Stroke genetics: a review and update.

Lindgren, Arne

Published in:
Journal of Stroke

DOI:
10.5853/jos.2014.16.3.114

2014

Citation for published version (APA):
Stroke Genetics: A Review and Update

Arne Lindgren

Department of Clinical Sciences Lund, Neurology, Lund University, Lund, Sweden
Department of Neurology and Rehabilitation Medicine, Skane University Hospital, Lund, Sweden

Stroke genetics includes several topics of clinical interest, including (1) molecular genetic variations affecting risk of monogenic stroke syndromes; (2) molecular genetic variations affecting risk of common stroke syndromes, sometimes with specific effects on risk of specific main types of stroke or subtypes of ischemic and hemorrhagic stroke; (3) genetics of conditions associated with stroke risk e.g. white matter hyperintensities, atrial fibrillation and hypertension; (4) hereditary causes of familial aggregation of stroke; (5) epigenetic impact on protein expression during acute brain injury; (6) genetic influence on stroke recovery; and (7) pharmacogenetics. Genetic research methods include candidate gene studies; Genome Wide Association Studies; family studies; RNA and protein analyses; and advanced computer-aided analytical methods to detect statistically significant associations. Several methods that could improve our knowledge of stroke genetics are being developed e.g.: Exome content analysis; Next-generation sequencing; Whole genome sequencing; and Epigenetics. During 2012-2014, several Single Nucleotide Polymorphisms (SNPs) have been related to common ischemic stroke risk. Certain SNPs have been associated with risk of specific ischemic stroke subtypes such as large vessel disease and cardiac embolism, particular subtypes of intracerebral hemorrhage (ICH), especially lobar ICH, and with prognosis after ICH. Large international studies on stroke recovery and exome content are ongoing. Advanced mathematical models have been used to study how several SNPs can act together and increase stroke risk burden. Such efforts require large numbers of patients and controls, which is achieved by co-operation in large international consortia such as the International Stroke Genetics Consortium. This overview includes an introduction to genetics, stroke genetics in general, and different genetic variations that may influence stroke risk. It presents some of the latest reports on stroke genetics published in high impact journals. The role of pharmacogenetics, the current clinical situation, and future prospects will also be discussed.

Keywords Stroke; Genetics; Pharmacogenetics; Genetic Association Studies; Humans

Introduction

The human genome

The human genome consists of 23 chromosome pairs i.e. in total 46 chromosomes. The DNA of these chromosomes contain about 6 billion base pairs. The genome can be divided into the exon part, called the exome. This is the 1.5% of DNA that contains approximately 25,000 genes coding for about 100,000 proteins in humans. The intron part is not coding for proteins but has regulatory properties. In addition, there is also mitochondrial DNA containing approximately 16,000 base pairs with 13 protein coding genes. Each human cell contains hundreds or thousands of sets of this mitochondrial DNA.
Genetic variation

Genetic variation in humans may have several causes. The DNA in the genome itself can be altered. Epigenetic factors can influence the gene expression through variations in DNA methylation and variations of histones - proteins controlling the DNA string formation. DNA changes range from large alterations of a part of or of a whole chromosome - cytogenetic changes that are possible to detect on microscopic chromosomal examination and sometimes not compatible with life. Alterations of intermediate size are on a submicroscopic level. The smallest molecular changes are confined to variation of one single base pair - a Single Nucleotide Polymorphism (SNP) where a nucleotide has been exchanged for another (Figure 1). Even though there are about 6 billion base pairs, only about 78 million - i.e. a small proportion of these base pairs have variations described as SNPs. Today's clinical molecular research is often heavily focused on SNP analyses and there is a risk that other alterations therefore remain unnoticed. It should therefore be kept in mind that other molecular DNA changes such as copy number variants (CNVs), repeats, insertions, deletions and microsatellites may also account for clinical genetic variation. Another question is how several variations may interact with each other and there is a need for advanced mathematical analytical methods utilizing powerful computer techniques to investigate such associations.

It should also be considered how common a genetic variation is and how common the disease studied is. For a common stroke phenotype the genetic variant can be common, rare, or even private i.e. confined to one individual or family. The impact of the genetic variant may differ considerably: with a high penetrance and importance it may cause a monogenic syndrome, with less influence it may still contribute to commonly occurring disorders with a more complex heritability.

Molecular genetic studies of stroke risk

One to two decades ago, molecular studies were often conducted as linkage studies where markers such as microsatellites were used to identify areas related to risk. As an example, one study using microsatellite markers and SNP analyses reported a possible association between stroke and variations in the PD-E4D gene.

An era of candidate gene studies followed the linkage studies. The idea is that by a logical and educated guess a candidate gene can be suggested as possibly related to variation in stroke risk. Numerous studies on this have been published, but many studies have been unsuccessful or not been possible to repeat. Genes of interest have included e.g. PDE4D and genes associated with cardiovascular disease. Many of the candidate gene studies have been on specific SNPs in the area of interest in the genome.

During the last decade many genome wide association studies (GWAS) have been performed. In a GWAS, an agnostic approach is used. A large number of SNPs, often in the range of 500,000 to 5,000,000 are examined throughout the whole chromosomal genome. Because so many SNPs are examined at the same time an adjustment for multiple testing has to be done. Therefore a p value threshold of $5 \times 10^{-8}$ - corresponding to a Bonferroni correction for 1,000,000 tests - is often set as a sig-

---

[Figure 1. Examples of DNA variations in the human genome.]
nificance level for GWAS investigations. It could be questioned whether this level should always be the same or if it should be adjusted because of e.g. the actual number of SNPs being analyzed. GWAS examinations have now yielded several very interesting results for ischemic stroke risk (Table 1). More studies are ongoing. Several overviews on stroke genetics have recently been published.10,11

Molecular genetic variations affecting risk of monogenic stroke syndromes

Today, there are several monogenic stroke syndromes that have been related to molecular genetic variation. Examples of monogenic stroke syndromes are given in Table 2. Comprehensive accounts of have been published.11,12,23 Information on the Internet is available at e.g. Genetics Home Reference at http://ghr.nlm.nih.gov/.

Molecular genetic variations affecting risk of common stroke syndromes, sometimes with specific effects on specific main types of stroke or subtypes of ischemic and hemorrhagic stroke

Ischemic stroke

The three main ischemic stroke syndromes: large vessel disease, cardioembolic stroke and small vessel disease have been studied separately regarding genetic risk. Several molecular genetic variations have been reported to be related to large artery disease and to cardioembolic stroke (Table 1). Some findings have been reported in several studies e.g. HDAC9 related to large vessel disease,13,15 and PITX2 and ZFHX3 related to cardioembolic stroke.13,14 The effect sizes of the identified SNPs have been modest; e.g. for a HDAC9 variant and stroke related to large vessel disease: odds ratio (OR) = 1.42 (95% confidence interval [CI] = 1.28-1.57); and a PITX2 variant and cardioembolic stroke OR1.32 (95% CI = 1.20-1.46).15 One observation is that the GWAS detected variations related to common ischemic stroke caused by small vessel disease have been fewer or absent in GWAS studies of ischemic stroke. The explanation for this is unknown but it has been suggested that small vessel disease may represent several different phenotypes and thus be a more heterogeneous condition.16 and also be subject to different definitions,15 whereas large artery disease and cardioembolic stroke may be less heterogeneous. The relation of genetic variants to overall ischemic stroke has also been reported but with less consistency. There is a trade-off between very specific phenotyping e.g. defined types of ischemic stroke and the number of subjects that can be included in the genetic studies. It seems clear today that ischemic stroke is not caused by one common pathogenetic factor, however there seems to be considerable overlap between the different subtypes of ischemic stroke regarding risk factors such as e.g. hypertension. It has been shown that a risk score taking several genetic variations related to ischemic stroke risk into account is associated with ischemic stroke overall.26

Intracerebral hemorrhage (ICH)

Also hemorrhagic stroke risk has been related to genetic variations (Table 3). The risk of lobar intracerebral hemorrhage has been firmly related to variations in the APOE gene, especially the ε2 or ε4 alleles. There are indications that although variations in the COL4A1 region have been related to monogenic related ICH, other variations in the same region may be related to a somewhat increased risk of sporadic ICH.27 A detailed description of genetic risk of sporadic ICH has recently been published.11

Genetics of conditions associated with stroke risk - intermediate phenotypes - e.g. white matter hyperintensities, atrial fibrillation, and hypertension

Several intermediate phenotypes are related to stroke risk and it is therefore of interest if genetic risk for these phenotypes is also associated with increased stroke risk either independently or through the intermediate phenotype. As mentioned above, a genetic risk score considering genetic variations linked to intermediate phenotypes related to stroke has been associated with overall risk of ischemic stroke.26 The same reference contains a comprehensive supplemental table listing genes related to intermediate phenotypes indicating stroke risk. Another study reported that a risk score including genetic variations related to stroke and its risk factors could improve the prediction of future stroke compared
Table 2. Examples of monogenetic stroke syndromes

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Abbreviation</th>
<th>Chromosome</th>
<th>Gene region</th>
<th>Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerebral autosomal dominant subcortical infarcts and leukoencephalopathy</td>
<td>CADASIL</td>
<td>19p13.2-p13.1</td>
<td>NOTCH3</td>
<td>Migraine, cognitive problems, depression, seizures, stroke</td>
</tr>
<tr>
<td>Cerebral autosomal recessive arteriopathy with subcortical infarcts and</td>
<td>CARASIL</td>
<td>10q26.3</td>
<td>HTRA1</td>
<td>Spasticity, stroke, cognitive problems, scalp hair loss, back pain</td>
</tr>
<tr>
<td>leukoencephalopathy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fabry disease</td>
<td>X</td>
<td>GLA</td>
<td></td>
<td>Episodes of pain in hands and feet, angiokeratomas, corneal opacity, renal affection, heart affection, stroke</td>
</tr>
<tr>
<td>Sickle cell disease</td>
<td>11p15.5</td>
<td>HBB</td>
<td></td>
<td>Anemia, pain episodes, infections, affection of lungs including pulmonary hypertension, kidneys, spleen, and brain including stroke</td>
</tr>
<tr>
<td>Hereditary endotheliopathy with</td>
<td>HERS</td>
<td>3p21.31</td>
<td>TREX1</td>
<td>Visual loss, cognitive problems, stroke-like episodes, renal dysfunction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marfan syndrome</td>
<td>15q21.1</td>
<td>FBN1</td>
<td></td>
<td>Lens dislocation, cataract, myopia, aortic aneurysm, aortic dissection, cerebral aneurysms, cerebral hemorrhage, arthritis, tall habitus, pectus excavatum, dural ectasia</td>
</tr>
<tr>
<td>Ehlers Danlos syndrome type IV</td>
<td>2q31</td>
<td>COL3A1</td>
<td></td>
<td>Joint hypermobility, cerebral aneurysm, arterial dissection, short stature, thin skin that easily bruises, intestinal and uterine fragility, joint subluxation and pain</td>
</tr>
<tr>
<td>Pseudoxanthoma elasticum</td>
<td>16p13.1</td>
<td>ABCC6</td>
<td></td>
<td>Papules in flexor areas of skin, visual loss, hypertension, arterial dissection</td>
</tr>
<tr>
<td>Homocystinuria</td>
<td>21q22.3, 1q26.3</td>
<td>CBS, MTHFR, and other</td>
<td></td>
<td>Varies. E.g. cognitive problems, myopia, lens dislocation, osteoporosis, thromboembolic events</td>
</tr>
<tr>
<td>Neurofibromatosis type 1 (von Recklinghausen’s disease)</td>
<td>17q11.2</td>
<td>NF1</td>
<td></td>
<td>Café-au-lait skin spots, neurofibromas, optic glioma, cerebral ischemia, intracranial aneurysm</td>
</tr>
<tr>
<td>von Hippel-Lindau syndrome</td>
<td>3p25.3</td>
<td>VHL</td>
<td></td>
<td>Hemangioblastoma in brain, spinal cord, retina. Intracerebral hemorrhage, pheochromocytoma, hearing loss</td>
</tr>
<tr>
<td>COL4A1-related brain small vessel disease</td>
<td>13q34</td>
<td>COL4A1</td>
<td></td>
<td>Hemorrhagic stroke, white matter changes, seizures, migraine</td>
</tr>
<tr>
<td>Hereditary hemorrhagic telangiectasia (Oster-Weber-Hendu disease)</td>
<td>Several</td>
<td></td>
<td></td>
<td>Telangiectasia, arteriovenous malformations in lungs, brain, liver intestines. Intracerebral hemorrhage. Ischemic stroke</td>
</tr>
<tr>
<td>Hereditary cerebral amyloid angiopathy</td>
<td>21q21.3 and other</td>
<td>APP and other</td>
<td></td>
<td>Lobar intracerebral hemorrhage, cerebral microbleeds, cognitive problems</td>
</tr>
<tr>
<td>Familial cerebral arterial aneurysms</td>
<td>Several</td>
<td></td>
<td></td>
<td>May be associated with other syndromes eg Marfan syndrome, Poly cystic kidney disease</td>
</tr>
<tr>
<td>Cerebral cavernous malformations</td>
<td>7q21.2, 7p13 and other</td>
<td>KRIT1, CCM2 and other</td>
<td></td>
<td>Cerebral hemorrhage, seizures, brainstem symptoms, cranial nerve symptoms</td>
</tr>
<tr>
<td>Mithochondrial encephalopathy lactic acidosis and stroke like episodes</td>
<td>MELAS</td>
<td>Mitochondrial DNA</td>
<td>Several</td>
<td>Muscle weakness, headache episodes, seizures, stroke like episodes</td>
</tr>
</tbody>
</table>

Table 3. Examples of SNPs related to ICH risk

<table>
<thead>
<tr>
<th>SNP in chromosome</th>
<th>Gene region</th>
<th>Relation to</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>19q13</td>
<td>APOE</td>
<td>Lobar ICH</td>
<td>28</td>
</tr>
<tr>
<td>1q22</td>
<td>PMF1/SLC25A44</td>
<td>Non-lobe ICH</td>
<td>29</td>
</tr>
<tr>
<td>13q34</td>
<td>COL4A1</td>
<td>ICH</td>
<td>27</td>
</tr>
<tr>
<td>6q21</td>
<td>KCNK17</td>
<td>ICH</td>
<td>30</td>
</tr>
</tbody>
</table>

SNP, single nucleotide polymorphism; ICH, intracerebral hemorrhage.

with using a risk score based only on clinical information.31 Some intermediate phenotypes are discussed in more detail below.

White matter hyperintensities (WMH)

The presence of white matter lesions has been related to stroke risk.32,33 Therefore it is possible that genetic changes resulting in WMH may also result in increased risk of stroke. The heritability of cerebral white matter hyperintensities is high.34 An association between the 17q25 locus and white matter hyperintensity volume has been reported.35,36 This association has also been reported in subjects with stroke although there was not a clear relation to the presence of small vessel disease manifested as lacunar infarction.37 It is likely that several different genetic variations contribute to the risk of WMH and a recent study of patients with CADASIL reported a polygenic risk score for WMH volume, illustrating that examination of a subgroup of individuals with high likelihood of a condition can be used to detect additional traits contributing to risk of the condition.38 ApoE ε4 carriers have been reported to have higher subcortical white matter lesion volume.39 However, the effect of the ApoE ε4 allele on white matter integrity is uncertain.40 Variations in the NOTCH3 gene may also influence the risk of WMH or small vessel disease in individuals without the typical CADASIL syndrome.41
Atrial fibrillation (AF)

Several genes have been related to AF.\(^{42-44}\) The mechanisms through which these genes contribute to AF risk are largely unknown although a recent study showed a relation between some of these genes and prolonged atrial action potential duration in an animal model.\(^{45}\) It seems as if adding a genetic risk score of gene variations related to AF in patients with AF improves the risk assessment for stroke in addition to the often used CHADS\(_2\) score in these patients.\(^{46}\)

Hypertension

Hypertension is one of the most important risk factors for stroke, both ischemic and hemorrhagic. Blood pressure has heritability estimates of 30%-50%.\(^{46}\) Several genes have been reported to be related to hypertension in GWAS studies of patients without or with stroke. A risk score of 29 such SNPs was related to stroke, not further subtyped, as well as to hypertension and to coronary heart disease.\(^{47}\) It is of interest that an age related effect of genes related to a certain phenotype e.g. blood pressure has been reported\(^{48}\) and it is possible that such age related effects are of importance also for other phenotypes including stroke.

In another study, a risk score consisting of 39 SNP associated with blood pressure levels was related to ICH, especially deep ICH rather than lobar ICH, supporting the concept that elevated blood pressure may be more prone to cause deep ICH.\(^{49}\) It was discussed if gene variations influencing blood pressure may be of importance for stroke risk also in individuals with seemingly “normal” blood pressure.\(^{50}\)

Ischemic heart disease

The situation for ischemic heart disease (IHD) is similar to that for hypertension. Many traditional risk factors are shared between stroke and IHD. Several genetic variations have been associated with ischemic heart disease.\(^{50}\) A variation in the chromosome 9p21 region has been related to ischemic heart disease.\(^{51}\) A subsequent follow-up study was not able to find additional variants from the Cardiogram study to be associated with stroke overall or main ischemic stroke subtypes although the number of patients in the individual subtypes were small.\(^{52}\) A much larger GWAS study could detect that several genetic variations were shared between ischemic stroke - especially the large artery disease subtype, and coronary heart disease.\(^{53}\)

It can be expected that in the near future new reports on stroke risk scores including several intermediate phenotypes for stroke risk, also other than discussed above, will be published using more elaborate statistical methods, larger number of genetic variants as well as larger number of individuals. Another method to consider for future studies is that if patients with a very specific phenotype can be detected this may decrease the number of subjects needed to detect genetic influence on cerebrovascular risk as illustrated in the study on CADASIL and WMH volume mentioned above.\(^{54}\)

Hereditary causes of familial aggregation of stroke

Apart from molecular analyses, family studies and twin studies are important tools to study heritability of stroke. Such studies clearly indicate heritability of stroke. One study reported a prevalence of stroke or TIA of 12.3% among first degree relatives of stroke patients compared with 7.5% among first degree relatives of control subjects.\(^{55}\) Another study showed that occurrence of stroke in a parent by 65 years of age was associated with a 3-fold increase in risk of stroke in their offspring.\(^{56}\) Twin studies suggest that a genetic component of stroke risk is present.\(^{54,55}\) However, subtyping of ischemic stroke and other types of stroke in twin studies would be useful to increase the possibility to better understand heritability of ischemic stroke.\(^{57}\)

Also GWAS studies have through statistical analyses shown evidence that there is a hitherto unexplained heritability component in the risk of ischemic stroke of about 38%, and that this may vary between different subtypes of ischemic stroke.\(^{8}\) Future family studies of stroke should preferably include stroke subtypes and also focus on not only first-degree relatives but also somewhat more distant relatives of the probands.

Epigenetic impact on expression of different proteins before, during and after acute brain injury

The genetic expression can be influenced by other causes than changes of the DNA content.\(^{57}\) Such influences may be referred to as epigenetic mechanisms (Figure 2). One such mechanism is that gene transcription can be regulated by e.g.:

- Methylation of DNA\(^{58}\)
- Histone modifications\(^{59}/\)Histone deacetylases (HDACs)
- Micro-RNA\(^{59}\)
- Other

It is possible to influence the above mechanisms with pharmacological agents. E.g. valproate is a HDAC inhibitor and has been suggested to perhaps inhibit atherosclerosis. However, much more studies are needed to examine if and how it is possible to treat patients by modifying these mechanisms.

The role of non-protein coding RNA (ncRNA) is intriguing. The DNA coding for ncRNA is in the intronic portion of the genome, which is the vast majority of the DNA. Several classes
of non-protein coding RNA exist, among these micro-RNA. Micro-RNA can regulate gene expression and is considered to be an epigenetic regulator. Micro-RNA has been suggested to regulate several mechanisms in brain ischemia and therefore be of importance for recovery after stroke. These regulatory mechanisms may also be involved in the situation of cerebral ischemia with influence on cell death as well as on regeneration after stroke.

The gene expression for coding of different proteins during pathological conditions is of importance for the response of the individual subject. Indeed microarray analyses indicate that there is a very dynamic response varying both with time after stroke as well as between the core and the perifactor areas of the ischemic area of the brain. An additional therapeutic epigenetic method suggested is to regulate endogenous or exogenous stem cells to respond to cerebral injury in stroke.

**Genetic influence on functional outcome and recovery after stroke**

Recovery after stroke begins immediately after the stroke onset. Many different biological responses are involved after ischemic stroke and these vary in time and between different areas of the affected brain. Several of these responses may be of interest from a genetic point of view. These include epigenetic mechanisms - discussed above - that can be targeted for treatment in the varying temporal phases after ischemic stroke onset.

It is also of interest whether genetic variation may influence the possibility and degree of functional outcome after stroke. Different drug therapies have been tried as treatments after stroke but the responses to these therapies vary between patients. E.g. genetic polymorphisms affect the response to L-dopa treatment. The brain derived neurotrophic factor (BDNF) is involved in brain repair and plasticity and a variation of a SNP (Val66Met) in the BDNF gene has been shown to be related to improved recovery although the early response was in the opposite direction. Other examples of genes related to functional outcome include IGF1, COX-2 and GPIIIa. After lobar ICH, APOE ε2 has been related to poorer out-
come. But also other genes are of interest for outcome after ICH: an heritability estimate of 90-day ICH mortality for non-APOE loci using genomewide complex trait analysis has been calculated to about 41%.

Apart from the study by Devan et al., all the above mentioned studies on functional outcome and recovery have been candidate gene studies. No large GWAS examining the genetic effect on functional outcome after stroke has been published yet. However, such a study is now ongoing - the Genetics of Ischemic Stroke Functional Outcome Study (GISCOME) and results are expected within the coming year.

Pharmacogenetics

Pharmacogenetics is a research area holding large promise to be of importance both for stroke and other diseases in the years to come. As has been discussed above, both epigenetics and SNP variations may be used for therapeutic considerations in stroke. Two additional examples are discussed below: Thrombolytic therapy with tissue-type plasminogen activator (tPA) and anticoagulation therapy with warfarin or dabigatran.

Thrombolytic therapy

A study examined 140 candidate SNPs in 497 tPA-treated ischemic stroke patients and showed that IL1B and vWF variants were associated with early recanalization. The vWF variant was also related to FVIII activity in a subsequent functional study.

Anticoagulation therapy with warfarin or dabigatran

The treatment with anticoagulants to prevent cardioembolic stroke is highly efficient in a population of individuals with atrial fibrillation and increased risk. However the metabolism of the anticoagulant administered may vary for several reasons where a genetic variation may be one of these. Cytochrome P-450 enzyme CYP2C9 gene variants as well as variants in VKORC1, coding for vitamin K epoxide reductase (VKOR) are related to warfarin metabolism but the usefulness of genetic testing regarding these variants for guidance on initiation of warfarin treatment has been debated. The future may lead to other conclusions - in a very recent study association with APOE ε2 and AP-OE ε4 for lobar warfarin related ICH was reported.

In the Randomized Evaluation of Long-term Anticoagulation Therapy (RE-LY) study a GWAS was performed in 2944 RE-LY patients and showed that the CES1 rs2244613 minor allele was associated with lower active dabigatran metabolite levels. This minor allele was associated with a lower risk of any bleeding in the dabigatran treated patients but there was no reported association with ischemic events. The value of genetic testing and clopidogrel treatment is also of interest. Clopidogrel is metabolized to its active metabolite by CYP2C19 but the clinical utility for genetic testing to detect influence on CYP3C19 activity is still under debate.

Conclusion

Stroke genetics today is involved in many fields, including risk, outcome and pharmacogenetics. The research on stroke genetics is progressing with high pace and is expected to continue to do so during the next decade. Stroke subtyping is very important for all these areas. New methods are emerging including detailed exome content analysis, exome sequencing, whole genome sequencing and advanced statistical analytical methods. Large numbers of subjects are often needed in genetic studies and therefore co-operation in international consortia such as the International Stroke Genetics Consortium (ISGC) - www.strokegenetics.org is necessary.

References

2. Topol EJ. The genetics of heart attack. Heart 2006;92:855-861.


34. Atwood LD, Wolf PA, Heard-Costa NL, Massaro JM, Beiser A, D’Agostino RB, et al. Genetic variation in white matter hy-


