

Figure 5 Group best model feature overlap. Overlap of all five data platforms between groups shown in upset plot. The molecular profiles of features common across >>2 groups along with hallmark associations are shown in heatmap. ARLIMATE min-max normalized feature importance scores were calculated across all data platforms together. All other teams show presence/absence in feature set. + indicate PAM50 membership. Gene symbols labeled for features selected by >=3 croups.

## Feature Overlap Between Top LGGGBM Models 60-5251 Feature Set Size 40-Platform CNVR METH GEXP 20-1211 AKLIMATE SubSCOPE 91 Cloud Forest 100 JADBIO 20 ScikitGrid 60 Set Size 90 ò 30 Methylation Features Selected by ≥ 2 Teams (n=62) METH Samples (n=782) AKLIMATE Subtype

EZF\_TARGETS (m-200)
C3M\_CHECKFORT (m-200)
INTERFERON\_GAMMA\_RESPONSE (m-200)
ENTHELIAL\_MESENCHMAL\_TRANSITION (m-200
INTERFERON\_ALPHA\_RESPONSE (m-97)

Feature Overlap Between Top COADREAD Models 50-40-Feature Set Size Platform 30-MUTA CNVR METH 20-GEXP MIR 10-221 AKLIMATE SubSCOPE 74 Cloud Forest 100 JADBIO 24 ScikitGrid 36 90 30 60 Set Size Methylation Features Selected by ≥ 2 Teams (n=58) 0.8 0.4 Samples (n=449) AKLIMATE min-max 0 Subtype 4 5 6 INTERFERON\_GAMMA\_RESPONSE (n=200) INTERFERON\_ALPHA\_RESPONSE (n=97)

INFLAMMATORY\_RESPONSE (#=200) ALLOGRAFT\_REJECTION (#=200) TNFA\_SIGNALING\_VIA\_NFKB (#=200)

## Pipeline Highlights

Updated on 5/17/21

Plots created based on:

Heatmaps only show features: found in at least 2 teams

## Data Used

- TMP v12 20210228.tar.oz (svn25007874)
- classifier metrics 20210514.tar.gz(syn25722573)
- aklimate feature importance scores 20200807.tar.gz (syn22294880)

Scaled data: GEXP MIR

AKLIMATE importance score normalized across all data platforms together, then we look at each data type separately on heatmaps. Thus some platforms may only have features with low importance scores.

Temporarily skipped cohorts: KIRCKICH, LIHCCHOL, UCEC