Blood Genomics in Human Stroke

Alison E. Baird, FRACP, PhD

Abstract—Advances in microarray technology and the sequencing of the human genome are opening up new possibilities for applying genomic information in clinical medicine, using information about structural polymorphisms in DNA and changes in gene expression as measured by mRNA. Gene expression profiling studies in cancer samples have led to the identification of clinical signatures that are already being applied in some centers. In stroke, it may be possible to use peripheral blood as a source of mRNA to study gene expression. After stroke, there is a selective recruitment and migration of white blood cells to the ischemic focus in the brain. This appears to involve all white cell types and is believed to impact significantly on tissue and clinical outcome through the exacerbation of ischemic injury, particularly after reperfusion, on the one hand, and conversely contributing to tissue remodeling and repair days to weeks after stroke. The first results from clinical studies in ischemic stroke suggest that a gene expression signature can be demonstrated from peripheral white blood cells and that this represents at least a partial adaptation to the altered cerebral microenvironment. Further studies are indicated to see whether these methods may lead to new management approaches for stroke. (Stroke. 2007;38[part 2]:694-698.)

Key Words: diagnostic methods ■ genetics ■ inflammation ■ stroke

The sequencing of the human genome1,2 holds great promise for advancing our current knowledge about the causes and mechanisms involved in stroke. Over the past decade, rare new monogenic causes of stroke have been identified such as cerebral autosomal-dominant arteriopathy with subcortical infarcts and leukoencephalopathy.3 Recent work from Iceland also indicates that several genes may be associated with an increased risk of stroke.4,5 However, it is likely that in the majority of patients, stroke will prove to be a complex and polygenic disorder resulting from a mixture of genetic and environmental factors. Along with the sequencing of the human genome,1,2 advances in microarray technology6,7 permit large-scale studies and application of genomic information in clinical medicine. This technology permits thousands of genes to be studied simultaneously and can be applied to the study of DNA polymorphisms and to combinations of polymorphisms that may lead to increased stroke risk (for example, by the use of genome-wide scanning). These advances can also be used to study changes in DNA gene expression, measured by mRNA, in a methodology known as gene expression profiling or transcriptional profiling. Gene expression profiling captures dynamic changes that occur with time along with alterations in gene expression resulting from structural variations in DNA as well as from environmental factors.

As the recent literature shows, new information about stroke that can improve diagnosis and management is much needed8–11; up to 30% of initial diagnoses of stroke may be incorrect with the associated costs of added investigations and management that these misdiagnoses entail. Stroke remains a leading cause of death and disability despite advances in treatment and prevention over the past 10 years, and there is still only one treatment available (recombinant tissue plasminogen activator [rt-PA]), which reaches only a small percentage of patients. The addition of biomarker information from brain imaging has proven invaluable with the use of CT and MRI. The use of genomic information from gene expression profiling could provide additional and personalized biomarker information. The first studies of gene expression profiling in human clinical stroke have been reported in 2005 and 2006 and are showing promise.12,13 In this brief overview of the potential of gene expression profiling in stroke treatment and management, the scientific rationale for blood profiling in stroke is first outlined followed by a short description of the methodology, early results in human studies, and concluding with future clinical and scientific directions.

Rationale for Stroke Genomics in Humans Using Peripheral Blood

In gene expression profiling, changes in DNA gene expression that occur as normal biologic functions or in response to disease are determined through the measurement of mRNA. mRNA levels reflect DNA gene expression because RNA is directly transcribed from DNA. In turn, mRNA is translated into protein. The sequencing of the human genome revealed that there are around 25 000 genes, although it is estimated that only around 11 000 are active in a given tissue under physiological conditions. Different genes are active in different tissues, although there is much overlap. Under physiological and pathologic conditions, gene expression changes; for example, a gene for an inflammatory marker in the kidney...
that is normally quiescent may be upregulated in response to infection. Because mRNA expression is dynamic and varies from moment to moment, a source of live cells is needed for gene expression studies, unlike DNA polymorphism studies. Gene expression studies in cancer tissue specimens have led the way and have indicated the potential of this methodology for clinical applications such as tumor subtype reclassification and the determination of clinical outcome and metastatic risk. In The Netherlands, gene expression signatures are used in clinical decision-making for breast cancer management.

An important factor in trying to apply gene expression methodology to stroke and other neurologic disorders is that brain tissue is rarely available or accessible for study. Therefore, investigators have used peripheral blood as a most practical and widely available source of RNA. The rationale for using peripheral blood is that after stroke, there is a well-characterized inflammatory response consisting of the selective migration and infiltration of bloodborne white cells into the brain involving all white cell types. Initially, polymorphs migrate into the ischemic tissue followed by a more prolonged and extensive monocyte–macrophage infiltration. Recent evidence suggests that lymphocytes, the key regulatory immune cells within the body, may also play an important role as early as 5 hours postexperimental stroke mediating microvascular dysfunction, perhaps by activating other circulating blood cells or resident extravascular cells such as resident macrophages in the brain. This white cell recruitment is believed to impact on the ischemic area both adversely (for example, exacerbation of reperfusion injury) as well as in a beneficial way (for example, tissue repair and angiogenesis). Therefore, investigators have used peripheral blood as a source of RNA in the clinical setting after stroke, even in the early post-stroke period.

**Methodology of Microarray Studies**

A microarray is typically a piece of glass or plastic or a porous membrane onto which thousands of single-stranded segments of complementary DNA (as large as 25,000 bases, known as “cDNA microarrays”) or oligonucleotides (small segments of 25 or fewer bases, known as “oligonucleotide microarrays”) have been affixed at high density. The typical experiment consists of a comparison of two test conditions, for example, disease versus control or pre- and posttreatment, and involves the following series of steps:

1. Isolation of mRNA from the tissue of interest.
2. Conversion of mRNA to complementary DNA (cDNA).
3. Labelling of cDNA or cRNA with a fluorescent probe (cDNA is used for hybridization to cDNA microarrays, whereas cRNA is used for hybridization to oligonucleotide arrays after it has retranscribed from cDNA).
4. Hybridization of the labeled cDNA or cRNA to a microarray (“chip”)—the labeled cDNA or cRNA binds in proportion to its concentration (as measured by fluorescent intensity), (5) Scanning, and (6) then very complex statistical analyses to compare the test and control conditions involving stringent corrections for multiple statistical comparison. The results are usually displayed in lists and in cluster diagrams that show up- and downregulated genes. Verification of a subset of upregulated genes may be performed by real-time polymerase chain reaction (PCR) to confirm RNA measurements as well as in independent patient and referent populations. Apart from changes in the expression of individual genes, biological network analyses are increasingly being used to study interactions between genes within and between biologic pathways.

**Can Transcriptional Profiling Be Translated Into Human Stroke Diagnosis and Management?**

The first human stroke gene expression profiling study was reported in 2005. Moore et al studied the gene expression profile of peripheral blood mononuclear cells (PBMCs) in 20 ischemic stroke patients and 20 control subjects similar in age and sex. PBMCs consist of T cell lymphocytes (60%), B cell lymphocytes (10%), monocytes (15%), and natural killer cells (15%), and were chosen because of their potential to show specific adaptive changes to stroke and because polymorphs were difficult to isolate, being unstable and prone to rapid enzymatic breakdown. The stroke patients in this study had higher Framingham stroke risk scores than the controls (16.2 versus 9.8, respectively), and the stroke patients were studied at a mean time of 32.4 hours poststroke with seven patients being studied <24 hours. Stroke severity tended to be mild with a median National Institutes of Health Stroke Scale score of 3.7. The commercially available and robust Affymetrix HU133A arrays (containing 22,283 gene probes) were used along with state-of-the-art bioinformatics approaches and stringent corrections for multiple comparisons. In all patients, the ischemic stroke was confirmed on neuroimaging studies (MRI or CT) as well as by clinical evaluation.

In this study of Moore et al, a gene expression profile of ischemic stroke was demonstrated in PBMCs. At varying levels of stringency for correction for multiple comparisons, the number of gene probes that were different between the patient and referent groups ranged from 78 (the most stringent permutation data set) to 231 (Bonferroni data set) to 771 (false discovery rate data set). There was a predominant upregulatory response in acute ischemic stroke with ap-
proximately 90% of genes being upregulated over all of the data sets (Figure). Pathophysiological gene classes that were differentially expressed in acute ischemic stroke included genes associated with white blood cell activation and differentiation, those with a response to hypoxic stress (eg, adrenomedullin and a number of genes inducible by hypoxia-inducible factor), those associated with vascular repair, and those suggestive of a response to the altered cerebral microenvironment through inhibition of neuronal apoptosis, enhanced neurotransmitter degradation, and promotion of neurite growth (ie, neuronal apoptosis inhibitory protein, glutamine ligase, and growth arrest specific 7) together with
some unidentified genes. mRNA changes were confirmed with real-time PCR in nine upregulated genes from the Bonferroni listing. To further validate the findings from this study, a panel of 22 genes was developed using the shrunken centroid algorithm—termed the prediction algorithm for microarrays (PAM) list—that consisted of 22 genes and that could possibly form the basis of a future diagnostic test for acute stroke. In an independent cohort of 20 subjects, the PAM list proved to be 78% sensitive and 80% specific for the diagnosis of stroke.

Although the gene listings in this study suggested at least a partial adaptive response to the altered cerebral microenvironment, an ongoing question is how specific are these gene lists for ischemic stroke? To address the specificity of the results in this study, the Bonferroni listing in stroke was compared with lists from PBMCs in two other clinical disorders. Sickle cell disease—a hematologic disorder characterized by chronic inflammation and ischemic crises, and multiple sclerosis—an inflammatory brain disorder. There was minimal overlap between the gene lists. In a further analysis, the effect of vascular risk conditions was examined, because atherosclerosis is an inflammatory disorder and as the Framingham scores were higher in the stroke subjects; it was found that there was a partial dependence of the stroke listing on coexisting vascular risk conditions after adjustment for the Framingham stroke risk score. In a subsequent time course analysis by Moore et al, blood samples were taken at admission (up to 72 hours), at 1 week, and then at 3 months posts ischemic stroke in 26 subjects and a longitudinal data analysis was performed. Gene expression changes were documented over time, but the majority of the changes occurred between 1 week and 3 months with little change during the first 1 week poststroke. Real-time PCR studies are in progress to see whether these results can be confirmed. These results suggested at least a partial adaptive response to the altered cerebral microenvironment, but also that some of the gene expression changes could be attributable to coexisting vascular risk conditions.

Tang et al recently reported a second series of gene expression changes in acute ischemic stroke in 15 patients and three control subjects. Patients had three blood draws, within 3 hours of onset and at 24 hours and 5 days poststroke. Patients were treated with an investigational reperfusion therapy after the first blood draw, which could be a confounding factor in their time course study. When the gene listings of Tang et al and Moore et al were compared, there was an overlap of 156 genes between the false discovery rate 771 gene probe panel of Moore et al and that of Tang et al using the Benjamini-Hochberg correction method (equivalent to the false discovery rate) indicating a 20% overlap between listings. This is despite the fact that Tang et al used a different procedure for RNA isolation and different white blood cell populations (inclusion of polymorphs in addition to PBMCs using the PAXgene RNA stabilization platform), obtained blood samples earlier after stroke onset, and used a different array (Affymetrix 2.0 Plus arrays compared with the HU133A arrays used by Moore et al). Furthermore, when the PAM list of Moore et al was applied to Tang et al’s data, it was 85% accurate in classifying stroke (reported by Dr Sharp at the 25th Princeton Conference on Cerebrovascular Disease in Oregon). The pathophysiological gene classes were also similar between the two lists, showing genes associated with white cell activation, hypoxic stress, vascular repair, and adaptive responses to the altered cerebral microenvironment. Tang et al reported that neutrophils and monocytes may be the source of many of the gene expression changes in this study but that further studies are needed.

These two studies highlight some important points that seem to be emerging in clinical gene expression profiling studies of stroke. First, this work provides strong evidence that using the peripheral blood gene expression signatures of disease states can be reproducibly demonstrated at least in part, although the overlap between gene listings was far from perfect and also highlighted the need for strict and consistent laboratory methods. This portends well for the future of the methodology, a major concern and criticism up until now being that this methodology did not give reproducible results. Second, time may not be as critical as first believed in looking at gene expression changes in the first hours to days post-stroke. Both Moore et al and Tang et al have found that gene expression appears to remain relatively constant over the first week of stroke. It may be that in the clinical setting only the most prolonged or robust gene expression changes will be reproducibly demonstrated. Third, despite the methodological differences between the studies, the results do seem to be largely complementary. The issue of predominant neutrophil changes awaits further study. Tang et al did not address the issue of specificity, but this is also a factor that requires further clarification and validation studies.

**Future Directions**

Gene expression profiling has the potential to give useful information for diagnosis and management of acute ischemic stroke. The work needs further refining in relation to the most applicable cell type or combination of cell types to use as well as for the best and easiest methods. The specificity of gene signatures for stroke also needs to be further determined and the clinical potential could be further explored, for example, in terms of applications such as predicting outcome and for the determination of hemorrhagic risk after thrombolytic therapy with rt-PA. Correlative proteomic studies could also be valuable. The potential clinical use of this new methodology awaits further study; it may well be in the provision of additional information that may stratify strokes into different groups with different therapeutic and prognostic implications. The potential significance of blood derived gene expression changes for stroke diagnosis will require further studies of the time course of gene changes after stroke. The extent to which the stroke-derived blood-derived gene expression signatures reflect the extent of atherosclerotic vascular disease also requires further investigation.

In summary, early studies are demonstrating that a gene expression profile of human acute ischemic stroke can be demonstrated with results being validated and partly replicable. The gene profiles suggest at least a partial adaptive response to the altered cerebral microenvironment in addition to a partial dependence on coexisting vascular risk conditions. Gene panels may have a role in stroke diagnosis and management as well as in vascular risk prediction and the use of preventive therapies. Further studies are in progress. It seems likely that the inclusion
of genomic information could increasingly individualize and improve stroke diagnosis, management, treatment, and prevention. The cost and complexity of this work should not be underestimated along with the need for a highly integrated team of investigators with expertise in multiple clinical and scientific disciplines. Based on these early reports, it appears that further studies are worth pursuing.

Sources of Funding
This work was supported by the Intramural Program of the National Institutes of Health.

Disclosures
None.

References
Blood Genomics in Human Stroke
Alison E. Baird

Stroke. 2007;38:694-698
doi: 10.1161/01.STR.0000250431.99687.7b
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2007 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/38/2/694

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org/subscriptions/