



## TCGA.SampleSelection Documentation

**Description:** Retrieve TCGA data from Broad FireBrowse and perform sample selection on the basis of expression levels for specific genes of interest for analysis using GSEA tools.

**Author:** Anthony S. Castanza, Barbara A. Hill

**Contact:** [genepattern.org/help](http://genepattern.org/help)

**Summary:** Queries cBioPortal for TCGA samples meeting criteria of mRNA expression z-scores relative to all samples (log RNA Seq V2 RSEM) greater than and less than user supplied thresholds. Outputs a GCT file containing TPM (Transcripts per Million) normalized RNA-seq quantifications suitable for ssGSEA, and a CLS file annotating samples as High or Low expression of the gene of interest.

**Parameters:**

Name	Description
TCGA Collection*	TCGA study cohort (tumor types) to query for sample selection.
Gene Symbol*	The HGNC Gene Symbol to use for classifying samples as high or low expression
High Expression*	mRNA expression is greater than or equal to this threshold for standard deviations above the mean will be classified as "high" expression of the selected gene. (mRNA expression z-scores relative to all samples) Default = 1
Low Expression*	mRNA expression is less than or equal to this threshold for standard deviations below the mean will be classified as "low" expression of the selected gene. (mRNA expression z-scores relative to all samples) Default = -1
Output Type*	Type of RSEM quantifications to output: TPM (transcripts per million, within sample normalization, useful for ssGSEA) Raw counts (unnormalized counts, usable with DESeq2 or other DEG calculations).
MSigDB Version*	MSigDB version to use for Gene Symbol lookup. This version should match the version of the gene sets intended for all downstream analysis.  Supports MSigDB versions 7.1 and higher.  <i><b>Note</b> that the default is 'latest' which queries <a href="https://www.gsea-msigdb.org/gsea/msigdb">https://www.gsea-msigdb.org/gsea/msigdb</a> to determine the current latest version of MSigDB. At the time this documentation was written, the latest version was 7.4.</i>  <i>A current listing of all MSigDB versions can be found here:</i> <a href="https://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/Release_Notes">https://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/Release_Notes</a>

\* = required

**Output File(s):** GCT file containing gene expression values in the selected output type for samples that pass the specified thresholds. A CLS file indicating the sample group assignments (high or low expression).

**Module Language:** R

**Source Repository:** <https://github.com/genepattern/TCGA.SampleSelection/tree/v0.10>

**Docker image:** genepattern/tcga-sampleselection:beta

Version	Date	Comment
---------	------	---------

# GenePattern

0.10	2021-05-04	Initial beta release.
------	------------	-----------------------