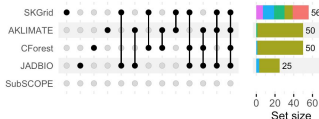
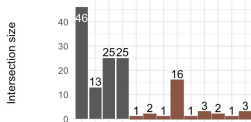
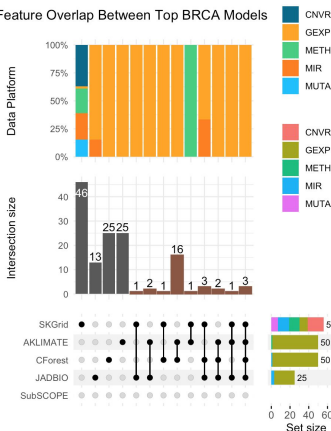


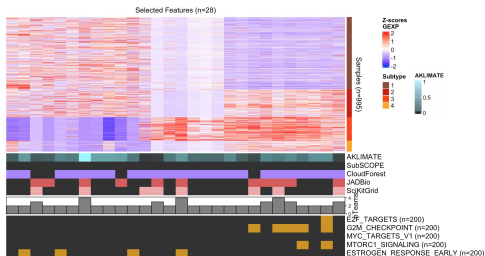
# Figure 5

Jordan Lee  
3/12/21

# Feature Overlap Between Top BRCA Models

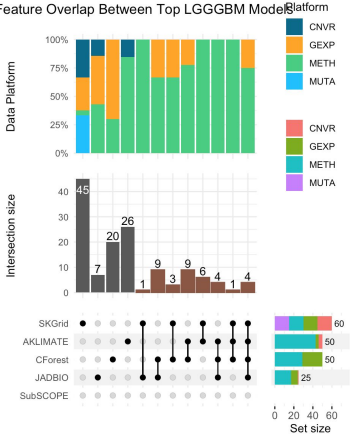


TODO: update order of data type legend, one color code for data platforms, rotate nTeams axis label

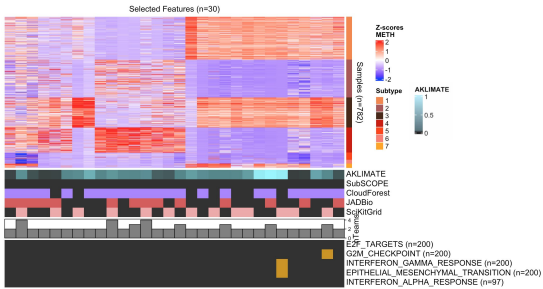




Feature Overlap Between Top LGGGBM Models



TODO: update order of data type legend, one color code for data platforms, rotate nTeams axis label, beta values



# Pipeline Highlights

Plots created based on:

1. Best model selected by: overall weighted F1
2. 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.gz (syn22337110)
- collected\_features\_matrix\_20200722.tsv.gz (syn22271992)
- akimate\_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