Figure 1: Overview

Data **collection**

Proteomics · Proteomics — *Patel et. al (2020)*

Su et. al (2020)

Metabolomics · Metabolomics — *Shen et. al (2020)*

Genomics · Genomics — *Overmyer et. al (2020)*

Data **processing**

Genecards

Protein name · ProtID

Combine on common features

Log2 transform, Remove batch effect, Median normalise

Stratified train/test split

Feature and classifier selection (for each platform)

50-repeated cv

Train fold · Test fold

Moderated t-statistic · logFC · feature loading in first PC — Feature selection method

10 . . . 50 — Select different sets of top features

+

KNN
Non-linear SVM
SVM
DLDA
Random Forest — Build models on each set

cv accuracy and recall

Choose optimal (model, features) combination based on cv accuracy and recall

Prevalidation

Prevalidated vectors from top three models

Platform vote · Platform vote · Platform vote

Final Models

Clinical Data + Logistic Regression → Predicted outcome (moderate/severe)