Evidence for a genetic basis for stroke comes from twin and family studies and from the occurrence of a number of uncommon monogenic disorders, but the contribution of genetic factors identified for stroke so far is small. Advances in genetics and genomics may permit new insights. In recent genome-wide association studies, a number of single-nucleotide polymorphisms have been associated with specific stroke subtypes and major stroke risk factors such as diabetes and atrial fibrillation. These await replication. Studies of messenger ribonucleic acid expression have also shown promise for the development of genomic signatures for stroke classification. Stroke and coronary heart disease share some features of pathophysiology, risk, and treatment, and their genetic and genomic bases also appear to overlap. (J Am Coll Cardiol 2010;56:245–53) © 2010 by the American College of Cardiology Foundation

When the human genome was sequenced in 2003, the implications for cardiovascular and stroke medicine—in terms of the new diagnostic, therapeutic, and preventive strategies that may ultimately result—were immediately recognized (1,2). Genetic and genomic factors appear to contribute to all stages of the development of vascular disease, from predisposition, through risk and subclinical disease, to overt disease. Although the contribution of genetic factors identified for stroke is small at present, it is very possible that the sequencing of the genome may result in novel clinical applications, including the development of biomarkers for stroke.

Strokes result from focal reductions of blood flow to the brain and are second to coronary heart disease (CHD) in terms of vascular disease incidence, morbidity, and mortality. Currently, there are about 795,000 new and recurrent strokes per year in the U.S. compared with 1,350,000 new and recurrent coronary events (3). There is a substantial additional burden of asymptomatic cerebrovascular disease. About 83% of strokes are due to arterial vascular occlusion (ischemic stroke), and about 17% are due to vascular rupture (hemorrhagic stroke) (4–6). Many scientific advances have occurred in stroke diagnosis, treatment, and prevention over the past 10 to 20 years (7,8), such as advances in neurovascular imaging and intravenous tissue plasminogen activator therapy, but important questions remain unanswered. Modifiable risk factors account for only about 60% of the population-attributable risk (PAR) for stroke (9,10), as opposed to risk factors identified for CHD, which may account for more than 90% of the attributable risk (11). The mechanisms for over 30% of ischemic strokes are not known, even after extensive workup (12–14). Stroke diagnosis is inaccurate in up to 30% of patients acutely, and it is not possible to reliably distinguish between ischemic and hemorrhagic stroke clinically. No blood-based diagnostic marker has yet been developed for stroke, unlike for acute coronary syndromes; this might be because of issues related to the blood-brain barrier. There is also no reliable way to predict which patients will develop hemorrhagic transformation in the brain after thrombolytic therapy (8).

This review begins with a description of the currently recognized genetic factors involved in stroke risk and etiology: family history, single-nucleotide polymorphisms (SNPs), and monogenic disorders. This is followed by the first results that have recently been obtained from genome-wide association and gene expression and protein expression profiling studies—new genomic methodologies that have been developed over the past decade (Online Appendix). The genetic and genomic overlap between stroke and CHD and other forms of vascular disease is becoming increasingly apparent, and some examples are provided in the final section.

Genetic Factors in Stroke Risk and Etiology

Genetic factors that are associated with increased stroke risk are a family history of stroke and a number of SNPs. A number of monogenic disorders cause stroke, either as a primary manifestation or a secondary manifestation. The contribution of these factors has been increasingly recognized over the past 10 to 20 years from findings in twin
Polymorphisms in 2 genes in-crease the risk for cerebral venous thrombosis. A polymor-
phism in the gene coding for coagulation factor V results in
a form of factor V variant, factor V Leiden, that is resistant
to degradation by activated protein C, resulting in a hyper-
coagulable state (19). This polymorphism involves a single
base pair substitution. The risk for cerebral venous throm-

The importance of the accu-
rate classification of stroke type and subtype needs to be appreci-
ated, because ischemic and hem-
orrhagic strokes differ in their
causes, pathophysiology, treatments, and outcomes (12,13)
(Table 1). Careful classification of stroke type and subtype permitted the identification of genetic factors associated with CADASIL and with cerebral venous thrombosis (16,19).

Family history of stroke. In twin studies, the concordance
rates for stroke are 17.7% in monozygotic twins and 3.6% in
dizygotic twins (15,20). The heritability of stroke has varied in family studies, but allowing for some methodological
weaknesses, it appears that a family history of stroke increases a subject’s stroke risk by 2 to 3-fold (21,22). In the Framingham study, using information obtained across 3

generations, including the original and offspring cohorts, a
parental history of stroke was associated with an approxi-
mately 2-fold increase in stroke risk: a paternal history of
parental history of stroke was associated with an approxi-

In other monogenic disorders—such as Fabry disease
(X-linked recessive inheritance), sickle-cell disease (autosomal dominant inheritance), and mitochondrial myopathy, eencephalopathy, lactic acidosis, and strokelike episodes—stroke is a secondary manifestation. In Table 2, monogenic
disorders associated with stroke and vasculopathy are listed,
along with disorders associated with premature atheroscle-
rosis, arterial dissection, and monogenic forms of ICH. The
reader is referred to several excellent reviews for specific
details of these uncommon disorders (26,32,33).

Table 1 Major Stroke Types and Subtypes

<table>
<thead>
<tr>
<th>Type</th>
<th>Subtype/Presumed Major Mechanism</th>
<th>Approximate Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ischemic stroke</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Large artery atherosclerosis</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>Cardioembolism</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Small-vessel disease</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Other determined</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Undetermined</td>
<td>3</td>
</tr>
<tr>
<td>Hemorrhagic stroke</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intracerebral hemorrhage</td>
<td>Hypertension</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Amyloid angiopathy</td>
<td>7</td>
</tr>
<tr>
<td>Subarachnoid hemorrhage</td>
<td>Intracranial aneurysm</td>
<td>3</td>
</tr>
<tr>
<td>Cerebral venous thrombosis</td>
<td>&lt;1</td>
<td></td>
</tr>
</tbody>
</table>

The TOAST Trial of Org 10172 in Acute Stroke Treatment (TOAST) is a common approach to classify ischemic stroke (14). Common subtypes of hemorrhagic stroke are intracerebral hemorrhage and subarachnoid hemorrhage. A minority of strokes result from vascular occlusion in the cerebral venous system, termed cerebral venous thrombosis (4–6).
Genomewide Association Studies

As a result of the sequencing of the human genome, genomewide association studies may permit the detection of common variant alleles (>5%) of mild to moderate effect (34,35) associated with common disorders such as stroke. It is currently believed that common stroke may result from a number of genes of modest effect that in combination exert a significant effect on stroke risk (36). Commercial arrays containing a huge number of SNPs are now available, but the power of a typical study of several thousand patients is of the order of only 10% to 25% (34,35), so it is most important that adequate sample sizes be studied and that notable findings be validated and replicated in independent cohorts. Significant amounts of time, energy, and resources were expended in linkage studies and association studies of candidate genes over the past decade (37,38); many early findings turned out to be false positives, with studies often being limited by insufficient sample sizes.

The first genomewide association study in stroke was reported from the ISGS (Ischemic Stroke Genetics Study) in 2007 (39). No genetic locus was specifically and robustly associated with stroke, but there were only about 250 patients and controls. This group went on to study the 9p21.3 region (i.e., a region on the short arm [p] of chromosome 9) that had already been associated with CHD (40,41). An association was found with stroke, although it was not determined by genotyping that this association involved the same variants (40). Another group has since described modest associations between variants in the 9p21 region (rs2383207 and rs10757274) and ischemic stroke (42). However, in another study, 6 SNPs in the 9p21 region (including rs2383207) were found to be associated with the large artery atherosclerotic subtype of ischemic stroke, independent of CHD and vascular risk factors (43). The lead SNP, rs1537378, had a pooled odds ratio for stroke of 1.21 and a point estimate of the PAR for atherosclerotic ischemic stroke of 20.1% (43).

In a case-control study reported in 2008 (44), the 4q25 region was found to be associated with the cardioembolic subtype of ischemic stroke (with an odds ratio of about 1.5). This locus had already been found to be associated with atrial fibrillation. An association, although to a lesser degree, was found with all ischemic stroke (44); this latter finding was attributed to possible undetected cases of atrial fibrillation in patients with cryptogenic strokes. Another locus for cardioembolic stroke, on 16q22, has been recently reported (45). These studies demonstrate the value of having accurate stroke subtype classification, although having a standardized methodology in multicenter studies is a major challenge. For intracranial aneurysms, initial and seemingly quite robust associations have been found with loci on 2q, 8q, and the 9p21 region (the rs10757278 variant that is also associated with CHD) (46,47).

The first prospective genomewide association study in stroke was reported in 2009 from the CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology)
Consortium (48). Consisting of 4 prospective epidemiological cohorts, this large study of nearly 19,600 subjects with 1,544 incident strokes identified 2 SNPs on chromosome 12, in the region of 12p13, although replication was obtained only for the rs12425791 SNP; the hazard ratios were 1.3 for all stroke and 1.0 for ischemic stroke. This SNP was located close to the ninjurin 2 gene, which encodes an adhesion molecule expressed in glia that shows increased expression after nerve injury. Without genotyping, it is not known if this finding is associative or causative. The International Stroke Genetics Consortium, established in 2007, is currently validating the results of the CHARGE study. The International Stroke Genetics Consortium is also collaborating with the Wellcome Trust Case Control Consortium–2 in a multistage genome-wide association study of ischemic stroke. Plans are under way for the development of a similar genomewide survey in the U.S. (49).

Aside from unaccounted stroke risk, genomewide association studies may reveal genetic bases for major stroke risk factors themselves, which account for up to 60% of the PAR for stroke (9,10,50) (Fig. 2). These include hypertension (the leading modifiable stroke risk factor), which has a PAR of 30% to 40%; diabetes; smoking; atrial fibrillation; and CHD itself (with a PAR of about 4% to 5%). Loci have been associated with atrial fibrillation (4q25 and 6q22) (44,45), diabetes (loci in and around the genes cyclin-dependent kinase inhibitor 2A and cyclin-dependent kinase inhibitor 2B in the 9p21 region, among others) (51), and hypertension (serine threonine kinase 39) (52).

There has clearly been a remarkable amount of information coming to light, and a number of loci of interest are emerging, but it is too early to tell how major an impact genomewide association studies will turn out to have. In studies of other disease states, the functions of some genes have not conformed to known disease biology or not coded for proteins, as expected, but rather for ribonucleic acids (RNAs), highlighting the importance of gene expression and of micro-RNAs as future players (34–36). It is also vital that these findings be replicated and that their PARs be evaluated further; loci found for some other disease states have been of very small effect or have had little predictive value after adjustment for known risk factors (36), as may be the case for the rs10757274 polymorphism on chromosome 11.

**Figure 1** Pathophysiology of CADASIL

Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) is an autosomal dominant disorder resulting from a number of mutations in the Notch 3 gene. This gene is located on chromosome 19, on the p (short) arm, at 19p13.2 to p13.1. This gene has 33 exons; the majority of mutations occur on exons 2 to 6 and are single-point or missense mutations. There is an accumulation of osmophilic granules in smooth muscle of blood vessels throughout the vasculature, particularly in the brain, the retina, and the skin.
9p21.3 and prediction of CHD risk (53). In the future, it is likely that the focus will be on genotyping studies to look for less common variant alleles of more significant effect and in studying copy number variants as additional genetic factors associated with common disorders such as stroke.

**Gene Expression Profiling, Proteomics, and Metabolomics**

A number of important methodologies and technologies have been developed in tandem to those for investigating deoxyribonucleic acid variations, based on the fact that deoxyribonucleic acid is transcribed to messenger RNA, which in turn is translated into protein and metabolites. These methods can be used to investigate the entire transcriptome (gene expression profiling or RNA expression profiling), the proteome (proteomics), and the metabolome (metabolomics). They capture not only functional changes induced by genetic variations but also changes induced by environmental factors, risk factors, and the effects of interventions, and they can be used to monitor functional changes occurring over time. The interest of these methods for stroke is that they offer potential for identifying biomarkers (e.g., from the blood) that can be used in the clinic to diagnose and monitor disease and changing risk associated with mutations and treatment interventions. In genome-wide association studies, they are used to study if and to what degree genetic polymorphisms are functional.

**Gene expression profiling and signatures for stroke classification.** Gene expression profiling has been most widely applied in the field of oncology. Preliminary studies have been performed in carotid plaque specimens removed during endarterectomy (54).

Messenger RNA profiling of the peripheral blood has also been used to investigate a number of neurological disorders, including stroke, despite initial concerns over the specificity of signatures obtained. Blood-derived mononuclear cells (monocytes and lymphocytes) are involved in the development of atherosclerotic lesions associated with many forms of vascular disease, including stroke (55). In stroke, peripheral blood mononuclear cells are also of significance because blood-borne leukocytes, first polymorphonuclear cells and then subsequently mononuclear cells, are key players in the evolution of brain infarcts. The cells selectively migrate to and infiltrate the ischemic brain tissue. These inflammatory and immune cells may not only exacerbate ischemic reperfusion injury but may also be involved in tissue repair and remodeling after stroke (56). Changes in functional gene expression in peripheral blood mononuclear cells in brain disorders may occur in response to exposure to brain antigens (57).

Early human studies in stroke using oligonucleotide microarrays have been promising. Using peripheral blood mononuclear cells, a validated and replicated gene expression signature of ischemic stroke within the first 72 h has been identified (58). The gene expression changes seemed to reflect an adaptive response, at least in part, to the ischemic stroke, with up-regulation of genes involved in cell adhesion, inhibition of neuronal apoptosis, response to hypoxia, and vascular repair (Table 3). One example is ectonucleoside triphosphate diphosphohydrolase 1 (cluster of differentiation 39), a prominently up-regulated gene that is of interest for its cerebroprotective and cardioprotective properties. A panel of 22 genes was about 80% accurate for the detection of ischemic stroke in an external sample (58), while another group found this panel to be over 85% accurate, using whole-blood profiling (59). Signatures have since been developed for ICH (60) (prominently up-regulated genes being amphiphysin and interleukin-1 receptor 2), stroke recovery (61), and ischemic stroke subtypes (62).

Given the increasing recognition of the role of lymphocytes in stroke and vascular disease (63), and the improved charac-
characterization of lymphocyte subpopulations, ongoing studies are investigating gene expression within leukocyte populations, along with circulating stem cells that may be involved in tissue repair and recovery after stroke. Real-time polymerase chain reaction studies are proving practical in ongoing application and validation studies. This approach offers hope for the development of a rapid diagnostic test for stroke if a suitable technology platform could be identified. This work also offers hope for the identification of gene panels for vascular disease risk and for prediction of response to treatment.

Proteomics and metabolomics. Blood-based protein biomarkers (64) could be of great value for the early detection of stroke because they could permit more accurate diagnoses to be made at the initial point of care. Currently, stroke evaluation is dependent on the patient’s being taken to the hospital and undergoing a brain scan, which can delay the commencement of therapy for up to 60 to 90 min. Using a candidate protein approach, a panel of 4 proteins showed modest accuracy for the diagnosis of acute ischemic and hemorrhage stroke (65). The proteins were a marker of glial activation (S100beta), a marker of inflammation (matrix metalloproteinase-9), brain natriuretic factor, and P-dimer. Using surface-enhanced laser desorption/ionization technology, apolipoprotein C-I and apolipoprotein C-III were identified as potential plasmatic markers to distinguish between ischemic and hemorrhagic stroke (66). The results were validated using enzyme-linked immunoabsorption assays; further findings on clinical utility are awaited.

In another study that used surface-enhanced laser desorption/ionization and logical analysis-of-data classification modeling, 3 biomarkers were identified from mass spectral peaks that could detect ischemic stroke with an accuracy of 75% (67). The results were upheld in a validation sample, although the proteins were not definitively identified. The correlation and root mean absolute error of the logical analysis-of-data predictive model were superior to those of the other statistical algorithms used. Metabolomics has not yet been applied to stroke.

Pharmacogenomics. Pharmacogenomic approaches have included the evaluation and identification of genetic modifiers of clinical responses to warfarin, the statin drugs, clopidogrel, and antihypertensive therapy (68). The goals of these approaches are to personalize dosing and to minimize side effects.

Heart–Brain Linkages

Coincident with the aforementioned findings of genetic loci and gene expression changes of interest in stroke have been dramatic advances in the genetics of heart and vascular disease. A number of genetic and genomic factors already appear to overlap between stroke and heart and vascular disease, contributing to all stages of disease development: familial predisposition (21,22), vascular risk, monogenic vasculopathies, and in genomewide association studies, for example, variants in and around the 9p21 region, as previously described, and the leukotriene pathway (58,69) (Fig. 3). These

<table>
<thead>
<tr>
<th>Pathophysiological Class</th>
<th>Examples</th>
</tr>
</thead>
</table>
| White cell activation and differentiation (65%) | Toll-like receptor 2  
Ectonucleoside triphosphate diphosphohydrolase 1 (CD39)  
CD14  
Interleukin-13 receptor, alpha 1  
FcGR2A |
| Response to hypoxia (15%) | Adrenomedullin  
Leukotriene A4 hydrolase  
Cytochrome b-245 |
| Vascular repair (10%) | CD36 antigen (thrombospondin receptor)  
Thrombomodulin |
| Response to the altered cerebral microenvironment (5%) | Neuronal apoptosis inhibitory protein (apoptosis inhibition)  
Catechol-O-methyl transferase (neurotransmitter degradation)  
Glutamine ligase (neurotransmitter degradation)  
Sortilin (neuronal apoptotic cell death)  
Phospholipid scramblase 1 (neurite outgrowth in neural development)  
Growth arrest specific 7 (neurite outgrowth) |
| Unknown (5%) | KIAA0146 protein |

Classes of genes showing increased expression in peripheral blood mononuclear cells in the first 72 h of ischemic stroke (58). These changes were suggestive of at least a partial adaptive response to the stroke.

CD = cluster of differentiation; FcGR2A = Fc fragment of IgG, low affinity IIa, receptor.
are deserving of further study. Furthermore, the vasculopathy of CADASIL has not been thought to involve the heart, but in 1 series, high rates of myocardial infarction were reported in patients with CADASIL (70). Higher rates of sudden unexplained death in patients with CADASIL have also been reported and have been suggested to be due to autonomic derangement, placing patients at higher risk for life-threatening arrhythmias (71).

**Conclusions**

At present, the contribution of genetic factors in stroke is small, involving familial predisposition, a small number of monogenic disorders such as CADASIL, and polymorphisms associated with cerebral venous thrombosis. New associations are now being reported in genomewide association studies, but these need confirmation. Future work may well focus on genotyping, looking for rarer high-risk variant alleles and studying copy number variants. It is especially important that initial results from expression profiling of peripheral blood leukocytes be further explored. Genomic methods already show particular promise for improving stroke classification. They could ultimately permit the development of biomarkers that could be valuable for the diagnosis, treatment, management, and assessment of not only stroke but many other vascular disorders.

**REFERENCES**


Key Words: genomics stroke cerebrovascular disorders genetics.

APPENDIX

For a glossary of terms, please see the online version of this article.