Name: Kwek Dan Yi Germaine

Matriculation Number: U1722265L

**The discovery of novel biomarkers using a multivariate regression in a lithium treated bipolar cohort**

**Introduction**

There has been growing interest in the inclusion of integrative analysis of multi-omics data in the study of psychiatric disorders. This allows researchers to better understand how different layers of omics interact to cause psychiatric disorders such as Bipolar Disorder (BD) and Schizophrenia. Additionally, comprehensive understanding of integrated multi-omics data provides clinicians and researchers a holistic appreciation of these disorders in the context of the biological mechanisms underpinning the pathology which can be appreciated complementarily to clinical observations. Biological descriptors obtained through integrative analysis of multi-omics data may reveal disorder subtypes and monitor treatment response thus returning valuable insight to clinicians for the purpose of personalised medicine.

Previous studies have primarily used supervised learning techniques to analyse multi-omics data. They focused on single-layer networks from one omics type and used simple integration methods. This fails to consider the interactions between multiple layers of different omics.

Unsupervised learning techniques such as Joint Dimensionality Reduction (JDR) techniques can address this problem. Regularised Generalised Canonical Correlation Analysis (RGCCA) is one such technique that this project will mainly use. JDR techniques such as RGCCA jointly convert multiple high-dimensional datasets into a lower-dimensional space, capturing biological and technical sources of common variability and disentangling heterogeneities across different omics types. Cluster analysis can then be performed to stratify patients. Further downstream applications such as identification of disease subtypes and patient subgroups and prediction of clinical outcomes can be done.

**Project Objective**

Single- and multi-omics data face the problem of high dimensionality relative to the number of samples. In this study, we will be using the JDR technique RGCCA to reduce the dimensionality of our Bipolar Disorder and Schizophrenia datasets to address this problem. Then, we will perform the relevant downstream analyses to uncover subtypes of patients who exhibit intra-group similarities based on multi-omics variance and cluster analysis on multi-omics derived representations, and any novel biomarkers for treatment outcome.

**Project Stages**

Stage 1: Perform a literature review to understand Bipolar Disorder and Schizophrenia, look at past research, and understand JDR techniques with a focus on RGCCA.

Stage 2: Source for appropriate datasets, for both healthy controls and patients treated with lithium.

Stage 3: Apply JDR technique RGCCA on the datasets.

Stage 4: Perform cluster analysis to stratify patients.

Stage 5: Perform further downstream analyses such as identification of disease subtypes and patient subgroups, novel biomarkers for treatment outcome, comorbidities associated with the disorders, etc.