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 even 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
   - 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 Rankine 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.

2. SNP validation
   - Validation of the pooling strategy was performed by individual genotyping of 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.

3. Association analysis in a larger sample
   - Association analysis with stroke outcome 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. 40 SNPs (approximately 41%) were validated, showing significant differences between patients with extremely good and extremely poor outcome at three months (1.7x10⁻⁴ < P<0.0049). 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 results with 56.8% and 57.1% of SNPs 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 26.8% of the SNPs were validated.

15 out of the 36 validated SNPs were associated with stroke outcome (4.3x10⁻⁴ < P < 0.014) 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 predictors (0.002 < P < 0.039) (Table 1). Two of these SNPs, rs10276384 and rs10974334, were located within the Bardet-Biedl syndrome 9 (BBBS) and GLIS family zinc finger 3 (GLIS3) genes, rs290916 is located downstream from a novel processed transcript (RP11-428L9.1-0.01), for which there is little information, and the other three SNPs were significantly associated with stroke outcome ( P<0.0049, false discovery rate [FDR] q < 0.024).

**DISCUSSION & CONCLUSIONS**

Our preliminary results highlight two unexpected genes, BBBS 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.

- BBBS encodes different isoforms of the PTH1B protein, which are expressed in different 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 most clinical manifestations of BBS. Interestingly, it was observed that mice with a targeted deletion in dietary energy restriction had smaller infant volumes and less neurological impairment after injury, 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 RP11-428L9.1-0.01 and of the three intergenic SNPs (which may be influencing the expression levels of distantly located genes), and their potential relation to stroke outcome.