Using Peripheral Blood Mononuclear Cells to Determine a Gene Expression Profile of Acute Ischemic Stroke

A Pilot Investigation

David F. Moore, MD, PhD, DIC; Hong Li, MD; Neal Jeffries, PhD; Violet Wright, RN, DNSc; Ronald A. Cooper, Jr, BA; Abdel Elkahloun, PhD; Monique P. Gelderman, PhD; Enrique Zudaire, PhD; Gregg Blevins, MD; Hua Yu, MD; Ehud Goldin, PhD; Alison E. Baird, FRACP, PhD

Background—Direct brain biopsy is rarely indicated during acute stroke. This study uses peripheral blood mononuclear cells (PBMCs) to determine whether a systemic gene expression profile could be demonstrated in patients with acute ischemic stroke.

Methods and Results—Using oligonucleotide microarrays, we compared the gene expression profile of an index cohort of 20 patients with confirmed ischemic stroke on neuroimaging studies with that of 20 referent subjects. Validation studies used quantitative real-time polymerase chain reaction to measure the levels of 9 upregulated genes in the index cohort, and an independent cohort of 9 patients and 10 referent subjects was prospectively studied to determine the accuracy of the Prediction Analysis for Microarrays list to classify stroke. After correction for multiple comparisons with the Bonferroni technique, 190 genes were significantly different between the stroke and referent groups. Broad classes of genes included white blood cell activation and differentiation (≈60%), genes associated with hypoxia and vascular repair, and genes potentially associated with an altered cerebral microenvironment. Real-time polymerase chain reaction confirmed increased mRNA expression in 9 of 9 upregulated stroke-associated genes in the index cohort. A panel of 22 genes derived from the Prediction Analysis for Microarrays algorithm in the index cohort classified stroke in the validation cohort with a sensitivity of 78% and a specificity of 80%. Control for the Framingham stroke risk score revealed only a partial dependence of the stroke gene expression profile in PBMCs on vascular risk.

Conclusions—This study demonstrated an altered gene expression profile in PBMCs during acute ischemic stroke. Some genes with altered expression were consistent with an adaptive response to central nervous system ischemia. (Circulation. 2005;111:212-221.)

Key Words: cerebral infarction, genes, ischemia, stroke

Recent scientific and technological advances, such as the sequencing of the human genome and gene microarrays consisting of probes to the functionally defined genome, make it possible to study gene expression in small tissue samples in humans. Thousands of genes can be studied simultaneously and serially. As opposed to single gene mutations and DNA polymorphisms, gene expression profiling involves the study of mRNA levels in a tissue sample to determine the expression levels of genes that are expressed or transcribed from genomic DNA. Gene expression profiling in the study of cancer tissue samples has facilitated the identification and refinement of tumor subtypes. Other studies have allowed the distinction between good-prognosis and poor-prognosis tumors and the prediction of response to treatment.

Gene expression profiling has not yet been applied to clinical stroke, partly because human brain biopsy and tissue samples are usually inaccessible and rarely justified. Animal experiments in focally ischemic brain tissue, however, have been reported, indicating substantial alteration in gene expression. In stroke and cardiovascular disease, it may be possible to use the peripheral blood mononuclear cells (PBMCs) around the time of the acute event to examine gene expression. PBMCs are of significance because blood-borne white blood cells, first polymorphonuclear cells and then subsequently mononuclear cells (ie, PBMCs), selectively

Received May 26, 2004; revision received September 11, 2004; accepted October 20, 2004.
From the Section of Neurology, University of Manitoba, Winnipeg, Manitoba, Canada (D.F.M.); Stroke Neuroscience Unit (H.L., V.W., H.Y., A.E.B.); Biostatistics Branch (N.J.), Micro-Array Core Facility (R.A.C., A.E.), Neuroimmunology Branch (G.B.), and Developmental and Metabolic Neurology Branch (E.G.), National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Md; Laboratory of Cellular Hematology, CBER, FDA, Rockville, Md (M.P.G.); and Cell and Cancer Biology Branch, National Cancer Institute, NIH, Bethesda, Md (E.Z.).

The online-only Data Supplement, which contains 8 appendices, can be found with this article at http://www.circulationaha.org.
Correspondence to Alison E. Baird, FRACP, PhD, Stroke Neuroscience Unit, NINDS/NIH, 10 Center Dr, MSC1294, Room 3N258, Bethesda, MD 20892-1294. E-mail bairda@ninds.nih.gov
© 2005 American Heart Association, Inc.

Circulation is available at http://www.circulationaha.org DOI: 10.1161/01.CIR.0000152105.79665.C6
migrate to and infiltrate the ischemic brain tissue. This early cellular inflammation and its consequences are recognized as important factors in the evolving cerebral infarct, particularly on reperfusion, and are being investigated as potential therapeutic targets. Proinflammatory cytokines and chemokines not only may activate endothelial and inflammatory cells and contribute to microvascular thrombosis but also may be directly neurotoxic. Other investigators have proposed that inflammatory cells may have a beneficial effect in recovery and repair from cerebral ischemia and/or may result in tissue remodeling. Changes in functional gene expression in PBMCs in brain disorders may occur in response to leaking brain antigens, and promising results have come from rodent studies and clinical studies of patients with multiple sclerosis (MS).

In this work, we undertook a large-scale pilot gene expression analysis to determine whether it was possible to use advances in genomics (especially mRNA levels in PBMCs) to improve the knowledge and understanding of acute ischemic stroke in humans. Specifically, we aimed to determine (1) which genes were expressed in PBMCs during acute ischemic stroke; (2) whether any genes were expressed more predominantly or significantly in ischemic stroke and, if so, attempt to obtain a workable panel of relevant genes for further study; (3) whether any of the genes in this panel had special significance to ischemic stroke, eg, nonspecific or specific for cerebral disorders such as infarction (bearing in mind that the functions of many genes are still not fully established); and (4) whether a larger study was justified.

Methods

The purpose of this study was to compare PBMC gene expression between a cohort of patients with acute ischemic stroke and a cohort of referent nonhospitalized subjects without stroke. RNA was extracted from PBMCs, followed by RNA labeling, microarray hybridization, and statistical analyses. The conduct and reporting of this study are in accordance with the MIAME criteria.

Ethical Considerations

The Institutional Review boards of the National Institute of Neurological Disorders and Stroke (NINDS) and Suburban Hospital (Bethesda, Md) approved the protocol. All patients or their designated healthcare proxy and referents gave full informed written consent before enrollment.

Ischemic Stroke Patients and Referents in the Index Cohort

Acute Ischemic Stroke Patients

Twenty acute ischemic stroke patients were recruited from the National Institutes of Health (NIH) Stroke Program at Suburban Hospital. Inclusion criteria were age >21 years and willingness to participate in the study after informed consent was given. Exclusion criteria were active medical problems, current symptomatic infection, and current severe allergic disorders.

For both the stroke and referent subjects, stroke risk factors were recorded according to the Framingham risk profile. The Framingham score is a composite score of age, history of hypertension, systolic blood pressure, smoking, cardiovascular disease, diabetes, atrial fibrillation, and left ventricular hypertrophy. Stroke and referent subjects underwent aseptic antecubital fossa venipuncture, followed by withdrawal of 30 mL blood into acid citrate dextrose A tubes (22.0 g/L trisodium citrate, 8.0 g/L citric acid, 24.5 g/L dextrose; BD). In the ischemic stroke subjects, blood samples were drawn as soon after hospitalization as possible. Blood was drawn at <24 hours (n=7), 24 to 48 hours (n=10), and >48 hours (n=3).

Total RNA Isolation

We extracted 5 to 15 μg total RNA from PBMCs within 2 hours of collection. PBMCs were separated from whole blood with a density gradient tube (Uni-Sep, Novamed) as follows: 20 to 30 mL ACD anticoagulated blood was diluted with an equal volume of PBS and added to the density gradient tube, followed by centrifugation at 1000 g for 30 minutes. At the end of centrifugation, the PBMC layer was carefully removed. The PBMC proportions were ~60% T-cell lymphocytes, 15% monocytes/macrophages, 10% B-cell lymphocytes, and 15% natural killer cells. RNA was extracted with the RNeasy Mini Kit (Qiagen) according to the manufacturer’s protocol.

Microarray Labeling

RNA labeling and cleanup steps were performed according to Affymetrix guidelines for human genome 133A arrays. To ensure the quality of the initial isolated total RNA, DNase was used to remove contaminant DNA from the sample. Additional quality control included Northern blot, followed by optical density analysis to determine the concentration of the RNA band. If the total RNA concentration was >5 μg, the RNA was used for subsequent gene chip hybridization as per the manufacturer’s protocol.

Microarray Hybridization

Coded mRNA samples were analyzed at the NINDS Microarray Facility with Affymetrix Human Genome U133A microarrays that contain 22 283 gene probes (~19 000 genes) of the best-characterized human genes (www.affymetrix.com/products/gc_U133_content.html). All samples were hybridized in an inter-leaved fashion so that systematic errors resulting from chip lot variation, laboratory reagent preparation, and machine drift between stroke patients and referents were minimized. Microarrays were scanned (Axon scanner, Axon Instruments Inc), and images were analyzed with GenePix image analysis software (Axon Instruments Inc). Gene spot fluorescent quantification followed subtraction of the surrounding background fluorescent signal within the Affymetrix MASS gene chip analysis suite, with production of CEL and DAT output files. The CDF file or annotation file for the Affymetrix Hu133A array and the CEL file containing the scanned gene expression information were the only data files used in all subsequent analyses. Data sets in which the Affymetrix-derived parameter percent present was <30% and/or the array background intensity was >100 were not used in further data analysis. The average percent present calls for the arrays in this study was 45%.

Data Normalization and Analyses

A flow diagram of the investigational design and statistical analyses is given in Figure 1. For the data analyses, CEL files of 19 patients and 19 referents were used after exclusion from analysis of 1 chip in each of the index patient and referent groups because of unsatisfactory hybridization. A glossary of terms used in the analyses is provided in Data Supplement Appendix 1, which, like all the appendices, can be found online at http://www.circulationaha.org. We followed the technique of Irizarry and colleagues for analyzing gene expression data. The analysis was completed using the Bioconductor applications of the R programming language (http://cran.
r-project.org) and implemented on a 64-bit operating system (SGI
Octane 14 000-MIPS 600-MHz CPU running IRIX 6.5.15). This was
required because of the large data set for analysis in the present
study.20 Sample RNA degradation during processing was found to be
tightly distributed and uniform across all chips (Data Supplement
Appendix 2). Quantile normalization was performed on the CEL data
sets from the index cohort and, after completion of the validation
study, on both the index and validation cohorts (Data Supplement
Appendix 3). After normalization, expression levels for each gene
were calculated with the perfect-match array probes and a robust
median polish technique after background correction and log
transformation. The resulting expression set was compared in a univar-
iate manner between the stroke patients and referent group through
the use of t-tests. The uncorrected (ie, raw) probability values were
assigned a cutoff threshold value of significance of <0.05. Subse-
quent multiple comparison corrections (MCCs) were performed by
use of false discovery techniques (eg, the method of Benjamini and
Yekutieli21) and Bonferroni corrections. Other MCC procedures,
including the Sidak SS procedure, the Sidak SD procedure, Holm’s
procedure, and Hoschberg’s procedure (Data Supplement Appendix
4), were evaluated. The uncorrected significant gene expression set
was further analyzed using the permutation analysis of Westfall and
Young.22

A further analysis used the Prediction Analysis for Microarrays
(PAM) algorithm (http://www-stat.stanford.edu/~tibs/PAM/), which
MCC procedures, including the Sidak SS procedure, the Sidak SD procedure, Holm’s
procedure, and Hoschberg’s procedure (Data Supplement Appendix
4), were evaluated. The uncorrected significant gene expression set
was further analyzed using the permutation analysis of Westfall and
Young.22

A further analysis used the Prediction Analysis for Microarrays
(PAM) algorithm (http://www-stat.stanford.edu/~tibs/PAM/), which
can be used to classify samples of unknown type. This classification
method uses the shrunken centroid method to distinguish between
the stroke and referent groups. To develop a classification model on
a data set, the algorithm23 essentially uses a threshold to select a
subset of genes that show differential expression above the threshold.

The algorithm then classifies an unknown case as the type that has
average values most similar to the unknown sample for the subset of
genes. The threshold (and hence subset of genes) is chosen by
fitting a distance matrix after background correction and log trans-
formation. The distance matrix then formed the basis for hierarchi-
cal clustering. Gene annotation and ontology were determined with the
Affymetrix online NetAffix suite, together with subsequent literature
searches and searches of Online Mendelian Inheritance in Man and
LocusLink; this allowed classification of the genes on the lists into
molecular function, cellular localization, and biological function.

To assess the dependence of the initial results on vascular risk
conditions, the genes were reanalyzed with both the Framingham
score and class designation (ie, stroke/referent) as covariates.
Adjustment of overall significance values between the stroke and
referent groups in the index cohort by overall vascular risk (as
measured by the Framingham stroke risk score) was also performed
according to the method of Goeman and colleagues25 (a single test of
the hypothesis that there is no differential expression in any gene
after controlling for Framingham score).

Reverse-Transcription and Real-Time Polymerase
Chain Reaction
To further validate the results obtained in the chip analysis, the levels
of several genes were quantified by real-time polymerase chain
reaction (PCR) using samples from 10 patients and 9 referents in the
index set. These samples were selected at random from the index set
without knowledge of individual gene expression changes. Total
RNA (2 µg) from patients and referents was retro-transcribed in a
final volume of 42 µL with the SuperScript First-Strand Synthesis
System (Invitrogen) following manufacturer’s instructions. The
quantitative real-time PCR reaction was run in an Opticon cycler
(MJ Research) with the Sybr Green PCR master mix (Applied Biosys-
tems) following manufacturer’s instructions. Thermocycling was
performed in a final volume of 25 µL containing 2 µL cDNA (1:10
dilution) and 400 nmol/L primers (see Data Supplement Appendix 5
for primer sequences). We ran 18S rRNA and HPRT for every
sample as normalizing housekeeping genes. For every sample, both
the housekeeping and target genes were amplified in triplicate in the
same run using the following cycle scheme: After initial denaturation
of the samples at 95°C for 2 minutes, 40 cycles of 95°C for 30
seconds, 60°C for 30 seconds, and 72°C for 45 seconds were
performed. Fluorescence was measured in every cycle, and a melting curve was run after the PCR by increasing the temperature from 50°C to 96°C (0.5°C increments). A defined single peak was obtained for all amplicons, confirming the specificity of the amplification. PCR results between patients and referents were compared with the stroke gene listing.

**Independent (Test) Validation Data Set**

The PAM listing was used to test the ability of the index set to classify the prospectively obtained samples from 10 stroke patients and 10 referents. Inclusion and exclusion criteria were the same for both stroke patients and referents as for the index cohort. The arrays of 9 patients and 10 control subjects were used; in 1 stroke case, the hybridization was not of sufficient quality to be included. The arrays of 9 patients and 10 control subjects were used; in 1 stroke case, the hybridization was not of sufficient quality to be included. The arrays of 9 patients and 10 control subjects were used; in 1 stroke case, the hybridization was not of sufficient quality to be included.

**Comparison With MS**

A cohort of subjects with active inflammatory MS was obtained from the Neuroimmunology Branch of the NIH under an NINDS Institutional Review Board–approved protocol to represent a positive disease control in which to evaluate the effects of nonspecific stress on gene expression in PBMCs. This group was chosen because in the index cohort, the subjects were day patients of the NIH Clinical Center and had active inflammatory status diagnosed from multiple enhancing brain lesions on neuroimaging studies. A total of 12 patients with active MS were compared with 10 independently gathered referent subjects. The PBMC samples of the MS data set were analyzed with the same Affymetrix chip type (HU133A), in the same microarray facility, and by the same statistical analyses used for the stroke samples. The gene listing from the MS cohort was then compared with the stroke gene listing.

**Results**

**Demographics**

Table 1 lists the demographic features of the patients and referents in the index cohort (n=38) and patients and referents in the validation (test) cohort (n=19). The patients in the index cohort were older, had a higher rate of hypertension, and had higher Framingham stroke risk scores than the referent group. The Framingham stroke risk score was higher in the stroke group in the test (validation) cohort than in the referent group.

**Gene Expression Profile**

We found that 5060 gene probes were significantly different in the data set without MCC. The significantly different gene probes by the different MCC methods are shown in Table 2. After Bonferroni correction, 231 gene probes of 190 genes were found to be significantly different, with a predominantly upregulatory pattern (195 gene probes of 154 genes were upregulated). The hierarchical cluster diagram of the Bonferroni listing is shown in Figure 2. A full listing of the
TABLE 2. Genes With Altered Expression in Acute Ischemic Stroke by Statistical Algorithm

<table>
<thead>
<tr>
<th>Statistics (algorithm)</th>
<th>Genes, n</th>
<th>Gene Probes, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonferroni correction set</td>
<td>190</td>
<td>231</td>
</tr>
<tr>
<td>Benjamini and Yekutieli set</td>
<td>637</td>
<td>771</td>
</tr>
<tr>
<td>Shrunken centroid algorithm</td>
<td>22</td>
<td>28</td>
</tr>
<tr>
<td>Permutation analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Westfall and Young data set</td>
<td>82</td>
<td>101</td>
</tr>
</tbody>
</table>

There are 22,283 gene probes on the microarray, coding for ~19,000 genes. We found that 5060 gene probes were different between stroke and referent groups in the raw probability value listing. The most conservative gene listing was the PAM data set, followed by the Westfall and Young, Bonferroni, and Benjamini and Yekutieli data sets.

Significantly different genes as determined by the various MCC methods can be found in Data Supplement Appendixes 6, 7, and 8 as XLS files (Microsoft Excel), including a breakdown of the genes into cellular localization (extracellular matrix, plasma membrane, cytoplasm, or nucleus), molecular function (transcription factor or membrane receptor), and biological function (cell adhesion or carbohydrate metabolism). The Bonferroni gene listing is given in Data Supplement Appendix 7. The PAM classification algorithm applied to the index set resulted in 22 genes, as shown in Table 3.

Classes of Gene Expression

With the information currently available, different broad classes of gene expression were found from the gene listings. The predominant group was genes that indicated differentiation of monocytes into macrophages and lymphocyte activation (~60%), eg, CD14, toll-like receptor 2, and FcGR2A, ie, an inflammatory response. Concomitantly, a number of genes for cell cycle arrest were also upregulated (eg, growth arrest–specific 7). Some other upregulated genes were for cytoskeletal proteins involved in anchoring cells to tissue (α-catenin and galectin 3).

There was a group related to hypoxia (dual-specificity phosphatase 1, cytochrome b-245, leukotriene A4 hydrolase, adrenomedullin). Another class of genes, eg, thrombomodulin and CD36 antigen (thrombospordin receptor), may be related to vascular repair. Upregulation of ectonucleoside triphosphate diphosphohydrolase 1 may result in decreased platelet interaction and aggregation. A further class of genes may be related to the altered cerebral microenvironment; such genes include catechol-O methyl transferase, glutamine ligase, neuronal apoptosis inhibitory protein, sortilin, phospholipid scramblase 1, and glioma pathogenesis related protein (Data Supplement Appendix 7).

Confirmation of Expression Measurements With Real-Time PCR

To confirm the measurement of mRNA concentrations, expression values derived from the microarrays were correlated with real-time PCR for 9 upregulated genes (Table 4 and Figure 3A and 3B) from the Bonferroni list of 231 gene probes. In real-time PCR, higher values for all 9 genes were found in patients compared with referents, with statistically significantly higher values for 7 of 9 genes. A negative control was also included (gene not upregulated or downregulated), and there was no difference between patients and referents.

Validation of the Classification Model

Leave-1-out cross validation of the index set accurately classified 17 of the 19 stroke patients (sensitivity, 89%) and 18 of the 19 referent cases (specificity, 95%), indicating that the microarrays strongly differentiated the 2 groups. When the model based on the index set was applied to the second cohort of 9 patients and 10 referents, the sensitivity and specificity were 78% and 80%, respectively (Table 5).

Dependence of Stroke PBMC Gene Expression Profile on Vascular Risk Conditions

After adjustment for Framingham risk score, there was still a substantial difference in gene expression between the stroke and referent groups. Adjustment for stroke risk by the global test revealed a value of P < 0.001, also indicating the persistence of strong differences in gene expression between the stroke and referent groups. In the global test analysis, stroke vascular risk measured by the Framingham score accounted for 28% of the variance of the gene expression profile in acute ischemic stroke.

Dependence of Gene Expression Profile on a Nonspecific Stress Response

Comparison of the active MS group with their independent referent group gave 218 significantly different gene probes after MCC (Bonferroni type). Comparison of this gene list with the stroke Bonferroni gene list (231 gene probes) demonstrated that 1 gene overlapped between the 2 lists: interferon gamma receptor 1 (IFNGR1).

Discussion

In this pilot study, a gene expression profile during acute ischemic stroke using PBMCs was demonstrated and validated. The significance and potential applications of these findings await further determination. The genes identified could serve as a basis for further studies of the vascular biology and neurobiology of ischemic stroke. The 22-gene panel identified by statistical analysis could also form the basis for further predictive, diagnostic, and prognostic tests for ischemic stroke.

Acute ischemic stroke was found to be associated with a predominantly upregulatory gene response in PBMCs. A predominant class of genes was involved in activation and differentiation of white blood cells (ie, inflammatory response, ~60%). Some genes in this group are involved in cell adhesion, and some code for enzymes that are involved in the cell membrane remodeling, allowing preparation for change to a more differentiated state. Some are related to cell–cell interactions; others are related to cell cycle arrest.

Other classes of genes may represent a response to hypoxia, vascular repair, and a possible central nervous system adaptation. The finding of genes related to hypoxia in the PBMCs (eg, cytochrome b-245, adrenomedullin, leukotriene...
Figure 2. Cluster diagram. Eisengram of Bonferroni-corrected hierarchical clustered analysis of index cohort of stroke patients and referents. From left to right, initial 19 columns represent referent samples; next 19 columns indicate stroke patient samples. Gene clusters are defined by dendrogram to left of heat plot. Affymetrix gene identifier for each gene found to be significantly altered is found in last column. Upregulated genes are shown in red; downregulated genes are shown in green. Increased intensity of color correlates with increased expression. Clear pattern of upregulation and downregulation is seen, with predominantly upregulatory gene expression associated with acute ischemic stroke.
A4 hydrolase) was somewhat surprising because the PBMCs were not hypoxic, although the ischemic brain tissue was. It is possible that this was a consequence of local or systemic signaling from the brain. Another class of genes may relate to repair of the vascular system, including thrombomodulin and CD36 antigen.

The gene expression profile may even provide a snapshot of the pattern of gene expression associated with stroke. However, it remains to be determined how specific these findings are for ischemic stroke and whether the findings could be due to nonstroke factors. We addressed these issues in several ways, first by looking at the effect of vascular risk conditions. From our results, it appeared that there is at least a partial contribution of vascular risk to the listing by as much as 28%. This could be an interesting topic for future study.

We further tried to address whether the results could be due to a nonspecific stress response by studying an MS cohort as a positive disease control. That the PBMC gene expression response demonstrated in our work may indeed be specific to stroke is suggested by the lack of concordance of the gene expression list with active MS (overlap of only 1 of 218 genes). By comparing only the final gene listing between the MS and stroke groups, we avoided problems with differing chip batches, noninterleaving of referent samples, and non-concurrence of the study samples.

We also have compared our results with the oxidative, hemolytic, and inflammatory stress gene expression in PBMCs associated with sickle cell disease (385 genes) that has recently been reported and found an overlap of 8 genes between that list and the stroke Bonferroni listing of 231 gene probes. Sickle cell disease is frequently associated with ischemic crises and with a chronic inflammatory state. We also found some different functional gene responses to those seen in the sepsis literature and in other studies of MS. Neutrophils, a major part of the nonspecific stress response, are not included in PBMCs.

### TABLE 3. PAM Gene List

<table>
<thead>
<tr>
<th>Rank</th>
<th>UniGene ID</th>
<th>Gene Name</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>White blood cell activation/differentiation</td>
</tr>
<tr>
<td>1</td>
<td>74076</td>
<td>CD163</td>
</tr>
<tr>
<td>2</td>
<td>178470</td>
<td>Hypothetical protein FLJ22662 Laminin A motif (probable adhesion)</td>
</tr>
<tr>
<td>5</td>
<td>279518</td>
<td>Amyloid β(A4) precursor-like protein 2</td>
</tr>
<tr>
<td>6</td>
<td>64896</td>
<td>N-acetylneuraminate pyruvate lyase</td>
</tr>
<tr>
<td>7</td>
<td>25647</td>
<td>v-fos FBJ murine osteosarcoma viral oncogene homolog</td>
</tr>
<tr>
<td>8</td>
<td>439608</td>
<td>Toll-like receptor 2</td>
</tr>
<tr>
<td>12</td>
<td>444105</td>
<td>Ectonucleoside triphosphate diphosphohydrolase 1</td>
</tr>
<tr>
<td>13</td>
<td>434488</td>
<td>Chondroitin sulfate proteoglycan 2 (versican)</td>
</tr>
<tr>
<td>15</td>
<td>285115</td>
<td>Interleukin 13 receptor, α1</td>
</tr>
<tr>
<td>18</td>
<td>75627</td>
<td>CD14 antigen</td>
</tr>
<tr>
<td>19</td>
<td>169998</td>
<td>Bone marrow stromal cell antigen 1/CD157</td>
</tr>
<tr>
<td>20</td>
<td>97199</td>
<td>Complement component 1, q subcomponent, receptor 1</td>
</tr>
<tr>
<td>21</td>
<td>122591</td>
<td>Paired immunoglobin-like type 2 receptor α</td>
</tr>
<tr>
<td>22</td>
<td>77424</td>
<td>Fc fragment of IgG, high-affinity la, receptor for (CD64)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hypoxia</td>
</tr>
<tr>
<td>3</td>
<td>441047</td>
<td>Adrenomedullin</td>
</tr>
<tr>
<td>11</td>
<td>171695</td>
<td>Dual-specificity phosphatase 1</td>
</tr>
<tr>
<td>14</td>
<td>88974</td>
<td>Cytochrome b-245, β polypeptide (chronic granulomatous disease)</td>
</tr>
<tr>
<td>16</td>
<td>81118</td>
<td>Leukotriene A4 hydrolase</td>
</tr>
<tr>
<td>17</td>
<td>292477</td>
<td>v-ets Erythroblastosis virus E26 oncogene homolog 2 (avian)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vascular repair</td>
</tr>
<tr>
<td>10</td>
<td>443120</td>
<td>CD36 antigen (thrombospondin receptor)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Possible response to altered cerebral microenvironment</td>
</tr>
<tr>
<td>9</td>
<td>508565</td>
<td>Baculoviral IAP repeat-containing protein 1 (Neuronal apoptosis inhibitory protein)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unknown</td>
</tr>
<tr>
<td>4</td>
<td>381058</td>
<td>KIAA0146 protein</td>
</tr>
</tbody>
</table>

The PAM list was generated from the shrunken centroid approach in the index cohort and used as a stroke classifier in the prospective independent cohort. The ranking was obtained from the statistical evaluation of the individual genes.
Further relating to the specificity of the results for ischemic stroke, some genes may specifically relate to the altered cerebral microenvironment, including those potentially associated with enhanced neurotransmitter degradation (catechol-o-methyl transferase and glutamine ligase). There is current interest in inhibition of apoptosis as a target for neuroprotective therapies, and 1 prominently upregulated gene was involved in the inhibition of neuronal apoptosis (neuronal apoptosis inhibitory protein).29,30 Sortilin,31 phospholipid scramblase 1,32 and Ets233 may also be involved in neuronal apoptotic cell death. Some genes may relate to neurite growth in neuronal development (growth arrest–specific 7). 34 A glioma-related protein, GLI pathogenesis-related 1, that is highly expressed in the human brain tumor glioblastoma multiforme/astrocytoma but not in normal fetal or adult brain tissue was also identified, although its function is not known.35

It is of interest that there was only a small overlap of our results with prior studies in experimental stroke models involving both brain tissue and peripheral blood,6,13 even allowing for methodological differences. Many experimental trials of stroke therapies have failed to translate to human clinical trials; the reasons for this failure need to be further explored, for example, through comparative genomics approaches.

This pilot study justifies further investigation of this method of obtaining a genomic fingerprint of ischemic stroke. Issues that could be addressed in subsequent studies would be the inclusion of larger numbers of subjects (eg, hospitalized patients with acute ischemic states) or subjects matched on stroke risk factors. Even though the index and test groups

### TABLE 4. Correlation of Gene Expression Data With Real-Time PCR Values

<table>
<thead>
<tr>
<th>Gene Name/Symbol</th>
<th>Genbank ID</th>
<th>Median Patients (n=10)</th>
<th>Median Referents (n=9)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adrenomedullin</td>
<td>NM_001124</td>
<td>1.295</td>
<td>0.39</td>
<td>0.0015</td>
</tr>
<tr>
<td>CD14</td>
<td>NM_000591</td>
<td>2.207</td>
<td>1.094</td>
<td>0.0003</td>
</tr>
<tr>
<td>CD36</td>
<td>M24795</td>
<td>2.08</td>
<td>1.23</td>
<td>0.02</td>
</tr>
<tr>
<td>Caspase 1</td>
<td>NM_033292</td>
<td>14.24</td>
<td>6.62</td>
<td>0.0041</td>
</tr>
<tr>
<td>a-Catenin</td>
<td>BC000385</td>
<td>2.559</td>
<td>1.5487</td>
<td>0.0789</td>
</tr>
<tr>
<td>FcGR2A</td>
<td>NM_021642</td>
<td>0.58</td>
<td>0.26</td>
<td>0.003</td>
</tr>
<tr>
<td>Cathepsin B</td>
<td>NM_147781</td>
<td>0.9</td>
<td>0.32</td>
<td>0.0041</td>
</tr>
<tr>
<td>Toll-like receptor 2</td>
<td>BC033756</td>
<td>0.4939</td>
<td>0.1561</td>
<td>0.0021</td>
</tr>
<tr>
<td>INFGR1</td>
<td>BC005333</td>
<td>0.985</td>
<td>0.64</td>
<td>0.1128</td>
</tr>
<tr>
<td>FcER1A</td>
<td>NM_002001</td>
<td>2.655</td>
<td>2.87</td>
<td>0.9048</td>
</tr>
</tbody>
</table>

On real-time PCR, higher values for 9 of 9 genes upregulated in the Bonferroni list (n=190 genes, 231 gene probes) were found in patients vs referents. A negative control was also included, and there was no difference between patients and referents.

### Figure 3. Real-time PCR. Box extends from 25th to 75th percentile, with horizontal line at median (50th percentile). Whiskers extend down to smallest value and up to largest value. A, Relative amounts of CD14 antigen mRNA between stroke and referent groups. This was highly statistically significant (P=0.0003). B, Relative amounts adrenomedullin (AM) mRNA between stroke and referent groups. This was again highly statistically significant (P<0.0015).

### TABLE 5. Accuracy of Prediction of Stroke in an Independent Prospective Cohort

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive Predictive Value</th>
<th>Negative Predictive Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>7/9</td>
<td>8/10</td>
<td>7/9</td>
</tr>
<tr>
<td>%</td>
<td>78</td>
<td>80</td>
<td>78</td>
</tr>
</tbody>
</table>

An independent cohort of 9 stroke patients and 10 referents was studied. With the PAM list of 22 genes, stroke was classified with a sensitivity of 78% and a specificity of 80%.
were not entirely homogeneous, the 22-gene PAM classification algorithm achieved high levels of sensitivity and specificity in the validation study; the validation study is considered consistent with “real-world data” in which patient samples are less highly controlled. Other aspects that could be addressed include more uniform and earlier times of blood draw and evaluation of the potential effects of treatments (in this study, no drug was uniformly administered to the stroke patients during hospitalization), clinical stroke severity, and ischemic lesion size. Blood has the advantage that it is easily available, can be drawn serially, and is a practical way to use this technology in patients. A specific blood marker of acute ischemic stroke would be of significant clinical value and utility but is not yet available. After stroke, leaking brain factors can be detected in the blood (S100B, neuron specific enolase, glial fibrillary acid protein, and myelin basic protein), although S100B and glial fibrillary acid protein are of low sensitivity for early stroke diagnosis and neuron specific enolase and myelin basic protein are nonspecific.\(^{36,37}\) More recently, a panel of 4 soluble factors in the plasma—2 markers of inflammation (matrix metalloproteinase-9 and vascular cell adhesion molecule), 1 marker of glial activation (S100\(\beta\)), and 1 thrombosis marker (von Willebrand factor)—demonstrated moderately high sensitivity and specificity for the diagnosis of stroke.\(^{36}\)

In this work, several gene lists obtained from a number of different MCC methods were generated because it is not yet known which is the most representative. If we are too stringent, genes that may have a real biological effect could be removed, and if we are too liberal, too many false-positives may be included. Further studies that should be undertaken include the determination of gene expression in PBMCs in other ischemic events, a comparison of gene expression changes with serum (or plasma) material (ie, proteomics correlations of the findings), and a comparison with gene expression changes in human brain tissue or other tissues (eg, heart) or in animal models (eg, rat). Gene expression profiling is a rapidly evolving field. The function of all genes has not yet been fully elucidated, and some of the genes identified in this study may in the future prove to be better fitted to other functional classes.

**Conclusions**

In this pilot study, we have defined and validated a PBMC gene expression response to acute ischemic stroke. If this particular method using PBMCs as a genomic marker is confirmed, it holds the promise to develop a blood test for ischemic stroke; furthermore, innumerable new areas of research, especially in relation to brain disorders, may open up.

**Acknowledgments**

This work was supported by the Intramural Program of the NINDS, NIH, Bethesda, Md. Dr Li performed the RNA extraction and labeling.

**References**


Using Peripheral Blood Mononuclear Cells to Determine a Gene Expression Profile of Acute Ischemic Stroke: A Pilot Investigation
David F. Moore, Hong Li, Neal Jeffries, Violet Wright, Ronald A. Cooper, Jr, Abdel Elkahloun, Monique P. Gelderman, Enrique Zudaire, Gregg Blevins, Hua Yu, Ehud Goldin and Alison E. Baird

Circulation. 2005;111:212-221; originally published online January 3, 2005;
doi: 10.1161/01.CIR.0000152105.79665.C6
Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2005 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/111/2/212

Data Supplement (unedited) at:
http://circ.ahajournals.org/content/suppl/2005/01/18/01.CIR.0000152105.79665.C6.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation 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 Circulation is online at:
http://circ.ahajournals.org/subscriptions/