

Download mutation data from TCGA



47 .maf files
23 cancer types
6,406 samples

Select non-synonymous mutations
Correct for gene length
Select primary tumors, merge duplicates



5,992 samples
19,065 genes

Select cancer-associated genes
Filter for samples without mutations



5,884 samples
2,219 genes

De-sparsify data using biological pathways
Filter samples with no pathway mutations



5,805 samples
2,219 genes
1,135 pathways

Correct for sample's mutation rate