Published in final edited form as:

J Child Neurol. 2011 September ; 26(9): 1131–1136. doi:10.1177/0883073811408093.

RNA Expression Profiles from Blood for the Diagnosis of Stroke and its Causes

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

Department of Neurology, University of California at Davis, Sacramento, California

Abstract

A blood test to detect stroke and its causes would be particularly useful in babies, young children, and patients in intensive care units, and for emergencies when imaging is difficult to obtain or unavailable. Using whole genome microarrays, we first showed specific gene expression profiles in rats 24 hours after ischemic and hemorrhagic stroke, hypoxia, and hypoglycemia. These proof-of-principle studies revealed that groups of genes (called gene profiles) can distinguish ischemic stroke patients from controls 3 hours to 24 hours after the strokes. In addition, gene expression profiles have been developed that distinguish stroke due to large-vessel atherosclerosis from cardioembolic stroke. These profiles will be useful for predicting the causes of cryptogenic stroke. Our results in adults suggest similar diagnostic tools could be developed for children.

Keywords

Gene expression profiling; blood; RNA; stroke; brain ischemia

Introduction

The diagnosis of myocardial infarction is currently made using a combination of a blood test for troponin and an electrocardiogram (ECG). This can be supplemented with additional evaluations, but the blood test and ECG are rapid and accurate. A similar tool would be useful for the diagnosis of stroke.

The diagnosis of stroke is made primarily based upon history, physical examination, and brain imaging. Brain imaging may be challenging in children, in patients in intensive care units, and in many emergencies. Indeed, the development of a stroke blood test has been identified as a priority by the National Institutes of Health. Many studies have now been performed searching for proteins in blood that could be used to diagnose stroke and its causes.^{1–3} Most of these protein biomarkers, however, have not been sufficiently sensitive and specific nor did they change early enough after stroke to be clinically useful. To address this, several recent studies have evaluated panels of proteins that have improved sensitivity and specificity.³

We adopted a different strategy by examining RNA expression in blood. The development of microarrays made it possible to assess all known genes on a single array. Since most

Address correspondence to: Frank R Sharp, MD, Department of Neurology, MIND Institute Wet Labs Room 2416, 2805 50th Street, Sacramento, CA 95817; Phone: 916-703-0368; Fax: 916-703-0369; frank.sharp@ucdmc.ucdavis.edu.

Presented at the Neurobiology of Disease in Children Symposium: Cerebrovascular Disease, in conjunction with the 39th Annual Meeting of the Child Neurology Society, Providence, Rhode Island, October 13, 2010.

RNA in blood is found in leukocytes, RNA expression from blood would mainly represent changes in immune cells related to brain injury. Another reason for studying RNA was that it is induced very rapidly and should be an earlier marker of injury than newly synthesized proteins. Finally, new arrays can assess individual exons, and next generation sequencing (RNAseq) can do the same to assess alternative splicing in immune cells in response to brain injury, including stroke.

Animal Studies

At the time we began these studies, it was not known whether gene expression would change in peripheral leukocytes in response to an isolated injury in brain. In addition, even gene expression changed, would there be enough cells to increase gene expression to a degree that microarray or reverse transcription polymerase chain reaction could detect it? Thus, we designed a study where adult rats were subjected to (1) ischemic stroke, (2) intracerebral hemorrhage produced by blood injections into brain, (3) kainic acid-induced status epilepticus, (4) hypoglycemia induced using insulin, and (5) systemic hypoxia (8% oxygen) and compared with sham controls and untouched control rats.⁴ After 24 hours, blood samples were drawn while the subjects were under deep anesthesia. RNA was isolated from peripheral blood monocytes and processed on early generation rat microarrays.⁴

Results of these studies are shown in Figure 1. An analysis of co-variance was performed for all the groups. A number of genes were differentially expressed between the groups. However, no single gene could differentiate one group from all of the others. Thus, we performed a cluster analysis of the regulated genes, as shown in Figure 1. Animal groups are shown on the y-axis and genes are on the x-axis. Up regulated gene expression is in red and down regulation is in bright green. Note that 2 groups of genes appear to be similarly regulated in all of the groups (shown with 2 arrows). We proposed that these genes related to surgical, handling, or other stresses that were common to the groups and that differed from those in the untouched control animal group.

The primary finding, however, was that although no single gene could differentiate the groups, a composite set of genes – called a gene expression profile – did separate the groups. If one looks down each row for each treatment condition and animal group, gene expression patterns differ for each condition. That is, there is a bar code of gene expression that is different for each group. This led us to hypothesize that there would be RNA expression profiles in the blood of humans that would be specific for ischemic stroke, hemorrhagic stroke, and other types of brain injury.⁴

Ischemic Stroke Studies in Humans

Following our proof of principle study in rodents,⁴ the first human study looking at whole-genome RNA expression profiles stroke patients was published by the Baird group.⁵ They isolated RNA from peripheral blood monocytes of patients several days after ischemic strokes and RNA from blood of healthy controls. Comparison of the RNA expression levels in the two groups showed 190 differentially expressed genes, and these predicted a second cohort with a sensitivity of 78% and a specificity of 80%.⁵ This study provided the first proof-of-concept in humans, though the accuracy of diagnosis was probably not sufficient for clinical purposes.

Therefore, we performed a separate study in humans but with some modifications that we thought would improve predictions. Instead of isolating peripheral blood monocytes from patients following stroke, we drew blood into PAXgene tubes. Isolating peripheral blood monocytes can potentially affect gene expression, particularly if there were differences in the time when the cells were isolated or variations in their processing. PAXgene tubes

offered the advantage that blood cells within the PAXgene tubes were immediately lysed and the RNA was stabilized by a proprietary reagent in the tubes. In addition, the PAXgene tubes were standard vacutainer tubes that could be drawn by untrained personnel and stored frozen until processed.

For our first human study, blood was collected into PAXgene tubes from 15 ischemic stroke patients at <3 hours, 5 hours, and 24 hours after their strokes and compared with 14 control samples.⁶ These patients were part of the CLEAR clinical trial, where patients received intravenous tissue plasminogen activator or tissue plasminogen activator plus eptifibatide after the first blood sample but before 3 hours. Blood was collected in PAXgene tubes and stored frozen at -70°C . RNA was isolated and processed on human Affymetrix microarrays (Santa Clara, CA). An ANOVA showed that 1355 genes were regulated in the blood of the ischemic stroke patients compared with healthy controls, using a false discovery rate of 5% and a fold change cut-off of 1.2. A cluster analysis of these 1355 genes showed they separated controls from ischemic stroke patients (Figure 2, lower panel). We then used a mathematical algorithm (Prediction Analysis of Microarrays) to derive the minimum number of genes that best predicted ischemic stroke versus controls (Figure 2). A cluster analysis of these 25 genes again showed that they separated ischemic stroke patients from healthy control subjects (Figure 2). These 25 genes were able to predict 100% of the stroke samples compared with control samples at 24 hours using cross validation, and they correctly predicted most patients who had strokes at 3 hours and 5 hours.⁶

To address the issue of reproducibility due to the problem of multiple comparisons, we have recently repeated our initial study in a larger cohort.⁷ Thus, patients with ischemic stroke ($n = 70$, 199 samples) were compared with control subjects who were healthy ($n = 38$), controls who had vascular risk factors ($n = 52$), and “controls” who had myocardial infarction ($n = 17$). Whole blood was drawn into PAXgene tubes at 3 hours, 5 hours and 24 hours after stroke onset. RNA was isolated and processed on microarrays. The genes from our original study predicted the new ischemic strokes with 93.5% sensitivity and 89.5% specificity. To obtain profiles that would distinguish ischemic stroke from all control subjects, we derived 60- and 46-gene profiles that differentiated control groups from 3-hour and 24-hour ischemic stroke samples, respectively (Table). A 97-probe set correctly classified 86% of ischemic strokes (3 hour+24 hour), 84% of healthy subjects, 96% of vascular risk factor subjects (Table), and 75% with myocardial infarction (not shown).⁷ 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.⁷

Causes of Ischemic Stroke in Humans

Ischemic strokes in adults are due to large-vessel atherosclerosis, cardioembolism, lacunar strokes, and a variety of medical and genetic causes. However, nearly a third of all ischemic strokes in adults do not have a cause identified in spite of intensive medical and neurological investigations. These strokes are referred to as “cryptogenic strokes” and are generally treated with platelet inhibitors if tolerated.^{8–13} However, if the cryptogenic strokes were known to be due to cardioembolism they would be treated with anticoagulants. As an example, patients may have a stroke due to atrial fibrillation, but the atrial fibrillation may have resolved by the time an ECG is performed. The identification of this paroxysmal atrial fibrillation may not be made until a second cardioembolic stroke.¹¹ Thus, we reasoned that if we could develop gene expression profiles for the known causes of ischemic stroke, then these profiles could be used to predict the causes of cryptogenic strokes.

We therefore performed the first human study to determine if there were RNA expression profiles in blood that could differentiate cardioembolic from large-vessel atherosclerotic

causes of ischemic stroke. In our first study, peripheral whole blood samples were drawn from acute ischemic stroke patients (<3, 5, and 24 hours) and healthy controls. RNA was isolated and processed on expression arrays. Expression profiles in the blood of cardioembolic stroke patients differed from large-vessel atherosclerotic stroke patients. Of the 77 genes that differed between the 2 groups (fold change >1.5, $P < .05$), a minimum number of 23 genes differentiated the 2 types of stroke with >90% specificity and sensitivity.¹⁴ Notably, some of the genes that distinguished cardioembolic from atherosclerotic stroke displayed little change over time. These might be genes expressed differentially prior to stroke – and perhaps indicate risk of stroke. Other genes displayed significant change over time, suggesting time-dependent alterations within the differential gene expression of immune cells when the stroke was caused by cardioembolism or atherosclerotic lesions and/or treatment effect.¹⁴ A caveat to the interpretation of these studies is that the changes of gene expression could be related to the treatments that the patients received at 3 hours. However, even with the <3 hours blood samples before any treatment, gene profiles differentiated cardioembolic from large-vessel atherosclerotic causes of stroke.¹⁴

We have recently confirmed these initial findings using 194 samples (<3 hours, 5 hours, and 24 hours after stroke) from 76 acute ischemic stroke patients.¹⁵ RNA was isolated from PAXgene tubes and processed on whole-genome human Affymetrix U133 expression microarrays. A 40-gene profile differentiated cardioembolic stroke from large-vessel stroke with >90% sensitivity and specificity (Figure 3). A separate 37-gene profile differentiated cardioembolic stroke due to atrial fibrillation from non-atrial fibrillation causes with >90% sensitivity and specificity (not shown). 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.¹⁵ 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 stroke in those patients where the causes cannot be definitely determined based upon current methodologies.

Conclusions

The animal and human studies to date suggest that RNA expression profiles obtained from blood have the potential to diagnose ischemic stroke and its causes. It should be emphasized that these findings have not been translated to clinical practice and further research is necessary before the profiles could be adapted for use in the intensive care unit, emergency room, and other venues where diagnosing pediatric stroke and its causes can be challenging. Even if the profiles are very predictive it takes a minimum of 2 to 3 days to obtain results, so more rapid analysis methods will be required. Furthermore, once all these technical issues are resolved, large multicenter trials will be required to validate any potential profiles for diagnosing ischemic stroke and/or identifying the causes of ischemic stroke. Future applications of this technology will include identifying patients with intracerebral hemorrhage, sepsis, and many other conditions where immune cells in blood respond to systemic factors or to focal brain injury.

Acknowledgments

Supported by grants from the National Institutes of Health (5R13NS040925-09), the National Institutes of Health Office of Rare Diseases Research, the Child Neurology Society, and the Children's Hemiplegia and Stroke Association. This work was supported by National Institutes of Health NS056302 (F.R.S.); and the American Heart Association Bugher Foundation (F.R.S.). Dr. Glen Jickling is a fellow of the Canadian Institutes of Health Research (CIHR). Dr. Bradley Ander and Dr. Yingfang Tian are fellows of the AHA-Bugher Foundation. This publication was also made possible by Grant Number UL1 RR024146 from the National Center for Medical Research to the

CTSC at UC Davis. Its contents are the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH. The authors wish to acknowledge Melanie Fridl Ross, MSJ, ELS, for editing assistance.

References

1. El Hussein N, Laskowitz DT. Clinical application of blood biomarkers in cerebrovascular disease. *Expert Rev Neurother.* 2010; 10:189–203. [PubMed: 20136376]
2. Foerch C, Montaner J, Furie KL, Ning MM, Lo EH. Invited article: Searching for oracles? Blood biomarkers in acute stroke. *Neurology.* 2009; 73:393–399. [PubMed: 19652144]
3. Saenger AK, Christenson RH. Stroke biomarkers: Progress and challenges for diagnosis, prognosis, differentiation, and treatment. *Clin Chem.* 2010; 56:21–33. [PubMed: 19926776]
4. Tang Y, Lu A, Aronow BJ, Sharp FR. Blood genomic responses differ after stroke, seizures, hypoglycemia, and hypoxia: Blood genomic fingerprints of disease. *Ann Neurol.* 2001; 50:699–707. [PubMed: 11761467]
5. Moore DF, Li H, Jeffries N, Wright V, et al. Using peripheral blood mononuclear cells to determine a gene expression profile of acute ischemic stroke: A pilot investigation. *Circulation.* 2005; 111:212–221. [PubMed: 15630028]
6. Tang Y, Xu H, Du X, et al. Gene expression in blood changes rapidly in neutrophils and monocytes after ischemic stroke in humans: A microarray study. *J Cereb Blood Flow Metab.* 2006; 26:1089–1102. [PubMed: 16395289]
7. Stamova B, Xu H, Jickling G, et al. Gene expression profiling of blood for the prediction of ischemic stroke. *Stroke.* 2010; 41:2171–2177. [PubMed: 20798371]
8. Amarenco P. Underlying pathology of stroke of unknown cause (cryptogenic stroke). *Cerebrovasc Dis.* 2009; 27(Suppl 1):97–103. [PubMed: 19342838]
9. Bang OY, Lee PH, Joo SY, et al. Frequency and mechanisms of stroke recurrence after cryptogenic stroke. *Ann Neurol.* 2003; 54:227–234. [PubMed: 12891675]
10. Belvis R, Santamaria A, Marti-Fabregas J, et al. Diagnostic yield of prothrombotic state studies in cryptogenic stroke. *Acta Neurol Scand.* 2006; 114:250–253. [PubMed: 16942544]
11. Eljovich L, Josephson SA, Fung GL, Smith WS. Intermittent atrial fibrillation may account for a large proportion of otherwise cryptogenic stroke: A study of 30-day cardiac event monitors. *J Stroke Cerebrovasc Dis.* 2009; 18:185–189. [PubMed: 19426887]
12. Guercini F, Acciarresi M, Agnelli G, Paciaroni M. Cryptogenic stroke: Time to determine aetiology. *J Thromb Haemost.* 2008; 6:549–554. [PubMed: 18208534]
13. Tayal AH, Tian M, Kelly KM, et al. Atrial fibrillation detected by mobile cardiac outpatient telemetry in cryptogenic tia or stroke. *Neurology.* 2008; 71:1696–1701. [PubMed: 18815386]
14. Xu H, Tang Y, Liu DZ, et al. 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.* 2008; 28:1320–1328. [PubMed: 18382470]
15. Jickling GC, Xu H, Stamova B, et al. Signatures of cardioembolic and largevessel ischemic stroke. *Ann Neurol.* 2010; 68:681–692. [PubMed: 21031583]

RNA expression Profiling of Blood from Rats

Unique expression profiles in blood 24 hours after injury

Figure 1.

Gene expression in the blood of rats 24 hours after injury is compared with gene expression in controls using a cluster analysis. Genes are on the x-axis and animal groups are on the y-axis. The groups included untouched controls, sham-operated controls, stroke produced using the suture/thread model, intracerebral hemorrhage produced by direct injections of blood into the striatum, status epilepticus produced by systemic kainic acid injections, 8% oxygen for 3 hours, and insulin-induced hypoglycemia. Up regulated genes are bright red, down regulated genes are bright green, and unchanged genes are yellow. The 2 arrows point to groups of genes that appear to be up regulated in all of the groups compared with untouched controls. This Figure is derived from Tang et al, published in 2001 in *Annals of Neurology*.⁴

RNA Expression in Blood after Stroke in Humans

Figure 2. RNA expression in the blood of humans at 3 hours, 5 hours, and 24 hours after ischemic stroke (on left) compared with healthy controls (on right) shown using cluster analyses. Subjects are on the x-axis and genes are on the y-axis. Up regulated genes are in red, down regulated genes are in green, and unchanged genes are in yellow. An ANOVA with a false discovery rate of 5% and fold change of 1.2 yielded 1355 genes that are shown in a cluster analysis in the bottom panel. Using the 1355 genes as the input into Prediction Analysis of Microarrays, a panel of 25 genes separated ischemic stroke samples from control samples as shown in the cluster analysis in the top panel. This Figure is derived from Tang et al, published in 2006 in the *Journal of Cerebral Blood Flow and Metabolism*.⁶

Gene Expression Prediction of Causes of Ischemic Stroke

Figure 3.

Genes that differentiate strokes due to large-vessel atherosclerosis compared with strokes due to cardioembolism. An ANOVA with a FDR of 5% and a fold change cut-off of 1.2 for atherosclerotic stroke patients compared with cardioembolic stroke patients yielded 40 genes. A cluster analysis of these 40 genes is shown in the upper left panel, a principal components analysis in the upper right panel, and a probability plot in the lower panel. Note that all of the subjects separate on the cluster and all but 2 separate on the principal components analysis. One cardioembolic and one atherosclerotic stroke patient are misclassified on the probability plot showing that the gene profile has a greater than 90% sensitivity and specificity. This Figure is derived from Jickling et al, published in 2010 in *Annals of Neurology*.¹⁵

Table

Sensitivity and Specificity for Predicting Stroke at 3 to 24 hours

Number of Genes	60 genes	46 genes	97 genes
	3 hour stroke vs controls	24 hour stroke vs controls	3 and 24 hour stroke vs controls
	Prediction Analysis of Microarrays	Prediction Analysis of Microarrays	Prediction Analysis of Microarrays
Ischemic Stroke	85%	91%	86%
Controls with Risk Factors	92%	92%	96%
Healthy Controls	84%	89%	84%

An ANOVA with a false discovery rate correction of 5% and a 1.2 fold change cut-off yielded 60 genes for 3 hour stroke versus controls, 46 genes for 24 hour stroke versus controls, and 97 genes for combined 3 hour and 24 hour stroke versus controls. These genes were then used as an input into Prediction Analysis of Microarrays software to predict ischemic stroke, controls with vascular risk factors, and healthy controls in a separate cohort. This table is derived from Stamova et al, published in 2010 in *Stroke*.⁷