

## Review Article

# Molecular markers and mechanisms of stroke: RNA studies of blood in animals and humans

Frank R Sharp, Glen C Jickling, Boryana Stamova, Yingfang Tian, Xinhua Zhan, DaZhi Liu, Beth Kuczynski, Christopher D Cox and Bradley P Ander

Department of Neurology, MIND Institute, University of California at Davis, Sacramento, California, USA

**Whole genome expression microarrays can be used to study gene expression in blood, which comes in part from leukocytes, immature platelets, and red blood cells. Since these cells are important in the pathogenesis of stroke, RNA provides an index of these cellular responses to stroke. Our studies in rats have shown specific gene expression changes 24 hours after ischemic stroke, hemorrhage, status epilepticus, hypoxia, hypoglycemia, global ischemia, and following brief focal ischemia that simulated transient ischemic attacks in humans. Human studies show gene expression changes following ischemic stroke. These gene profiles predict a second cohort with >90% sensitivity and specificity. Gene profiles for ischemic stroke caused by large-vessel atherosclerosis and cardioembolism have been described that predict a second cohort with >85% sensitivity and specificity. Atherosclerotic genes were associated with clotting, platelets, and monocytes, and cardioembolic genes were associated with inflammation, infection, and neutrophils. These gene profiles predicted the cause of stroke in 58% of cryptogenic patients. These studies will provide diagnostic, prognostic, and therapeutic markers, and will advance our understanding of stroke in humans. New techniques to measure all coding and noncoding RNAs along with alternatively spliced transcripts will markedly advance molecular studies of human stroke.**

*Journal of Cerebral Blood Flow & Metabolism* (2011) 31, 1513–1531; doi:10.1038/jcbfm.2011.45; published online 20 April 2011

**Keywords:** brain; gene expression profiling; ischemia; microarrays

## Introduction—blood and blood vessels in stroke

Ischemic stroke, in most cases, is a disease of the vasculature and blood, involving platelets, red blood cells, clotting factors, inflammatory cells, and endothelium. Though the result is ischemic brain damage, the primary disease process does not generally relate to the brain tissue itself except in some diseases such as mitochondrial diseases. Most strokes in humans are due to large-vessel atherosclerotic disease, cardioembolic disease (blood clots), and lacunar small vessel disease (Amarenco *et al*, 2009). Atherosclerosis is an inflammatory

disease due to a complex interaction between lipids, endothelium, and vascular risk factors (Chalela, 2009; Chamorro and Hallenbeck, 2006; Hansson, 2009). Atherosclerotic plaques cause stroke by thromboembolism of clot/platelets formed at the plaque or following plaque fragmentation (Chalela, 2009; Wu and Grotta, 2010). Cardioembolic stroke represents a group of cardiac disorders with a propensity to form blood clots in the heart that embolize to brain (Babarro *et al*, 2009). The third major stroke subtype, lacunar, is due to ‘lipohyalinosis’ of small penetrating vessels in the brain (Nah *et al*, 2010; Stevenson *et al*, 2010).

Thus, studies of blood cells and vessels are very relevant for understanding all subtypes of ischemic stroke. The role of the immune system has also gained more attention based upon human and animal studies. The importance of the immune system in human stroke was emphasized by the Enlimomab trial, which worsened outcome. In a follow-up animal study it was found that the deleterious effect of the murine anti-intercellular adhesion molecule 1 (ICAM-1) antibody was probably due to the augmentation of ischemic brain injury by the immunogenicity of the

Correspondence: Dr FR Sharp, Department of Neurology, MIND Institute, University of California at Davis Medical Center, Wet Labs Room 2416, 2805 50th Street, Sacramento, CA 95817, USA. E-mail: frank.sharp@ucdmc.ucdavis.edu

This study was supported by NS056302 (FRS), PO21040N635110 (JPB), and the American Heart Association Bugher Foundation (FRS).

Received 11 January 2011; revised 9 March 2011; accepted 10 March 2011; published online 20 April 2011

Enlimolab antibody (Furuya *et al*, 2001). Though considered a 'negative' trial, this trial does point out the great importance of the immune system involvement in human ischemic stroke (Furuya *et al*, 2001). An increasing number of animal studies demonstrate the role of inflammatory cytokines, chemokines, neutrophils, lymphocytes, monocytes, and a variety of other immune molecules including toll receptors in modulating stroke outcome (Becker, 2010; Chamorro and Hallenbeck, 2006; del Zoppo, 2010; Downes and Crack, 2010; Elkind, 2010; Hallenbeck, 2010; Marsh *et al*, 2009). These studies prompted our studies of gene expression in blood to evaluate cerebrovascular diseases.

## Markers for stroke assessed from blood

A benefit of studying blood is that one could potentially derive biomarkers of stroke to facilitate diagnosis, determine cause, offer prognostic information, and provide insight into pathogenesis and prevention. A number of potential biomarkers have been examined including proteins, peptides, cytokines, chemokines, metabolites, leukocytes, platelets, stem or progenitor cells, microparticles, and others (Jickling *et al*, 2009; Reid *et al*, 2010). Most studies have examined serum or plasma. The goal of many studies has been to search for proteins released from injured neurons, glia, or endothelial cells that could be used as indicators of tissue damage. These molecules are the subject of a number of recent reviews (Chavez *et al*, 2009; Dassan *et al*, 2009; El Husseini and Laskowitz, 2010; Elkind, 2009; Foerch *et al*, 2009; Jensen *et al*, 2009; Montaner, 2009; Saenger and Christenson, 2010; Whiteley *et al*, 2009a, b).

This review will focus primarily on RNA expression in blood following stroke (Sharp *et al*, 2006). The rationale for examining RNA is outlined below, followed by proof-of-principle studies in animals that formed the basis for subsequent human studies. This review will mainly summarize our own studies and those most related to them. This is done because it is difficult to compare between platforms and using different approaches for isolating cells and isolating RNA. We recommend other studies as a contrast to our own since most support the same proof-of-principles even if the same molecules were not always identified (Baird, 2006, 2007; Barr *et al*, 2010; Carmichael, 2003; Freedman *et al*, 2010; Grond-Ginsbach *et al*, 2008, 2009; Kassner *et al*, 2009; Li *et al*, 2010; Ridder *et al*, 2009).

## Advantages of RNA and microarrays

RNA was studied for several reasons. RNA is induced extremely rapidly (within minutes) well before events could be detected using protein markers. Second, almost all of the currently known coding RNAs have been described and are available

on a single platform (Alizadeh *et al*, 2000). This provides a major advantage over many protein biomarker studies. Using arrays of all known RNAs allows one to derive 'the best' markers rather than prospectively guessing at what the best biomarkers might be (Tang *et al*, 2001, 2002, 2005).

The approaches for identifying proteomic markers are quite different (Zhang *et al*, 2008). The most common approach uses candidate protein biomarkers, which may or may not be the best. Alternative proteomics approaches use mass spectrometry and some sort of preselection, which usually assesses only part of the proteome (Yao *et al*, 2009). However, protein biomarkers for vasospasm after subarachnoid hemorrhage have been discovered using this approach (King *et al*, 2010; Maurer *et al*, 2007, 2008). Mass spectrometry offers the advantage of being able to examine protein modifications and isoforms. However, newer RNA technologies are making it possible to also examine alternative splicing using exon arrays or high-throughput RNA sequencing.

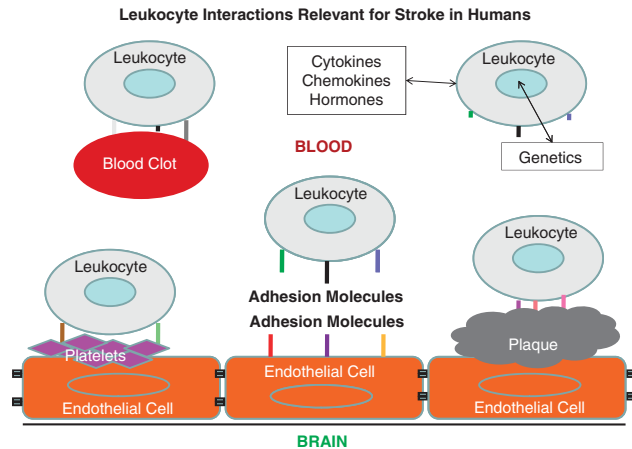
One potential advantage of identifying RNAs that are regulated in stroke is that these might serve to guide the search for similarly regulated proteins. However, this may not be desirable. First, some RNAs do not code for proteins. Second, the synthesis of some RNAs does not result in a corresponding change of protein synthesis. Third, the RNAs usually represent intracellular molecules, with the possible exception of some microRNAs, most of which are not secreted from the leukocytes, platelets, and red blood cells. Thus, RNA itself might be preferable for some indications like studying the immune response, and proteins would be preferred for other indications like injury to endothelium, glia, and neurons.

## Disadvantages of RNA and microarrays

Studying RNA in blood of stroke patients has been challenging. Fold changes of expression are usually less than twofold, making detection of biological effects difficult. At present, studies of whole blood are the most feasible and thus lose specificity related to responses of specific cell types. This is because we use special vacutainer tubes to obtain whole blood via venipuncture, which lyse all cells and immediately stabilize released RNA. This is essential since RNA is unstable and rapidly degrades unless stabilized. The other disadvantage of RNA is that it can be slow to measure—requiring arrays, reverse transcriptase polymerase chain reaction (PCR), or blots. This may be resolved in part by attempts at developing point of care PCR (Lee *et al*, 2010).

## Why study leukocyte RNA in blood of stroke patients

A conceptual issue in the field has been how leukocytes can report on or respond to brain



**Figure 1** Leukocyte interactions with intravascular elements relevant for stroke in humans. Leukocytes (neutrophils, lymphocytes, monocytes), platelets, red blood cells, and other cells interact with each other and with endothelial cells in normal vessels. Leukocyte interactions with endothelial cells, platelets, atherosclerotic plaque, blood clots, and intravascular molecules (cytokines, chemokines, hormones, others) likely account for some changes of gene expression following stroke. Leukocytes also signal to these other cell types, which also account for changes of gene expression.

infarction. Though the answer is still unclear, leukocytes patrol the body, interact with cells from every tissue, endothelial cells of the vasculature, foreign organisms or cells, injured cells and every element within blood (Franks *et al*, 2010). In addition, leukocytes have a complement of expressed RNAs that reflect a combination of the genetics of these cells as well as the interactions of those cells with their environment.

In Figure 1, we show an overview of how leukocytes might respond to cues relevant to stroke in humans. Leukocytes of all types express various adhesion molecules on their surface and interact with normal or inflamed endothelium (Figure 1) (Pries and Kuebler, 2006). Thus, injured brain signals to endothelium to express different adhesion molecules on the luminal side of the vessel, which in turn signal to leukocytes (Beck *et al*, 1997). Leukocytes in blood are known to interact with both platelets and elements of the atherosclerotic plaque and may contribute to disease (Caplan and Fisher, 2007; Franks *et al*, 2010; Htun *et al*, 2006). Thus, endothelial cell–platelet interactions signal to leukocytes and endothelial cell–atherosclerotic plaque interactions signal to leukocytes (Clark *et al*, 1993). Moreover, blood clots that form in the heart or other sites are detected by circulating leukocytes. In addition, the platelets and clots that form on atherosclerotic plaques interact with leukocytes, which may contribute to instability of the potentially embolic elements (Akopov *et al*, 1996). Finally, leukocytes independent of the above influences respond to cytokines, chemokines, hormones, and

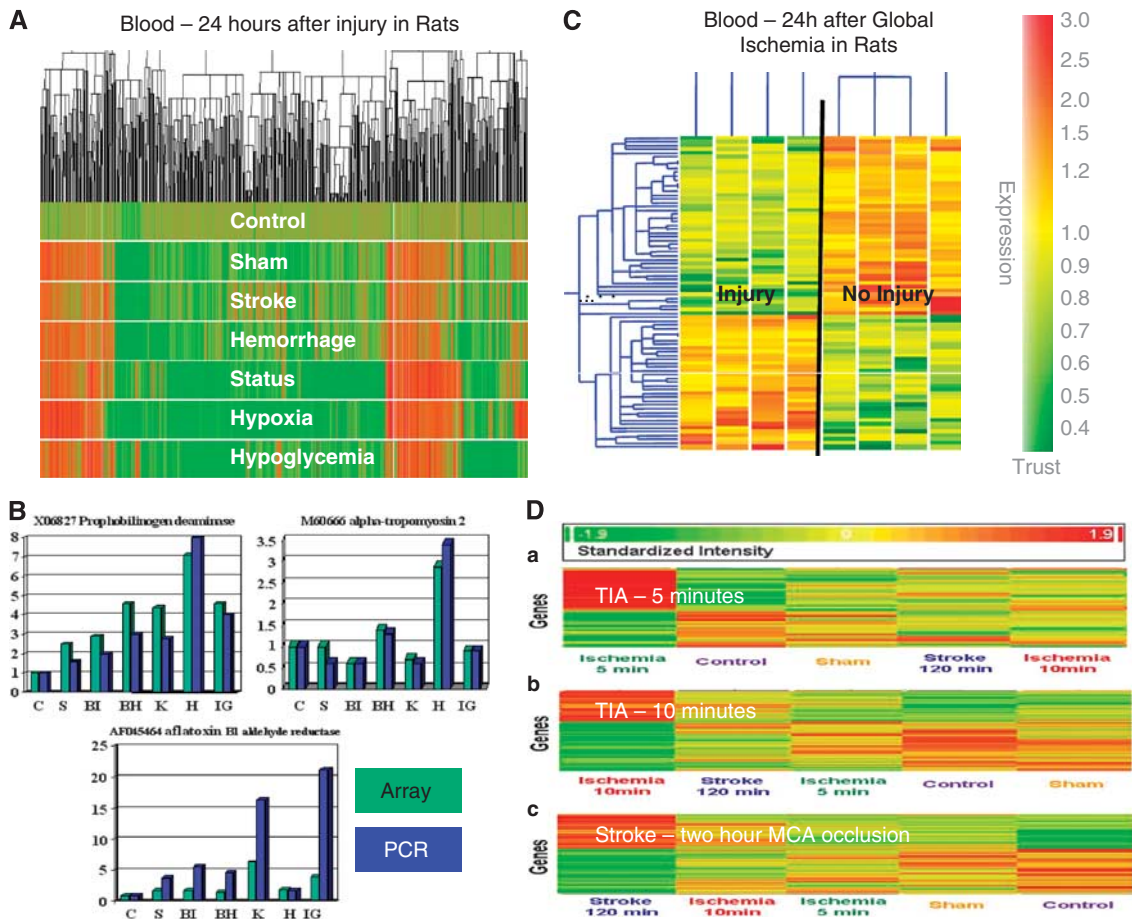
other molecules in blood, which differ as a function of the genetics of each individual cell and their environment (Figure 1). Indeed, each of the above leukocyte interactions results from a complex interplay of the individual environmental interactions of each cell and the genetic make up of the cells (Elneihom *et al*, 1996). Thus, though leukocytes themselves do not cause strokes, they can sense and likely have specific intracellular signaling related to the main causes of stroke. The rapidly growing literature on the immune response to stroke and associated signaling events is beyond the scope of this review.

### Animal studies—focal ischemia, hemorrhage, seizures, hypoxia, hypoglycemia

Even assuming leukocytes detect all of the events pictured in Figure 1, we initially had no idea whether there would be enough reactive leukocytes in blood that could be detected by taking a single blood sample from an animal or human at a single point in time. After all, there are hundreds of thousands of leukocytes in peripheral blood only a fraction of which might respond to factors related to stroke, potentially making it difficult or impossible to detect RNA changes in the few leukocytes collected in a single blood sample. To address this question, we performed the first study of its kind using several different experimental brain injury models in rats (Tang *et al*, 2001).

Adult rats were subjected to experimental ischemic strokes, hemorrhagic strokes, kainic acid-induced status epilepticus, insulin-induced hypoglycemia, or hypoxia and compared with sham-operated controls and to untouched, naive controls (Tang *et al*, 2001). At 24 hours, RNA from peripheral blood monocytes (PBMCs) was processed on Affymetrix microarrays. There were hundreds of upregulated and downregulated genes for each condition compared with sham or untouched controls (Figure 2A). This study demonstrated that (1) there were detectable changes of gene expression in blood 24 hours after each injury; (2) no single gene was specific for a given injury; and (3) there were groups or ‘profiles’ of genes that distinguished each condition from the other (Tang *et al*, 2001) (Figure 2A). Though there were genes shared by every injury, perhaps related to stress or similar mechanisms of injury, there were profiles specific for the injury (Figure 2B). Comparisons of expression using arrays (Figure 2B, teal) and PCR (Figure 2B, blue) showed very similar patterns for each even in these early studies (Figure 2B), which has been validated in many studies since that time (Git *et al*, 2010; Mieczkowski *et al*, 2010).

These studies provided the first proof-of-principle that gene expression in blood changed 1 day



**Figure 2** Animal studies of gene expression in blood following cerebral ischemia and other injuries. **(A)** Cluster plot of gene expression in blood of adult rats 24 hours following sham operation (S), suture-induced brain ischemia stroke (BI), brain hemorrhage (BH), kainate-induced status epilepticus (K), hypoxia (H), and insulin-induced hypoglycemia (IG) compared with controls (C). Increased gene expression is shown in red and decreased gene expression is shown in deep green. **(B)** Comparison of gene expression in the experiments in 'A' using microarrays (teal) and reverse transcriptase polymerase chain reaction (RT-PCR) (blue). Data for three genes are shown for the seven different experimental conditions: C, S, BI, BH, K, H, and IG. Abbreviations are given in '(A)'. **(C)** Adult rats were subjected to global cerebral ischemia using the two-vessel occlusion model. RNA from blood was examined 24 hours later and brain injury was assessed using both terminal deoxynucleotidyl transferase-mediated 2'-deoxyuridine 5'-triphosphate-biotin nick end labeling (TUNEL). There were genes differentially regulated in the two groups, which completely separated these animals on this cluster analysis. **(D)** Gene expression was assessed in blood of adult rats 24 hours following 5, 10, and 120 minutes of focal cerebral ischemia using the suture/thread technique. Genes were discovered that were distinct for each duration of occlusion—5, 10, and 120 minutes. Thus, the 5- and 10-minute duration of focal ischemia that mimic human transient ischemic attacks (TIAs) produce unique changes of gene expression in blood 24 hours after the events. This figure is adapted from several publications (Tang *et al*, 2001, 2003; Zhan *et al*, 2010).

following brain injury and following systemic metabolic stresses like hypoxia and hypoglycemia (Tang *et al*, 2001). The studies also showed that different types of brain injury were associated with specific gene expression profiles in blood. In a follow-up study we demonstrated that there were specific gene profiles in brain for each of these injuries as well (Tang *et al*, 2002). Just as in the blood, there were genes that were common to all of the injuries and could represent responses to stress, neuronal injury, or death and other factors common to each condition. Though many genes expressed in blood were also expressed in brain, the majority were different (Tang

*et al*, 2002). Thus, one cannot necessarily use blood to infer changes of gene expression in brain.

### Animal studies—*injury (neuronal cell death) versus no injury*

We next determined whether there might be a profile for 'neuronal injury.' Global ischemia was produced by bilateral carotid occlusion in rats, which resulted in no cell death in some animals and cell death in hippocampus and cortex documented with terminal

deoxynucleotidyl transferase-mediated 2'-deoxyuridine 5'-triphosphate-biotin nick end labeling (Tang *et al*, 2003). Blood obtained from these animals 24 hours later showed 37 upregulated and 67 downregulated genes in the 'injury' animals compared with the 'no injury' animals (Figure 2C). These data were important for showing that gene expression changes were detectable even when there was selective neuronal cell death (Tang *et al*, 2003).

## Animal studies—transient focal cerebral ischemia

Recent studies have also shown changes of gene expression in blood of rats following brief periods of focal ischemia that mimic human transient ischemic attacks (Zhan *et al*, 2008, 2010). Adult rats were subjected to sham operations or 5 minutes, 10 minutes, or 2 hours of middle cerebral artery ischemia using the suture model. Hsp70 protein was induced 24, 48, and 72 hours later in neurons throughout the middle cerebral artery territory. Following 5 and 10 minutes middle cerebral artery occlusions, 9 of 32 animals (28%) had microinfarcts in striatum (Zhan *et al*, 2008). These studies showed that brief ischemia in animals can be associated with microinfarcts in some animals. Moreover, there was a stress gene response in the 'penumbra' in brain as manifested by induction of Hsp70 protein in the middle cerebral artery distribution. We postulated that these molecular responses in brain would be associated with molecular responses of circulating leukocytes (Zhan *et al*, 2008).

To test this, the identical study was performed except that whole blood was obtained at 24 hours following 5, 10, or 120 minutes of focal cerebral ischemia (Zhan *et al*, 2010). These studies showed genes regulated in blood that were specific for 5, 10, and 120 minutes of focal ischemia (Figure 2D). Moreover, there were 103 genes common to brief focal ischemia and ischemic stroke (Zhan *et al*, 2010). The data confirm immune responses to brain that may not be associated with cell death. The nature of this 'sublethal' signaling is less clear but might include cytokines and chemokines that could signal from brain to blood and blood to brain.

## Human studies—first study assessing diagnosis of ischemic stroke (2005)

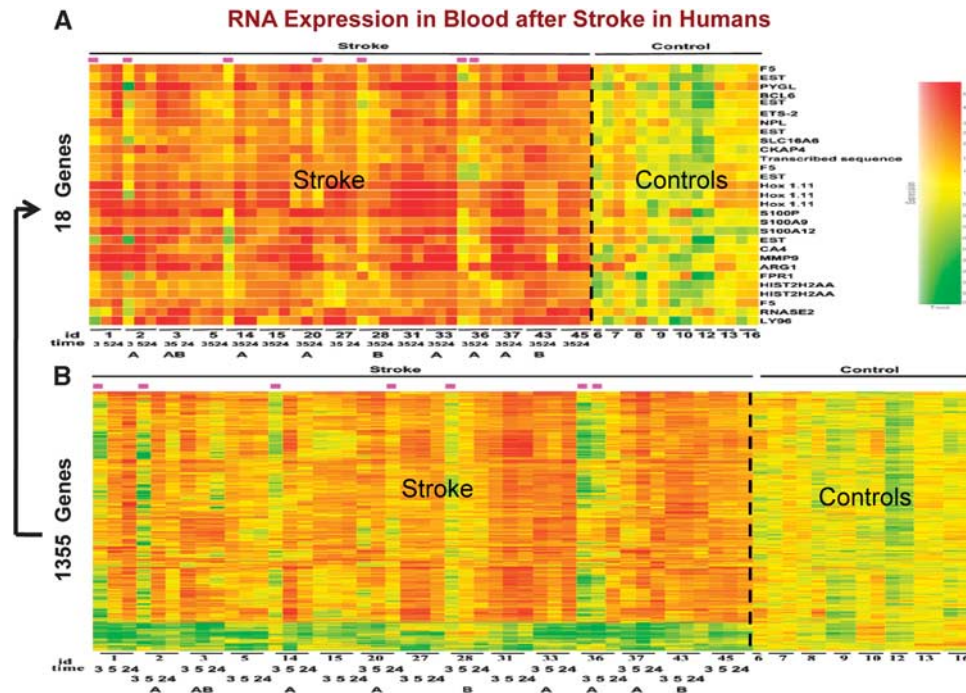
The first human study to assess RNA expression in stroke was published by the Baird group (Moore *et al*, 2005). Using PBMCs obtained 1 to 4 days following ischemic stroke, 190 genes were significantly regulated in 20 stroke compared with 20 control subjects (Moore *et al*, 2005). A panel of 22 genes derived from the prediction analysis for microarrays algorithm in

the index cohort ( $n=40$ ) classified stroke in the validation cohort ( $n=20$ ), with a sensitivity of 78% and a specificity of 80%. These findings were extremely important because they used the genes in one cohort to predict stroke in a second cohort—supporting the validity of the findings within this study (Moore *et al*, 2005); supporting the proof-of-principle in our previous rodent studies (Tang *et al*, 2001). Notably, even though virtually identical methods were used in our prior rodent study (Tang *et al*, 2001) compared to this human study (both used PBMCs and Affymetrix arrays), very few genes were similarly regulated in the blood of rats (Tang *et al*, 2001) compared to the humans with ischemic strokes (Moore *et al*, 2005). The explanation for this is unclear but is reviewed elsewhere (Turner *et al*, 2011a). It is possible that the experimental methods of producing ischemia (suture methods, anesthesia, young animals with no vascular disease) in animals simply results in different gene expression responses compared with stroke in humans; or the immune responses in rodents may be different in humans following ischemic stroke. These and other possibilities need further study since they are relevant to understanding how well animal stroke studies 'model' human stroke.

## Human studies—whole genome, whole blood studies of ischemic stroke

In our initial rodent study (Tang *et al*, 2001), and in the first human study (Moore *et al*, 2005) blood was drawn and then PBMCs were separated from the blood using a Ficoll gradient and centrifugation. This procedure in and of itself could affect gene expression in blood. If performed at different times after stroke, or if the methods were not identical, this could affect gene expression.

We therefore utilised PAXgene tubes to address this issue. These commercial vacutainer tubes lyse cells and stabilize RNA, and have proven to be reliable for many clinical studies (Chai *et al*, 2005; Thach *et al*, 2003; Vartanian *et al*, 2009; Yamamoto *et al*, 2006). Blood from 15 patients was drawn into PAXgene tubes at < 3, 5, and 24 hours after ischemic stroke ( $n=45$  samples) and compared with 14 control samples (Tang *et al*, 2006). RNA processed on Affymetrix U133 microarrays showed that over 1,000 genes were upregulated or downregulated in the blood of ischemic stroke compared with control subjects (Figure 3). Most of the genes expressed at 2 to 3 hours after stroke (before treatment) were also expressed at 5 and 24 hours after the strokes (Tang *et al*, 2006). Prediction analysis of microarrays derived the 25-probe sets for 18 genes that were most predictive of stroke (Figure 3). The fold change of these genes varied from 1.6 to 6.8 and these genes correctly classified 10/15 patients at 2.4 hours, 13/15 patients at 5 hours, and 15/15 patients at 24 hours



**Figure 3** Gene expression in whole blood of humans following ischemic stroke. Patients had whole blood drawn before 3 hours after an ischemic stroke and before treatment with tissue plasminogen activator (tPA) with or without eptifibatid. Blood samples from the same patients were drawn again at 5 and 24 hours. RNA from these samples was processed on whole genome microarrays. When comparing ischemic stroke to control patients, over 1,000 genes were regulated in blood (B). Using prediction analysis of microarrays, the optimal set of genes ( $n = 18$ ) were derived that best distinguished ischemic stroke patients from controls (A). This figure is adapted from Tang *et al* (2006).

**Table 1** Sensitivity and specificity of diagnosing ischemic stroke ( $n = 199$ ) versus controls ( $n = 90$ )

Group	60 Probe sets 3 hours ischemic stroke versus controls		46 Probe sets 24 hours ischemic stroke versus controls		97 Probe sets 3 and 24 hours ischemic stroke combined versus controls	
	PAM	SVM	PAM	SVM	PAM	SVM
IS	85%	94%	91%	94%	86%	95%
SAVVY	92%	96%	92%	96%	96%	96%
Healthy	84%	68%	89%	84%	84%	68%

IS, ischemic stroke; PAM, prediction analysis of microarrays; SAVVY, vascular risk factor controls; SVM, support vector machine prediction algorithm. This is taken from Stamova *et al* (2010).

after stroke (Tang *et al*, 2006). When the results of this study were compared with the previous study from Moore *et al*, however, there were very few genes that were common to both. The explanations for these differences probably included bloods were drawn at different times after the stroke; differences in treatment; the use of PBMCs for the Moore *et al* study and the use of whole blood/PAXgene tubes for the Tang *et al* study; and the use of different RNA isolation and labeling methods and the use of different arrays for the two studies.

To address the issue of reproducibility, we have recently repeated our initial study in a larger cohort (Stamova *et al*, 2010). Thus, patients with ischemic stroke ( $n = 70$ , 199 samples) were compared with

control subjects who were healthy ( $n = 38$ ), controls with vascular risk factors ( $n = 52$ ), and to subjects who had myocardial infarction ( $n = 17$ ). Whole blood was drawn into PAXgene tubes at  $\leq 3$ , 5, and 24 hours after stroke onset and RNA processed on whole genome Affymetrix U133 microarrays. The 25-probe sets previously reported in our study by Tang *et al* predicted a new set of ischemic strokes with 93.5% sensitivity and 89.5% specificity. In order to derive profiles that would distinguish ischemic stroke from all control subjects, we derived 60- and 46-probe sets that differentiated control groups from 3 and 24 hours ischemic stroke samples, respectively (Table 1). Thus, this study replicated our previously reported gene expression profile in a larger cohort

and identified additional genes that discriminate ischemic stroke from relevant control groups (Stamova *et al*, 2010).

Finally, another recent study from Barr supports the above findings. Whole blood was obtained in PAXgene tubes from 39 ischemic stroke patients and 25 healthy control subjects (Barr *et al*, 2010). RNA was processed on Illumina HumanRef-8v2 bead chips. Among a large number of regulated genes, they identified a nine-gene profile that separated ischemic stroke patients compared with controls (Barr *et al*, 2010). Moreover, five of these nine genes were identified in our previous study (Tang *et al*, 2006). Thus, another group has confirmed at least a core set of genes, and we have replicated our own gene expression studies following ischemic stroke. All of these studies, however, are confounded to some degree by various treatments, comparisons to healthy controls, variations in time after stroke, different risk factors between groups and differences of age, race, and gender. Nonetheless, the first test of the technology has been achieved: replication of results and independent validation by at least two different groups. Such promising results provide strong support for further study.

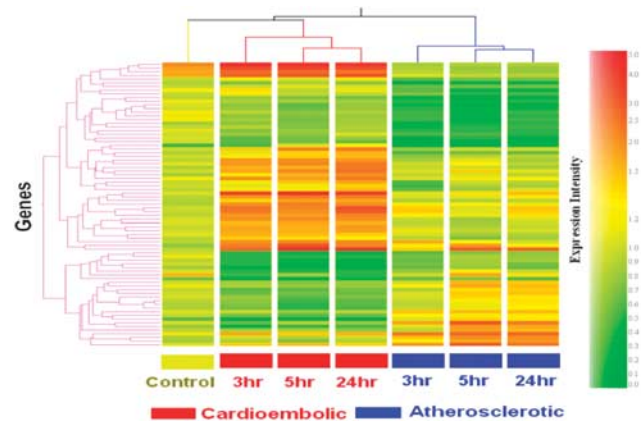
## Causes of stroke in humans—large vessel and cardioembolic

Early on it was apparent that developing a ‘diagnostic test for stroke’ would be difficult, and perhaps not of practical use unless it could be performed within the first few hours of stroke. In addition, a diagnostic test would not only have to diagnose ischemic stroke but rule out hemorrhagic stroke if it were used to guide acute stroke treatments such as tissue plasminogen activator (tPA).

Thus, we have approached different questions that could be rapidly translated to the care of stroke patients. We asked whether gene profiles in blood exist that are specific for the different causes of ischemic stroke. The main reason for developing such profiles would be to use them to diagnose the cause of ischemic stroke in those patients with ‘cryptogenic stroke’ with no known cause who represent approximately one third of all ischemic strokes.

In the first study, whole blood was collected in PAXgene tubes from acute ischemic stroke patients (< 3, 5, and 24 hours) and healthy controls. RNA was isolated and processed on Affymetrix Human U133 Plus 2.0 Arrays. Expression profiles in the blood of cardioembolic stroke patients differed from large-vessel atherosclerotic stroke patients (Figure 4). Of the 77 genes that differed between the two groups (fold change > 1.5,  $P < 0.05$ ), a minimum number of 23 genes differentiated the two types of stroke with > 90% specificity and sensitivity (Xu *et al*, 2008). Notably, some of the genes that distinguished

## RNA Expression in Blood Differentiates Different Causes of Stroke in Humans



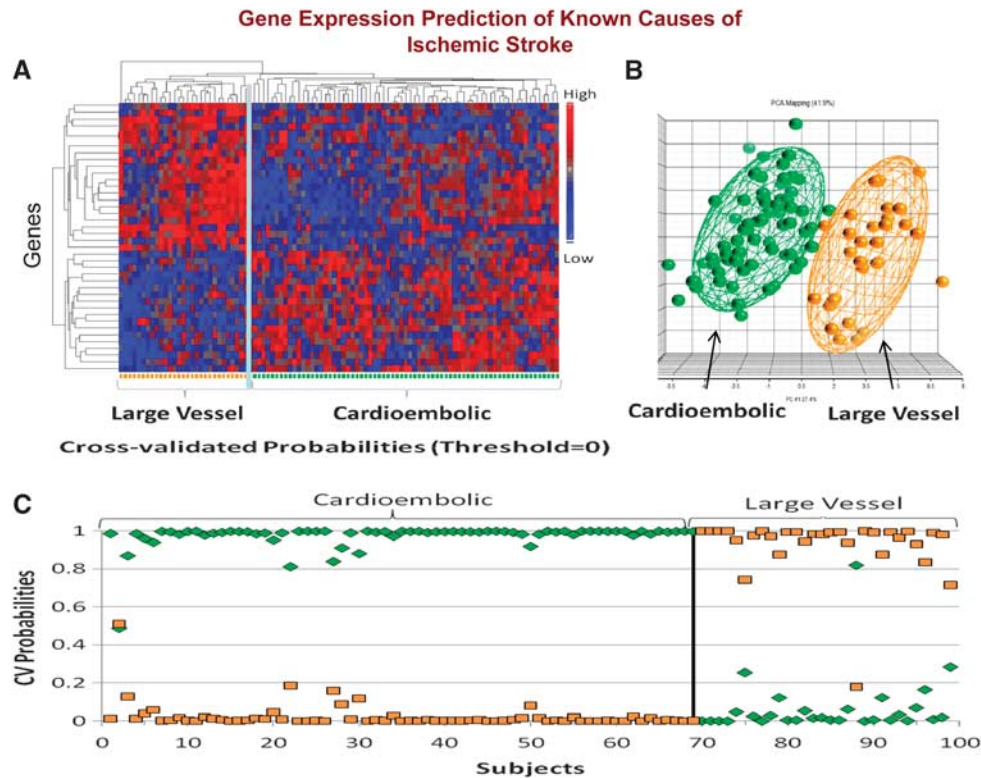
**Figure 4** Cluster analysis of the 77 genes (analysis of variance (ANOVA),  $P < 0.05$ , fold change > 1.5) that were differentially expressed between cardioembolic and atherosclerotic stroke. X axis shows each condition at 3, 5, and 24 hours after the ischemic stroke. Y axis shows individual genes. The color coding indicates gene expression intensity, with red being high and green being low. This figure is adapted from Xu *et al* (2008).

cardioembolic from atherosclerotic stroke displayed little change over time (Figure 4). These might be genes expressed differentially prior to stroke—and perhaps indicate risk of stroke. Other genes displayed significant change over time, suggesting that these time-dependent alterations in gene expression were associated with differential gene expression of immune cells due to the strokes caused by cardioembolism compared with atheroembolism (Figure 4) (Xu *et al*, 2008).

We have recently confirmed these initial findings using identical methods to study 194 samples from 76 acute ischemic stroke patients (Jickling *et al*, 2010). A 40-gene profile differentiated cardioembolic stroke from large-vessel stroke with > 90% sensitivity and specificity (Figure 5). A separate 37-gene profile differentiated cardioembolic stroke due to atrial fibrillation from nonatrial fibrillation causes with > 90% sensitivity and specificity (Figure 6). When these profiles were applied to patients with cryptogenic stroke, 17% were predicted to be large vessel and 41% to be cardioembolic stroke. Of the cryptogenic strokes predicted to be cardioembolic, 27% were predicted to have atrial fibrillation (Jickling *et al*, 2010). Thus, we have demonstrated the feasibility of using gene expression to demonstrate the causes of ischemic stroke, and to use these profiles to predict the causes of cryptogenic stroke.

## Markers versus mechanisms

The above descriptions outline the approach for deriving biomarkers. However, for each analysis we obtain a short list of genes for prediction and a long



**Figure 5** (A) Hierarchical cluster plot of the 40 genes that were significantly different between cardioembolic stroke and large-vessel stroke. Genes are shown on the Y axis and subjects are shown on the X axis. Red indicates a high level of gene expression and blue indicates a low level of gene expression. Subjects cluster by diagnosis. (B) Principal components analysis of the 40 genes that differentiated cardioembolic stroke from large-vessel stroke. Each sphere represents a single subject. The ellipsoid surrounding the spheres represents 2 s.d. from the group mean. (C) Leave-one-out crossvalidation prediction analysis of the 40 genes found to differentiate cardioembolic stroke from large-vessel stroke. The probability of the predicted diagnosis is shown on the Y axis. The actual diagnosis is shown on the X axis, where patients with known cardioembolic stroke are shown on the left (top bracket), and patients with known large-vessel stroke are shown on the right (top bracket). The probability of a predicted diagnosis of cardioembolic stroke is indicated with green diamonds, and the probability of a predicted diagnosis of large-vessel stroke is indicated with orange squares. Subjects with known cardioembolic stroke were predicted to have cardioembolic stroke for 69 out of 69 samples (100% correct prediction). Subjects with known large-vessel stroke were predicted to have large-vessel stroke for 29 out of 30 samples (96.7% correct prediction). A sample is considered misclassified if the predicted class does not match the known class with a probability  $> 0.5$ . This figure is adapted from Jickling *et al* (2010).

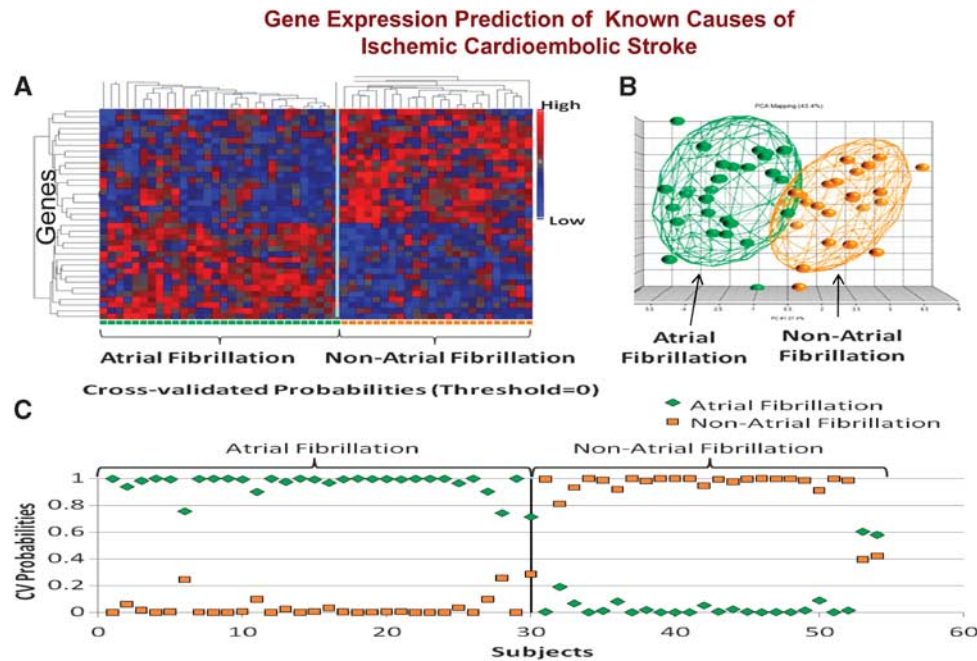
list of genes for functional analyses (Figure 7). For example, after obtaining data from an experimental data set (stroke) compared with controls, different statistical approaches are used. First, the least number of genes that best distinguish the groups are determined using strict statistical cutoffs. Second, the maximal number of genes/probes that distinguish the two groups are derived using less stringent statistical cutoffs (Figure 7). The least number of genes is used for biomarkers and prediction (diagnosis) (Figure 7). The large list of regulated genes is used to derive the gene ontologies and the known pathways and networks that differentiate the two groups (Figure 7). These changes may represent causes or effects related to the stroke, but nonetheless provide information regarding pathogenesis.

One of the important and difficult to grasp features of these analyses is that there may be no best number of classifier/biomarker/diagnostic genes. For example,

a panel of 25, 30, or 35 genes may all have a similar sensitivity and specificity. This can explain why different gene lists might be quite predictive in different studies. Generally, too few genes or too many genes can significantly decrease predictive power.

A different approach is used for pathway analyses. For most studies, a given level of significance is set (corrected  $P < 0.05$  with a false-discovery rate of 5%—see below) coupled with a fold change cutoff, which is usually  $> 1.2$  for blood studies (Jickling *et al*, 2010; Stamova *et al*, 2010). Using this cutoff for most of our stroke studies, one obtains hundreds of genes/probe sets. These can then be used for most pathway/gene ontology programs to determine if there are more genes than expected by chance expressed in a particular pathway. If there are too many genes, the approach does not work well because pathways are saturated. In addition, if there





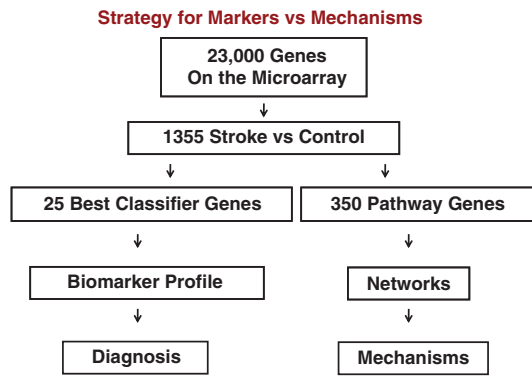
**Figure 6** (A) Cluster analysis of the 37 genes that were significantly different in subjects with cardioembolic stroke due to atrial fibrillation compared with subjects with nonatrial fibrillation causes. Genes are shown on the Y axis and subjects are shown on the X axis. Red indicates a high level of gene expression and blue indicates a low level of gene expression. (B) Principal components analysis of the 37 genes that differentiated cardioembolic stroke due to atrial fibrillation from nonatrial fibrillation causes. Each sphere represents a single subject. The ellipsoid surrounding the spheres represents 2 s.d. from the group mean. (C) Leave-one-out crossvalidation prediction analysis of the 37 genes found to differentiate cardioembolic stroke due to atrial fibrillation from nonatrial fibrillation causes. The probability of the predicted diagnosis is shown on the Y axis. The actual diagnosis is shown on the X axis, where patients with known atrial fibrillation are on the left (top bracket), and patients with known nonatrial fibrillation stroke are on the right (top bracket). The probability of a predicted diagnosis of stroke due to atrial fibrillation is indicated with green diamonds, and the probability of a predicted diagnosis of nonatrial fibrillation stroke is indicated with orange squares. Subjects with known cardioembolic stroke due to atrial fibrillation were predicted to have atrial fibrillation as a cause of stroke in 30 out of 30 samples (100% correct prediction). Subjects with cardioembolic stroke due to nonatrial fibrillation causes were correctly predicted in 22 out of 24 samples (91.7% correct prediction). A sample is considered to be misclassified if the predicted class does not match the known class with a probability > 0.5. The figure is adapted from Jickling *et al* (2010).

are too few it also does not work well since there are too few genes per pathway. There are probably statistical approaches for deriving the optimal number of genes for these pathway analyses that the authors are not aware of.

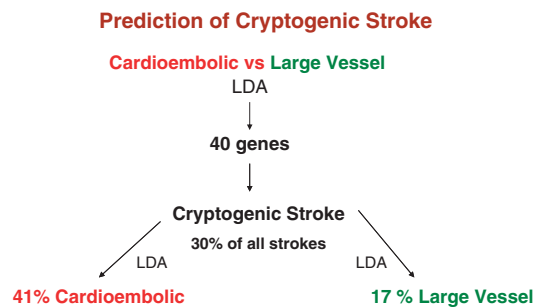
Figures 8–11 provide examples of markers and mechanisms derived from several of our microarray studies. Figure 8 shows the derivation of markers and their use to diagnose cryptogenic stroke (Jickling *et al*, 2010). Using linear discriminant analysis (or other methods like support vector machines or *k*-nearest neighbors/prediction analysis of microarrays) one can derive the 40 genes that best differentiate cardioembolic stroke and large-vessel stroke (Figure 8). Using these 40 genes and applying them only to the patients with cryptogenic strokes, one can again utilise linear discriminant analysis to predict which cryptogenic strokes were cardioembolic (41%) and which were due to large-vessel disease (17%). This shows how biomarkers are used for prediction (Jickling *et al*, 2010). The goal of a biomarker prediction study is to derive the least

number of the best predictors to derive the best sensitivity and specificity for the prediction in question. This can be applied to any problem with stroke—or any other medical condition for that matter.

Figure 9 shows how the genes regulated following ischemic stroke change in the interleukin-1A and 1B signaling pathways from 2 to 3 hours (upper panel) compared with 5 hours following stroke (lower panel). Interleukin-1 worsens stroke in most animal models (Allan and Rothwell, 2003; Rothwell and Luheshi, 2000). In humans, it is one of the mediators of the febrile response, which may worsen stroke in and of itself. There are few genes regulated between 2 and 3 hours after stroke whereas by 5 hours virtually every gene in the pathway is regulated (Figure 9; red = upregulated; green = downregulated). This is a little misleading since changes were occurring before 3 hours but these were too small for detection. The data suggest that drugs acting on interleukin-1 pathways could be acting within the therapeutic window for acute human ischemic stroke, i.e., 0 to 4.5 hours.



**Figure 7** Strategy for deriving gene lists for prediction (biomarker profile and diagnosis) versus lists to study pathways, networks and mechanisms. Stroke versus controls produces a large gene list using loose statistical criteria. Prediction algorithms (e.g., linear discriminant analysis, support vector machine and others) are used to derive the best classifier/prediction genes, which are then applied to a second test set for prediction. An intermediate statistical cutoff is used to generate a list of hundreds of genes that can then be used for pathway, gene ontology, and other analyses to assess pathways, networks, and mechanisms.



**Figure 8** An example of the derivation and use of genes for prediction. Using linear discriminant analysis (LDA), the least number of genes that best differentiate cardioembolic and large-vessel atherosclerotic stroke are derived ( $n = 40$  genes). Using these 40 genes in linear discriminant analysis, the causes of strokes in cryptogenic stroke patients are predicted. Of the cryptogenic stroke patients, 41% were predicted to be cardioembolic and 17% were predicted to be large vessel. These data are from Jickling *et al* (2010).

Figure 10 shows how genes that were expressed differentially in cardioembolic versus large-vessel stroke relate to different mechanisms (Xu *et al*, 2008). For cardioembolic stroke, the main functions were immune response, defense response, response to biotic stimuli, and response to pathogens (Xu *et al*, 2008). This suggests that inflammation and perhaps even infection might be important in these cardioembolic strokes. In contrast, hemostasis, chemokine activity, and related pathways were the highest ranked functions for large-vessel atherosclerotic stroke. This would be consistent with the observation that antiplatelet agents are useful in decreasing risk from strokes due to large-vessel disease.

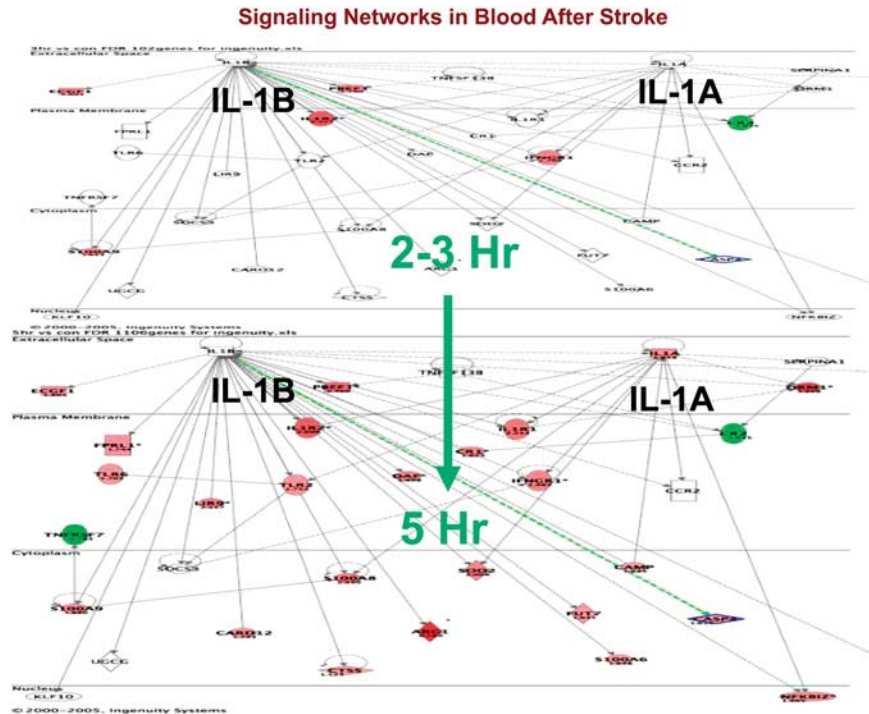
## White-matter hyperintensities and brain ischemia

RNA expression studies also offer the opportunity to compare with other disease entities and mechanisms shared or not shared with brain ischemia. For example, we recently studied a group of elderly patients with white-matter hyperintensities (WMH) (Xu *et al*, 2010). White-matter hyperintensities are areas of high signal detected by T2 and fluid-attenuated inversion recovery sequences on brain magnetic resonance imaging. Although associated with aging and vascular risk factors, the pathogenesis of WMH remains unclear. One postulated cause is chronic 'brain ischemia.' To address this, 20 subjects with extensive WMH (WMH+), 45% of whom had Alzheimer disease, were compared with 18 with minimal WMH (WMH-), 44% of whom had Alzheimer disease. RNA was processed on HU133 Plus 2.0 microarrays. In all, 241 genes were differentially regulated (fold change > 1.2,  $P < 0.005$ ) between the groups. A cluster analysis separated the group with extensive WMH from the group with minimal WMH (Figure 11). Function analyses suggested that WMH-specific genes were associated with oxidative stress, inflammation, detoxification, and hormone signaling, and included genes associated with oligodendrocyte proliferation, axon repair, long-term potentiation, and neurotransmission (Xu *et al*, 2010). Notably, only 4 of the 241 WMH genes were previously identified in our studies outlined above of ischemic stroke (Stamova *et al*, 2010; Xu *et al*, 2010). This does not exclude ischemia as a causal factor in WMH, but it does suggest that there must be differences between acute ischemic stroke and any 'ischemia-related changes' that might be associated with WMH. The WMH study was important for showing that the stroke-related genes probably are not related to coincidental WMH; the WMH and Alzheimer and ischemic stroke genes identified in peripheral blood were different; and that WMH may in fact be a discrete entity since a specific gene expression profile can be associated with it in peripheral blood.

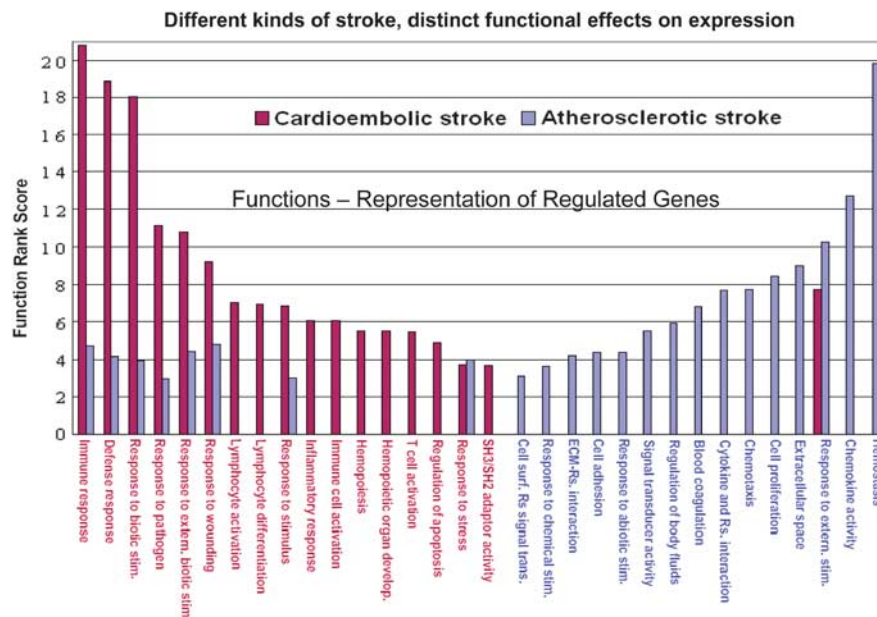
## Technical challenges associated with RNA studies

The technical difficulties and multiple methodological and analytical pitfalls associated with these studies cannot be overemphasized. Moreover, the chances for failure are high given the small groups usually being studied, small biological changes that require great technical rigor, and different statistical approaches that can influence study findings.

For those beginning in the field it is critical to read the many reviews that are now available and make critical decisions based upon as much knowledge as possible. Some of the decisions that we have made



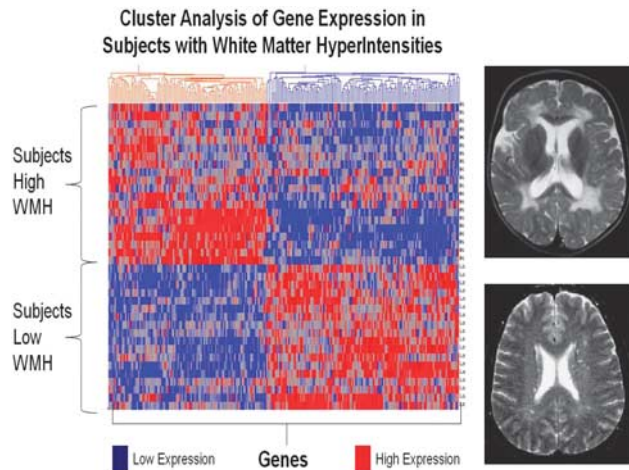
**Figure 9** Signaling pathways for interleukin (IL)1A and IL1B at 2 to 3 hours (upper panel) and at 5 hours (lower panel) following ischemic stroke in humans. Data are from Tang *et al* (2006).



**Figure 10** Functional comparisons of cardioembolic stroke-specific genes versus atherosclerotic stroke-specific genes. The function rank score (Y axis) is a log-transformed *P* value. The higher the score, the more significant is the functional pathway (shown on the Y axis). The functions are on the X axis. Cardioembolic functions are in red and large-vessel atherosclerotic functions are in blue. This figure is adapted from Xu *et al* (2008).

and why follow. We started with and continue to use Affymetrix arrays. Though expensive, the company has remained in business where others have not. This will be much less of an issue once RNA

sequencing is generally used (see below). There have been numerous protocols for RNA isolation and labeling. We have kept with large companies who maintain and support their products with the highest



**Figure 11** Cluster analysis of white-matter hyperintensity (WMH)-associated genes for normal control subjects and Alzheimer subjects. The 241 genes that were differentially expressed in extensive WMH subjects (high WMH) versus minimal WMH subjects (low WMH) were used for an unsupervised Pearson cluster analysis ( $P < 0.005$  and fold change  $> 1.2$ ). Individual genes ( $n = 241$ ) are shown on the X axis and individual subjects ( $n = 38$ ) are shown on the Y axis. Genes showing high expression are indicated in red and genes with low expression are indicated in blue. Note that all of the subjects with high WMH are clustered separately from low WMH subjects; and that there is a specific gene expression profile for each. Half of the high WMH subjects had Alzheimer disease (AD) and half were cognitively normal. Half of the low WMH subjects had AD and half were cognitively normal. This figure is adapted from Xu *et al* (2010).

standards. Given the expense of these studies one does not cut corners on reagents. We have avoided major protocol changes unless absolutely necessary. Changes in the protocols that have been made included a DNase step in RNA isolation, and the use of NuGEN reagents so that very small amounts of RNA can be used for each array (Jickling *et al*, 2010; Stamova *et al*, 2010; Xu *et al*, 2010). If one is contemplating a preliminary study followed by a confirmatory one, one must standardize all methods and arrays across those two studies. Ideally, methods used previously by others should be used if confirmation across studies is desired.

Though now well known in the microarray field, the ‘batch effects’ problem is still a major one. This is a problem where one set of samples is run together and then compared with separate samples that are also run together—often separated over some time—or performed in different laboratories. The universal experience is that it is difficult and can be impossible to compare two different ‘batches.’ To combat the batch problem, experimental and control samples must be intermixed, and the samples should be run in the same laboratory and preferably all run at a single time. If samples are to be run at different times or in different locations, there are only a few available approaches.

One approach is to use a reference or standard RNA sample with each batch (Walker *et al*, 2008). The standard RNA sample is run each time the other samples or groups of samples are run. Spiked RNAs have been used in the past. However, this does not solve the problem. For the control RNA samples, it is critical that the source of the RNA be the same as the source of the target samples (Walker *et al*, 2008). For example, if ischemic stroke blood samples are compared with control blood samples, then the ‘standard or reference RNA sample’ must be a human blood sample that is identical for every batch and that has been handled and processed in a way identical to that used for the stroke/control samples. This approach allows one to correct across samples on a gene-by-gene basis using a Bayes approach we have published (Walker *et al*, 2008). The within batch version of this approach is called COMBAT and is publicly available.

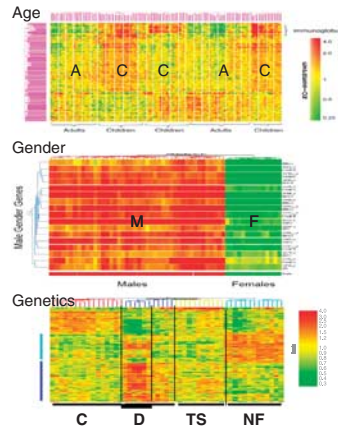
Another approach to the batch problem is to use the reference gene approach that we have developed (Stamova *et al*, 2009). This approach identifies a group of genes that are expressed at similar levels across control and experimental subjects. The control gene approach is similar to the principle applied to reverse transcriptase PCR where a reference gene is run at the same time as the target. The problem with the control gene approach on arrays, however, appears to be that even ‘within array variance’ can be considerable (unpublished data). Rigorous internal reference approaches need to be developed before the results of the data described in this review can be applied in the clinic.

Analytical issues also go beyond the correction for batch effects. In the early gene expression studies of blood, no attempt was made to account for covariates. It has become quite clear, however, that age and gender have large effects on gene expression in blood (Figure 12) (Tang *et al*, 2004a, b, c, 2005). If there are genetic differences in the cohort, then this alone may segregate subjects based upon gene expression in peripheral blood (Figure 12) (Tang *et al*, 2004c). Because of the need to account for covariates, analysis of variance approaches are more commonly used now for the statistical analyses. Earlier statistics were based upon permutation approaches, which were excellent at distinguishing groups but often did not account for covariates.

The next analytical task is to design the study so that marker and mechanism analyses are clear-cut and robust. The first tasks are to identify the ‘regulated genes’ between groups. This always requires some sort of within and across chip normalization—RMA, GCRMA, and others. The choice of the normalization method depends upon the array type, group preferences, study goals, and other considerations (Mieczkowski *et al*, 2010; Millenaar *et al*, 2006; Shedden *et al*, 2005; Zakharkin *et al*, 2005). The statistical criteria for selecting genes are also goal dependent. Some studies have used fold change cutoffs only (Shi *et al*, 2010), and others have

### Future of RNA Expression in Stroke

- Exon Expression–Alternative Splicing
- Long(regulatory) non-coding RNAs (lncRNAs)
- Small interfering RNAs (siRNAs)
- microRNAs (miRNAs)
- PIWI-interacting RNAs (piRNAs)
- Promoter-associated RNAs (PARs)
- Small nucleolar RNAs (snoRNAs)
- X-inactivation RNAs (xiRNAs)
- Sno-derived RNAs (sdRNAs)
- microRNA-offset RNAs (moRNAs)
- tRNA-derived RNAs
- MSY2-associated RNAs (MSY-RNAs)
- Telomere small RNAs (tel-sRNAs)
- Centriosome-associated RNAs (crasiRNAs)



**Figure 12** Future studies of RNA expression in blood of subjects with stroke will investigate the effects of age, gender, race, underlying genetic differences in coding and noncoding RNAs. The role of alternative splicing is likely to revolutionize these analyses. Age: A, adults; C, children. Gender: F, females; M, males. Genetics: C, controls; D, Down syndrome; NF, neurofibromatosis; TS, tuberous sclerosis. Figures are adapted from Tang *et al* (2004b, c).

used *P* value cutoffs only (Shi *et al*, 2010; Tusher *et al*, 2001). Both of these have weaknesses when applied to studies of blood. Generally, some combination of fold change and a reasonable statistical cutoff result in the most reliable changes of gene expression for studies of blood (Guo *et al*, 2007; Shi *et al*, 2010; Tibshirani *et al*, 2002; Tusher *et al*, 2001). A great many issues revolve around the choice of cutoffs, which include everything from a very strict Bonferroni (correction for total number of comparisons) to a less stringent Benjamini–Hochberg false-discovery rate (Benjamini *et al*, 2001; Hsueh *et al*, 2003; Reiner *et al*, 2003; Shi *et al*, 2010; Tan *et al*, 2006). The accepted ‘standard in the field’ is to use a false-discovery rate of 5%, which means that a given list of genes contains no more than 5% false positives. Thus, for preliminary microarray studies most investigators focus on pathways rather than single genes. The only way to confirm a given gene in a whole genome microarray study is to confirm it in a second independent study.

Once regulated genes are identified, the tools for deriving biomarkers are excellent. These include prediction analysis of microarrays, support vector machine, linear discriminant analyses, machine learning, and a host of others (Shi *et al*, 2010). Methods of assessing sensitivity and specificity using receiver operating curves and other approaches are generally used by many and are readily available in various public and private software packages. What is less clear is how ‘to do’ the pathway and network analyses. There are many versions of these approaches—from known protein–protein interaction databases, to information processing-relational databases. One can also assess whether given lists of

genes are associated with specific pathways, gene ontologies (molecular function, biological process, cellular compartment), have overrepresentation of transcription factor binding sites, regulated by specific microRNAs, have specific chromosomal distribution, or have other biological features that make them particularly interesting or relevant for a given question (Bammler *et al*, 2005; Draghici *et al*, 2003; Hedegaard *et al*, 2009). Even identifying genes associated with a specific cell type in blood may provide enormous mechanistic information, for example T cells versus B cells versus platelets (Du *et al*, 2006; Xu *et al*, 2008). At the present time, these approaches help the researcher distill a very large amount of data into a single word or phrase. They still need to be supplemented by an in-depth review of the literature for each question and study.

Most microarray studies, particularly those of blood, have not replicated single genes. This may relate to the possibility that there may not be a perfect set of predictors for any condition. That is, in our recent replication studies compared with our previously published studies, two different sets of genes predicted the previous study with >80% sensitivity and specificity, whereas because of the statistical design predicted the current study with >90% sensitivity and specificity (Jickling *et al*, 2010; Stamova *et al*, 2010). Finally, there may be hundreds of predictive genes, and the best predictors might actually vary over different populations at different times.

There are several approaches to validating microarray studies. First, virtually every array study now derives a set of gene predictors and then uses these to perform ‘crossvalidation’ on the same cohort. If this does not produce >80% sensitivity and specificity, then the prospects for these predictors being useful and replicable are poor. In addition, this is not an acceptable method for accounting for multiple comparisons since it tests the predictors on the gene set from which it was derived.

In order to account for multiple comparisons, the traditional approach is to develop a gene expression profile in a ‘training cohort’ and then validate it in a ‘test cohort’ that is independent of the first. This basic principle is applied to all whole genome studies or any study where multiple parameters are studied.

Currently, the most common way of performing validation studies is to design a study with two cohorts. In the first cohort, the disease of interest is compared with controls and a gene profile derived from this comparison. This gene profile is then used on the second cohort in the same study to predict who has the disease of interest (or other variable) or who is a control. This approach I will refer to as a ‘within study with two cohorts’ approach. Though this approach demonstrates that the predictors are reliable and not false positives, the approach does not necessarily account for the problem of batch or between study comparisons. For the within study

with two cohorts approach, most laboratories would intermix the two cohorts and the control and experimental samples to eliminate or decrease batch effects. Thus, this approach does not address the question of whether the predictors will work across different studies performed on different cohorts at different sites or at different points in time.

The only way to prove predictors is to use a 'two study two cohort approach.' For this approach, the first study uses two cohorts: a gene profile is derived from the training cohort and this profile is then used to predict the individuals in the second 'test cohort.' Once done, an independent study is performed: the gene profile from the previous study is used to predict the individuals in this second study, blinded to the clinical disease group. What if the second study fails? Though it is not entirely clear how to solve this, potentially taking half of the individuals in each study to derive a profile for the training set could be used to predict the other half of the individuals. Alternatively, combining the gene profiles in both studies might be successful for predicting subjects in a third study. The reason for performing a two study–two cohort study is to absolutely ensure that the predictors were not dependent upon an unrecognized batch effect or unrecognized bias.

### Role of reverse transcriptase polymerase chain reaction and future of isolating single cell types

The majority of initial preliminary studies have used microarrays because they can survey all of the known RNAs. It is still not clear what platform is best for confirming a predictor set of genes derived from a 'training set' of individuals on the 'test set' of individuals. This could be done using microarrays, reverse transcriptase PCR, Nanostring, next generation sequencing or other emerging technologies. Using microarrays has the advantage of using the same platform to replicate using the same platform the genes were derived from. However, PCR has the advantage of providing a technical replication, focusing the study just on the predictor set of genes to markedly decrease multiple comparison corrections, and precisely identify the RNA of a given gene using specific primers for a gene or exon. The eventual platform that might be used for clinical assessments is unclear since PCR on a large number of genes is still costly, microarrays continue to evolve and new technologies like Nanostring and Next Generation Sequencing are providing viable alternatives.

Most of our own work has focused on whole blood obtained using PAXgene tubes because of the ease of obtaining samples, storing them frozen and processing afterwards. One disadvantage of this approach is that a large portion of the RNA is accounted for by

globin so that low abundance RNA transcripts tend to be lost. There are globin reduction methods, which have improved over the years and are commercially available. We have not used these because our preliminary studies had not used globin reduction and early globin reduction methods produced variable results. If globin reduction is to be used, reproducibility and reliability of results needs to be confirmed.

The major disadvantage of PAXgene tubes, however, is that whole blood is assessed. This may be adequate for biomarkers, but is woefully far from understanding mechanisms related to different cell types in blood. The future will require isolation of individual cell types including neutrophils, B and T lymphocytes, monocytes and their many subtypes. We and others have published gene expression profiles for these cell types (Du *et al*, 2006; Kobayashi *et al*, 2007; Kotz *et al*, 2010; Robbins *et al*, 2008; Wong *et al*, 2010). The potential problem with these studies is that isolation may affect gene expression (Holmes *et al*, 2009; Kotz *et al*, 2010; Lee *et al*, 2010; Shim *et al*, 2010). This may be solved by microfluidic devices that use cell-specific antibodies that isolate specific cell types at the bed side (Kotz *et al*, 2010).

### Future of RNA profiling—exon arrays, splicing, next gen RNA sequencing

Even given the above problems, gene expression studies at the very least will provide enormous insights into the pathophysiology of stroke. Though not reviewed here because of space, microRNA studies of blood and brain, are providing novel insights into how large groups of genes are regulated following brain ischemia (Dharap *et al*, 2009; Jeyaseelan *et al*, 2008; Liu *et al*, 2010; Reid *et al*, 2010; Tan *et al*, 2009). A number of studies of gene expression related to atherosclerotic plaques have been published and support the inflammatory nature of these lesions, and provide insights into additional pathways associated with high-risk lesions (Faber *et al*, 2002; Gagarin *et al*, 2005; Vemuganti and Dempsey, 2005). Ongoing studies in our laboratory and others are assessing subarachnoid hemorrhage, arteriovenous malformations, intracerebral hemorrhage, lacunar stroke (Jickling *et al*, submitted for publication), transient ischemic attacks (Zhan *et al*, submitted for publication), and risk of stroke from transient ischemic attacks, effects of gender on gene expression, and the correlation between gene expression and CIMT (carotid intima-media thickness), which is an index of risk of vascular disease including stroke (Turner *et al*, 2011b).

Not considered in this review is the potential impact of exon arrays and next generation RNA sequencing (Taub *et al*, 2010). Current exon arrays allow one to assess expression of individual exons,

and to predict alternative splicing of some genes (Tian *et al*, 2010, 2011). This is of great interest since exon splicing accounts in part for diversity among cell types and organs. Since there appear to be types of alternative splicing that can be specific for certain tumor types (Werner, 2010), it is possible that the immune response to brain injury might be associated with organ and perhaps even disease-specific alternative splicing due to the cell signaling that is specific for immune cells' response to brain ischemia (Courtney *et al*, 2010). These studies have yet to be done in stroke in part because they require even more complex and expensive resources.

Microarrays may eventually be replaced by next generation sequencing methods (RNA sequencing) to evaluate gene expression (Hawkins *et al*, 2010). With this technology, the entire transcriptome (RNAs) of a given individual can be sequenced (Werner, 2010). In theory, alternative splice variants of a given RNA can be quantified, as can the expression at the individual exon level (Hawkins *et al*, 2010; Werner, 2010). In addition, it will be possible to examine noncoding RNAs, which include the recently discovered micro-RNAs (Werner, 2010). The names of the many more noncoding RNAs are listed in Figure 12—with the realization that most of the genome codes for RNAs that are not translated into proteins, but instead likely serve regulatory functions many of which are still unknown. This exploding field will likely fuel disease-related research (Hawkins *et al*, 2010). This may also apply to different types of brain injuries where immune-specific splicing or types of noncoding RNA regulation will be specific for each type of brain injury (Hawkins *et al*, 2010).

## Disclosure/conflict of interest

The authors declare no conflict of interest.

## References

- Akopov S, Sercombe R, Seylaz J (1996) Cerebrovascular reactivity: role of endothelium/platelet/leukocyte interactions. *Cerebrovasc Brain Metab Rev* 8:11–94
- Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson Jr J, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403:503–11
- Allan SM, Rothwell NJ (2003) Inflammation in central nervous system injury. *Philos Trans R Soc Lond B Biol Sci* 358:1669–77
- Amarenco P, Bogousslavsky J, Caplan LR, Donnan GA, Hennerici MG (2009) Classification of stroke subtypes. *Cerebrovasc Dis* 27:493–501
- Babarro EG, Rego AR, Gonzalez-Juanatey JR (2009) Cardioembolic stroke: call for a multidisciplinary approach. *Cerebrovasc Dis* 27(Suppl 1):82–7
- Baird AE (2006) Blood genomic profiling: novel diagnostic and therapeutic strategies for stroke? *Biochem Soc Trans* 34:1313–7
- Baird AE (2007) Blood genomics in human stroke. *Stroke* 38:694–8
- Bammler T, Beyer RP, Bhattacharya S, Boorman GA, Boyles A, Bradford BU, Bumgarner RE, Bushel PR, Chaturvedi K, Choi D, Cunningham ML, Deng S, Dressman HK, Fannin RD, Farin FM, Freedman JH, Fry RC, Harper A, Humble MC, Hurban P, Kavanagh TJ, Kaufmann WK, Kerr KF, Jing L, Lapidus JA, Lasarev MR, Li J, Li YJ, Lobenhofer EK, Lu X, Malek RL, Milton S, Nagalla SR, O'malley JP, Palmer VS, Pattee P, Paules RS, Perou CM, Phillips K, Qin LX, Qiu Y, Quigley SD, Rodland M, Rusyn I, Samson LD, Schwartz DA, Shi Y, Shin JL, Sieber SO, Slifer S, Speer MC, Spencer PS, Sproles DI, Swenberg JA, Suk WA, Sullivan RC, Tian R, Tennant RW, Todd SA, Tucker CJ, Van Houten B, Weis BK, Xuan S, Zarbl H (2005) Standardizing global gene expression analysis between laboratories and across platforms. *Nat Methods* 2:351–6
- Barr TL, Conley Y, Ding J, Dillman A, Warach S, Singleton A, Matarin M (2010) Genomic biomarkers and cellular pathways of ischemic stroke by RNA gene expression profiling. *Neurology* 75:1009–14
- Beck J, Stummer W, Lehmborg J, Baethmann A, Uhl E (1997) Leukocyte-endothelium interactions in global cerebral ischemia. *Acta Neurochir Suppl* 70:53–5
- Becker KJ (2010) Modulation of the postischemic immune response to improve stroke outcome. *Stroke* 41:S75–8
- Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I (2001) Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 125:279–84
- Caplan LR, Fisher M (2007) The endothelium, platelets, and brain ischemia. *Rev Neurol Dis* 4:113–21
- Carmichael ST (2003) Gene expression changes after focal stroke, traumatic brain and spinal cord injuries. *Curr Opin Neurol* 16:699–704
- Chai V, Vassilakos A, Lee Y, Wright JA, Young AH (2005) Optimization of the PAXgene blood RNA extraction system for gene expression analysis of clinical samples. *J Clin Lab Anal* 19:182–8
- Chalela JA (2009) Evaluating the carotid plaque: going beyond stenosis. *Cerebrovasc Dis* 27(Suppl 1):19–24
- Chamorro A, Hallenbeck J (2006) The harms and benefits of inflammatory and immune responses in vascular disease. *Stroke* 37:291–3
- Chavez JC, Hurko O, Barone FC, Feuerstein GZ (2009) Pharmacologic interventions for stroke: looking beyond the thrombolysis time window into the penumbra with biomarkers, not a stopwatch. *Stroke* 40:e558–63
- Clark WM, Coull BM, Briley DP, Mainolfi E, Rothlein R (1993) Circulating intercellular adhesion molecule-1 levels and neutrophil adhesion in stroke. *J Neuroimmunol* 44:123–5
- Courtney E, Kornfeld S, Janitz K, Janitz M (2010) Transcriptome profiling in neurodegenerative disease. *J Neurosci Methods* 193:189–202
- Dassan P, Keir G, Brown MM (2009) Criteria for a clinically informative serum biomarker in acute ischaemic stroke: a review of S100B. *Cerebrovasc Dis* 27:295–302
- del Zoppo GJ (2010) Acute anti-inflammatory approaches to ischemic stroke. *Ann N Y Acad Sci* 1207:143–8

- Dharap A, Bowen K, Place R, Li LC, Vemuganti R (2009) Transient focal ischemia induces extensive temporal changes in rat cerebral microRNAome. *J Cereb Blood Flow Metab* 29:675–87
- Downes CE, Crack PJ (2010) Neural injury following stroke: are Toll-like receptors the link between the immune system and the CNS? *Br J Pharmacol* 160:1872–88
- Draghici S, Khatri P, Martins RP, Ostermeier GC, Krawetz SA (2003) Global functional profiling of gene expression. *Genomics* 81:98–104
- Du X, Tang Y, Xu H, Lit L, Walker W, Ashwood P, Gregg JP, Sharp FR (2006) Genomic profiles for human peripheral blood T cells, B cells, natural killer cells, monocytes, and polymorphonuclear cells: comparisons to ischemic stroke, migraine, and Tourette syndrome. *Genomics* 87:693–703
- El Husseini N, Laskowitz DT (2010) Clinical application of blood biomarkers in cerebrovascular disease. *Expert Rev Neurother* 10:189–203
- Elkind MS (2009) Inflammatory markers and stroke. *Curr Cardiol Rep* 11:12–20
- Elkind MS (2010) Inflammatory mechanisms of stroke. *Stroke* 41:S3–8
- Elneihoum AM, Falke P, Axelsson L, Lundberg E, Lindgarde F, Ohlsson K (1996) Leukocyte activation detected by increased plasma levels of inflammatory mediators in patients with ischemic cerebrovascular diseases. *Stroke* 27:1734–8
- Enlimomab Acute Stroke Trial Investigators (2001) Use of anti-ICAM-1 therapy in ischemic stroke: results of the Enlimomab Acute Stroke Trial. *Neurology* 57:1428–34
- Faber BC, Heeneman S, Daemen MJ, Cleutjens KB (2002) Genes potentially involved in plaque rupture. *Curr Opin Lipidol* 13:545–52
- Foerch C, Montaner J, Furie KL, Ning MM, Lo EH (2009) Invited article: searching for oracles? Blood biomarkers in acute stroke. *Neurology* 73:393–9
- Franks ZG, Campbell RA, Weyrich AS, Rondina MT (2010) Platelet-leukocyte interactions link inflammatory and thromboembolic events in ischemic stroke. *Ann N Y Acad Sci* 1207:11–7
- Freedman JE, Vitseva O, Tanriverdi K (2010) The role of the blood transcriptome in innate inflammation and stroke. *Ann N Y Acad Sci* 1207:41–5
- Furuya K, Takeda H, Azhar S, McCarron RM, Chen Y, Ruetzler CA, Wolcott KM, DeGraba TJ, Rothlein R, Hugli TE, del Zoppo GJ, Hallenbeck JM (2001) Examination of several potential mechanisms for the negative outcome in a clinical stroke trial of enlimomab, a murine anti-human intercellular adhesion molecule-1 antibody: a bedside-to-bench study. *Stroke* 32:2665–74
- Gagarin D, Yang Z, Butler J, Wimmer M, Du B, Cahan P, McCaffrey TA (2005) Genomic profiling of acquired resistance to apoptosis in cells derived from human atherosclerotic lesions: potential role of STATs, cyclinD1, BAD, and Bcl-XL. *J Mol Cell Cardiol* 39:453–65
- Git A, Dvinge H, Salmon-Divon M, Osborne M, Kutter C, Hadfield J, Bertone P, Caldas C (2010) Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA* 16:991–1006
- Grond-Ginsbach C, Hummel M, Wiest T, Horstmann S, Pflieger K, Hergenbahn M, Hollstein M, Mansmann U, Grau AJ, Wagner S (2008) Gene expression in human peripheral blood mononuclear cells upon acute ischemic stroke. *J Neurol* 255:723–31
- Grond-Ginsbach C, Horstmann S, Hummel M, Wiest T, Honold C, Pflieger K, Hergenbahn M, Hollstein M, Weninger A, Knyazev Y, Mansmann U, Wagner S, Grau AJ (2009) Increased expression of cell-cell signaling genes by stimulated mononuclear leukocytes in patients with previous atherothrombotic stroke. A whole genome expression profile study. *Eur Neurol* 62:30–9
- Guo Y, Hastie T, Tibshirani R (2007) Regularized linear discriminant analysis and its application in microarrays. *Biostatistics* 8:86–100
- Hallenbeck J (2010) How inflammation modulates central nervous system vessel activation and provides targets for intervention—a personal perspective. *Ann N Y Acad Sci* 1207:1–7
- Hansson GK (2009) Inflammatory mechanisms in atherosclerosis. *J Thromb Haemost* 7(Suppl 1):328–31
- Hawkins RD, Hon GC, Ren B (2010) Next-generation genomics: an integrative approach. *Nat Rev Genet* 11:476–86
- Hedegaard J, Arce C, Bicciato S, Bonnet A, Buitenhuis B, Collado-Romero M, Conley LN, Sancristobal M, Ferrari F, Garrido JJ, Groenen MA, Hornshoj H, Hulsege I, Jiang L, Jimenez-Marin A, Kommadath A, Lagarrigue S, Leunissen JA, Liaubet L, Neerinx PB, Nie H, van der Poel J, Prickett D, Ramirez-Boo M, Rebel JM, Robert-Granie C, Skarman A, Smits MA, Sorensen P, Tosser-Klopp G, Watson M (2009) Methods for interpreting lists of affected genes obtained in a DNA microarray experiment. *BMC Proc* 3(Suppl 4):S5
- Holmes D, Pettigrew D, Reccius CH, Gwyer JD, van Berkel C, Holloway J, Davies DE, Morgan H (2009) Leukocyte analysis and differentiation using high speed microfluidic single cell impedance cytometry. *Lab Chip* 9:2881–9
- Hsueh HM, Chen JJ, Kodell RL (2003) Comparison of methods for estimating the number of true null hypotheses in multiplicity testing. *J Biopharm Stat* 13:675–89
- Htun P, Fateh-Moghadam S, Tomandl B, Handschu R, Klinger K, Stellos K, Garlich C, Daniel W, Gawaz M (2006) Course of platelet activation and platelet-leukocyte interaction in cerebrovascular ischemia. *Stroke* 37:2283–7
- Jensen MB, Chacon MR, Sattin JA, Levine RL, Vemuganti R (2009) Potential biomarkers for the diagnosis of stroke. *Expert Rev Cardiovasc Ther* 7:389–93
- Jeyaseelan K, Lim KY, Armugam A (2008) MicroRNA expression in the blood and brain of rats subjected to transient focal ischemia by middle cerebral artery occlusion. *Stroke* 39:959–66
- Jickling G, Salam A, Mohammad A, Hussain MS, Scozzafava J, Nasser AM, Jeerakathil T, Shuaib A, Camicioli R (2009) Circulating endothelial progenitor cells and age-related white matter changes. *Stroke* 40:3191–6
- Jickling GC, Xu H, Stamova B, Ander BP, Zhan X, Tian Y, Liu D, Turner RJ, Mesias M, Verro P, Khoury J, Jauch EC, Pancioli A, Broderick JP, Sharp FR (2010) Signatures of cardioembolic and large-vessel ischemic stroke. *Ann Neurol* 68:681–92
- Kassner SS, Kollmar R, Bonaterra GA, Hildebrandt W, Schwab S, Kinscherf R (2009) The early immunological response to acute ischemic stroke: differential gene expression in subpopulations of mononuclear cells. *Neuroscience* 160:394–401
- King MD, Laird MD, Ramesh SS, Youssef P, Shakir B, Vender JR, Alleyne CH, Dhandapani KM (2010) Elucidating novel mechanisms of brain injury following subarachnoid hemorrhage: an emerging role for neuroproteomics. *Neurosurg Focus* 28:E10



- Kobayashi SD, Sturdevant DE, DeLeo FR (2007) Genome-scale transcript analyses in human neutrophils. *Methods Mol Biol* 412:441–53
- Kotz KT, Xiao W, Miller-Graziano C, Qian WJ, Russom A, Warner EA, Moldawer LL, De A, Bankey PE, Petritis BO, Camp II DG, Rosenbach AE, Goverman J, Fagan SP, Brownstein BH, Irimia D, Xu W, Wilhelmy J, Mindrinos MN, Smith RD, Davis RW, Tompkins RG, Toner M (2010) Clinical microfluidics for neutrophil genomics and proteomics. *Nat Med* 16:1042–7
- Lee D, Chen PJ, Lee GB (2010) The evolution of real-time PCR machines to real-time PCR chips. *Biosens Bioelectron* 25:1820–4
- Li S, Overman JJ, Katsman D, Kozlov SV, Donnelly CJ, Twiss JL, Giger RJ, Coppola G, Geschwind DH, Carmichael ST (2010) An age-related sprouting transcriptome provides molecular control of axonal sprouting after stroke. *Nat Neurosci* 13:1496–504
- Liu DZ, Tian Y, Ander BP, Xu H, Stamova BS, Zhan X, Turner RJ, Jickling G, Sharp FR (2010) Brain and blood microRNA expression profiling of ischemic stroke, intracerebral hemorrhage, and kainate seizures. *J Cereb Blood Flow Metab* 30:92–101
- Marsh BJ, Stevens SL, Hunter B, Stenzel-Poore MP (2009) Inflammation and the emerging role of the toll-like receptor system in acute brain ischemia. *Stroke* 40:S34–7
- Maurer MH, Haux D, Sakowitz OW, Unterberg AW, Kuschinsky W (2007) Identification of early markers for symptomatic vasospasm in human cerebral microdialysate after subarachnoid hemorrhage: preliminary results of a proteome-wide screening. *J Cereb Blood Flow Metab* 27:1675–83
- Maurer MH, Haux D, Unterberg AW, Sakowitz OW (2008) Proteomics of human cerebral microdialysate: from detection of biomarkers to clinical application. *Proteomics Clin Appl* 2:437–43
- Mieczkowski J, Tyburczy ME, Dabrowski M, Pokarowski P (2010) Probe set filtering increases correlation between Affymetrix GeneChip and qRT-PCR expression measurements. *BMC Bioinformatics* 11:104
- Millenaar FF, Okyere J, May ST, van Zanten M, Voesenek LA, Peeters AJ (2006) How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results. *BMC Bioinformatics* 7:137
- Montaner J (2009) Blood biomarkers to guide stroke thrombolysis. *Front Biosci (Elite Ed)* 1:200–8
- Moore DF, Li H, Jeffries N, Wright V, Cooper Jr RA, Elkahloun A, Gelderman MP, Zudaire E, Blevins G, Yu H, Goldin E, Baird AE (2005) Using peripheral blood mononuclear cells to determine a gene expression profile of acute ischemic stroke: a pilot investigation. *Circulation* 111:212–21
- Nah HW, Kang DW, Kwon SU, Kim JS (2010) Diversity of single small subcortical infarctions according to infarct location and parent artery disease: analysis of indicators for small vessel disease and atherosclerosis. *Stroke* 41:2822–7
- Pries AR, Kuebler WM (2006) Normal endothelium. *Handb Exp Pharmacol* 176:1–40
- Reid G, Kirschner MB, van Zandwijk N (2010) Circulating microRNAs: association with disease and potential use as biomarkers. *Crit Rev Oncol Hematol* 7:7
- Reiner A, Yekutieli D, Benjamini Y (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 19:368–75
- Ridder DA, Bulashevskaya S, Chaitanya GV, Babu PP, Brors B, Eils R, Schneider A, Schwaninger M (2009) Discovery of transcriptional programs in cerebral ischemia by in silico promoter analysis. *Brain Res* 1272:3–13
- Robbins SH, Walzer T, Dembele D, Thibault C, Defays A, Bessou G, Xu H, Vivier E, Sellars M, Pierre P, Sharp FR, Chan S, Kastner P, Dalod M (2008) Novel insights into the relationships between dendritic cell subsets in human and mouse revealed by genome-wide expression profiling. *Genome Biol* 9:R17
- Rothwell NJ, Luheshi GN (2000) Interleukin 1 in the brain: biology, pathology and therapeutic target [in process citation]. *Trends Neurosci* 23:618–25
- Saenger AK, Christenson RH (2010) Stroke biomarkers: progress and challenges for diagnosis, prognosis, differentiation, and treatment. *Clin Chem* 56:21–33
- Sharp FR, Xu H, Lit L, Walker W, Apperson M, Gilbert DL, Glauser TA, Wong B, Hershey A, Liu DZ, Pinter J, Zhan X, Liu X, Ran R (2006) The future of genomic profiling of neurological diseases using blood. *Arch Neurol* 63:1529–36
- Shedden K, Chen W, Kuick R, Ghosh D, Macdonald J, Cho KR, Giordano TJ, Gruber SB, Fearon ER, Taylor JM, Hanash S (2005) Comparison of seven methods for producing Affymetrix expression scores based on false discovery rates in disease profiling data. *BMC Bioinformatics* 6:26
- Shi L, Campbell G, Jones WD, Campagne F, Wen Z, Walker SJ, Su Z, Chu TM, Goodsaid FM, Pusztai L, Shaughnessy Jr JD, Oberthuer A, Thomas RS, Paules RS, Fielden M, Barlogie B, Chen W, Du P, Fischer M, Furlanello C, Gallas BD, Ge X, Megherbi DB, Symmans WF, Wang MD, Zhang J, Bitter H, Brors B, Bushel PR, Bylesjo M, Chen M, Cheng J, Cheng J, Chou J, Davison TS, Delorenzi M, Deng Y, Devanarayan V, Dix DJ, Dopazo J, Dorff KC, Elloumi F, Fan J, Fan S, Fan X, Fang H, Gonzaludo N, Hess KR, Hong H, Huan J, Irizarry RA, Judson R, Juraeva D, Lababidi S, Lambert CG, Li L, Li Y, Li Z, Lin SM, Liu G, Lobenhofer EK, Luo J, Luo W, McCall MN, Nikolsky Y, Pennello GA, Perkins RG, Philip R, Popovici V, Price ND, Qian F, Scherer A, Shi T, Shi W, Sung J, Thierry-Mieg D, Thierry-Mieg J, Thodima V, Trygg J, Vishnuvajjala L, Wang SJ, Wu J, Wu Y, Xie Q, Yousef WA, Zhang L, Zhang X, Zhong S, Zhou Y, Zhu S, Arasappan D, Bao W, Lucas AB, Berthold F, Brennan RJ, Buness A, Catalano JG, Chang C, Chen R, Cheng Y, Cui J, Czika W, Demichelis F, Deng X, Dosymbekov D, Eils R, Feng Y, Fostel J, Fulmer-Smentek S, Fuscoe JC, Gatto L, Ge W, Goldstein DR, Guo L, Halbert DN, Han J, Harris SC, Hatzis C, Herman D, Huang J, Jensen RV, Jiang R, Johnson CD, Jurman G, Kahlert Y, Khuder SA, Kohl M, Li J, Li L, Li M, Li QZ, Li S, Li Z, Liu J, Liu Y, Liu Z, Meng L, Madera M, Martinez-Murillo F, Medina I, Meehan J, Miclaus K, Moffitt RA, Montaner D, Mukherjee P, Mulligan GJ, Neville P, Nikolskaya T, Ning B, Page GP, Parker J, Parry RM, Peng X, Peterson RL, Phan JH, Quanz B, Ren Y, Riccadonna S, Roter AH, Samuelson FW, Schumacher MM, Shambaugh JD, Shi Q, Shippey R, Si S, Smalter A, Sotiriou C, Soukup M, Staedtler F, Steiner G, Stokes TH, Sun Q, Tan PY, Tang R, Tezak Z, Thorn B, Tsyganova M, Turpaz Y, Vega SC, Visintainer R, von Frese J, Wang C, Wang E, Wang J, Wang W, Westermann F, Willey JC, Woods M, Wu S, Xiao N, Xu J, Xu L, Yang L, Zeng X, Zhang J, Zhang L, Zhang M, Zhao C, Puri RK, Scherf U, Tong W, Wolfinger RD (2010) The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of

- microarray-based predictive models. *Nat Biotechnol* 28:827–38
- Shim JS, Browne AW, Ahn CH (2010) An on-chip whole blood/plasma separator with bead-packed microchannel on COC polymer. *Biomed Microdevices* 12:949–957
- Stamova B, Xu H, Jickling G, Bushnell C, Tian Y, Ander BP, Zhan X, Liu D, Turner R, Adamczyk P, Khoury JC, Pancioli A, Jauch E, Broderick JP, Sharp FR (2010) Gene expression profiling of blood for the prediction of ischemic stroke. *Stroke* 41:2171–7
- Stamova BS, Apperson M, Walker WL, Tian Y, Xu H, Adamczyk P, Zhan X, Liu DZ, Ander BP, Liao IH, Gregg JP, Turner RJ, Jickling G, Lit L, Sharp FR (2009) Identification and validation of suitable endogenous reference genes for gene expression studies in human peripheral blood. *BMC Med Genomics* 2:49
- Stevenson SF, Doubal FN, Shuler K, Wardlaw JM (2010) A systematic review of dynamic cerebral and peripheral endothelial function in lacunar stroke versus controls. *Stroke* 41:e434–42
- Tan KS, Armugam A, Sepramaniam S, Lim KY, Setyowati KD, Wang CW, Jeyaseelan K (2009) Expression profile of MicroRNAs in young stroke patients. *PLoS One* 4:e7689
- Tan YD, Fornage M, Fu YX (2006) Ranking analysis of microarray data: a powerful method for identifying differentially expressed genes. *Genomics* 88:846–54
- Tang Y, Gilbert DL, Glauser TA, Hershey AD, Sharp FR (2005) Blood gene expression profiling of neurologic diseases: a pilot microarray study. *Arch Neurol* 62:210–5
- Tang Y, Hershey AD, Powers SW, Gilbert DL, Glauser TA, Sharp FR (2004a) Genomic abnormalities in patients with migraine and chronic migraine: blood gene expression evidence of platelet abnormalities. *Headache* 44:994–1004
- Tang Y, Lu A, Aronow BJ, Sharp FR (2001) Blood genomic responses differ after stroke, seizures, hypoglycemia, and hypoxia: blood genomic fingerprints of disease. *Ann Neurol* 50:699–707
- Tang Y, Lu A, Aronow BJ, Wagner KR, Sharp FR (2002) Genomic responses of the brain to ischemic stroke, intracerebral haemorrhage, kainate seizures, hypoglycemia, and hypoxia. *Eur J Neurosci* 15:1937–52
- Tang Y, Lu A, Ran R, Aronow BJ, Schorry EK, Hopkin RJ, Gilbert DL, Glauser TA, Hershey AD, Richtand NW, Privitera M, Dalvi A, Sahay A, Szaflarski JP, Ficker DM, Ratner N, Sharp FR (2004b) Human blood genomics: distinct profiles for gender, age and neurofibromatosis type 1. *Brain Res Mol Brain Res* 132:155–67
- Tang Y, Nee AC, Lu A, Ran R, Sharp FR (2003) Blood genomic expression profile for neuronal injury. *J Cereb Blood Flow Metab* 23:310–9
- Tang Y, Schapiro MB, Franz DN, Patterson BJ, Hickey FJ, Schorry EK, Hopkin RJ, Wylie M, Narayan T, Glauser TA, Gilbert DL, Hershey AD, Sharp FR (2004c) Blood expression profiles for tuberous sclerosis complex 2, neurofibromatosis type 1, and Down's syndrome. *Ann Neurol* 56:808–14
- Tang Y, Xu H, Du X, Lit L, Walker W, Lu A, Ran R, Gregg JP, Reilly M, Pancioli A, Khoury JC, Sauerbeck LR, Carrozzella JA, Spilker J, Clark J, Wagner KR, Jauch EC, Chang DJ, Verro P, Broderick JP, Sharp FR (2006) Gene expression in blood changes rapidly in neutrophils and monocytes after ischemic stroke in humans: a microarray study. *J Cereb Blood Flow Metab* 26:1089–102
- Taub MA, Corrada Bravo H, Irizarry RA (2010) Overcoming bias and systematic errors in next generation sequencing data. *Genome Med* 2:87
- Thach DC, Lin B, Walter E, Kruzlock R, Rowley RK, Tibbetts C, Stenger DA (2003) Assessment of two methods for handling blood in collection tubes with RNA stabilizing agent for surveillance of gene expression profiles with high density microarrays. *J Immunol Methods* 283:269–79
- Tian Y, Apperson ML, Ander BP, Liu D, Stamova BS, Jickling GC, Enriquez R, Agius MA, Sharp FR (2010) Differences in exon expression and alternatively spliced genes in blood of multiple sclerosis compared to healthy control subjects. *J Neuroimmunol* 2:2
- Tian Y, Liao IH, Zhan X, Gunther JR, Ander BP, Liu D, Lit L, Jickling GC, Corbett BA, Bos-Veneman NG, Hoekstra PJ, Sharp FR (2011) Exon expression and alternatively spliced genes in Tourette syndrome. *Am J Med Genet B Neuropsychiatr Genet* 156:72–8
- Tibshirani R, Hastie T, Narasimhan B, Chu G (2002) Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci USA* 99:6567–72
- Turner R, Jickling GC, Sharp FR (2011a) Are underlying assumptions of current animal models of human stroke correct: from STAIRs to high hurdles? *Trans Stroke Res*; advance online publication, 12 February 2011.
- Turner RJ, Bushnell CD, Register TJ, Sharp FR (2011b) Gender dependent correlations of carotid intima media thickness with gene expression in blood. *Trans Stroke Res* (in press)
- Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 98:5116–21
- Vartanian K, Slotke R, Johnstone T, Casale A, Planck SR, Choi D, Smith JR, Rosenbaum JT, Harrington CA (2009) Gene expression profiling of whole blood: comparison of target preparation methods for accurate and reproducible microarray analysis. *BMC Genomics* 10:2
- Vemuganti R, Dempsey RJ (2005) Carotid atherosclerotic plaques from symptomatic stroke patients share the molecular fingerprints to develop in a neoplastic fashion: a microarray analysis study. *Neuroscience* 131:359–74
- Walker WL, Liao IH, Gilbert DL, Wong B, Pollard KS, McCulloch CE, Lit L, Sharp FR (2008) Empirical Bayes accommodation of batch-effects in microarray data using identical replicate reference samples: application to RNA expression profiling of blood from Duchenne muscular dystrophy patients. *BMC Genomics* 9:494
- Werner T (2010) Next generation sequencing in functional genomics. *Brief Bioinform* 2010:25
- Whiteley W, Chong WL, Sengupta A, Sandercock P (2009a) Blood markers for the prognosis of ischemic stroke: a systematic review. *Stroke* 40:e380–9
- Whiteley W, Jackson C, Lewis S, Lowe G, Rumley A, Sandercock P, Wardlaw J, Dennis M, Sudlow C (2009b) Inflammatory markers and poor outcome after stroke: a prospective cohort study and systematic review of interleukin-6. *PLoS Med* 6:e1000145
- Wong HR, Freishtat RJ, Monaco M, Odoms K, Shanley TP (2010) Leukocyte subset-derived genomewide expression profiles in pediatric septic shock. *Pediatr Crit Care Med* 11:349–55
- Wu TC, Grotta JC (2010) Stroke treatment and prevention: five new things. *Neurology* 75:S16–21
- Xu H, Stamova B, Jickling G, Tian Y, Zhan X, Ander BP, Liu D, Turner R, Rosand J, Goldstein LB, Furie KL, Verro P, Johnston SC, Sharp FR, Decarli CS (2010) Distinctive

- RNA expression profiles in blood associated with white matter hyperintensities in brain. *Stroke* 41:2744–9
- Xu H, Tang Y, Liu DZ, Ran R, Ander BP, Apperson M, Liu XS, Khoury JC, Gregg JP, Pancioli A, Jauch EC, Wagner KR, Verro P, Broderick JP, Sharp FR (2008) Gene expression in peripheral blood differs after cardioembolic compared with large-vessel atherosclerotic stroke: biomarkers for the etiology of ischemic stroke. *J Cereb Blood Flow Metab* 28:1320–8
- Yamamoto T, Sekiyama A, Sekiguchi H, Yoshida T, Miyagi Y (2006) Examination of stability of bone marrow blood RNA in the PAXgene tube. *Lab Hematol* 12:143–7
- Yao H, Nakahara T, Nakagawa N, Hashimoto K, Kuroki T (2009) Regional and temporal changes in proteomic profile after middle cerebral artery occlusion with or without reperfusion in rats. *Neurochem Res* 34:1999–2007
- Zakharkin SO, Kim K, Mehta T, Chen L, Barnes S, Scheirer KE, Parrish RS, Allison DB, Page GP (2005) Sources of variation in Affymetrix microarray experiments. *BMC Bioinformatics* 6:214
- Zhan X, Ander BP, Jickling G, Turner R, Stamova B, Xu H, Liu D, Davis RR, Sharp FR (2010) Brief focal cerebral ischemia that simulates transient ischemic attacks in humans regulates gene expression in rat peripheral blood. *J Cereb Blood Flow Metab* 30:110–8
- Zhan X, Kim C, Sharp FR (2008) Very brief focal ischemia simulating transient ischemic attacks (TIAs) can injure brain and induce Hsp70 protein. *Brain Res* 1234:183–97
- Zhang X, Guo T, Wang H, He W, Mei H, Hong M, Yu J, Hu Y, Song S (2008) Potential biomarkers of acute cerebral infarction detected by SELDI-TOF-MS. *Am J Clin Pathol* 130:299–304



This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>