A genome-wide association study using a DNA pooling strategy identifies *BBS9* and *GLIS3* as novel loci influencing patient's outcome after stroke

H. Manso[1,2,3], Paulos-Pinheiro S[,2], T. Krug[1,4], J. Sobral[1,2,3], I. Albergaria[2], G. Gaspar[2], J.M. Ferro[10], S.A. Oliveira[1,4], A.M. Vicente[1,2,3]

1 – Instituto Gulbenkian de Ciência, Oeiras, Portugal, 2 – Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisbon, Portugal, 3 – Center for Biodiversity, Functional & Integrative Genomics, Lisbon, Portugal, 4 – Instituto de Medicina Molecular, Lisbon, Portugal, 5 – Serviço de Neurologia, Hospital de Santa Maria, Lisbon, Portugal

INTRODUCTION

Stroke is a leading cause of morbidity and mortality in developed countries, with a significant proportion of stroke survivors requiring institutional care and/or remaining permanently disabled [1]. Finding adequate treatments to promote patient's recovery is therefore a priority task, requiring the elucidation of the molecular pathways influencing brain recovery. Family and animal studies suggest that stroke recovery is influenced by genetic factors [2] but, except for the apolipoprotein E (*APOE*) gene [3, 4], few candidate genes have been tested for association with stroke outcome and no genome-wide association study (GWAS) has been reported.

Performing GWAS in DNA pooled samples is a cost-effective alternative to traditional GWAS and has successfully been used to identify genes associated with several traits [6-8]. With this technique, DNA from different individuals are pooled together and the SNP allele frequencies from each DNA pool are estimated using single nucleotide polymorphism (SNP) microarrays, a strategy known as allelotyping. SNPs associated with the phenotype in an initial phase can then be confirmed by individual genotyping.

This study describes a pilot GWAS to identify genetic factors contributing to patient's outcome, using a DNA pooling design.

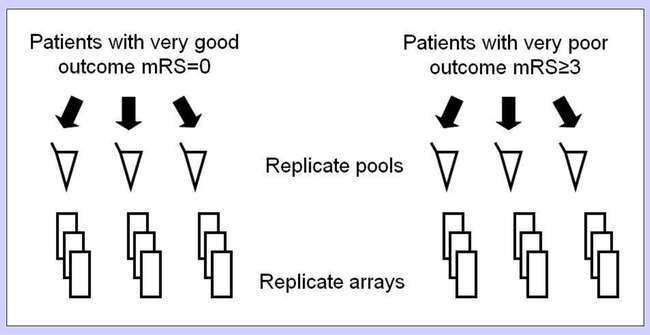
METHODS

This work was conducted in three stages:

1) Pooling-based association analysis

A – Pool construction and allelotyping (estimation of SNP allele frequencies) in each pool, using genotyping arrays

- Pooling-based association analysis of two pools of patients classified in the extremes of a clinical outcome assessment instrument, the modified Rankin Scale (mRS). A pool of 87 patients with very good outcome (mRS=0 – no disability symptoms) was constructed and compared with a pool of 100 patients with very poor outcome (mRS≥3 – moderate to severe disability and death).
- Both pools were allelotyped using the 250K Affymetrix GeneChip® Mapping Assay – Nsp I that allows the analysis of 262,264 SNPs.



B – Selection of SNPs for individual genotyping

Since there is no consensus on the best strategy for SNP selection, we chose the 100 most interesting markers based on four plausible strategies:

a) SNPs with the largest allele frequency differences between the two pools of extremely good and poor outcome [9] (N=46),

b) SNPs with the lowest Student's *t*-test p-values for the differences between allele frequency estimates [7, 10] (N=34),

c) SNPs that were clustered in 3 or more consecutive markers within 100kb from each other (N=14), and

d) SNPs that were clustered in 3 or more consecutive markers within the same gene (according to RefSeq database) (N=15).

For strategies c and d, SNPs were selected amongst the top 1000 SNPs with larger allele frequency differences between pools and the top 1000 SNPs with lower *t*-test p-values.

2) SNP validation

Validation of the pooling strategy was performed by individual genotyping of the 100 most interesting SNPs, in the same individuals (87 patients with very good outcome and 100 patients with very poor outcome), using the Sequenom MassARRAY system.

3) Association analysis in a larger sample

Association analysis with stroke outcome of validated SNPs was carried out in a larger sample of stroke patients using a more clinically sensible mRS cut-off for good and poor recovery. 230 patients with mRS≤1 (no symptoms/some symptoms but able to perform usual activities) were compared with 184 patients with mRS>1 (unable to perform usual activities to bedridden and death).

RESULTS

- 12 SNPs were excluded due to failure of quality control measures; 88 SNPs were analyzed. 36 SNPs (approximately 41%) were validated, showing significant differences between patients with extremely good and extremely poor outcome at three months ($1.7x10^{-4}<P<0.049$). Of these, 13 SNPs had p-values below 0.005.
- SNP selection strategies were evaluated by comparing the percentage of true-positive/validated markers obtained for each strategy. Selecting SNPs according to allele frequency difference between groups and selecting consecutive SNPs showed better performances, with 56.8% and 57.1%, respectively, of SNP validation. Choosing SNPs according to the *t*-test p-values showed a poor performance, with a SNP validation of 20.7%. Selecting SNPs within the same gene showed an intermediate performance, as 28.6% of SNPs were validated.
- 15 out of the 36 validated SNPs were associated with stroke outcome ($4.3x10^{-4}<_{uncorrected}P<0.047$) in a larger sample of 230 patients including the whole range of mRS scores (Table 1).
- Six SNPs remained associated with stroke outcome after adjusting for stroke severity parameters ($0.002<_{uncorrected}P<0.039$) (Table 1). Two of these SNPs, rs10273634 and rs10974334, are located within the Bardet-Biedl syndrome 9 (*BBS9*) and GLIS family zinc finger 3 (*GLIS3*) genes. rs290916 is located downstream from a novel processed transcript (RP11-428L9.1-001), for which there is little information, and the other three SNPs are located in intergenic regions and far from other known loci.
- In addition, we found a *GLIS3* haplotype (T-A-G, rs7024250-rs1000128-rs10974334) significantly associated with stroke outcome ($_{uncorrected}P=0.004$, false discovery rate [FDR] $q=0.024$).

Table 1 – Association results for the 15 SNPs associated with stroke outcome in the sample of 230 patients

SNP ID	Chr	position	gene	freq Good Recov (mRS≤1)	freq Poor Recov (mRS>1)	$_{unadjusted}P$ [a]	$_{adjusted}P$ [b]	OR [95% CI][b]	FDR q [c]
rs290916	10	9000349		68.5	79.2	**$4.3x10^{-4}$**	**0.016**	0.59 [0.39-0.92]	0.150
rs10974334	9	4092339	*GLIS3*	69.6	79.7	**$7.4x10^{-4}$**	**0.038**	0.64 [0.42-0.98]	0.155
rs2587163	4	186457029	*SNX25*	58.1	48.9	**0.008**	0.337		
rs811322	3	139812552	*FAIM*	56.9	49.7	**0.047**	0.545		
rs6590261	11	126806806		51.1	58.2	**0.040**	0.703		
rs9293983	6	75618311		58.4	65.2	**0.047**	**0.039**	0.68 [0.46-0.98]	0.155
rs7131780	12	101722143		76.1	69.1	**0.028**	0.070		
rs17627020	5	54203597		98.7	95.1	**0.005**	0.090		
rs7664979	4	95544420		59.8	67.4	**0.027**	**0.025**	0.65 [0.45-0.95]	0.155
rs1295826	14	69186662	*KIAA0247*	86.2	91.3	**0.026**	0.078		
rs1243659	14	20151347		80.0	87.1	**0.005**	**0.008**	0.50 [0.29-0.84]	0.109
rs17300340	1	178588341	*ACBD6*	81.3	86.7	**0.037**	0.169		
rs7024250	9	3823480	*GLIS3*	91.0	89.6	**0.020**	0.485		
rs10273634	7	33318154	*BBS9*	86.5	78.0	**$7.7x10^{-4}$**	**0.002**	2.21 [1.32-3.7]	0.057
rs3027232	17	7962790	*ALOXE3*	73.1	80.9	**0.009**	0.095		

Odds Ratio (OR) >1 indicates increased probability of poor outcome at three months for the carriers of the minor allele.
Chr – Chromosome, CI – confidence interval, FDR – false discovery rate, freq Good Recov – allele frequency (%) in patients with good outcome, freq Poor Recov – allele frequency (%) in patients with poor outcome.
[a] *P* for the log-additive model.
[b] OR [95% CI] and *P* for the log-additive genetic model after adjustment for significant covariates (history of hypertension, and occurrence of aphasia, paresis, altered consciousness and medical complications during hospitalization).
[c] FDR *q*values.

DISCUSSION & CONCLUSIONS

- Our preliminary results highlight two unexpected genes, *BBS9* and *GLIS3*. Additional studies are required to validate this hypothesis and to understand their connection to stroke-induced disability and/or stroke recovery processes. Association analyses will be conducted with haplotype tagging SNPs to fully cover the genetic variability within these genes, and the results will need to be replicated in independent population samples, which are currently being recruited by several research groups.
- *BBS9* encodes different isoforms of the PTHB1 protein, which are expressed in a variety of tissues, including the brain [11]. Mutations in this gene were identified in patients with Bardet-Biedl syndrome (BBS) (MIM ID: 209900). Obesity is one of the major clinical manifestations of BBS. Interestingly, it was observed that mice maintained in dietary energy restriction had smaller infarct volumes and less neurological impairment after stroke, which suggests that excessive energy intake or obesity can negatively influence stroke outcome [12].
- *GLIS3* encodes different isoforms of the zinc finger protein GLIS3, a transcription factor that can act as a transcriptional activator and repressor [13]. This protein is expressed in the brain, among other tissues [13]. Polymorphisms within *GLIS3* have been associated with type 1 diabetes [14], and with glycemic traits and type 2 diabetes [15]; and it was observed that diabetes is associated with severe disability after stroke [16].
- Further studies are needed to investigate the role of the new processed transcript RP11-428L9.1-001 and of the three intergenic SNPs (which may be influencing the expression levels of distantly located genes), and their potential relation to stroke outcome.

REFERENCES

1. Asplund K, Stegmayr B, Peltonen M. From the twentieth to the twenty-first century: a public health perspective on stroke. In: Ginsberg MD, Bogousslavsky J, eds. Cerebrovascular Disease Pathophysiology, Diagnosis, and Management. Oxford, UK: Blackwell Science, 1998: 901-918.
2. Atochin DN, Yuzawa I, Li Q, et al. Soluble guanylate cyclase alpha1beta1 limits stroke size and attenuates neurological injury. Stroke. 2010 41: 1815-1819.
3. Jeffs B, Clark JS, Anderson NH, et al. Sensitivity to cerebral ischaemic insult in a rat model of stroke is determined by a single genetic locus. Nat Genet. 1997 16: 364-367.
4. Sarzynska-Dlugosz I, Gromadzka G, Baranska-Gieruszczak M, Ciesielska A, Czlonkowska A. APOE does not predict poor outcome 1 year after ischemic stroke. Neurol Res. 2007 29: 64-69.
5. Treger I, Froom P, Ring H, Friedman G. Association between apolipoprotein E4 and rehabilitation outcome in hospitalized ischemic stroke patients. Arch Phys Med Rehabil. 2003 84: 973-976.
6. Brown KM, Macgregor S, Montgomery GW, et al. Common sequence variants on 20q11.22 confer melanoma susceptibility. Nat Genet. 2008 40: 838-840.
7. Diergaarde B, Brand R, Lamb J, et al. Pooling-based genome-wide association study implicates gamma-glutamyltransferase 1 (GGT1) gene in pancreatic carcinogenesis. Pancreatology. 2010 10: 194-200.
8. Shifman S, Johannesson M, Bronstein M, et al. Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. PLoS Genet. 2008 4: e28.
9. Abraham R, Moskvina V, Sims R, et al. A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. BMC Med Genomics. 2008 1: 44.
10. Baum AE, Akula N, Cabanero M, et al. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. Mol Psychiatry. 2008 13: 197-207.
11. Adams AE, Rosenblatt M, Suva LJ. Identification of a novel parathyroid hormone-responsive gene in human osteoblastic cells. Bone. 1999 24: 305-313.
12. Arumugam TV, Phillips TM, Cheng A, Morrell CH, Mattson MP, Wan R. Age and energy intake interact to modify cell stress pathways and stroke outcome. Ann Neurol. 2010 67: 41-52.
13. Kim YS, Nakanishi G, Lewandoski M, Jetten AM. GLIS3, a novel member of the GLIS subfamily of Krüppel-like zinc finger proteins with repressor and activation functions. Nucleic Acids Res. 2003 31: 5513-5525.
14. Senée V, Chelala C, Duchatelet S, et al. Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. Nat Genet. 2006 38: 682-687.
15. Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet. 2009 41: 703-707.
16. Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010 42: 105-116.

Acknowledgments: This work was supported by the grant PTDC/SAU-GMG/64426/2006, Fundação para a Ciência e Tecnologia (FCT). Helena Manso and Tiago Krug were supported by FCT fellowships.