

# Figure 5

Jordan Lee  
3/25/21

# Pipeline Highlights

Plots created based on:

1. Best model selected by: overall weighted F1
2. Best model selection limited to:  $\leq 100$  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.gz (syn22337110)
- collected\_features\_matrix\_20200722.tsv.gz (syn22271992)
- aklimite\_feature\_importance\_scores\_20200807.tar.gz (syn22294880)

**Best model per team** - mean overall weighted F1 score

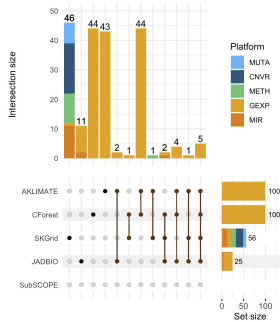
Scaled data: GEXP, MIR

Temporarily skipped cohorts: KIRCKICH, LIHCCHOL, UCEC

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

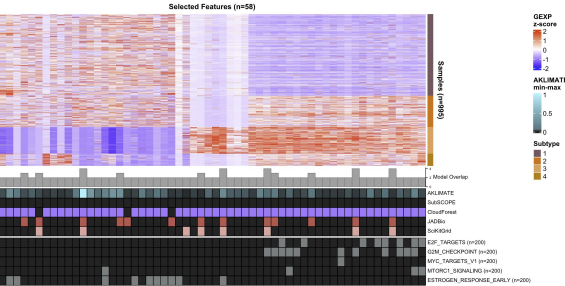
As a group, finalize and post team color codes?

Feature Overlap Between Top BRCA Models

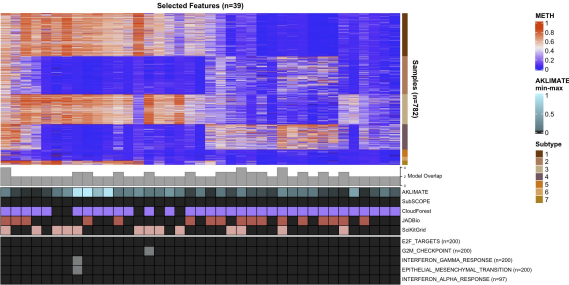
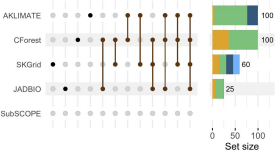
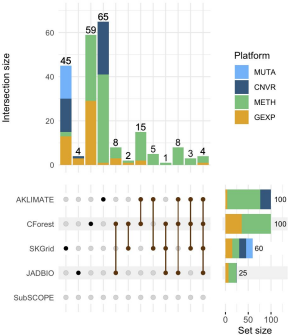


Should the upset plot be wider?

Do we still want to keep the upset plot



Feature Overlap Between Top LGGBM Models



# Feature Overlap Between Top ACC Models

