Published in final edited form as:

Ann Neurol. 2010 November ; 68(5): 681–692. doi:10.1002/ana.22187.

Signatures of cardioembolic and large vessel ischemic stroke

Glen C Jickling, MD¹, Huichun Xu, MD, PhD¹, Boryana Stamova, PhD¹, Bradley P Ander, PhD¹, Xinhua Zhan, MD, PhD¹, Yingfang Tian, PhD¹, Dazhi Liu, PhD¹, Renée J Turner, PhD¹, Matthew Mesias, BSc¹, Piero Verro, MD¹, Jane Khoury, PhD², Edward C Jauch, MD³, Arthur Pancioli, MD⁴, Joseph P Broderick, MD⁴, and Frank R Sharp, MD¹

¹ Departments of Neurology and the MIND Institute, University of California at Davis, Sacramento, California, USA

² Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA

³ Division of Emergency Medicine, Medical University of South Carolina, Charleston, South Carolina, USA

⁴ Department of Neurology, University of Cincinnati, Cincinnati, Ohio, USA

Abstract

Objective—The cause of stroke remains unknown or cryptogenic in many patients. We sought to determine whether gene expression signatures in blood can distinguish between cardioembolic and large vessel causes of stroke, and whether these profiles can predict stroke etiology in the cryptogenic group.

Methods—A total of 194 samples from 76 acute ischemic stroke patients were analyzed. RNA was isolated from blood and run on Affymetrix U133 Plus2.0 microarrays. Genes that distinguish large vessel from cardioembolic stroke were determined at 3, 5, and 24 hours following stroke onset. Predictors were evaluated using cross-validation and a separate set of patients with known stroke subtype. The cause of cryptogenic stroke was predicted based on a model developed from strokes of known cause and identified predictors.

Results—A 40 gene profile differentiated cardioembolic stroke from large vessel stroke with >95% sensitivity and specificity. A separate 37 gene profile differentiated cardioembolic stroke due to atrial fibrillation from non-atrial fibrillation causes with >90% sensitivity and specificity. The identified genes elucidate differences in inflammation between stroke subtypes. When applied to patients with cryptogenic stroke, 17% are 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.

Interpretation—Gene expression signatures distinguish cardioembolic from large vessel causes of ischemic stroke. These gene profiles may add valuable diagnostic information in the management of patients with stroke of unknown etiology though they need to be validated in future independent, large studies.

Keywords

Gene expression; ischemic stroke; biomarker

Corresponding author: Dr. Glen C Jickling, University of California at Davis, M.I.N.D. Institute, 2805 50th Street, Sacramento, CA 95817, USA. Ph: 916-703-0449; FAX: 916-703-0369; gcjickling@ucdavis.edu.

Potential Conflicts of Interest

Nothing to report.

INTRODUCTION

Ischemic stroke is commonly classified by stroke etiology into cardioembolic, large vessel, small vessel lacunar, other and cryptogenic causes¹. Etiologic classification can guide treatment when a known cause is clearly identified^{2, 3}. However, in about 30% of patients the cause of stroke remains unknown or cryptogenic despite extensive investigation. Thus, better tools to identify the cause of stroke are warranted⁴.

Blood based biomarkers represent a potential tool to determine the cause of stroke. A number of protein biomarkers have been associated with stroke subtypes. Cardioembolic stroke is associated with brain natriuretic peptide and D-dimer; large vessel stroke is associated with C-reactive protein; and small vessel lacunar stroke is associated with homocysteine, ICAM-1, and thrombomodulin^{5–8}. However, biomarkers of ischemic stroke subtype currently lack sufficient sensitivity and specificity to be used in clinical practice.

In this discovery based study we sought to determine whether gene expression signatures in blood can distinguish cardioembolic from large vessel ischemic stroke, and whether these gene expression signatures can be used to predict cardioembolic or large vessel causes in patients with cryptogenic stroke. Our preliminary data suggests this is feasible, though further investigation of the concept and gene profiles is required⁹. The rationale for RNA expression changes in the blood of patients with ischemic stroke include inflammatory and prothrombotic changes associated with acute cerebral ischemia, symptomatic atherosclerosis and thromboembolism^{9–11}. Using whole genome microarrays, we identified a 40 gene profile that distinguished cardioembolic stroke from large vessel stroke, and a separate 37 gene profile that distinguished cardioembolic stroke due to atrial fibrillation from non-atrial fibrillation causes. These genes play roles in inflammation and represent a step toward better determining the cause in cryptogenic stroke.

SUBJECTS AND METHODS

1. Study Patients

Patients with acute ischemic stroke were enrolled from the CLEAR trial: a multicenter, randomized, double-blind safety study of recombinant tissue-plasminogen activator (rt-PA) and eptifibatide as previously described¹² (NCT00250991 at Clinical-Trials.gov). The institutional review board of each site approved the study protocol and written informed consent was obtained from each patient prior to study entry. Eligible patients had a diagnosis of acute ischemic stroke, therapy initiated within 3 hours of stroke onset, a National Institutes of Health Stroke Scale (NIHSS) >5, and were 18–80 years of age. All patients had standardized evaluations, including clinical examination, brain imaging, and investigations to determine cause as described below. Blood samples were drawn into PAXgene tubes (PreAnalytiX, Hilden, Germany) at ≤3 hours, 5 hours, and 24 hours after stroke onset for use in gene expression analysis. A total of 194 samples were obtained from 76 patients over the three time points.

Patients with cardioembolic stroke, large vessel stroke, and cryptogenic stroke (undetermined etiology) were included for study. Cause of stroke was determined using medical history, blood tests, brain imaging, Doppler and vascular angiography, and cardiac investigations. Patients with atrial fibrillation were identified using electrocardiogram, echocardiogram, and 24–48 hour cardiac monitoring. Cardioembolic stroke required at least one source of cardiac embolus to be identified as well as the exclusion of large vessel or small vessel causes of stroke. Cardioembolic sources included atrial fibrillation, acute myocardial infarction, prosthetic valve, and cardiomyopathy. Patients with endocarditis,

atrial myxoma and patent foramen ovale with atrial septal aneurysm were not included for study. Large vessel stroke required >50% stenosis of ipsilateral extracranial or major intracranial artery (MCA, PCA, BA) presumed to be due to atherosclerosis, and exclusion of cardioembolic and small vessel causes of stroke. Small vessel strokes were not included in the study and were defined as symptoms corresponding to a subcortical infarction less than 15mm in longest diameter on brain imaging. Stroke of other determined causes such as arterial dissection, vasculitis, and hypercoagulable states were excluded from the study. Cryptogenic stroke was defined as undetermined cause despite complete medical evaluation. Control blood samples were drawn from 23 healthy control subjects similar in age, gender and race to stroke subjects. These subjects were either healthy relatives of stroke patients, patients seen in outpatient clinics, or healthy subjects with no known medical disease. They had no history of ischemic stroke or cardiovascular disease, no recent infection, and no hematological disease.

2. Sample Processing

Whole blood was collected from the antecubital vein into PAXgene tubes (PreAnalytiX, Germany). PAXgene tubes were frozen at -80°C after 2 hours at room temperature. All samples were processed in the same laboratory. Total RNA was isolated according to the manufacturer's protocol (PAXgene blood RNA kit; Pre-AnalytiX). RNA was analyzed using Agilent 2100 Bioanalyzer for quality and Nano-Drop (Thermo Fisher) for concentration. Samples required A_{260}/A_{280} absorbance ratios of purified RNA ≥ 2.0 and 28S/18S rRNA ratios ≥ 1.8 . Reverse transcription, amplification, and sample labeling were carried out using Nugen's Ovation Whole Blood Solution (Nugen Technologies, San Carlos, CA). Each RNA sample was hybridized according to the manufacturer's protocol on Affymetrix Human U133 Plus 2.0 GeneChips (Affymetrix Santa Clara, CA), which contain 54,697 probe sets. The arrays were washed and processed on a Fluidics Station 450 and then scanned on a Genechip Scanner 3000. Samples were randomly assigned to microarray batch stratified by cause of stroke.

3. Gene Expression Profile Analyses

Raw expression values (probe level data) were imported into Partek software (Partek Inc., St. Louis, MO). They were log transformed and normalized using RMA (Robust Multichip Average) and our previously reported internal gene normalization method¹³. Statistical analysis, principal components analysis, and hierarchical unsupervised clustering analysis were performed using Partek Genomics Suite 6.04. The fidelity of genetic biomarker subsets as class prediction tools was established using k-nearest neighbor and 10-fold leave-one-out cross-validation in PAM (Prediction Analysis of Microarrays)¹⁴. Leave-one-out cross-validation provides a relatively unbiased estimate of the generalization ability of the genetic classifier. A model is generated on 90% of the samples and used to predict the remaining 10% of samples. This procedure is repeated 10 times to compute the overall error in the model. Ingenuity Pathway Analysis (IPA, Ingenuity Systems®, www.ingenuity.com) was used to determine whether the numbers of genes regulated within given pathways or cell functions were greater than expected by chance (Fisher's exact test).

4. Statistical Analyses

Differences in demographic data between groups were analyzed using Fisher's exact test and a two-tailed t-test where appropriate. All data is presented as mean $\pm$ standard error. To identify the gene expression profiles that distinguish cardioembolic stroke from large vessel stroke, a mixed effects model was used and included stroke etiology, time, the interaction between stroke etiology and time, and within subject variance. Unsupervised hierarchical clustering and principal components analysis (PCA) were used to evaluate the relationships

between cardioembolic stroke and large vessel stroke. Gene probes with a p value ≤ 0.005 and a fold change $\geq |1.2|$ were considered significant.

A similar analysis was used to identify the gene expression profiles that distinguish cardioembolic stroke due to atrial fibrillation from non-atrial fibrillation causes. A mixed effects model was used and included cardioembolic stroke etiology, time, the interaction between stroke etiology and time, and within subject variance. Unsupervised hierarchical clustering and PCA were used to evaluate the relationships between cardioembolic stroke caused by atrial fibrillation and non-atrial fibrillation. Gene probes with a p value ≤ 0.005 and a fold change $\geq |1.2|$ were considered significant.

Functional analysis was performed by comparing subjects with cardioembolic stroke and large vessel stroke to control subjects. A one-way analysis of covariance was used adjusting for age and gender. Gene probes with a p value ≤ 0.005 and a fold change $\geq |1.2|$ were considered significant and analyzed in IPA.

RESULTS

Cardioembolic versus Large Vessel Ischemic Stroke

Demographic and clinical characteristics of subjects used for the comparison of cardioembolic stroke to large vessel stroke are shown in Table 1. Atrial fibrillation was the only variable significantly different between groups ($p < 0.05$). There were 69 samples from twenty-three patients with cardioembolic stroke and 30 samples from ten patients with large vessel stroke.

Initially we evaluated the ability of our previously published 77 gene list to distinguish cardioembolic stroke from large vessel stroke⁹. This gene list was based on the first eleven patients enrolled in the CLEAR trial: seven with cardioembolic stroke and four with large vessel stroke. Using a k-nearest neighbor prediction model, the preliminary 77 gene list was used to predict the completed CLEAR trial patient population. Cardioembolic stroke was correctly predicted in 82.6% of samples, and large vessel stroke was correctly predicted in 80.0% of samples. However, on 10-fold leave-one-out cross-validation, 56.5% were correctly predicted as cardioembolic stroke and 60% were correctly predicted as large vessel stroke, with the probability of predicted diagnosis being below 90% in most samples. These results suggests that gene expression profiles in blood can distinguish cause of stroke, though further refinement is required to better generalize genomic predictors to a larger patient population.

Analysis of the complete CLEAR trial patients was thus performed. A mixed effects model identified 40 genes significantly different between cardioembolic stroke and large vessel stroke at all three time points (Supplementary Table 1). A hierarchical cluster plot of the 40 genes is shown in Figure 1a, and a Principal Component Analysis (PCA) in Figure 1b. The 40 genes separate cardioembolic stroke from large vessel stroke by at least 2 standard deviations (Figure 1b). The hierarchical cluster plot demonstrates a group of genes that are up-regulated in cardioembolic stroke and down-regulated in large vessel stroke. There is also a group of genes that are down-regulated in cardioembolic stroke and up-regulated in large vessel stroke. The 40 genes separate cardioembolic from large vessel stroke at each of the three time points studied (≤ 3 hours, 5 hours and 24 hours) (Supplementary Figure 1).

Prediction of Cardioembolic and Large Vessel Stroke

The ability of the 40 genes to predict cardioembolic stroke from large vessel stroke was evaluated using 10-fold leave-one-out cross-validation model in PAM. Of the 99 samples, 100% of the 69 samples with cardioembolic stroke were correctly predicted, and 96.7% of

the 30 samples with large vessel stroke were correctly predicted (Figure 2). The probability of predicted diagnosis was >90% for the majority of samples (Figure 2). To further evaluate the 40 gene list, it was applied to a separate group of patients with known cardioembolic stroke. Of the 10 samples, 90% (9/10) were correctly predicted as cardioembolic stroke.

The 40 gene list was subsequently used to predict the cause of stroke in patients with cryptogenic stroke. There were 36 patients (85 samples) with cryptogenic stroke. To be considered classified by the prediction model, all samples from each patient were required to have a >90% probability of the same predicted diagnosis. A total of 15 patients (41%) were predicted to have a profile similar to cardioembolic stroke with a probability >90%, and a total of 6 patients (17%) were predicted to have a profile similar to large vessel stroke with a probability >90%. This represents a potential re-classification of 58% of cryptogenic strokes to either cardioembolic or large vessel stroke.

Functional Analysis

To determine the functional pathways associated with cardioembolic and large vessel stroke, subjects with cardioembolic and large vessel stroke were compared to controls. There were 731 genes that were significantly different between cardioembolic stroke subjects and controls, and 782 genes that were significantly different between large vessel stroke and controls ($p < 0.005$, fold change ≥ 1.2). The overlap of these two gene lists is shown in a Venn diagram in Supplementary Figure 2. There were 503 genes unique to cardioembolic stroke, 554 genes unique to large vessel stroke, and 228 genes common to cardioembolic stroke and large vessel stroke. The top canonical and molecular functions of these gene lists are shown in Tables 2–4.

Of the 503 cardioembolic stroke genes, specific genes that have been previously associated with three of the main cardiac causes of stroke include atrial fibrillation genes - CREM, SLC8A1, KNCH7, KCNE1, myocardial infarction genes - PDE4B, TLR2, and heart failure genes - MAPK1, HTT, GNAQ, CD52, PDE4B, RAF1, CFLAR, and MDM2 (Table 2). Cardioembolic stroke was associated with development of lymphocytes, inflammatory disorder, cardiomyocyte cell death, and phosphatidylinositol 4-phosphate modification. Top canonical pathways included renin-angiotensin signaling, thrombopoietin signaling, NF- κ B activation, cardiac hypertrophy, and B cell receptor signaling (Table 2).

Of the 554 large vessel stroke genes, specific genes that have been previously associated with atherosclerotic lesion and atherosclerotic plaque include MMP9, FASLG, CX3R1, RAG1, TNF, IRAG1, CX3CR, and THBS1 (Table 3). Large vessel stroke was associated with T cell and leukocyte development, inflammation, and invasion. Top canonical pathways include T cell activation and regulation, CCR5 signaling in macrophages, relaxin signaling, and corticotropin releasing hormone signaling (Table 3).

A total of 228 genes were common to both subtypes of ischemic stroke (Supplementary Figure 2). They were associated with leukocyte and phagocyte development and movement, cardiovascular processes, NF- κ B response element expression, and oxidative stress (Table 4). Top canonical pathways included p38 MAPK signaling, toll-like receptor signaling, IL-6 and IL-10 signaling, NK- κ B signaling, B-cell receptor signaling, and NRF-mediated oxidative stress (Table 4).

Atrial fibrillation versus Non-Atrial fibrillation Cardioembolic Stroke

There were 23 subjects with cardioembolic stroke, 10 with atrial fibrillation and 13 with no atrial fibrillation identified on routine investigation. Initially, we sought to exclude any subjects in the non-atrial fibrillation group who are more likely to have undetected paroxysmal atrial fibrillation. To do this, the 10 patients with stroke due to atrial fibrillation

were initially compared to the 10 patients with large vessel stroke. A mixed effects model identified a 39 gene profile for atrial fibrillation. This profile was then used to predict which of the 13 cardioembolic stroke subjects without atrial fibrillation identified on routine investigation had the highest probability of being similar to atrial fibrillation. There were 5 subjects who fell within 4 standard deviations of the mean predicted probability of patients with known atrial fibrillation. These patients were considered more likely to have paroxysmal atrial fibrillation and thus were excluded from further analysis as a conservative method to reduce the possibility of paroxysmal atrial fibrillation being present in the non-atrial fibrillation group. The remaining eight non-atrial fibrillation patients were compared to the ten patients with atrial fibrillation. The demographic and clinical characteristics are shown in Table 5. Atrial fibrillation was the only variable significantly different between the two groups ($p < 0.05$). A mixed effects model identified 37 genes that were significantly different between atrial fibrillation and non-atrial fibrillation causes of cardioembolic stroke (Supplementary Table 2). A hierarchical cluster plot of the 37 genes is shown in Figure 3a, and a PCA in Figure 3b. The 37 genes clearly separate atrial fibrillation from non-atrial fibrillation (Figure 3) and can separate atrial fibrillation from non-atrial fibrillation cardioembolic stroke at each of the three time points studied (3 hours, 5 hours and 24 hours) (Supplementary Figure 3). The 37 genes were applied to the five subjects excluded from analysis, with two being predicted to be atrial fibrillation, two being indeterminate, and one being predicted to be non-atrial fibrillation cardioembolic stroke.

Prediction of Atrial Fibrillation and Non-Atrial Fibrillation Cardioembolic Stroke

The ability of the 37 genes to predict atrial fibrillation from non-atrial fibrillation causes of cardioembolic stroke was evaluated using a 10-fold leave-one-out cross-validation model in PAM. In the 60 samples, 100% of the 30 samples with atrial fibrillation cardioembolic stroke were correctly predicted, and 91.7% of the 30 samples with non-atrial fibrillation cardioembolic stroke were correctly predicted (Figure 4). Additionally, the probability of predicted diagnosis was $>90\%$ for most samples.

The 37 gene list was used to predict a test set of ten samples with cardioembolic stroke who did not have atrial fibrillation identified on routine testing. Of these ten samples, three (30%) were predicted to have paroxysmal atrial fibrillation with $>90\%$ probability when compared to the gene expression profile of subjects with known symptomatic atrial fibrillation. The 37 gene list was also used to predict the cause of stroke in patients with cryptogenic stroke. There were eleven patients with cryptogenic stroke who were predicted to have cardioembolic stroke based on the 40 gene profile. Of these eleven patients, three patients (27%) were predicted to have paroxysmal atrial fibrillation with a probability $>90\%$ based on a gene expression profile that was similar to subjects with known atrial fibrillation stroke.

DISCUSSION

Determining the cause of ischemic stroke is of paramount importance to optimally implement stroke prevention treatments. Previous reports demonstrate that about 30% of stroke patients have unknown or cryptogenic cause when classified by TOAST criteria⁴. However, patients initially diagnosed as cryptogenic can later be found to have a detectable cause of stroke, such as paroxysmal atrial fibrillation. This suggests that improved tools to determine cause of stroke are warranted. We describe the use of gene expression signatures in blood to distinguish cardioembolic from large vessel stroke on a molecular level. A 40 gene expression profile can distinguish cardioembolic from large vessel stroke, and a separate 37 gene expression profile can distinguish cardioembolic stroke due to atrial fibrillation from non-atrial fibrillation causes. When applied to cryptogenic stroke, 58% of subjects can be reclassified as being either cardioembolic or large vessel stroke with a probability $>90\%$.

It is important to emphasize some of the limitations of large scale gene expression profiling before discussing the use of gene expression profiles to distinguish stroke subtypes.¹⁵ Large populations of patients are required to fully develop and validate gene expression profiles. In breast cancer, several discovery and validation studies were required to develop a profile for clinical use. Likewise, the identified profiles that distinguish cardioembolic from large vessel stroke will require further study in larger samples that better estimate variance of a stroke population. In addition, the analysis of many genes increases the chance of false discovery. Validation in a second independent cohort of patients is required to evaluate the genes that distinguish cardioembolic from large vessel stroke. Despite the limitations of large scale gene expression studies, comparable approaches have been applied in patients with malignancy that have translated to PCR based arrays for diagnostic purposes^{16, 17}

Gene Expression Signatures and Stroke Classification

We identified a gene expression profile able to differentiate cardioembolic stroke from large vessel stroke. This distinction is clinically important because treatment and diagnostic testing are different between the two subtypes. In general, cardioembolic strokes benefit from anticoagulation, whereas large vessel strokes benefit from antiplatelet therapy and vascular surgery. We suggest that gene expression profiles could be used to complement current diagnostic tests to determine cause of stroke. In many cases gene expression profiles could help target specific testing, particularly in cryptogenic stroke. As a result, costly resources could be focused on subjects where they will have the highest yield.

Cardiac monitoring for atrial fibrillation is critical given the proven benefits of anticoagulation in primary and secondary stroke prevention. However, electrocardiogram and cardiac monitoring for 24 to 48 hours do not detect all patients with paroxysmal atrial fibrillation^{18, 19}. A gene expression profile suggesting a patient has a high probability of atrial fibrillation may provide an additional tool to prevent such missed treatment opportunities. Stroke patients with a molecular signature similar to cardioembolic stroke due to atrial fibrillation may represent a group where long term cardiac monitoring can be focused^{18, 20–22}. In this study, 41% of cryptogenic stroke patients were suggested to have a cardioembolic cause, and of these 27% were suggested to have paroxysmal atrial fibrillation. This is consistent with previous studies of cryptogenic stroke where an additional 9–28% cases of paroxysmal atrial fibrillation can be identified using long term cardiac monitoring^{18, 19, 23, 24}.

Gene expression profiles may also aid in the diagnosis of large vessel stroke. Evaluation of large vessel atherosclerotic disease includes imaging of extracranial and intracranial vessels using magnetic resonance angiography (MRA), computed tomography angiography (CTA), ultrasound, and conventional angiography. Inconsistencies in the results of vascular imaging do occur. For example, the degree of carotid stenosis by ultrasound may not be consistent with the degree of stenosis by MRA or CTA. Supplementing imaging with a gene expression profile suggestive of symptomatic atherosclerotic disease could add confidence to the diagnosis of large vessel atherosclerotic disease. The presence of large vessel disease is largely based on a single factor, that being the degree of vascular stenosis. The TOAST criteria defines a stenosis less than 50% as being negative for large vessel disease¹. However, in our study 17% of the cryptogenic group was predicted to have large vessel stroke. This finding may represent a stenosis <50% that was symptomatic, though further study of this potentially important finding is required. Gene expression profiles provide an additional measure of factors associated with symptomatic atherosclerotic disease, particularly inflammation. This concept is similar to MRI methods used to determine atheroma inflammation²⁵. Although these proposed applications of gene expression profiles show promise as tools to be used in conjunction with current diagnostic methods to determine cause of stroke, they will require validation in larger cohorts.

Functional Analysis

The rationale for changes in blood gene expression in patients with ischemic stroke rests largely in differences in patterns of inflammation. The major source of RNA in the blood is immune cells including leukocytes, neutrophils, and monocytes¹¹. Immune cells provide an indirect reflection of a patient's disease state and subsequent response, such as the immune response to ischemic brain tissue and immune response to disease mediated by vascular risk factors. Though the majority of these responses remain unclear, it appears there are differences in the way the responses are orchestrated between patients with cardioembolic and large vessel stroke. This is evidenced by the 40 gene profile for cardioembolic and large vessel stroke, and the 37 gene profile for cardioembolic stroke due to atrial fibrillation and non-atrial fibrillation. The fact that different genes are associated with stroke of large vessel, cardioembolic and atrial fibrillation origin suggests specific immune responses in each condition. The precise cause for these differences, including immune cell and immune-endothelial interactions remain largely unknown, but should become clearer as each condition and cause is better studied.

This study has several limitations. The sample size is small and further study in larger cohorts is required. Biases in terms of sample selection and distribution of patient characteristics have larger effects in smaller patient samples. Though not significant, in the samples used to derive the genes that distinguish large vessel from cardioembolic stroke there was more diabetes in the large vessel group and more prior strokes in the cardioembolic group. Likewise, in the samples used to derive the genes that distinguish atrial fibrillation from non-atrial fibrillation cardioembolic stroke, there were fewer males in the atrial fibrillation group. Subjects in this study also received fibrinolytic treatment after the first blood draw (≤ 3 h). Investigation of the profiles in untreated subjects will be required. However, samples from the 3 hour time point were untreated and were correctly classified, suggesting the described gene profiles are independent of treatment effects. The samples analyzed were limited to the 3 hour to 24 hour time window following ischemic stroke. Further study is required at time points beyond the initial 24 hours from stroke onset.

In conclusion, we provide evidence that gene expression signatures can distinguish between cardioembolic and large vessel subtypes of ischemic stroke. With further study in larger cohorts, gene expression profiles show promise for the development of PCR or microfluidic chip based blood test in ischemic stroke. This test could aid in determining the cause of stroke, particularly in patients currently classified as cryptogenic, and thus improve delivery of treatments to prevent stroke.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institutes of Health [NS056302 to F.R.S., PO21040N635110 to J.P.B.]; 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. Huichun Xu, Dr. Bradley Ander and Dr. Yingfang Tian are AHA-Bugher Fellows. 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 NCMR or NIH. We thank the investigators of the SPOTRIAS Stroke Network involved in the CLEAR trial at the University of Cincinnati for supplying blood samples for analysis. We appreciate the support of the MIND Institute, the Genomics and Expression Resource at the MIND Institute, and the UCD Department of Neurology.

References

1. Adams HP Jr, Bendixen BH, Kappelle LJ, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*. 1993; 24:35–41. [PubMed: 7678184]
2. Goldstein LB, Jones MR, Matchar DB, et al. Improving the reliability of stroke subgroup classification using the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) criteria. *Stroke*. 2001; 32:1091–1098. [PubMed: 11340215]
3. Ay H, Benner T, Arsava EM, et al. A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. *Stroke*. 2007; 38:2979–2984. [PubMed: 17901381]
4. Ionita CC, Xavier AR, Kirmani JF, et al. What proportion of stroke is not explained by classic risk factors? *Prev Cardiol*. 2005; 8:41–46. [PubMed: 15722693]
5. Laskowitz DT, Kasner SE, Saver J, et al. Clinical usefulness of a biomarker-based diagnostic test for acute stroke: the Biomarker Rapid Assessment in Ischemic Injury (BRAIN) study. *Stroke*. 2009; 40:77–85. [PubMed: 18948614]
6. Shibazaki K, Kimura K, Iguchi Y, et al. Plasma brain natriuretic peptide can be a biological marker to distinguish cardioembolic stroke from other stroke types in acute ischemic stroke. *Intern Med*. 2009; 48:259–264. [PubMed: 19252345]
7. Montaner J, Perea-Gainza M, Delgado P, et al. Etiologic diagnosis of ischemic stroke subtypes with plasma biomarkers. *Stroke*. 2008; 39:2280–2287. [PubMed: 18535284]
8. Hassan A, Hunt BJ, O'Sullivan M, et al. Markers of endothelial dysfunction in lacunar infarction and ischaemic leukoaraiosis. *Brain*. 2003; 126:424–432. [PubMed: 12538408]
9. 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]
10. 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]
11. Du X, Tang Y, Xu H, et al. 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*. 2006; 87:693–703. [PubMed: 16546348]
12. Pancioli AM, Broderick J, Brott T, et al. The combined approach to lysis utilizing eptifibatide and rt-PA in acute ischemic stroke: the CLEAR stroke trial. *Stroke*. 2008; 39:3268–3276. [PubMed: 18772447]
13. Stamova BS, Apperson M, Walker WL, et al. Identification and validation of suitable endogenous reference genes for gene expression studies in human peripheral blood. *BMC Med Genomics*. 2009; 2:49. [PubMed: 19656400]
14. Tibshirani RJ, Efron B. Pre-validation and inference in microarrays. *Stat Appl Genet Mol Biol*. 2002; 1 Article1.
15. Schulze A, Downward J. Navigating gene expression using microarrays—a technology review. *Nat Cell Biol*. 2001; 3:E190–195. [PubMed: 11483980]
16. Hedenfalk I, Duggan D, Chen Y, et al. Gene-expression profiles in hereditary breast cancer. *N Engl J Med*. 2001; 344:539–548. [PubMed: 11207349]
17. Valk PJ, Verhaak RG, Beijnen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med*. 2004; 350:1617–1628. [PubMed: 15084694]
18. 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]
19. Ziegler PD, Glotzer TV, Daoud EG, et al. Incidence of newly detected atrial arrhythmias via implantable devices in patients with a history of thromboembolic events. *Stroke*. 41:256–260. [PubMed: 20044517]
20. Harloff A, Handke M, Reinhard M, et al. Therapeutic strategies after examination by transesophageal echocardiography in 503 patients with ischemic stroke. *Stroke*. 2006; 37:859–864. [PubMed: 16439702]

21. Sacco RL, Prabhakaran S, Thompson JL, et al. Comparison of warfarin versus aspirin for the prevention of recurrent stroke or death: subgroup analyses from the Warfarin-Aspirin Recurrent Stroke Study. *Cerebrovasc Dis.* 2006; 22:4–12. [PubMed: 16567931]
22. Mohr JP, Thompson JL, Lazar RM, et al. A comparison of warfarin and aspirin for the prevention of recurrent ischemic stroke. *N Engl J Med.* 2001; 345:1444–1451. [PubMed: 11794192]
23. 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]
24. Gaillard N, Deltour S, Vilotijevic B, et al. Detection of paroxysmal atrial fibrillation with transtelephonic EKG in TIA or stroke patients. *Neurology.* 74:1666–1670. [PubMed: 20498434]
25. Tang TY, Muller KH, Graves MJ, et al. Iron oxide particles for atheroma imaging. *Arterioscler Thromb Vasc Biol.* 2009; 29:1001–1008. [PubMed: 19229073]

Figure 1.

Figure 1A. Hierarchical cluster plot of the 40 genes that were significantly different for cardioembolic stroke compared to 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. One group of genes has a high level of expression in cardioembolic stroke and a low level of expression in large vessel stroke. A separate group of genes has a low level of expression in cardioembolic stroke and a high level of expression in large vessel stroke. The cardioembolic group may cluster into two subgroups.

Figure 1B. Principal Components Analysis (PCA) of the 40 genes found to differentiate cardioembolic stroke from large vessel stroke. Each sphere represents a single subject. The ellipsoid surrounding the spheres represents two standard deviations from the group mean.

Figure 2.

Leave one out cross-validation 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, and patients with known large vessel stroke are shown on the right. The probability of a predicted diagnosis of cardioembolic stroke is indicated by the green diamonds, and the probability of a predicted diagnosis of large vessel stroke is indicated by the 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 greater than 0.5.

Figure 3.

Figure 3A. Hierarchical cluster analysis of the 37 genes that were significantly different in subjects with cardioembolic stroke due to atrial fibrillation compared to those with non-atrial 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. Subjects cluster by diagnosis. One group of genes has a high level of expression in cardioembolic stroke due to atrial fibrillation and a low level of expression in non-atrial fibrillation causes. Another group of genes has a low level of expression in cardioembolic stroke due to atrial fibrillation and a high level of expression in non-atrial fibrillation causes. Figure 3B. Principal Components Analysis of the 37 genes found to differentiate cardioembolic stroke due to atrial fibrillation from non-atrial fibrillation causes. Each sphere represents a single subject. The ellipsoid surrounding the spheres represents two standard deviations from the group mean.

Figure 4.

Leave one out cross-validation prediction analysis of the 37 genes found to differentiate cardioembolic stroke due to atrial fibrillation from non-atrial 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, and patients with known non-atrial fibrillation stroke are on the right. The probability of a predicted diagnosis of stroke due to atrial fibrillation is indicated by the green diamonds, and the probability of a predicted diagnosis of non-atrial fibrillation stroke is indicated by the 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 non-atrial fibrillation causes were correctly predicted in 22 out of 24 samples (91.7% correct prediction). A sample is considered misclassified if the predicted class does not match the known class with a probability greater than 0.5.

Table 1

Demographic variables for subjects with cardioembolic stroke and large vessel stroke. p-values represent comparisons of subjects with cardioembolic to large vessel stroke using Fisher's exact test or two-tailed t-test where appropriate. (BP, blood pressure; CABG, coronary artery bypass graft)

Variables	Cardioembolic (n=23)	Large Vessel (n=10)	p value
Mean Age (years)	71.7±1.6	66.9±2.9	0.14
Sex, male (%)	12 (52.2%)	8 (80%)	0.25
Race, Caucasian (%)	15 (65.2%)	8 (80%)	0.68
Hypertension (%)	16 (69.6%)	8 (80%)	0.55
Mean Systolic BP	158.3±6.1	163.5±8.0	0.63
Mean Diastolic BP	80.6±3.6	88±6.7	0.30
Diabetes (%)	4 (17.4%)	4 (40%)	0.21
Hyperlipidemia (%)	6 (26.1%)	3 (30%)	1.00
Mean Weight (kg)	81.9±4.5	89.6±6.2	0.34
Atrial Fibrillation (%)	10 (43.4%)	0 (0%)	0.03
Myocardial Infarction (%)	4 (17.3%)	2 (20%)	1.00
Congestive Heart Failure	8 (34.8%)	2 (20%)	0.68
Coronary Artery Bypass	5 (21.7%)	1 (10%)	0.64
Carotid Endarterectomy	0 (0%)	2 (20%)	0.08
Femoral Popliteal Bypass	0 (0%)	1 (10%)	0.30
Prior Stroke	7 (30.4%)	1 (10%)	0.38
Eptifibatide (%) (5&24	17 (74%)	7 (70%)	1.00
Mean NIHSS 3 hours	11.9±1.7	12.7±1.0	0.69
Mean NIHSS 24 hours	11.2±1.8	13.9±3.0	0.44
Mean NIHSS 5 days	10.3±2.1	12.1±4.7	0.69

Table 2

Functional analysis of 503 genes found to be unique to Cardioembolic strokes when compared to controls (p <0.005, FC > |1.2|)

	CE	Genes	p-value
Canonical Pathways	Renin-angiotensin signaling	ADCY4, GNAQ, PAK1, MAPK14, PIK3C3, PIK3C2B, PRKARIA, PRKCZ, RAF1	2.6×10^{-4}
	Thrombopoietin signaling	MAPK1, PIK3C3, PIK3C2B, PRKCZ, RAF1, STAT5B*, CXCR5, ITGAL, MAPK1, PIK3C3, PIK3C2B, PRKCZ, RAF1	5.8×10^{-4}
	NF-κB activation	MAPK1, MDM2, PIK3C3, PIK3C2B, PRKCZ, PTEN*, RAF1, TFDP1	6.0×10^{-4}
	Cardiac Hypertrophy role of NFAT	ADCY4, GNAQ, MAPK1, MAPK14, MEF2A, PIK3C3, PIK3C2B, PRKARIA, PRKCZ, RAF1, SLC8A1	9.9×10^{-4}
	B cell receptor Signaling	BCL6, FCGR2C, MAP3K2, MAPK1, MAPK14, PIK3C3, PIK3C2B, PTEN*, RAF1	2.4×10^{-3}
Molecular Functions	Lymphocyte development	APC, BCL6, CARD11, CD55, CFLAR, CXCR5, DTX1, GATA3, HIST1H1C, HLA-DOA, IFNGR1, IL13RA1, IL27RA, ITGAL, KLF13, MAP3K2, MAPK1, MAPK14, MBP, MDM2, PRKCZ, PTEN, RAF1, RBPJ, SEMA4A, SMARCA4, SRGN, STAT5B, STK17B, TXN, XRCC5	1.4×10^{-5}
	Cardiomyocytes cell death	CREM, GHRL, GNAQ, MAPK1, MAPK14, MDM2, NAMPT, PTEN, RAF1, SLC8A1, SOD2	1.8×10^{-4}
	T lymphocyte development	APC, BCL6, CARD11, CD55, CFLAR, DTX1, GATA3, HIST1H1C, HLA-DOA, IFNGR1, IL27RA, KLF13, MAP3K2, MAPK1, MAPK14, MBP, MDM2, PRKCZ, PTEN, SEMA4A, SMARCA4, SRGN, STAT5B, STK17B, XRCC5	2.8×10^{-4}
	Inflammatory disorder	ABCB4, AHNK, AKAP13, ANXA3, AQP9, ARF1, ASPH, B4GALT1, BCL6, CARD11, CASC4, CD55, CDH26, CFLAR, CLEC4D, CMIP, CR1, CREM, CXCR5, DHX37, DIS3L2, DYNC1LI1, ENG, ENTPD1, ETV5, EXOC6, FBXL13, FGGY, GATA3, GEMIN5, GLG1, H3F3B, HCG27, HDGFRP3, HLA-DOA, HTT, IFNGR1, IL18RAP, IL27RA, ITGAL, KALRN, KIF13A, KLF13, LHX2, LYST, MAP4, MAPK14, MBP, MDM2, MED24, MEF2A, MTHFS, NAMPT, NAT10, NBAS, NCOA2, NUMB, NUP62, OSBPL1A, PDE4B, PHF15, PHRF1, PRKCZ, PTEN, RAF1, S100A9, S100A12, SBF2, SLC22A4, SLC7A11, SLC8A1, SOD2, SOX6, STAT5B, STK17B, TPST1, TRRAP, TSPAN2, TTYH2, TXN, USP11, USP15, VIM, VSIG4, YEATS2, ZFHX3, ZNF230, ZNF831	4.1×10^{-4}
	Phosphatidylinositol 4-phosphate modification	PI4KA, PIK3C3, PIK3C2B, PTEN	4.4×10^{-4}

Table 3

Functional analysis of the 554 genes unique to large vessel atherosclerotic stroke when compared to controls ($p < 0.005$, $FC > |1.2|$).

	LV	Genes	p-value
Canonical Pathways	Cytotoxic T lymphocyte target cell apoptosis	APAF1, CD247, FADD, FASLG, TRA@, TRD@	1.3×10^{-5}
	CCR5 signaling in macrophages	CD247, FASLG, GNAI1, GNG*, MAPK13*, PRKCZ, TRA@, TRD@	1.3×10^{-4}
	Relaxin signaling	GNAI1, GNG2, GUCY2D, MAPK1, MMP9, NF- κ B IE, NPR2, PDE2A, PIK3R3, PRKACA, PRKCZ	1.5×10^{-4}
	Corticotropin releasing hormone signaling	FASLG, GNAI1, GUCY2D, MAPK1, MAPK13, NPR2, NR4A1	1.0×10^{-3}
	T lymphocyte regulation	CD247, NR4A1, PRKCZ, TRA@, TRD@, ZAP70*	1.3×10^{-3}
Molecular Functions	T lymphocyte differentiation	ADA, TRA@, ZAP70	1.4×10^{-3}
	Leukocyte development, morphology	ADA, APAF1, CCR8, CD83, CD247, CSF1, CSF2RA, CXCL12, EZH2, F5, FADD, FASLG, HIVEP2, IL12RB1, IL21R, ITGB7, MAPK1, NR4A1, PDIA3, PRKCZ, RAG1, RNASEL, SMAD7, THBS1, TRA@, XIAP, ZAP70	1.5×10^{-3}
	Invasion of cells	MMP9, PLAUR	1.7×10^{-3}
	Inflammatory disorder	ADA, ADORA3, ADRB2, ALOX5AP, APAF1, ARHGDIB, ARHGEF17, ASPH, ATP4B, C20ORF43, CA4, CA13, CAPN10, CD83, CD247, CDK6, CLCN6, COL9A3, COLQ, CORIN, CORO2A, CSF1, CSF2RA, CX3CR1, CXCL12, EGFL8, F5, FAM101B, FASLG, FBF1, GNG2, GRB10, HIC1, HIVEP1, HP, KIAA1908, LIMD1, LTB4R, MAPK13, MDC1, MMP9, MPHOSPH9, MSRA, MYH3, NFIA, NR4A1, NUMA1, OLAH, PACSIN2, PADI4, PCNX, PDE2A, PDIA3, PER1, PFTK1, PGLYRP1, PGM1, PHF19, PIK3R3, PITPNA, PLAUR, PMF1, PPARGC1B, PRKCZ, PRR5L, PTGDR, PXX, RAB7A, RAG1, RAPH1, RARG, ROPN1L, SAMS1, SERPINE2, SEZ6L, SLC25A15, SLC26A8, SLC8A1, SLC40A1, SPRED1, SPTLC2, SRPK2, STK36, TAF7L, TBC1D1, TGFB3, THBS1, TKT, TNFSF8, TNK, TRA@, TTC7A, TUBA4, TUBA4A, VARS2, ZEB1	2.8×10^{-3}
	T cell development	ADA, APAF1, CCR8, CD83, CD247, CXCL12, F5, FADD, FASLG, HIVEP2, IL12RB1, IL21R, ITGB7, MAPK1, NR4A1, PDIA3, PRKCZ, RAG1, RNASEL, SMAD7, THBS1, TRA@, XIAP, ZAP70	5.2×10^{-3}

Table 4

Functional analysis of 228 genes common to cardioembolic and large vessel atherosclerotic stroke when compared to controls ($p < 0.005$, $FC > |1.2|$).

	LV-CE Common	Genes	p-value
Canonical Pathways	p38 MAPK Signaling	DUSP1*, IL1R2, IRAK3*, MAP2K6, MAPK14, MAX, MKNK1, TNF	3.6×10^{-6}
	Toll-like receptor signaling	IRAK3*, MAP2K6, MAPK14, TLR2, TOLLIP	1.2×10^{-4}
	IL-6 signaling	ABCB1, IL1R2*, MAP2K6, MAPK14, SOS2, TNF	2.4×10^{-4}
	NF- κ B Signaling	GSK3B, IL1R2, IRAK3, MAP2K6, MAP3K3, TLR2, TNF	4.5×10^{-4}
	B Cell Receptor Signaling	GSK3B, MAP2K6, MAP3K3, MAPK14, NFATC2, PTEN, SOS2	5.6×10^{-3}
	Role of Macrophages, fibroblasts and endothelial cells in RA	CEBPD*, GSK3B, IL1R2*, IRAK3*, MAP2K6, MAPK14, NFATC2, TLR2, TNF	2.8×10^{-3}
	IL-10 Signaling	IL1R2, MAP2K6, MAPK14, TNF	3.1×10^{-3}
	NRF2-mediated Oxidative Stress	DNAJC3, FKBP5, GSK3B, MAP2K6, MAPK14, TXN	6.6×10^{-3}
Molecular Functions	Neutrophil/Phagocyte/Leukocyte movement	CAMP, CD55, CSF2RA, DUSP1, FCAR*, LILRA6, MAPK14, PTEN, SLPI, TLR2, TNF	1.7×10^{-5}
	Leukocyte development, activation	BST1, CAMP, CD55, CD59, CEBPD, CFLAR, CSF2RA, F5, GATA3, GSK3B, IL2RB, LILRA6, MAPK14, MLL, NFATC2, PRKDC, PTEN, RGL4, TLR2, TNF, TXN	5.4×10^{-5}
	Cardiovascular process	BMX, GSK3B, IL18BP, MAP3K3, MAPK14, TLR2	1.2×10^{-4}
	NF-kappa B response element expression	GSK3B, MAPK14, TLR2, TNF	2.5×10^{-4}
	Leukocyte proliferation	CAMP, CD59, CFLAR, CSF2RA, GATA3, IL2RB, IRS2, MAPK14, MLL, NFATC2, PCYT1A, PTEN, SLPI, TLR2, TNF, TXN	2.9×10^{-4}
	Oxidative stress	TNF, TXN	3.0×10^{-4}

Table 5

Demographic variables for subjects with cardioembolic stroke due to atrial fibrillation and non-atrial fibrillation causes. p-values represent comparisons of subjects with atrial fibrillation to those with non-atrial fibrillation using Fisher's exact test or two-tailed t-test where appropriate. (BP, blood pressure; CABG, coronary artery bypass graft)

Variables	Atrial Fibrillation (n=10)	Non-Atrial Fibrillation (n=8)	p value
Mean Age (years)	72.9±2.3	68.5±3.1	0.26
Sex, male (%)	4 (40%)	6 (75%)	0.19
Race, Caucasian (%)	6 (60%)	5 (62%)	0.65
Hypertension (%)	8 (80%)	6 (75%)	1.00
Mean Systolic BP	158.3±9.3	160.4±10.7	0.88
Mean Diastolic BP	80.8±4.7	86.1±8.0	0.56
Diabetes (%)	1 (10%)	2 (25%)	0.56
Hyperlipidemia (%)	3 (30%)	3 (30%)	1.00
Mean Weight (kg)	86.9±8.2	84.9±5.7	0.85
Myocardial Infarction (%)	1 (10%)	3 (37%)	0.28
Congestive Heart Failure	5 (50%)	3 (37%)	0.28
Coronary Artery Bypass	1 (10%)	1 (10%)	1.00
Carotid Endarterectomy	0 (0%)	0 (0%)	-
Femoral Popliteal Bypass	0 (0%)	0 (0%)	-
Prior Stroke	3 (30%)	3 (37%)	1.00
Eptifibatide (5&24 hours)	7 (70%)	5 (62%)	1.00
Mean NIHSS 3 hours	17.3±3.1	13.6±2.4	0.39
Mean NIHSS 24 hours	11.6±3.5	10.6±3.0	0.83
Mean NIHSS 5 days	9.9±3.5	10.0±4.0	0.98