Report STM

During the short period as a host in the Stroke and Dementia Research Centre at St George's University of London (SGUL), I learned some new techniques for the statistical analysis of genotypic data and I applied them to the analysis of the case-control cohort of Italian individuals genotyped at genome-wide level and already under examination in the laboratory of computational biology at IGM-CNR in Pavia.

During the first part of the fellowship period, I tried to get new insight from the results already obtained in Pavia, taking advantage of the large number of individuals present in Metastroke, a collaborative meta-analysis of ischaemic stroke Genome Wide Association Study (GWAS) data (http://www.strokegenetics.com/members-area/meta-stroke) and available at SGUL, one of the leading centre in the analysis of Metastroke data.

In particular, at SGUL it is now ongoing the analysis of White Matter Hyperintensity (WMH) Volume as an intermediate phenotype of Ischemic Stroke. In fact, it has been showed that in stroke patients WMHs are more frequent compared with healthy individuals and are usually associated with small-vessel disease, a subtype of stroke comprising individuals whose strokes, also labelled as lacunar infarcts, result from occlusion of small perforating cerebral arteries. We tested if SNPs of the 9p21 region associated with cerebrovascular disease in our cohort were also associated with White Matter Hyperintensity Volume in the Metastroke cohorts.

First of all, we looked at the association results with WMH volume for the SNPs of the 9p21 region in the cohorts with data already available, namely Oxford, Munich, Edinburgh, Milan and St. George's, London.

The Milan cohort comprises some of the individuals affected by ischemic stroke (IS) whose genotype data are already in analysis in Pavia. This study population comprised 152 individuals affected by ischemic stroke treated at the Neurological institute Besta in Milan. During the clinical evaluation, the Magnetic Resonance Images (MRI) were acquired. These cases were genotyped using Illumina Human610-Quad or Human660W-Quad and the genotypic data analyzed both in Pavia and in London as part of the Metastoke project. The WMHs on MRI were evaluated at the Stroke and Dementia Research Centre in London. In the Milan cohort, we identified two SNPs slightly significantly associated with IS. In order to detect variants with smaller effect sizes, we decided to perform a metanalysis for all the cohorts whose MRI data were available. This technique allows to combine p-values across studies, taking sample size and direction of effect into account. The meta-analysis of all the groups was performed with METAL (Willer et al., 2010). We identified some SNPs with slightly significant p value of the association test.

| MarkerName | Allele1 | Allele2 | Weight | Zscore | P-value | Direction |
|---|---|---|---|---|---|---|
| rs75546694 | t | g | 1062 | -3.599 | 0.0003193 | ?--?--- |
| rs72652428 | c | g | 1265 | 2.832 | 0.004633 | ?++-+++ |
| rs72652413 | t | c | 1265 | -2.823 | 0.004757 | ?--++-- |

In the second part of the fellowship period, we moved toward the analysis of endophenotypes that we hypotesized being involved in stroke etiology. In particular, we decided to focus our attention on lipids because evidences suggest that the lipid levels could be risk factors for ischemic stroke and in particular for large artery stroke, the stroke subtype characterized by atherosclerosis, for which the lipids levels play an important role (Badimon and Vilahur, 2012).
In 2010 it has been published a meta-analysis of 46 lipids GWAS, including more than 100,000 individuals (Teslovic et al., 2010). This study identified 95 loci associated with the plasma levels of total cholesterol, low-density lipoprotein cholesterol and high-density lipoprotein cholesterol and triglycerides. The association results, comprising P value and effect size, are available at http://www.sph.umich.edu/csg/abecasis/public/lipids2010.
We hypothesized that the lipids levels act as intermediate phenotype in stroke onset and we tested the association of genetic variants associated with lipids on ischemic stroke disease risk.
We performed a risk score analysis to test the association with ischemic stroke of SNPs already demonstrated to be involved in variation in the plasma concentrations of lipids.
As a study population we used the Milan genome-wide ischemic stroke case-control cohort, comprising 366 cases and 407 controls.
We investigated the association of lipid SNPs with IS creating genetic risk scores using the genotypic data of SNPs already associated with lipids. Our strategy was to test the association using different risk scores comprising an increasing number of SNPs gradually lowering the pvalue threshold for the selection. We created the following six risk scores:

S1 0.00 0.00000005
S2 0.00 0.000001
S3 0.00 0.00001
S4 0.00 0.0001
S5 0.00 0.001
S6 0.00 0.05

The scores were weighted by the effect size reported in the original study.
As a lipid trait, in a preliminary analysis, we decided to investigate the total cholesterol.
The obtained score were tested for association with IS and the stroke subtypes using a logistic regression model. Moreover, we calculated also the Nagelkerke's R squared parameter which is a measure of the explanation power of the logistic model.
In the entire cohort of ischemic stroke cases none of the scores, was significantly associated with the disease.
We tried to analyze also the stroke subtypes but the sample sizes are too small to reach a sufficient power for the analysis. However, considering only the group of cases with large artery, atherosclerotic stroke we obtained results suggestive of a possible positive association (S3: P= 0.081).

We are now performing the analysis on a larger cohort of stroke cases whose genotypic data are available at St. George's University of London.