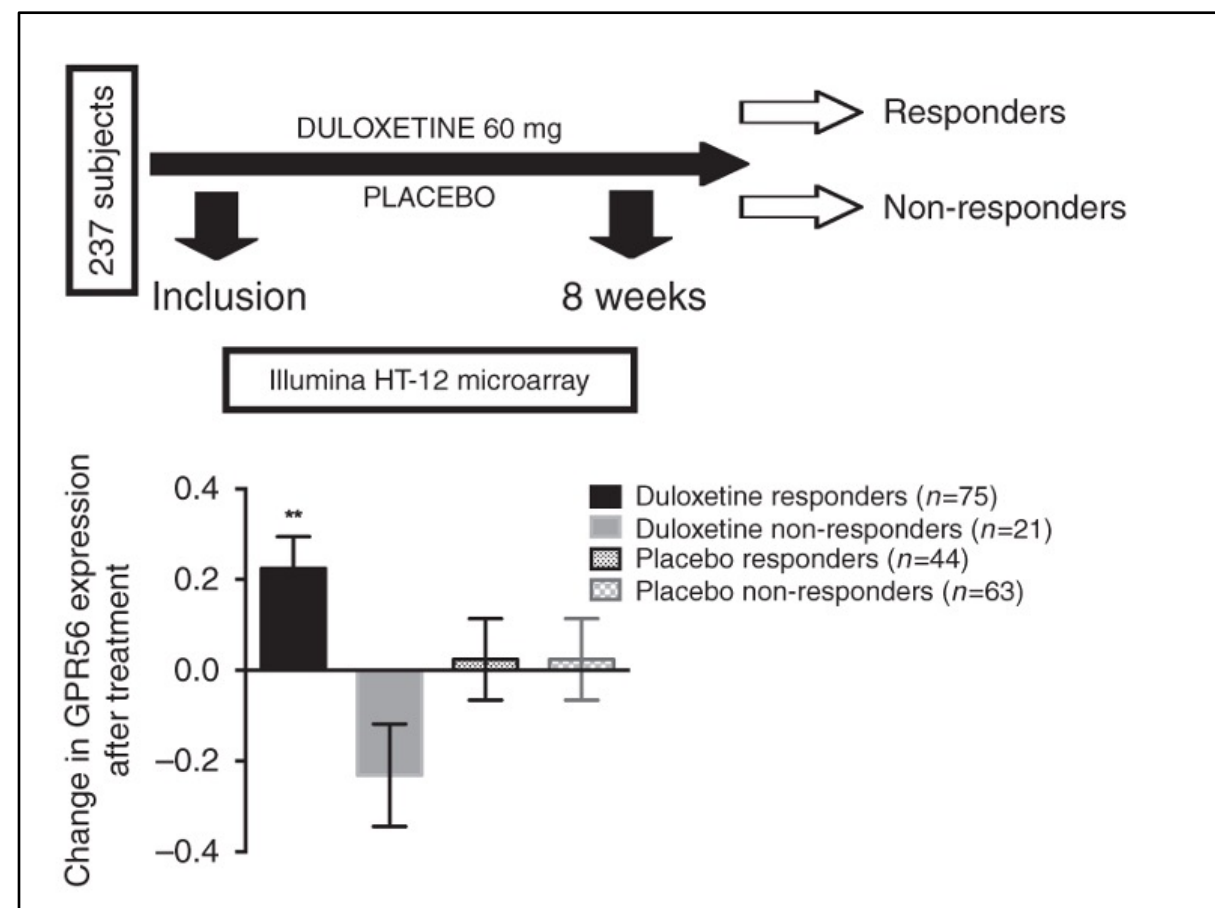
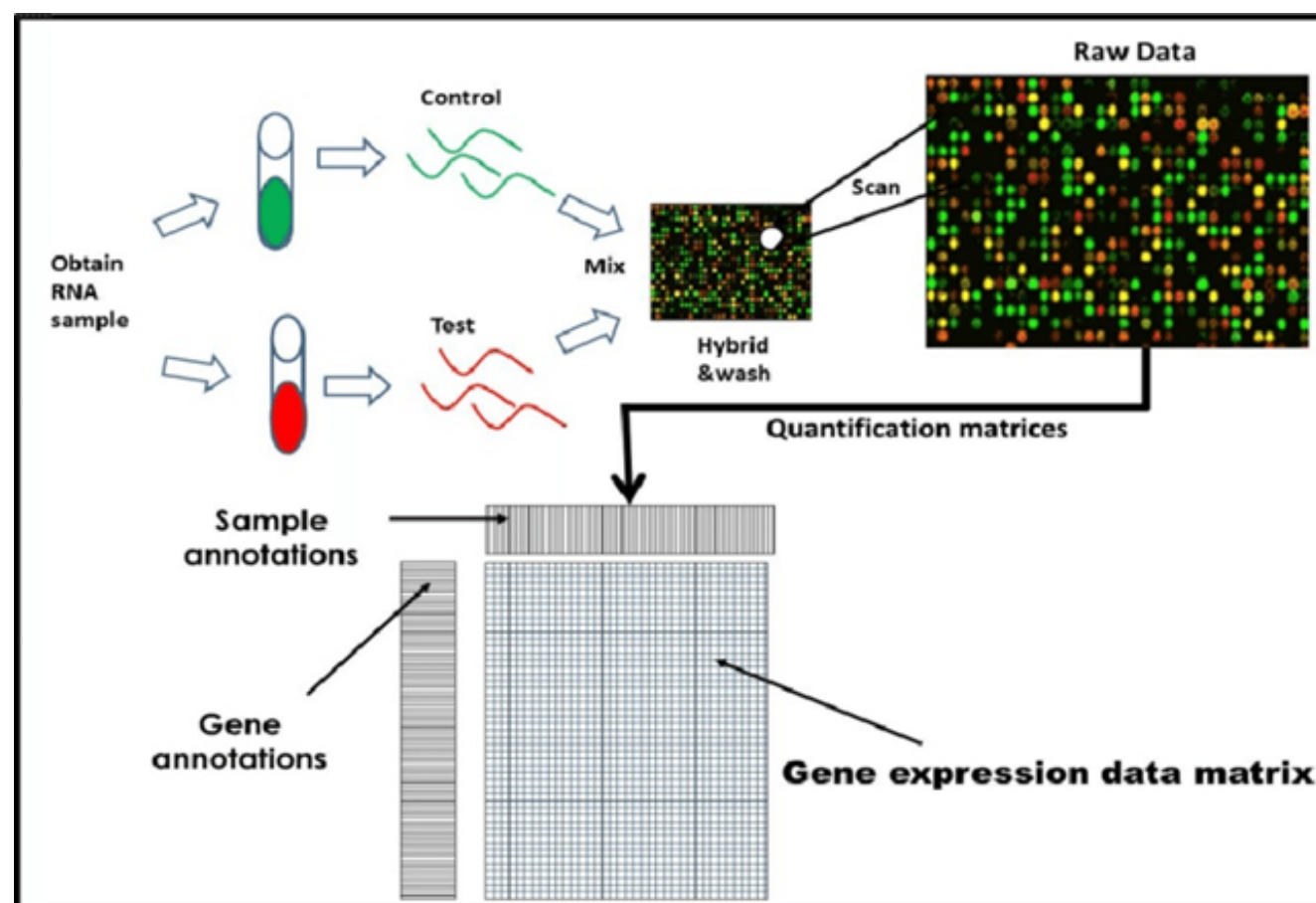


The discovery of novel biomarkers using multivariate modelling in Major Depressive Disorder

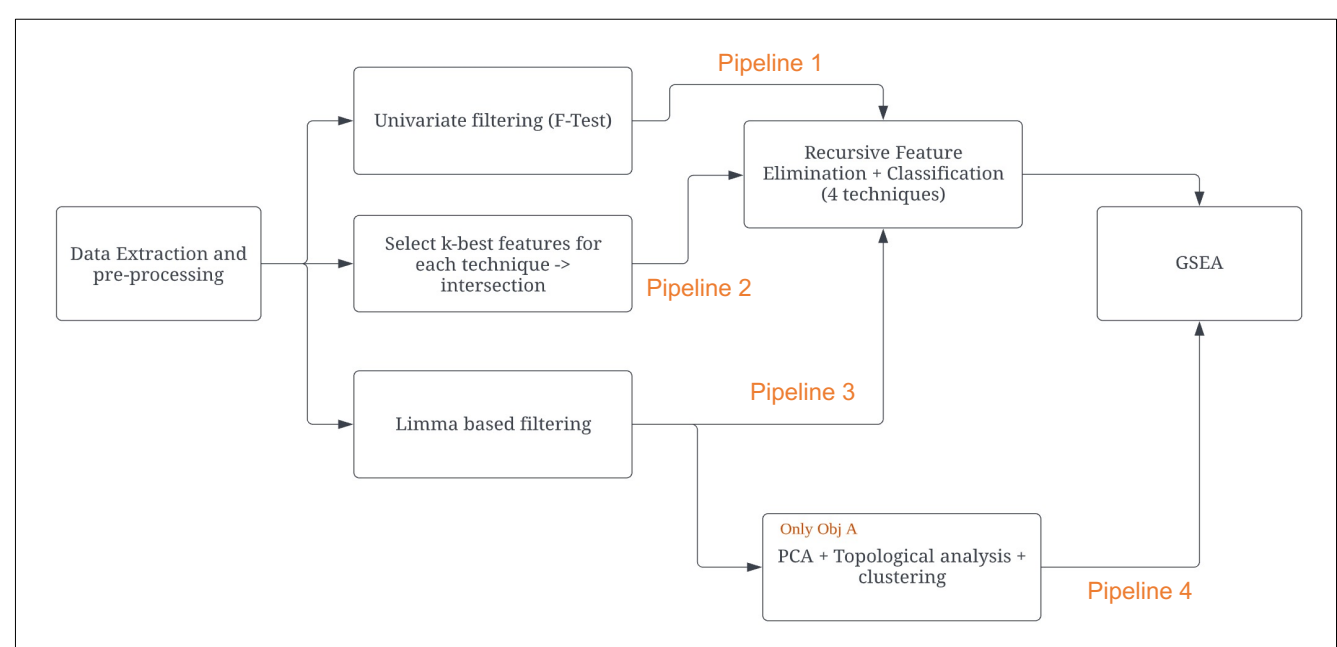
SCSE21-0415

Student: Simrita Janakiraman Supervisor: Prof Jagath Chandana Rajapakse



Project Objectives:

High dimensional transcriptomic data with insufficient number of data points makes it tougher to analyse and accurately classify the biological cohorts. Using ML techniques, we aim to narrow down transcriptomic biomarkers that have a significant impact in identifying major depressive disorder (MDD).



Patients with depression do not respond to antidepressant medication for unknown reasons. We aim to use ML techniques to classify gene expression data to understand the effects of psychiatric disorder better. This project compares performance of three pipelines combining a feature selection technique with recursive feature elimination (RFE) for classification. These pipelines are used for 2 objectives – a) predict the response to the treatment b) predict if the patient was on the treatment drug or placebo.

We found that SVM and p-LR performed better, while combining univariate filtering with RFE was found to give the most accurate results. A fourth pipeline was also used combining limma based filtering with principal component analysis and clustering, making use of topological properties of the data. This paves the way for future work to identify mental disorders on a larger dataset using univariate filtering with RFE and also the possibility of using topological properties to find the link between data points to validate the conclusions made in this project.