Using Alternatives to GWAS to show that the gene FOXF2 increases the risk of having a stroke

Understanding the change in gene expression levels after strokes is a difficult problem in biology. There is not enough information, because stroke onset is an unpredictable event, thus comparing the gene expression levels before and after the stroke is very difficult. However, in 2016 using GWAS analysis, researchers in Boston university discovered the gene FOXF2, which evidently increases the risk of having a stroke (1). In my project, I will use multi-locus mixed linear model methodology to see if the data will produce results of the same significance.

I will use python to conduct statistical analysis. The data will be obtained from the additional files section in this [paper](https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-021-00908-9#availability-of-data-and-materials). I have already built a network and found several instances of the FOXF2 gene. In addition, there is significant data published in the supplementary materials of the paper “Identifi cation of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies”, where the discovery of FOXF2 gene and stroke relation was put forth.

To conduct the analysis, I will use random-SNP-effect MLM (RMLM), as described in [Wang et al.](https://www.nature.com/articles/srep19444#:~:text=Multi%2Dlocus%20models%20are%20better,and%20empirical%20Bayes12%20methods) (2), which considers SNPs to be random events, but still uses the Bonferroni correction method to set the p value for significance tests. Wang et al. propose that RMLM has a higher accuracy than GWAS, thus my project will be benchmarking these two methods and comparing the significance of FOXF2 and strokes generated by the given data.

The anticipated results would be visualized in a web project using d3.js comparing the results generated by GWAS and RMLM. It is expected that RMLM will change the end result by a small margin, as it has been proven to be more accurate than GWAS.

References

1. [Identification of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies](https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(16)00102-2/fulltext).
2. [Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology](https://www.nature.com/articles/srep19444#:~:text=Multi%2Dlocus%20models%20are%20better,and%20empirical%20Bayes12%20methods).