Predicting response to cancer immunotherapy

In cancer immunotherapy, the patient's immune system is stimulated to destroy tumor cells, which has proven effective in numerous indications. However, only 20-40% of patients respond to immunotherapy and one of the challenges in cancer research is to predict which patients will be responders.

A dataset related to patients who underwent cancer immunotherapy is provided with patients' clinical information, RNA sequencing data and response status. The task is to build a model predicting the patient's probability of response to immunotherapy. The performance will be evaluated on an independent test set.

Datasets for model training are available in the tab-separated values (.tsv) format:

X_covariates.tsv - clinical information for all patients
X_genes.tsv - normalized RNA-sequencing data (TPM values for 31085 genes)
y.tsv - response status of patients (1-responder, 0-non-responder)

Note: information for patients in all datasets is provided in a row-wise manner (1 row per patient).

Datasets for model evaluation (will be made available 2 hours before the end of the competition):

X_covariates_test.tsv
X_genes_test.tsv

The participants should:

- read the above files in,
- generate predictions (probabilities of the patients being a responder/non-responder) to "predictions.tsv" (one row for each prediction, no header in the file)
- provide this file to the mentors before the end of the competition.

Description of some of the columns in X_covariates.tsv:

'FMOne mutation burden per MB' - Tumor Mutational Burden (TMB) measured with Foundation Medicine panel
'Neoantigen burden per MB' - Neoantigen Burden
'Enrollment IC' - PD-L1 measured at enrollment
'IC Level' - PD-L1 measured in tumor-infiltrating immune cells (IC)
'TC Level' - PD-L1 measured in tumor cells (TC)
'Immune phenotype'

'Sex'
'TCGA Subtype' - disease subtype according to The Cancer Genome Atlas (TCGA)
'Lund' - Lund genomic cancer subtype
'Lund2' - Lund2 genomic cancer subtype
'Received platinum' - whether patient received platinum prior to the treatment
'Met Disease Status' - metastatic disease status
'Sample age'
'Sample collected pre-platinum'
'Intravesical BCG administered' - whether intravesical Bacillus Calmette-Guerin (BCG) has been used
'Baseline ECOG Score' - Eastern Cooperative Oncology Group (ECOG) score (published by Oken et al. in 1982). Attempts to quantify cancer patients' general well-being and activities of daily life.
'Tobacco Use History'