The Genetics of Ischemic Stroke

Sunaina Yadav, PhD & Pankaj Sharma, MD PhD FRCP

Institute of Cardiovascular Research
Royal Holloway University of London (ICR2UL)
& Imperial College London

Correspondence: pankaj.sharma@rhul.ac.uk

Pankaj Sharma is Professor of Neurology at Royal Holloway University of London and Director of its Institute of Cardiovascular Research (ICR2UL) which is closely partnered with local hospitals Ashford & St Peters NHS Foundation Trust. PS is also consultant neurologist at Imperial College London, Editor-in-Chief of the *Journal of the Royal Society of Medicine Cardiovascular Disease* and Hon. Medical Director of the UK national charity, *Different Strokes*.

Sunaina Yadav undertook her PhD in stroke genetics under Professor Sharma at Imperial College London. She currently heads the Department of Research and Development at Lankem Ceylon PLC (Colombo, Sri Lanka), leading and driving forward national scientific efforts in cross-functional areas of Next Generation DNA sequencing and Nanotechnology.

Acknowledgements:
PS is funded by grants from the Qatar National Research Fund, British Council, UKIERI and the Royal Society.

Key learning points:

- There are a number of monogenic disorders for stroke of which CADASIL is the most common

- Large scale population studies have demonstrated that 'sporadic' ischaemic stroke has a genetic aetiology

- Studies on those of non-European descent are few and require more investigations to make reliable conclusions and comparisons

Nearly 15 million individuals suffer from stroke each year of which 5.5 million (10% of all global deaths) die. Stroke also consumes 2-4% of global health care costs and about 4% of direct health care costs in industrialized countries. Besides the economic effects of stroke there are several social implications of the disease. Stroke leaves over 60% of the survivors with moderate to severe disability, limiting their ability to gain employment and resulting in a decline in their social functioning. Stroke survivors are also known to suffer from psychological disorders such as post stroke depression which results in greater mortality rates than non-depressive survivors. Socio-economic status of stroke patients also affects their propensity to risk factors and mortality due to an individual's ability to access healthcare, take medication and maintain a healthy lifestyle.

Heritability of Stroke

90% of the population attributable risk for stroke rests with ten conventional stroke risk factors including hypertension, atrial fibrillation, cigarette smoking, diabetes mellitus and obesity. Management of these risk factors offers the exciting possibility of near complete elimination of stroke. However, stroke risk extends well beyond the boundaries of these risk factors and the disparity in stroke prevalence within a population that is uniformly exposed to environmental risk factors suggests that some other unknown mechanisms are at play. Some of this phenotypic variability has been attributed to genetic differences, with familial patterns of inheritance lending support.

Most family and twin studies suggest the genetic liability is greater in individuals aged younger than 70 years and varies with stroke subtype. Case-control studies suggest a 76% increase in the risk of ischemic stroke in the presence of a family history of stroke, although not all reports have demonstrated a positive relationship with family history possibly due to confounding factors such as blood pressure.

The genetic basis of stroke may, for practical purposes, broadly be divided between single gene (monogenic) and polygenic (complex/multifactorial, i.e. genes interacting with environmental determinants). The difference is clinically important as the monogenic diseases have a higher penetrance and larger effect size, while polygenic presence may have lower penetrance but likely be more prevalent in the population and

may be countered by managing modifiable environmental determinants (e.g. hypertension).

Monogenic stroke studies

Monogenic stroke provides the most convincing evidence for the genetic aetiology of human stroke and genes have been identified using solely the distribution of genotypes and phenotypes within narrowly delimited families to determine the location of disease loci. While stroke remains principally a common sporadic disorder, our understanding of monogenic forms of stroke has improved greatly in recent times. However, these rare forms of stroke account for only a small percentage of stroke incidence and while not useful for determining the incidence of sporadic or polygenic forms of ischemic stroke that affect the general population they may be extremely useful in improving our understanding of the underlying mechanisms involved in the more common disorder.

CADASIL

Described by Joutel et al in 1996, CADASIL is a Mendelian form of hereditary small-vessel disease and vascular dementia. Over 100 pathogenic mutations in the *NOTCH3* gene, an evolutionarily highly conserved transmembrane receptor protein regulating cell fate, are known to almost always lead to an odd number of cysteine residues in one of the 33 EGF like repeats in the extracellular domain of the Notch3 protein. These mainly missense mutations are thought to result in conformational changes of the Notch3 protein. Mutations have predominately been identified in individuals of European descent, although cases have been found in other populations such as South Asia. A recent sequencing study has shown the association between common variants in the *NOTCH3* gene and increase in the risk of age-related white matter hyperintensities in hypertensives, suggesting that *NOTCH3* may play an important role in sporadic stroke as well[1].

The prevalence of CADASIL is likely underestimated, as clinical suspicion along with laboratory diagnosis is required. There are few prevalence studies but UK estimates of prevalence rate of confirmed CADASIL cases are about 1.98/100,000. Genotype-phenotype correlations have been difficult to determine precisely, mainly because of the heterogeneous nature of the mutations, although some mutations are associated with a

worse prognosis [2 3]. Adding to this problem, CADASIL-like symptoms have also been observed in patients without *NOTCH3* mutations[4]. Phenotypic differences such as higher volume of white matter hyperintensities have also been observed in patients with mutations in the *NOTCH3* Delta/Serrate/LAG-2 (DSL) ligand-binding domain as compared to patients with mutations outside of the DSL-binding domain.

Studies investigating CADASIL in monozygotic twins with the *NOTCH3* Cys251Tyr mutation demonstrated significant phenotypic differences in the severity of disease. The study hinted at interplay of genes and environment, with the physically inactive-smoking twin suffering a stroke 14 years earlier than the twin who led an active and healthy lifestyle[5].

There is no cure for CADASIL with treatment mainly directed at aggressive vascular risk management.

CARASIL

CARASIL or Maeda syndrome is caused by mutations in *HTRA1* gene localized on Chr10q encoding HTRA1 that represses signalling mediated by Transforming Growth Factor β (TGF-β) family. Resultantly, CARASIL patients have unproteolized cellular proteins, which affect the signal transduction process. Brain MRI shows diffuse white matter changes and multiple lacunar infarctions in the basal ganglia and thalamus. Histopathologically, arteriosclerosis is seen in the penetrating arteries in the absence of granular osmiophilic or amyloid material. Compared to CADASIL, CARASIL patients are also less likely to have migraines and exhibit psychiatric disorders, such as euphoria and emotional liability.

Prevalence rates for CARASIL are lower than CADASIL, although it is probably more frequent than the few dozen currently reported cases, which to-date have only been described from Japan and China.

Fabry's Disease

Fabry disease is a congenital metabolic disorder caused by deficient activity of α-galactosidase A, resulting in a progressive accumulation of globotriaosylceramide and

related glycosphingolipids within vascular endothelial cells, myocardial cells and neurons. Prevalence rate of Fabry's is unclear with studies reporting different results. A German study by Rolf et al reported the prevalence of Fabry's in young male stroke patients as 4.9%[6] and suggested that Fabry's could be a common cause of cryptogenic ischemic stroke. Another multi-racial study refutes this finding suggesting reporting Fabry's disease in 0.18% of all strokes and 0.65% of cryptogenic strokes[7]. Although an X-linked lysosomal storage disorder, female carriers can develop symptoms that appear comparatively later in life as compared to males, at a median age of 45.7 years.

Treatment for Fabry's includes bi-weekly recombinant α-gal enzyme replacement therapy at a dose of 1mg/kg body weight, however, continued management of conventional stroke risk factors is important as well.

MELAS

MELAS is one of the most clinically prevalent and commonly encountered genetic disorder, 80% of which is accounted for by maternally transmitted mitochondrial tRNA (Leu) A3243G mutations. Another 10% of patients carry the T3271C mutation. The prevalence of MELAS varies from 7.9/100,000 in England to 236/100,000 in Australia with an age of onset ranging from 2 to 20 years.

Treatments for MELAS are varied and include the use of vitamin supplements (B complex, E and C) and enzyme co-factors (Q10, idebenone) that enhance mitochondrial metabolism and respiratory chain activity.

Other Monogenic Disorders

A number of other monogenic disorders have also been associated with stroke; Marfan syndrome, Sickle cell disease, homocystinuria and systemic lupus erythematosus. For a more detailed review of these disorders we direct the reader to the book 'Stroke Genetics' published in 2012 by Springer (Eds: Pankaj Sharma & James Meschia).

Strategies to Study Genetics of Stroke

With the emergence of large stroke consortia and developments in genotyping technology, statistical methods and computational power, researchers have finally

begun to address the genetics of ischemic stroke effectively. Advances in our knowledge of the molecular underpinnings of stroke will enable scientists and clinicians to better understand the mechanistic workings of stroke and design effective treatments for it.

Evidence for stroke genetics can come from two different platforms; study of individuals and population based studies. Study of individuals can help identify genetic variants that causally affect stroke and provides concrete evidence for the genetic risk of stroke. Individual studies usually identify rare genetic variants with large effect sizes and high penetrance. Although such studies are of immense value, they rarely contribute to the prevention of stroke at a population level since it involves a large number of individuals at a small risk of stroke which gives rise to more cases of disease than a small number who are at high risk. Individual based studies have led to the identification of several monogenic forms of stroke such as CADASIL and enabled clinicians to use this information in their everyday clinical practice.

Population based genetic association studies have found great popularity with genetic epidemiologists since the study samples are more representative of the general population and easier to recruit as compared to stroke families. Results from a large population study are useful in calculating population attributable risk (PAR) of a genetic variant, which can be extrapolated to the general population. Population studies also have greater power in detecting common genetic variants that affect >5% of the population.

Candidate gene based studies

The study of candidate genes are often based on a priori hypothesis, primarily driven by the choice of a candidate gene which is based on the investigators research interest in a particular biological pathway such as coagulation, lipid metabolism, inflammation and blood pressure regulation, or candidate genes derived from related vascular conditions such as MI or CAD. This is not surprising as the pathophysiology of stroke and coronary disease are similar. However, replication of such candidate genes in other phenotypes has not always been successful, with some candidates appearing to be organ-specific rather than pathophysiology-specific[8]. Candidate genes found to be associated with stroke in one ethnic population are also routinely replicated in other ethnicities. Genes

involved in lipid metabolism and enzymatic activities are the most widely studied candidates for association with stroke.

Recently, findings from candidate gene association studies in stroke and other vascular phenotypes such as CAD and AF were replicated using statistically robust GWAS models. Using >3500 stroke cases and 5700 controls from the WTCCC-2 (Wellcome Trust Case Control Consortium) ischemic stroke GWAS[9] association for 50 previously reported candidate genes were tested[10]. Of the 32 stroke associated genes tested, 4 genes *ALOX5AP (*CE), *APOA (LPA)* (SVD), *Fibrinogen* (all ischemic stroke), and *Paroxonase-1* (SVD)) survived Bonferroni correction but failed when the more stringent Nyholt correction was applied. The study also tested 18 genes associated with cardiovascular phenotypes and validated the association for 3 genes at the modified Nyholt threshold: *PHACTR1* in LVD (P=2.63x10^{-6}), *PITX2* in CE stroke (*P*=4.78x10^{-8}), and *ZFHX3* in CE stroke (*P*=5.50x10^{-7}). Given the failure to replicate most stroke associated genes, the study concluded that the risk association is likely to be sub-type specific and success in identifying risk variants would continue to evade researchers unless the study populations are larger and extensively sub-typed.

Candidate gene studies have also been applied to test the progression of stroke through its intermediate phenotypes. Adib-Samii et al examined the 17q25 locus that was previously found to be associated with white matter hyperintensities in stroke-free individuals and replicated the association with white matter hyperintensity volume in ischemic stroke patients to determine whether the 17q25 locus promotes small vessel arteriopathy. The study furnished evidence in support of an association between 17q25 and white matter hyperintensities[11].

Similarly, search for blood pressure genes by the International Consortium for blood pressure genome-wide association studies (ICGP 2011) in 200,000 individuals of European descent identified 16 novel loci who's cumulative genetic risk score (in addition to 13 other loci) was associated with stroke[12]. Hypertension being the biggest risk factor for stroke, rendered this an anticipated finding. Besides the 'routine' phenotypes of blood pressure such as systolic BP, diastolic BP, mean arterial pressure and pulse pressure, the genetics of long-term variability in blood pressure or episodic

hypertension have also been investigated. In a recent study, the ASCOT IR-UK cohort identified an association between NLGN1 gene and BP variability but could not replicate the association in a large ischemic stroke population comprising 8624 cases and 12722 controls[13].

Unique step-back approaches have also been implemented to test association of candidate genes with stroke. A study by Krug et al performed gene expression profiling in peripheral blood mononuclear cells of 20 stroke cases and 20 controls and examined the differentially expressed genes between the two groups. Sixteen differentially expressed genes were then mapped to GWAS-derived regions associated with various vascular disorders. Using this approach the group was able to identify a risk association between stroke and the *TTC7B* gene locus[14].

Over all, results from candidate gene studies suggest that common stroke has a genetic component with several genes exerting individual modest effects but no single gene having a major effect. Meta-analyses of these studies has allowed disease associated genes to be reliable identified and assigned odds ratios with much greater robustness (OR 1.1-1.8) depending on the gene of interest[15]. Candidate gene studies have also demonstrated that genetic risk associations for ischemic stroke are broadly similar across different ethnicities, with some notable exceptions[16]. Such studies have also implicated disparity in the genetic burden of stroke for different stroke subtypes long before this was discovered in large-scale GWA studies.

Candidate gene studies demonstrate that while the effect sizes per gene were small, the sum of the PARs across all associations is ~30% and given the relative frequency of stroke, translates to a large clinically observed effect, although publication bias may be a reason for this probably inflated size estimate. Using a Mendelian randomization methodology, some candidate genes (*MTHFR, Factor V Leiden, ACE, Prothrombin and PAI-1*) have gone on to be not just associated with stroke but causally linked[15]. Some of these genes are associated with an ischemic process per se (e.g. with stroke and ischemic heart disease) while others are stroke specific (Table 1)[15].

Genome wide association studies

The natural extension of studying single gene regions in the human chromosome to studying all regions (of millions of genetic variants in a single experiment), and thereby avoiding investigator bias, has led to an explosion in genome wide association studies. As these studies are conducted without an *a priori* hypothesis, these have the added advantage of potentially identifying new and unpredictable genes, which could eventually lead to the development of novel therapeutic targets. The emergence of genome wide approaches have also presented investigators with an alternative and more powerful method to test the productivity of the candidate gene based approached.

Powerful GWA studies have been made possible by the advent of the human genome project and the HapMap consortium. With completion of the Human Genome Project in 2003, scientists identified regions of variation between individuals, the most common form of which is the single nucleotide polymorphism or SNP. The human genome is believed to consist of over 10 million SNPs and, with the efforts of the International HapMap project, 3 million SNPs have been characterised. Information provided by HapMap has enabled the development of commercially available genotyping microarrays, which heralded the era of the GWA study. In recent times the 1000 Genomes project (http://www.1000genomes.org) has provided 4X deep sequencing data and added immensely to the knowledge base. As technology used to unravel the genetic basis of disease has advanced, our ability to rapidly and inexpensively search for susceptibility loci has dramatically improved. Individual candidate gene studies have predominantly been replaced by whole-genome screening, which has been successfully conducted in a variety of disorders including bipolar disorder, CAD, Crohn's disease, hypertension, rheumatoid arthritis and diabetes (NHGRI catalogue, http://www.genome.gov/gwastudies/).

One of the first major GWAS in stroke was published in 2003, which identified phosphodiesterase 4D (PDE4D) to be significantly associated with risk of ischemic stroke in an Icelandic population[17]. However several attempts to replicate these findings failed, while others reported conflicting results. These discrepancies were attributed to possible problems in study design, i.e. not accounting for stroke sub-type heterogeneity.

The WTCCC-2 and the ISGC (International Stroke Genetics Consortium) performed a GWAS involving 3,548 cases of ischemic stroke with replication of potential signals in 5,859 additional cases[9]. The study demonstrated, as others had done previously, associations for CE stroke near *PITX2* and *ZFHX3*, which are known risk loci for AF. The study also confirmed the association for LVD and 9p21 locus. A novel finding was an association for large vessel stroke within *HDAC9* on chromosome 7p21.1 (OR 1.42). In a recent GWAS, the evidence for a stroke sub-type specific genetic influence became more compelling with the association of the 6p21.1 locus with large artery stroke subtype[18]. The METASTROKE meta-analysis, (~12,000 cases and ~60,000 controls) further validated previous findings of genes *PITX2*, *ZFHX3*, and *HDAC9* suggesting that these were true associations[19]. All loci exhibited heterogeneous effect across subtypes, supporting distinct genetic architectures for each subtype. The largest and most recent GWAS consisting of 17,900 ischemic stroke cases failed to replicate the METASTROKE findings but identified a novel locus at 12q24.12[20].

Several other GWAS have been conducted in stroke mostly in those of European descent, with very little comparative data available in other ethnic populations. A few studies have been conducted in populations of Asian ancestry with broadly similar effect sizes (< 1.85). Many studies, however, have failed to replicate their findings.

Noting that some ischemic stroke has a maternal heritability, a GWAS of common mitochondrial sequence variants failed to find a genome significance threshold, although this study was underpowered for GWAS[21]. GWA studies on stroke twins found no significant hits but were able to demonstrate significant correlation of age at stroke within pairs of affected siblings (r=0.83, 95% CI 0.78–0.86, p=2.2x10^{-16}) and high concordance of stroke subtypes among affected pairs (33.8%, kappa=0.13, p=5.06x10^{-4}) which did not differ by age at stroke in the proband[22]. Some investigators have undertaken GWAS on surrogate markers such as white matter hypertensities intermediate phenotypes or intermediate phenotypes such as intima-media thinkness.

Reports of a new wave of GWA studies are underway, including the WTCCC-2 and NINDS Stroke Genetics Network, which will utilize the CCS classification system[23]. The studies will focus entirely on sub-typing large number of ischemic stroke cases. A

total of 24 genetic research centres across Europe and America will participate in this global consortium amassing over 14,549 stroke cases. Large-scale prospective case-control studies examining ethnic/racial variances in ICH are also in the making and are likely to extend to ischemic stroke in the near future.

GWA studies are not a panacea for identifying genetic loci, suffering several important limitations. Errors in genotyping, quality control, and choice of analytical methods can lead to false positive results. The European population is genetically stratified and admixture of populations with different ancestry can also lead to inflated statistics. Results from large GWA studies have implied that dissecting out the susceptibility genes for stroke needs to consider its subtypes as different entities. This should not be surprising, as stroke is a clinical syndrome encompassing any sudden focal neurological deficit from a vascular etiology. Notwithstanding the arguments about sub-typing, a recent GWAS provided evidence for a genetic influence on *all-cause* ischemic stroke[18]. GWA studies also provide a limited understanding of the gene-environment interaction, which may play a major role in the differential gene expression. Another major limitation of the GWA study model is its inability to identify common genetic variants (>5%) with large effect sizes that exert an effect measureable at the population level. Most published studies have identified common variants with small to modest effect sizes for dichotomous traits (OR <1.5) and variance of <1% for quantitative traits.

Genotyping platforms

DNA microarrays allow researchers the ability to undertake high-throughput gene expression. Rapid growth in microarray technology has been spearheaded by companies such as Illumina (San Diego, USA) and Affymetrix (Santa Clara, USA), which differ considerably in SNP selection strategy and hybridization chemistry.

These technologies are reliant on the availability of large, well-characterised bio banks. A number of stroke specific DNA repositories exist and our own biobank, Bio Repository of DNA in Stroke (BRAINS) is composed of samples from those of European descent in the UK, British Asians, Indians living in India and Middle Eastern's in Qatar

(www.BrainsGenetics.com). This international repository should allow a unique comparison between disparate ancestral stroke populations[24].

Next generation sequencing

Just as candidate genes were regarded as a stepping-stone to GWAS, the latter may be regarded as a stepping-stone to Next Generation Sequencing (NGS) which will allow deep sequencing of the human genome and detection of 'rare variant, common disease'. As the cost of whole genome sequencing has plummeted in the last decade (from $ 100 million per genome in 2001 to $ 8 thousand per genome in 2014) (www.genome.gov), NGS is likely to greatly advance our understanding in stroke - where bio-repositories will be well placed to take advantage.

Although HapMap database has some rare variants, it is mostly the common SNPs that are genotyped. The general perception is that the 'missing heritability 'of stroke lies with rare genetic variants, which are too infrequent to be picked up by commercially available genotyping platforms. The availability of the entire human genome via HapMap aided by advances in statistical computation makes it a promising strategy for studying genetics of stroke. The 1000 Genomes project with whole genomes of 1000 healthy individuals will further provide dense coverage of both common and rare variants and add important information to the current knowledge base.

Although large-scale NGS approaches are already in the pipeline for various disease traits, currently there have been no published NGS studies on common stroke.

Conclusion

The genetic aetiology of stroke is currently the subject of intense international collaborative efforts. It is unlikely that a single gene will be responsible for sporadic age-related stroke; rather multiple genes acting with environmental determinants will decide eventual susceptibility. This is an exciting time in stroke genetics with promises of understanding its molecular mechanisms likely to be honoured, potential novel therapeutic targets identified and pharmacological interventions being directed by genotyping in a more personalized medicine approach.

References

1. Ross OA, Soto-Ortolaza AI, Heckman MG, et al. Variants and Risk of Ischemic Stroke. PloS one 2013;**8**(9):e75035.
2. Monet-Lepretre M, Bardot B, Lemaire B, et al. Distinct phenotypic and functional features of CADASIL mutations in the Notch3 ligand binding domain. Brain : a journal of neurology 2009;**132**(Pt 6):1601-12.
3. Opherk C, Peters N, Herzog J, et al. Long-term prognosis and causes of death in CADASIL: a retrospective study in 411 patients. Brain : a journal of neurology 2004;**127**(Pt 11):2533-9.
4. Pantoni L, Pescini F, Nannucci S, et al. Comparison of clinical, familial, and MRI features of CADASIL and NOTCH3-negative patients. Neurology 2010;**74**(1):57-63.
5. Mykkanen K, Junna M, Amberla K, et al. Different clinical phenotypes in monozygotic CADASIL twins with a novel NOTCH3 mutation. Stroke; a journal of cerebral circulation 2009;**40**(6):2215-8.
6. Rolfs A, Bottcher T, Zschiesche M, et al. Prevalence of Fabry disease in patients with cryptogenic stroke: a prospective study. Lancet 2005;**366**(9499):1794-6.
7. Wozniak MA, Kittner SJ, Tuhrim S, et al. Frequency of unrecognized Fabry disease among young European-American and African-American men with first ischemic stroke. Stroke; a journal of cerebral circulation 2010;**41**(1):78-81.
8. Helgadottir A, Gretarsdottir S, Thorleifsson G, et al. Apolipoprotein(a) genetic sequence variants associated with systemic atherosclerosis and coronary atherosclerotic burden but not with venous thromboembolism. Journal of the American College of Cardiology 2012;**60**(8):722-9.
9. International Stroke Genetics C, Wellcome Trust Case Control C, Bellenguez C, et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. Nature genetics 2012;**44**(3):328-33.
10. Bevan S, Traylor M, Adib-Samii P, et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. Stroke; a journal of cerebral circulation 2012;**43**(12):3161-7.
11. Adib-Samii P, Rost N, Traylor M, et al. 17q25 Locus is associated with white matter hyperintensity volume in ischemic stroke, but not with lacunar stroke status. Stroke; a journal of cerebral circulation 2013;**44**(6):1609-15.
12. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 2011;**478**(7367):103-9.
13. Yadav S, Cotlarciuc I, Munroe PB, et al. Genome-wide analysis of blood pressure variability and ischemic stroke. Stroke 2013;**44**(10):2703-9.
14. Krug T, Gabriel JP, Taipa R, et al. TTC7B emerges as a novel risk factor for ischemic stroke through the convergence of several genome-wide approaches. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism 2012;**32**(6):1061-72.
15. Bentley P, Peck G, Smeeth L, et al. Causal relationship of susceptibility genes to ischemic stroke: comparison to ischemic heart disease and biochemical determinants. PloS one 2010;**5**(2):e9136.
16. Ariyaratnam R, Casas JP, Whittaker J, et al. Genetics of ischaemic stroke among persons of non-European descent: a meta-analysis of eight genes involving approximately 32,500 individuals. PLoS medicine 2007;**4**(4):e131.
17. Gretarsdottir S, Thorleifsson G, Reynisdottir ST, et al. The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. Nature genetics 2003;**35**(2):131-8.
18. Holliday EG, Maguire JM, Evans TJ, et al. Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. Nature genetics 2012;**44**(10):1147-51.

19. Traylor M, Farrall M, Holliday EG, et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE Collaboration): a meta-analysis of genome-wide association studies. Lancet neurology 2012;**11**(11):951-62.
20. Kilarski LL, Achterberg S, Devan WJ, et al. Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. Neurology 2014;**83**(8):678-85.
21. Anderson CD, Biffi A, Rahman R, et al. Common mitochondrial sequence variants in ischemic stroke. Annals of neurology 2011;**69**(3):471-80.
22. Meschia JF, Nalls M, Matarin M, et al. Siblings with ischemic stroke study: results of a genome-wide scan for stroke loci. Stroke; a journal of cerebral circulation 2011;**42**(10):2726-32.
23. Ay H, Benner T, Arsava EM, et al. A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. Stroke; a journal of cerebral circulation 2007;**38**(11):2979-84.
24. Cotlarciuc I KM, Maheshwari A,Yadav S, Khan FY, Al-Hail H, de Silva R, Gorthi SP, Gupta S, Sharma SR, Sylaja PN, Prasad K, Sharma P on behalf of the BRAINS collaborators. Bio-repository of DNA in stroke: a study protocol of three ancestral populations. J R Soc Med Cardiovasc Dis 2012;**1**(10).