8. Gulcher JR, Gretarsdottir S, Helgadottir A, Stefansson K. Genes contributing to risk for common forms of stroke. *Trends Mol Med* 2005;11:217-24.
9. Hassan A, Markus HS. Genetics and ischaemic stroke. *Brain* 2000;123:1784-812.
10. Kuller LH. Epidemiology and prevention of stroke, now and in the future. *Epidemiol Rev* 2000;22:14-7.
11. Lindgren A, Lovkvist H, Hallstrom B, et al. Prevalence of stroke and vascular risk factors among first-degree relatives of stroke patients and control subjects. A prospective consecutive study. *Cerebrovasc Dis* 2005;20:381-7.
12. Markus HS, Martin RJ, Simpson MA, Dong YB, Ali N, Crosby AH, Powell JF. Diagnostic strategies in CADASIL. *Neurology* 2002;59:1134-8.
13. Touzé E, Rothwell PM. Sex differences in heritability of ischemic stroke. A systematic review and meta-analysis. *Stroke* 2007. p. 45-55
14. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661-78.
15. Vink J.M.,. Boomsma D.I. Gene finding strategies Biological Psychology 61. 2002. p. 53-57.
16. www.molstroke.com.
17. Kruglyak, L. and Nickerson, D. A. (2001) Variation is the spice of life. *Nat Genet.*, 27, 234-236.

IDENTIFICATION OF GENETIC RISK TO ISCHEMIC STROKE – THE GENOME WIDE ASSOCIATION STUDY AND META-ANALYSIS (REVIEW)

Elena Mocan
Institute of Genetics and Plant Physiology, Academy of Sciences of Moldova

Summary

A GWAS is an approach that involves rapidly scanning markers of many samples across the complete genome, to find genetic variations associated with a particular disease. Such studies are particularly useful in finding genetic variations that contribute to common complex diseases such as ictus.

We have shown analysis of recent articles dedicated to GWA studies of stroke with scopes to demonstrate positive associations with ischemic stroke. Here we proposed next candidate genes and their polymorphisms such as factor V Leiden Gln506, ACE I/D, MTHFR C677T, prothrombin G20210A, PAI-1 5G allele, ACE I/D and glycoprotein IIIa Leu33Pro to use in research of patients with ischemic stroke from Moldavian population.

Rezumat

Identificarea riscului genetic la ictus – studii de asociere largă a genomului şi meta-analiză

GWAS (Genome wide association study sau Studii de asociere largă a genomului) este o metodologie care implică scanarea rapidă a markerelor de mai multe probe în genomul complet, pentru a găsi variaţii genetice associate cu o anumită boală. Aşa studii sunt utilizate particular în gasirea variaţiilor genetice care pot să descrie predispoziţia la bolile comune complexe, cum ar fi ictus cerebral.

Am arătat analiza articolelor recente dedicate studiului despre GWAS cu scopul de a demonstra asocierea pozitivă la ictusul ischemic. Aici ne-am propus următoarele gene candidate şi polimorfismul lor cum ar fi factor V Leiden Gln506, ACE I/D, MTHFR C677T, prothrombina G20210A, PAI-1 5G allele, ACE I/D şi glycoproteina IIIa Leu33Pro pentru a le utiliza în cercetarea pacienţilor cu ictus cerebral din populaţia Republicii Moldova.

Genome-wide association studies

The field of complex genetics has been revolutionized by the advent of the genome wide association study (GWAS)[8]. This can be thought of as a large series of candidate gene studies performed in a single experiment on an array based format. As many as 1.2 million polymorphisms at a time can now be studied in this manner. Crucially, these are spread throughout the entire genome and such experiments are thus non-hypothesis driven, overcoming a major limitation of the candidate gene study. Such a large number of experiments in a single study requires a large sample sizes to allow sufficient power, even after statistical correction for multiple comparisons. Also crucial to progress has been the realisation that careful phenotyping is important, and that associations should be replicated in a second population before publication [5].

As a consequence of this and other studies, the enormous potential of GWAS to identify common variants associated with common diseases became recognized, with perhaps the seminal GWAS publication by the Wellcome Trust Case Control Consortium 1 study making GWAS a mainstream technique in disease gene identification 23. This study examined 14,000 cases of seven common diseases and 3,000 shared controls in an effort to identify genetic variants in human disease. Investigating bipolar disorder, coronary artery disease, Crohn's disease, hypertension, rheumatoid arthritis, and type I and type II diabetes, this single study identified over 58 novel loci as potentially contributing genetic risk in these conditions. To date, the GWAS technique has identified over 1212 new genetic loci predisposing to common polygenic disease [1].

Novel genetic associations with a range of cardiovascular phenotypes including myocardial infarction, coronary artery disease, diabetes and hyperlipidaemia have been reported, but few variants have been confirmed for ischaemic stroke.

It should be noted that while GWAS is a powerful technique, it requires very large, well phenotyped case series – typically in the thousands of samples, and even with these sample sizes is powered only to detect modest risks, typically with odds ratios in the region of 1.2-1.5. Thus the contribution of each risk locus to overall disease incidence is likely to be minor, although these risks are additive and as such identification of multiple loci may allow individual risk profiles to be determined [1, 6, 14].

Identification of high risk individuals could be useful in early intervention to reduce conventional risk factors, more rigorous screening for early signs of disease and in investigating severity of disease at onset as well as associations with disease recurrence.

GWAS and ischaemic stroke

The last decade has seen tremendous advances in sequencing and genotyping technologies. This development has been a prerequisite for the completion of both the sequencing of the human genome and the mapping of human haplotypes. Now such technology offers the possibility of typing hundreds of thousands of SNPs simultaneously in genome-wide association (GWA) studies [5,14].

This approach relaxes the need for a prior hypothesis in case-control studies and allows genes and genetic regions with unknown function to be tested. The great challenge with this transition to GWA studies is to separate true associations from the huge amount of false positives that will be produced [2, 4, 5]. Recently, the Wellcome Trust Case Control Consortium showed that GWA studies are feasible. In a joint effort they examined 2,000 individuals for each of 7 major diseases and a shared set of 3,000 controls in the British population [1]. They were able to replicate some previous findings, and also discovered several new candidate loci for these diseases.

The collection of large, well phenotyped sample cohorts for genetic analysis in stroke presents major challenges. In particular phenotyping, which we now realize is essential, requires detailed and expensive investigations. As in other complex diseases, collection of sufficiently large sample sizes depends on larges scale international collaborations, and to address this the International Stroke Genetics Consortium (ISGC – http://www.strokegenetics.org) was

established. Currently an ischaemic stroke GWAS in 4000 cases is near completion as part of the Wellcome Trust Case Control Consortium 2 study (WTCCC2) in collaboration with the ISGC. GWAS studies in countries including the US and Australia are also ongoing with results expected in 2011 [2]. A lesson from other disease areas is that, even with sample sizes of several thousands, power is limited and meta-analysis of multiple GWAS studies has become standard practice. The Meta-stroke collaboration has been formed in ischaemic stroke to address this.

Some novel genetic variants initially associated with other cardiovascular diseases have recently been identified as risk factors in stroke populations. A Chromosome 9 variant associated with myocardial infarction and coronary artery disease [1, 2, 11] was found to also be associated with stroke across multiple populations, but this association was due to an association with large artery stroke, and no association was found with other stroke subtypes [12]. Two other variants identified as risk factors for atrial fibrillation have also been associated with stroke; here the association was primarily with cardioembolic stroke [11,12]. These findings emphasise that genetic risk factors may predispose to specific subtypes of stroke. Therefore, identifying such risk factors will depend on rigorous stroke phenotyping and large numbers of cases with each stroke subtype.

One notable exception to this has been in Iceland, where the DeCode group reported identification of the first genetic risk for common polygenic ischaemic stroke via such a familial linkage study, which they named *STRK1* [1]. This study used the unique national collection of genealogical samples and family structures tracked in the Icelandic population to retrospectively determine cause of death and provide material for genotyping. The *STRK1* locus was identified as overlying the gene phosphodiesterase4D (PDE4D), a cyclic AMP regulator which is a plausible biological candidate 19 [1, 2].

Subsequent replication in European cohorts failed to confirm these findings 20. This study was undertaken as large scale genome wide experiments were being developed as a mainstream technique. By current standards the DeCode finding would today be considered underpowered as it failed to exceed the currently agreed statistical threshold for such studies.

Ischemic Stroke Candidate Gene Meta-Analysis

Among the literature investigating the associations between candidate genes and stroke there also show inconsistent results, but most of these studies have been underpowered.

Here I would like to propose research of Paul Bentley and coworkers as a base to present studies of genetic polymorphisms in stroke patients of Moldavian population.

There were identified 187 candidate genetic polymorphism case-control studies, incorporating 37,481 ischemic stroke cases and 95,322 controls that fulfilled the inclusion criteria. Between them, 43 polymorphisms were interrogated in 29 genes, with the mean number of studies per candidate polymorphism being 6.6 (95% CIs 4.4 – 8.8). For 23 out of the 43 candidate polymorphisms (53%), the combined studies comprised >1000 cases (and >1000 controls) in aggregate. It is these that are focused on in the rest of the results [3].

Of the 23 genetic polymorphisms candidates tested in >1000 cases, six polymorphisms in six genes were found to show an overall significant effect, with no significant between-study heterogeneity. These were, in order of case-numbers: factor V Leiden Gln506, angiotensin converting enzyme (ACE) I/D, methylene tetrahydrofolate reductase (MTHFR) C677T, prothrombin G20210A, plasminogen activator inhibitor-1 5G allele and glycoprotein IIIa Leu33Pro. The summary ORs for these genes ranged from 1.15 (95% CI: 1.06 – 1.25) for ACE I/D, to 1.60 (95% CI: 1.28 – 2.00) for prothrombin G20210A [3, 8, 11-14]. The corresponding population attributable risks for the genes listed above are, respectively: 1.8%, 3.9%, 3.1%, 1.9%, 11.2% and 5.8% (total: 27.5%) [3]. The remainder 17 polymorphisms that were tested in >1000 pooled cases failed to demonstrate association with ischemic stroke. Within this group, ten polymorphisms showed between-study heterogeneity ($p>0.05$) [3].

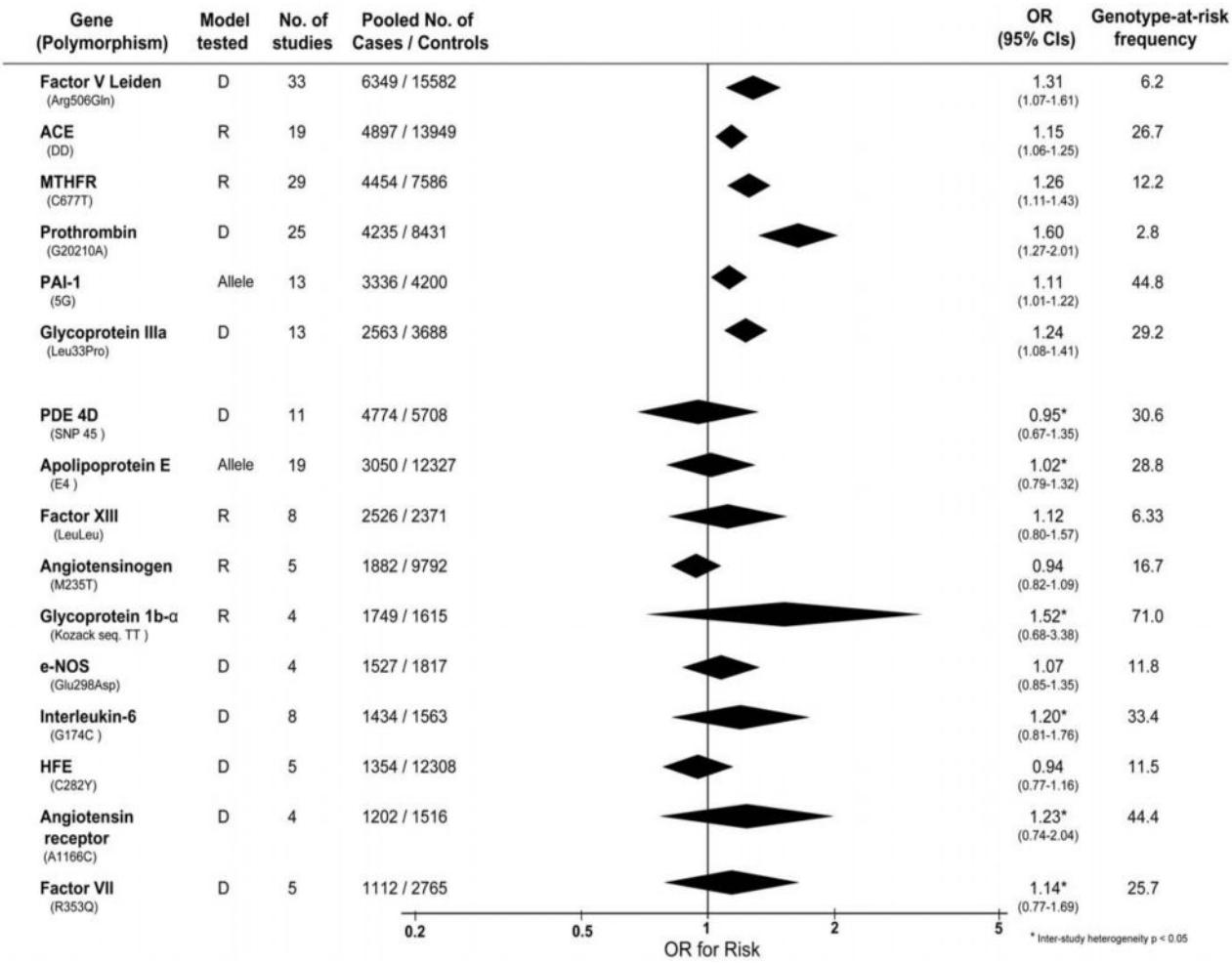

Figure 1. Summary of meta-analyses testing associations of candidate genetic polymorphisms with ischemic stroke [3].

Table reports genetic model tested (D – dominant; R – recessive); numbers of studies; numbers of pooled cases and controls; ORs with 95% confidence intervals, and at-risk genotype frequency

In order to evaluate the relative efficiency of the candidate gene method over time, they divided studies according to whether meta-analysis either identified, or failed to identify, a significant association, and plotted pooled case numbers for each of the largest polymorphisms against publication years. That shows that during the first decade of published studies (1993 – 2003) candidate polymorphisms were predominantly those found to be associated with stroke (according to our meta-analysis), whereas more recently (2004+), an increasing number of studied cases are for polymorphisms that show no association after pooling. Such declining success of candidate-gene studies is also seen by plotting the probability that cases were tested for polymorphisms that were found after meta-analysis to be associated, rather than unassociated, over time [3, 4, 5].

Future perspectives

The development of techniques and algorithms, and increase in publicly available data, is transforming the field of genetic epidemiology at a tremendous pace.

When the results of the human genome project were initially released in 2001, state-of-the-art technology was shotgun sequencing and genotyping technologies that analyzed one genotype at a time. The rather laborious restriction fragment length polymorphism (RFLP) technique, using different restriction endonucleases was losing ground to more recent techniques for SNP genotyping, based on clever use of enzymatic and detection methods. These novel techniques

relied on four general mechanisms for allelic discrimination: allele-specific hybridization, allele-specific primer extension, allele-specific oligonucleotide ligation, and allele-specific invasive cleavage [1,13].

Since then, driving down the genotyping cost and increasing throughput has meant increasing SNP content on arrays, and in general, a move to multiplex assays. Today, the most common platforms for whole-genome association studies, Illumina and Affymetrix, offer arrays that are capable of analyzing 1 million SNPs and copy-number variants using different techniques and populated with slightly different content. As regards sequencing, recently developed technology (such as the Illumina/Solexa 1G genome analyzer, Roche/454 Life Sciences Genome Sequencer FLX system, or the Applied Biosystems/APG SOLiD™ system) is capable of producing vast amount of sequence in few runs [1, 2].

It is perhaps just a question of time before platforms will be available that sequence the whole human genome in a single run. All these methods produce vast amounts of data and the development of algorithms, software and quality control systems are struggling to keep up. Certainly, there will be a high demand for skilled bioinformaticians to manipulate and keep track of statistical procedures to analyze this data.

In line with this development, there is a general tendency for association studies of complex diseases to move towards typing more SNPs on larger study.

Conclusions

Considerable evidence suggests genetic factors are important in ischaemic stroke risk. The advent of new techniques such as GWAS has contributed enormously to the understanding of the genetics of other complex disease and progress is just beginning to be made in stroke. For success large, well phenotyped case cohorts are required, and international collaborations are essential.

Acknowledgements

This work was supported in part by the state contract and Russian Foundation of Basic Research - individual Project № 08-04-90902, and U.S. Civilian Research & Development Foundation-Academy of Sciences of Moldova Moldovan Young Scientist Support Program MYSSP №1403

References

1. http://www.genome.gov/gwastudies
2. http://www.strokegenetics.org
3. Bentley P, Peck G, Smeeth L, Whittaker J, Sharma P Causal Relationship of Susceptibility Genes to Ischemic Stroke: Comparison to Ischemic Heart Disease and Biochemical Determinants. PLoS ONE 5(2), 2010: e9136, p. 1-15.
4. Bersano A., E. Ballabio, N. Bresolin, and L. Candelise. Genetic Polymorphisms for the Study of Multifactorial Stroke. Human mutation. 2008 29(6), p.776-795.
5. Casas J.P., Hingorani A.D., Bautista L. E. at al. Meta-analysis of Genetic Studies in Ischemic Stroke: Thirty-two Genes Involving Approximately 18 000 Cases and 58 000 Controls // Archives of Neurology, 2004; 61: 1652-1661.
6. Dichgans M. Genetics of ischaemic stroke. Lancet Neurol 2007; 6: 149–61.
7. Ikram MA, Seshadri S, Bis JC, Fornage M, DeStefano AL, et al. Genomewide association studies of stroke. N Engl J Med. 2009, 360: 1718–1728.
8. Johnston SC, Mendis S, Mathers CD (2009) Global variation in stroke burden and mortality: estimates from monitoring, surveillance, and modelling. Lancet Neurol 8: 345–354.
9. Hassan A, Markus HS. Genetics and ischaemic stroke. Brain 2000; 123: 1784–812.
10. James F Meschia, Thomas G Brott, Robert D Brown Jr, Richard JP Crook, Michael Frankel, John Hardy, José G Merino, Stephen S Rich, Scott Silliman and Bradford Burke Worrall. Study protocol. The Ischemic Stroke Genetics Study (ISGS) Protocol. *BMC Neurology* 2003, 3:4, p1-13.

11. Kim RJ, Becker RC. Association between factor V Leiden, prothrombin G20210A, and methylentetrahydrofolate reductase C677T mutations and events of the arterial circulatory system: A meta-analysis of published studies.Am Heart J 2003; 146: 948–57.
12. MacCallum Peter. Markrs of hemostasis and systematic inflammation in heart disease and atherosclerosis in smokers. Proceeding of the American yhoracic society., 2005, Vol. 2, p.34-43
13. Per-Gunnar Wiklund, Lennart Nilsson, Sofie Nilsson Ardnor, Per Eriksson, Lars Johansson, Birgitta Stegmayr, Anders Hamsten, Dan Holmberg and Kjell Asplund. Plasminogen Activator Inhibitor-1 4G/5G Polymorphism and Risk of Stroke: Replicated Findings in Two Nested Case-Control Studies Based on Independent Cohorts. Stroke 2005;36;1661-1665
14. Ye Z, Liu EH, Higgins JP, Keavney BD, Lowe GD, et al. Seven haemostatic gene polymorphisms in coronary disease: meta-analysis of 66,155 cases and 91,307 controls. Lancet. 2006. 367(9511): 651–8.