**Name:** Darren

**Supervisors**: Panos Parpas, Karim Malki

**Provisional Project Title**: Epigenetic Mechanisms Involved in Schizophrenia

**Background:** Schizophrenia is a complex psychiatric disorder encompassing a range of symptoms and etiology dependent upon the interaction of genetic and environmental factors. Epigenetics is the study of molecular variations that are independent of genetic sequence and have been shown to be influenced by an individual life experience and exposure to different pathogens; in many ways epigenetics is a window through which the environment can interact with an individual genetic make-up. The study uses amulti-stage epigenome-wide association study (EWAS), quantifying genome-wide patterns of DNA methylation in a total of 1,714 individuals from three independent sample cohorts.

**Aims and Objectives:**

1. To explore the potential of using epivariations to predict Schizophrenia cases.
2. To gain insight into which methylation sites show a stronger association with the disease and uncover relevant neurological pathways.
3. Gain further understanding in the number of features likely to be associated with the disorder and use growth curves to understand how detection power changes with sample size.
4. Provide a methodological comparison using machine learning against traditional linear methods that explore the association between each feature and outcome label. This may provides clues to the polygenic nature of the disease and account for some of the missing heritability.

**Data Availability:** Data is available on NCBI GEO.

www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84727

**Data delivered:** Yes, data can be downloaded and used without restriction.

**Data Description:** Data has been aggregated from 3 different studies. Phase 1 included 675 individuals 353 schizophrenia cases and 322 non-psychiatric controls. Phase 2 consists of 847 individuals (414 schizophrenia cases and 433 non-psychiatric controls. Phase 3 includes 96 MZ twins pairs discordant for the disorder. Beta coefficients are already normalised (.5 million per individual) can be used as predictive features and case control status as outcome labels.

**Provisional methodology:**

**Lables:**

Case-control (0,1) can be used as outcome label. This is consistent across the three different studies.

**Features:**

Approximately 450 000 features for each individual which can be reduced by half after initial QC. No missing data and the same feature is present across all cases. Features are effectively beta-coefficients which are normalised.

**Suggested Method**

SVM – with different methods for feature reduction including t.test, Fisher’s Test, RFE and RFE+mrm. Explore different kernels and parameter optimisation.

Potentially, use Random Forest or Elastic net for comparison purposes.

Conduct the analysis using the aggregated dataset. It is also possible to train in phase 1 data and test in phase 2. There may be a need to control for potential study effect using a surrogate variable analysis method when combining the datasets.

Extract SVM weights and use the most predictive features in a gene-enrichment analysis to gain biological insight into the associations between the genes and disease by uncovering relevant gene-pathways.

Optional:

There is the potential to explore the use of neural nets.

**References:**

Hannon, E., Dempster, E., Viana, J., Burrage, J., Smith, A.R., Macdonald, R., St Clair, D., Mustard, C., Breen, G., Therman, S. and Kaprio, J., 2016. An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biology*, *17*(1), p.176.

Malki, K., Tosto, M.G., Mouriño‐Talín, H., Rodríguez‐Lorenzo, S., Pain, O., Jumhaboy, I., Liu, T., Parpas, P., Newman, S., Malykh, A. and Carboni, L., 2016. Highly polygenic architecture of antidepressant treatment response: Comparative analysis of SSRI and NRI treatment in an animal model of depression. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*.