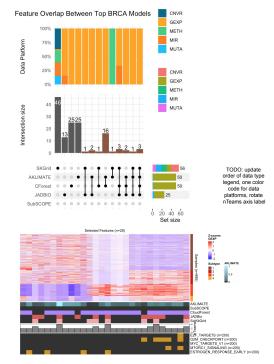
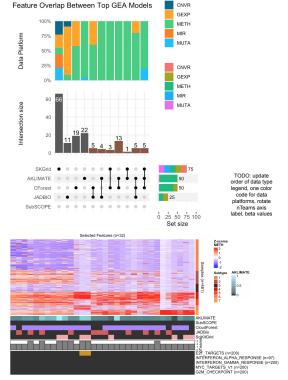
Figure 5

Jordan Lee 3/12/21





Feature Overlap Between Top LGGGBM Model Blatform CNVR 100% GEXP METH Data Platform 75% MUTA 50% CNVR GEXP 0% METH MUTA 40 Intersection size 30 26 20 10 SKGrid TODO: update AKLIMATE order of data type CForest legend, one color JADBIO code for data platforms, rotate SubSCOPE nTeams axis 20 40 60 label, beta values Set size Selected Features (n=30) AKLIMATE AKLIMATE SubSCOPE CloudForest JADBio SciKitGrid F/F TARGETS (n=200) G2M_CHECKPOINT (n=200) INTERFERON GAMMA RESPONSE (n=200) EPITHELIAL MESENCHYMAL TRANSITION (n=200) INTERFERON ALPHA RESPONSE (n=97)

Pipeline Highlights

Plots created based on:

- Best model selected by: overall weighted F1
- Best model selection limited to: <100 feature set size
- 3. Heatmaps only show features: found in at least 2 teams

Data Used

TMP_v8_20200203.tar.gz*

- feature_list_with_performance_with_subtype_names_20200828.tsv.g z (syn22337110)
- collected_features_matrix_20200722.tsv.gz (syn22271992)
 - aklimate_feature_importance_scores_20200807.tar.gz (syn22294880)

Best model per team - mean overall weighted F1 score

Scaled data: METH, GEXP, MIR

Temporarily skipped cohorts: KIRCKICH, LIHCCHOL, UCEC

*Will re-run analysis with final V10 results once all classifier teams submit results