chain polypeptide (MYL9) and other genes with platelet dysfunction and CBFA2/RUNX1 mutation: insights from platelet expression profiling. *J. Thromb. Haemost.* 5: 146-154.
- Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 39: D561-D568.
- Troyanskaya O, Cantor M, Sherlock G, Brown P, et al. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics* 17: 520-525.
- Wang L, Hauser ER, Crosslin D and Nelson S (2007). Abstract 3564: A Multi-Stage Evaluation of Genetic Association with Early-Onset Coronary Artery Disease in MYLK Gene. *Circulation* 116: 806.
- Wiebers DO, Whisnant JP, Huston J III, Meissner I, et al. (2003). Unruptured intracranial aneurysms: natural history, clinical outcome, and risks of surgical and endovascular treatment. *Lancet* 362: 103-110.
- Yu H, Greenbaum D, Xin LH, Zhu X, et al. (2004). Genomic analysis of essentiality within protein networks. *Trends Genet.* 20: 227-231.
- Zhu L, Vranckx R, Khau Van KP, Lalande A, et al. (2006). Mutations in myosin heavy chain 11 cause a syndrome associating thoracic aortic aneurysm/aortic dissection and patent ductus arteriosus. *Nat. Genet.* 38: 343-349.