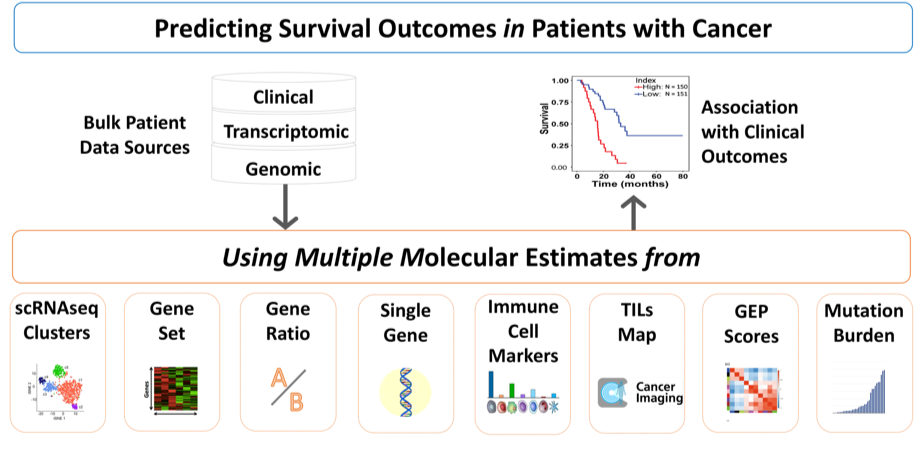
**Survival Genie**

A web portal for single-cell data, gene-ratio, and cell composition-based survival analyses

**Motivation**: The genomics data-driven identification of gene signatures, and pathways has been routinely explored for predicting cancer survival and making decisions related to targeted treatments. A large number of packages and tools have been developed to correlate expression or mutations of a gene to clinical outcome, but lack on performing such analysis based on pathways, gene sets, and gene ratios. Furthermore, in this single-cell omics era, the cluster markers from cancer single-cell transcriptomics studies remains an underutilized prognostic option. Additionally, no bioinformatics online tool evaluates associations between the enrichment of canonical cell types and survival outcome across cancers.

Here we developed Survival Genie, an interactive R shiny tool to correlate gene sets, pathways, cellular enrichment and single cell signatures to clinical outcome to assist in developing next generation prognostic and therapeutic biomarkers.



Survival Genie Overview

**Significance**: The analytical options and comprehensive collection of cancer datasets makes Survival Genie a unique resource to correlate gene sets, pathways, cellular enrichment and single cell signatures to clinical outcome to assist in developing next generation prognostic and therapeutic biomarkers.

**How it works:** User selects the analysis type (e.g., single cell data or gene sets), select the inputs, and select the dataset (e.g., TCGA-LAML, TARGET-AML), and partitioning method (i.e., mean, median, quartile, cutp) to determine effect of their levels on patient survival outcomes.

**Input:** Users are able to select a molecular input from a variety of options:

* Gene-based
  + single-cell RNA-seq data containing all markers from each cluster
  + list of genes representing a set or signature
  + input of two-genes
  + single-gene input
* Cell-based
  + proportion of tumor-infiltrating lymphocytes (TILs) using CIBERSOFT method
  + percentage of tumor-infiltrating lymphocytes (TILs) from H&E digital images
* Profile-based
  + gene expression profiles (e.g., derived from weighted gene co-expression network analysis)
* Mutation-based
  + tumor mutation burden (TMB) estimated by the total number of non-synonymous, exonic or all somatic mutations

**Output:** The tool provide comprehensive results including,

* Box plot distribution of high and low sample groups
* Correlation of molecular profile and cell composition
* Kaplan-Meier plot with log-rank p-value
* Hazard-Ratio plot with wald-test p-value
* Univariate Cox-Regression survival analysis table

**Datasets**: Survival Genie contains,

* 53 [datasets](Datasets.html) of 27 distinct malignancies from 11 different cancer [programs](Programs.html) for both adult and pediatric cancers

**Data collection and processing**:

* The clinical and omics data was downloaded from the [GDC data portal](https://portal.gdc.cancer.gov/) using The GenomicDataCommons Bioconductor R package. Only patients of available clinical survival and tumor mRNA-seq normalized FPKM expression or Whole Exome Seq derived somatic variants were processed.
* The H&E digital image-based quantification of Tumor-infiltrating lymphocytes (TILs) was downloaded from the [Cancer Imaging Archive](https://www.cancerimagingarchive.net/) for available 13 TCGA datasets. The proportion of six (LM6 signature) and 22 (LM22 signature) infiltrating immune cell types in tumors were estimated using [CIBERSORT](https://cibersort.stanford.edu/) algorithm.
* Gene expression profile (GEP) scores are calculated as weighted sum of the normalized FPKM expression values of the gene signature for each tumor. In the example data, the weightings of each gene in the signature are obtained from (claim # 21c) as published in the Patent filed under [“WO201609437”](https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2016094377&tab=PCTCLAIMS). Otherwise, the GEP scores are computed based on the user-uploaded gene signature weightings.
* The somatic mutations reported by Mutect2 pipeline from the [GDC data portal](https://portal.gdc.cancer.gov/) were downloaded and used to estimate the Tumor Mutation Burden (TMB) by the total number of non-synonymous, exonic, and all somatic mutations per Mb.

**Survival analysis**:

* Overall survival (OS) and Event-free survival (EFS) statistical analysis is performed using the ‘survival’ R package.
* Kaplan Meier survival curves are used to estimate the OS/EFS using survfit function and a log-rank test is done to compute differences in OS/EFS between the defined high and low groups. Univariate analysis with Cox proportional hazards regression model is performed on the data set using coxph function.
* The results are considered significant if the p-values from log rank and Wald test were below 0.05.