Molecular Genetics of Stroke

Manop Pithukpakorn, M.D.
Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand.

Siriraj Med J 2008;60:283-287
E-journal: http://www.sirirajmedj.com

Stroke is the most common and debilitating neurological disease. It is a leading cause of death and disability both nationally and worldwide. The incidence of stroke in different countries varies from 47/100,000 to 300/100,000 and all available evidences indicated that the incidence of stroke has been on the rise. In Thailand, a study in 1994-1996 found that the prevalence of cerebrovascular disease in the population over 60 years of age was 11.2/1,000. Regarding the outcome, stroke contributed over 3 million deaths globally each year. This made stroke the 5th leading cause of death in people with 15-59 years of age and became the 2nd leading cause of death in people older than 60 years old. The burden of stroke was enormous both socially and economically. Stroke is now the most common cause of major disability in the developed countries. Though there is no available data in Thailand, the burden was projected to be similar. For the above reasons, the National Research Council of Thailand ranked cerebrovascular disease as the nation’s forth priority for crucial medical problem that needs medical research attention.

Stroke is defined as a neurological deficit of a cerebrovascular cause that persists beyond 24 hours or is interrupted by death within 24 hours. The disease is well known to be clinically and etiologically diverse. However, advancement in biomedical research in cerebrovascular disorders concluded that genetic factors play a role in stroke pathogenesis.

Monogenic Disorders of Stroke

Single gene disorders of stroke are rare in the general population. However, those conditions are of interest in medical research trying to fully understand stroke pathogenesis. Uniqueness in family history and clinical presentation has made these disorders very important in diagnostic, therapeutic as well as research implication.

Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL): A Model of Ischemic Stroke

CADASIL is an autosomal dominant condition causing nonatherosclerotic arteriopathy due to progressive degeneration of arterial smooth muscle cells. The incidence of this condition is estimated to be 1.98 in 100,000. Onset of the disease usually occurs in early to middle adulthood. Phenotypes of CADASIL are clinically diverse. The majority of affected individuals presented with transient or permanent ischemic events. They typically appear between 30-50 years of age. Most of the events are lacunar infarcts in the absence of hypertension or other recognized atherosclerotic risk factors. Many patients suffer multiple recurrent subcortical infarcts, leading to a stepwise decline in cognitive function and dementia. However, progressive dementia with no history of previous strokes can also be a common presenting symptom of CADASIL. Interestingly, many affected individuals have initial symptoms of migraine with aura, preceding ischemic stroke for many years. Those auras are usually of a visual or sensory subtype. Most of the patients with migraine noticed an unusually high frequency of attacks. Another manifestation that is often seen includes psychiatric disturbance such as severe depression.

Brain magnetic resonance imaging (MRI) plays an important role in the diagnosis and evaluation of CADASIL. Though non-specific, its findings include diffuse hyperintensity on T2-weighted imaging that is found consistently in the white matter and is particularly frequent in the periventricular area. Occasionally the basal ganglia and pons involvement can occur, and hypointensity on T1-weighted imaging may also be found in those areas. However, T2 hyperintensity in the absence of T1 hypointensity does occur in up to one third of affected individuals.

Pathologically, CADASIL is characterized by the progressive degeneration of vascular smooth muscle cells and the accumulation of granular osmiophilic material (GOM) within the smooth muscle cell basement membrane. GOM accumulation in vascular smooth muscle cells is one of the most distinguishing features of CADASIL.

NOTCH Signaling and Molecular Genetics of CADASIL

The discovery of a genetic defect underlying CADASIL had brought advancement of knowledge about role of NOTCH signaling in vascular disease. CADASIL is caused by mutations in the NOTCH3 gene on chromosome 19q12. The product of the NOTCH3 gene is a 2,321 amino acid transmembrane receptor and
a component of an intercellular signaling pathway essential for controlling cell fate during development.10

In mammalian cells, the NOTCH receptor family consists of four transmembrane receptors (NOTCH1 to NOTCH4) that regulate cell fate through cell-cell interaction. NOTCH receptors are translated as a large (approximately 300 kDa) protein comprising an extracellular, a transmembrane, and an intracellular domain.11 The extracellular domain of NOTCH receptors contains 29–36 multiple epidermal growth factor (EGF)-like repeats, depending on the specific NOTCH receptor, and 3 lin-12/NOTCH (LNR) motifs. The EGF-like repeats are responsible for ligand interaction, while LNR motifs are responsible for preventing receptor activation in the absence of receptor-ligand engagement. There are five ligands for the NOTCH family (so-called DSL family); Jagged (JAG1), JAG2, Delta-like1 (DLL1), DLL3, and DLL4. These ligand proteins are themselves also transmembrane proteins, with an extracellular domain comprised of 7–16 EGF-like repeats and a DSL domain, which is unique to NOTCH ligands. JAG1 and JAG2 have an additional cysteine-rich domain and a von Willebrand factor type C domain in the extracellular region.

It had previously been believed that the primary factor that regulates vascular differentiation of arteries and veins was blood flow. The endothelial cells that line arteries are exposed to higher blood pressure, higher hemodynamic flow and higher oxygen tension than do the venous endothelial cells. However, recent data has established that genetic pre-patterning, largely mediated by the NOTCH pathway, plays a primary role in regulating arteriovenous differentiation.12 This genetically determined pre-pattern has already been established prior to the initiation of blood flow, though endothelial cells at this stage are not yet committed to an arterial or venous cell fate. In addition to regulating arterial specification of endothelial cells, NOTCH signaling also regulates arterial specification of vascular smooth muscle cells as it is found that the NOTCH3 gene is expressed exclusively in the vascular smooth muscle cells of arteries.13

Mechanical forces are one of several factors implicated in regulating vascular smooth muscle cell differentiation and physiology. Adult vascular smooth muscle cells are not terminally differentiated and can exhibit substantial plasticity in their phenotype in response to local environmental changes. The exposure of vascular smooth muscle cells to mechanical strain causes a significant reduction in NOTCH1 and NOTCH3 receptor expression, concomitant with increased expression of vascular smooth muscle cell differentiation markers. Vascular smooth muscle cells that are exposed to mechanical strain also exhibit reduced proliferation and increased apoptosis and the effects could be restored by overexpression of the NOTCH1 or NOTCH3 intracellular domains in vascular smooth muscle cells.14 These results indicate that mechanical strain inhibits vascular smooth muscle cell growth while increasing apoptosis, and that these effects are mediated, at least in part, via the modulation of NOTCH signaling.

The expression of NOTCH3, which is restricted to the vascular smooth muscle cells, correlates well with the smooth muscle defects seen in CADASIL. All NOTCH3 mutations associated with CADASIL occur in an extracellular domain and result in a gain or loss of a cysteine residue in 34 EGF-like repeats of the NOTCH3 receptor. The characteristic nature of these mutations, in addition to the absence of any deletions or loss-of-function mutations in CADASIL patients, strongly suggests that CADASIL associated mutations are not NOTCH3-null alleles. Moreover, the Notch3 transgenic mice are all viable and fertile.15 No characteristic pathologic description is observed from their brain specimen. In CADASIL patients, the ectodomain of the NOTCH3 protein accumulates in the cerebral microvasculature at the cytoplasmic membrane of vascular smooth muscle cells. This finding has brought the possibility of a dominant negative or a novel effect of a mutant NOTCH3 protein in the pathogenesis of CADASIL.

**Familial Cerebral Amyloid Angiopathy: A Model of Hemorrhagic Stroke**

Cerebral amyloid angiopathy (CAA) can occur both as spontaneous intracerebral hemorrhage in the elderly and as a rare early onset familial syndrome. Sporadic CAA is a well known cause of recurrent intracerebral hemorrhage and incidence increases with age. Most reported cases of sporadic CAA have a mean age over 70 years. Intracerebral hemorrhage secondary to CAA is not associated with hypertension. The predilection of vasculopathy towards cortical arteries may explain the usual location of bleeding in the lobar area, instead of the basal ganglia, in patients with CAA.16 There is no pathognomonic clinical feature of CAA. Headache, focal neurological deficit, seizures and alteration of consciousness occur depending on the size and location of hemorrhage, although headache and seizures are more common in lobar than in deep bleeding. The less frequency of coma is probably related to the peripheral location of hematoma and cerebral atrophy in older people.

In contrast to the sporadic form, familial CAA is very rare in the general population and is usually present with early onset recurrent bleeding. The condition is inherited as an autosomal dominant manner with age of presentation ranging from 44 to 58 years.17 The majority of patients eventually suffer from progressive dementia. Some families also develop leukoencephalopathy and striking calcification in the occipital area. Pathological findings of CAA are similar to the sporadic form. At histopathologic analysis of autopsy or biopsy tissue, CAA is identified by amyloid deposition in vascular walls and destruction of these walls of capillaries, arterioles and small and medium-sized arteries of the cerebral cortex, leptomeninges, and cerebellum. No amyloid deposit is found outside the brain.

Most familial forms of CAA involve mutations within the gene for the β-amyloid precursor protein (APP).18 The regional specificity of sporadic and APP-related CAA is such that vessels in other regions, including the deep hemispherical structures (e.g. thalamus and basal ganglia) and brain stem, are generally spared. Vascular amyloid, similar to the amyloid plaques in Alzheimer disease (AD), is composed chiefly of Aβ, a 39 to 43 amino acid proteolytic fragment of APP. Involvement ranges from mild, where amyloid accumulates at the border of the media and adventitia of the vessel, to severe, in which there is total replacement of the smooth muscle in tunica media with...
amyloid accompanied by vasculopathic changes that can include the formation of microaneurysms, chronic inflammatory infiltrates, and fibrinoid necrosis. All APP mutations associated with CAA cluster within the Aβ-coding region of the gene. In addition to point mutation within APP, duplication of the APP locus on chromosome 21 has also been identified in families with familial early-onset AD and CAA. An interesting observation is highly variable expressivity of the condition.

The other familial form of CAA is associated with loss-of-function mutations in the CST3 gene. The gene encodes a 120 amino acid peptide called cystatin C, which is a potent lysosomal proteinase inhibitor. The role of cystatin C in vascular integrity and cerebral hemorrhage is initially investigated through studies about the pathogenesis of abdominal aortic aneurysm (AAA). The hallmark of AAA involves breakdown of the elastic lamina in tunica media. Elastolytic cysteine proteases, including cathepsins S (CTSS) and K (CTSK) are overexpressed at sites of arterial elastin damage. In aneurysmal aortic lesions, a severe reduction in cystatin C levels, compared to normal vascular wall smooth muscle cells, is found. Among 122 AAA patients screened by ultrasonography, the increased abdominal aortic diameter correlated inversely with serum cystatin C levels. In vitro, cytokine-stimulated vascular smooth muscle cells secreted cathepsins (CTSS and CTSK) whose elastolytic activity could be blocked when cystatin C secretion was induced by treatment with TGFβ-1.

These findings highlighted a potentially important role for imbalance between cysteine proteases and cystatin C in arterial wall remodeling and established that cystatin C deficiency occurs in vascular disease. A mechanistic link between APP and cystatin C in CAA pathogenesis is shown by overexpression of human cystatin C in brains of APP transgenic mice that reduces cerebral Aβ deposition and that cystatin C binds β-amyloid and inhibits fibril formation. It is hypothesized that endogenous cystatin C is a carrier of soluble β-amyloid in cerebral spinal fluid, blood, and brain, where it inhibits β-amyloid aggregation into insoluble plaques.

**Stroke as a Complex Disease**

Unlike monogenic stroke disorders, stroke in the general population is mostly sporadic with enormous phenotypic variability. Diversity in the pathophysiology and clinical presentation has limited study to determine genetic factors of stroke. Heterogeneity in stroke presents at different levels including mechanism and location of stroke. The most obvious distinction in mechanism differentiates ischemic and hemorrhagic stroke. Even in the ischemic type, which constitutes the majority of strokes, the etiology could also be different; atherosclerosis, cardioembolism or small vessel occlusion. Because there is wide range of heterogeneity, it is expected that there would be a lot of genetic factors and each factor may contribute a low impact on stroke risk. Therefore, the goals of genetics studies in stroke are limited mainly in identifying genes that modify the predisposition of stroke and genes that modify the stroke outcome.

**TOAST Classification and Subtyping Ischemic Stroke**

To minimize heterogeneity, various categorizations have been implemented and the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) classification system is the most widely used. The system indicates that stroke can be caused by embolism and intrinsic disease of small and large cerebral vessels. These subtypes likely occur through distinct mechanisms.

The system classifies all ischemic strokes into five different categories: large-artery atherosclerosis, cardioembolism, small-vessel occlusion, stroke of other etiology, and stroke of undetermined etiology. Each subtype is further divided into possible or probable subgroups on the basis of the strength of the evidence supporting the diagnosis. The diagnosis of large-artery atherosclerosis stroke subtype requires physicians to have evidence by duplex ultrasound imaging or arteriography of occlusion or stenosis greater than 50% in an appropriate major artery or branch cortical artery. To diagnose cardioembolic stroke, physicians must identify a potential source of cardioembolism, such as atrial fibrillation. The small-vessel stroke subtype is diagnosed on the basis of clinical deficits with or without brain imaging showing an infarct less than 1.5 cm in maximum diameter. Stroke of other etiology is diagnosed when stroke is caused by an uncommon mechanism, such as cerebral vasculitis or anti-phospholipid syndrome. Stroke of undetermined etiology is subdivided based on the nature of the diagnostic uncertainty: an evaluation that identifies more than one potential cause, a complete but negative evaluation, or an incomplete evaluation.

While the TOAST classification has been generally accepted in stroke studies including stroke genetics, it is also well known that there is still significant overlap between the basic mechanisms of these stroke subtypes. Moreover, it is frequently impossible to use this classification to definitively differentiate the true etiology of stroke.

As mentioned previously, the influence of genes on stroke has never been high. Analysis of genetic studies prior to 2003 concluded that the best quality case-controlled and cohort studies yielded an odds ratio (OR) of 1.3 overall when patients from independent studies were systematically analyzed. Though several recent studies have identified stroke genes in specific populations, these genes have relatively modest influence on the susceptibility of humans to stroke, influencing stroke odds by two fold or less, while the carrier frequency for these genetic polymorphisms in control populations is also significant (at least 10% for most haplotypes). Given the low impact of specific genomic loci on overall stroke risk, the combination of the large number of genes are more likely to play a role in an extremely heterogeneous disease.

**Linkage Analysis of Stroke Genes: PDE4D and ALOX5AP**

The deCODE genetics program has investigated the Icelandic population, which is relatively small and whose genealogy has been well traced for over 10 centuries. A total of 2,854 patients with stroke (both ischemic and hemorrhagic) and their families were studied using microsatellite markers to uncover genetic loci that were linked to stroke. A locus on chromosome 5q was reported with a LOD score of 4.4. The LOD score increased to 4.9 when 6% of patients with hemorrhagic stroke were removed, suggesting that the
gene contributing this linkage peak was likely a risk factor for ischemic stroke. Further analysis of this location using a case-controlled analysis and additional markers demonstrated that the strongest association to stroke was for markers in the alternative-promoter region of one of the eight isoforms of the PDE4D gene. Interestingly, the subtypes with the highest risk were large-vessel occlusive stroke and cardioembolic stroke (with RR of 1.98), not small vessel occlusive disease. A mutually exclusive protective haplotype was also described that conferred an RR of 0.68. The function of PDE4D, a phosphodiesterase enzyme, is to degrade a second messenger cAMP, which is a signaling molecule that is important for the pathogenesis of atherosclerosis. Studies have shown that decreased cAMP levels cause an increased proliferation and migration of vascular smooth muscle cells, which is a hallmark of atherosclerosis. Further support comes from the inhibition of smooth muscle proliferation by PDE4 antagonists in animal carotid injury model. PDE4D is also expressed in activated macrophages and, therefore, the role of PDE4D might be in inflammation within atherosclerosis plaques and it could be involved in either atherogenesis or plaque instability, or both.

In the second study, the deCODE group mapped susceptibility loci for myocardial infarction (MI) and stroke in Icelandic families using microsatellite markers on chromosome 13. A candidate gene ALOX5AP was identified following analysis using SNPs in the region. Four SNPs composed a haplotype within the ALOX5AP gene that was associated with an RR of 1.7 for conferring stroke susceptibility. ALOX5AP encodes the arachidonate 5-lipoxygenase-activating protein. This protein is involved in the initial steps of leukotriene synthesis. In the leukotriene biosynthetic pathway, unesterified arachidonic acid is converted to leukotriene A4 (LTA4) by the action of 5-lipoxygenase and its activating protein. Other inflammatory and vasoactive mediators, including leukotrienes B4, C4, D4 and E4, are then produced from LTA4 by the action of leukotriene A4 hydrolase and leukotriene C4 synthase. The synthesis of LTB4 in stimulated neutrophils from patients with a history of MI is greater than from controls, supporting the hypothesis that increased activity of the leukotriene pathway may have a role in atherosclerosis. Elevated LTB4 level might contribute to atherogenesis or plaque instability by increasing inflammation of the atherosclerotic plaques. The role of an upregulation of the leukotriene pathway in atherosclerosis is further supported by studies on human atheromas that have shown an abundant expression of members of the 5-lipoxygenase pathway in the lesions, and the number of 5-lipoxygenase-positive cells (macrophages, dendritic cells, mast cells and neutrophils) is markedly increased in advanced lesions.

**Association Studies of Stroke Genes: Model for Complex Disease**

One major approach of multifactorial stroke investigation is the candidate-gene approach, which consists of identifying molecular variants within functionally relevant genes and establishing their function in stroke risk by association case-controlled or cohort studies. Similar vascular pathophysiology between ischemic stroke and coronary artery disease brought interest in studying candidate genes that underlie several components of clot formation such as platelet adhesion (platelet receptor glycoprotein genes), coagulation pathway (fibrinogen genes, coagulation factor genes), fibrinolytic system (plasminogen activator genes) and homocysteine (enzymes involved in methionine-homocysteine metabolism). While data from several studies demonstrate significant association between polymorphisms in those candidate genes and ischemic stroke, many of them showed uncertain or did not demonstrate significant association. These results support the complexity and difficulties in assigning a causality role to polymorphisms for the multifactorial nature of stroke pathogenesis, in which many genetic variants could contribute, together with environmental factors and the interactions of other genes.

The technologies for high-throughput genotyping are rapidly developing as well as the statistical methods to analyze increasingly complex data. The technical developments are likely to outpace the collection of large carefully phenotyped samples. The future of stroke genetics will therefore depend on the sample availability and on close collaborations between clinicians and geneticists. Association-based methods are a powerful instrument to identify small relative risks. In this respect they are particularly suited to address common multifactorial stroke. The candidate gene approach remains a valid strategy and there are several means by which the power of such studies can be improved. Critical issues include the selection of candidate genes and appropriate phenotypes, sample-size issues, and replication in independent studies. There is also a growing interest in systematic hypothesis-free approaches. Genome-wide studies using 100,000 to 500,000 single-nucleotide polymorphisms have become technically possible but are major challenges in terms of resources and study design. Such studies may require the collaborative effort of multiple centers and the construction of large databases containing clinical and epidemiological data.

**ACKNOWLEDGMENTS**

The author is a recipient of the Research Potential Development Grant for Young Academics of the Commission of Higher Education (CHE) and Thailand Research Fund (TRF). A part of the content in this article was presented at the TRF-Senior Research

<table>
<thead>
<tr>
<th>Genes</th>
<th>Gene product</th>
<th>Polymorphism</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE</td>
<td>Angiotensin converting enzyme</td>
<td>I/D</td>
</tr>
<tr>
<td>AGT</td>
<td>Angiotensin 1</td>
<td>M235T</td>
</tr>
<tr>
<td>MTHFR</td>
<td>Methylenetetrahydrofolate reductase</td>
<td>677C&gt;T</td>
</tr>
<tr>
<td>APOE</td>
<td>Apolipoprotein E</td>
<td>E4</td>
</tr>
<tr>
<td>F5</td>
<td>Factor V</td>
<td>1691G&gt;A</td>
</tr>
<tr>
<td>F2</td>
<td>Prothrombin</td>
<td>20120G&gt;A</td>
</tr>
<tr>
<td>F13A1</td>
<td>Factor XIII</td>
<td>143G&gt;T</td>
</tr>
<tr>
<td>FGB</td>
<td>Beta-fibrinogen</td>
<td>455G&gt;A</td>
</tr>
<tr>
<td>SERPINE1</td>
<td>Plasminogen activator inhibitor-1</td>
<td>675_676delinsG</td>
</tr>
<tr>
<td>ITGA2</td>
<td>Platelet glycoprotein Iα</td>
<td>807C&gt;T</td>
</tr>
<tr>
<td>GP1BA</td>
<td>Platelet glycoprotein Ib</td>
<td>HPA2,-5T&gt;C</td>
</tr>
<tr>
<td>ITGB3</td>
<td>Platelet glycoprotein IIIa</td>
<td>PLA1/2</td>
</tr>
<tr>
<td>PON1</td>
<td>Paraoxonase 1</td>
<td>107C&gt;T, 824G&gt;T</td>
</tr>
<tr>
<td>NOS3</td>
<td>Endothelial nitric oxide synthase</td>
<td>894G&gt;T</td>
</tr>
</tbody>
</table>

**TABLE 1. Genes and polymorphisms that demonstrated association with ischemic stroke.**
REFERENCES


