

Description

SomInaClust_det returns a prioritized list of putative driver genes and, if requested, saves this list and related files in the current (or specified) working directory.

Usage

```
SomInaClust_det(maf, database = "TCGA", save_output = TRUE,
  return_results = TRUE, convert_cosmic_to_tcga = FALSE,
  calculate_CDS = FALSE, convert_genenames_to_HGNC = TRUE,
  save_new_maf = TRUE,
  convert_cosmic_tcga_input = converttable_mutation_cosmic_tcga,
  convert_genenames_HGNC_input = converttable_genenames_HGNC,
  CDS_length = ref_v71_CDSlength, cluster_matrix_input = ref_v71_clusters,
  corr_factor_input = ref_v71_corr, save_location = getwd(),
  filename = "SomInaClust_results.txt", n_cores = NA,
  create_pyramidplot = FALSE, pyramidplot_n_genes = 20,
  create_summary_table = FALSE, CGC = CGC,
  colname_CGC_class = "Molecular.Genetics")
```

Arguments

<code>maf</code>	mutation file in mutation annotation format.
<code>database</code>	character string that specifies the maf file format. Must be one of "TCGA" or "COSMIC"
<code>save_output</code>	logical value indicating whether the result should be saved. Default is TRUE.
<code>return_results</code>	logical value indicating whether the results should be returned as an R object. Default is TRUE.
<code>convert_cosmic_to_tcga</code>	logical value indicating whether maf from Cosmic (database is "COSMIC") should be converted to TCGA format. Default is FALSE.
<code>calculate_CDS</code>	logical value indicating whether the CDS mutation position should be calculated. Default is FALSE.
<code>convert_genenames_to_HGNC</code>	logical value indicating whether genenames should be checked and if necessary converted to currently approved HGNC symbols. Default is TRUE.
<code>save_new_maf</code>	logical value indicating whether the processed maf files should be saved. Default is TRUE. By saving and reusing these maf files instead of the original, additional analyses will be significantly faster.
<code>convert_cosmic_tcga_input</code>	data frame that contains the information to convert mutation types from Cosmic to TCGA. By default, the predefined object <code>converttable_mutation_cosmic_tcga</code> is used.

<code>convert_genenames_HGNC_input</code>	data frame that contains the information to convert gene names to approved HGNC symbols. By default, the precalculated object <code>converttable_genenames_HGNC</code> is used.
<code>CDS_length</code>	data frame that specifies the gene's CDS length. By default, the precalculated object <code>ref_v71_CDSlength</code> is used.
<code>cluster_matrix_input</code>	data frame that specifies the gene-specific mutation cluster position as determined in a mutation reference database. By default, the precalculated object <code>ref_v71_clusters</code> is used.
<code>corr_factor_input</code>	data frame that specifies the gene-specific background correction factors as determined in a mutation reference database. By default, the precalculated object <code>ref_v71_corr</code> is used.
<code>save_location</code>	the working directory where the file should be saved. Defaults to current working directory.
<code>filename</code>	character string that specifies the name of the output file. Default is <code>SomInaClust_results.txt</code> .
<code>n_cores</code>	numeric value indicating the number of cores to be used. It is advised to use as much cores as possible to increase processing speed. Defaults to one less than the available number of cores. The check the available number use <code>detectCores</code>
<code>create_pyramidplot</code>	logical value indicating whether a pyramid plot figure of the highest ranked genes should be created. Default is <code>TRUE</code> .
<code>pyramidplot_n_genes</code>	numeric value indicating the number of genes to be plotted on the pyramid plot when <code>create_pyramidplot</code> is <code>TRUE</code> . Default value is 20
<code>create_summary_table</code>	logical value indicating whether a summary table of the results should be created. Default is <code>TRUE</code> .
<code>CGC</code>	cancer gene census (CGC) data frame. By default, the object <code>CGC</code> is used.
<code>colname_CGC_class</code>	character string specifying the column name of CGC that contains the molecular genetics information. Default is <code>"Molecular.Genetics"</code> .

Value

`SomInaClust_det` returns a dataframe containing a prioritized list of putative driver genes as an R object (when `return_results` is `TRUE`). Furthermore it saves the following files to the specified directory (for the default `filename`):

`SomInaClust_results.txt` Text file containing the results. Created when `save_output` is `TRUE`.

`SomInaClust_results_log.txt` Log file that contains the following information:

- Date and time information of the analysis.
- The parameter settings.
- The different steps followed by `SomInaClust_det`.
- All the files that are generated.

- The number of genes that were analyzed and the progress of the analysis. This information is useful to estimate the remaining time during of an analysis.

SomInaClust_results_pyramidplot.png This figure visualizes the top ranked putative driver genes with their respective TSG and OG score. Created when `create_pyramidplot` is TRUE.

SomInaClust_results_summary_table.txt This summary table contains all putative driver genes and their classification (OG or TSG). When `CGC` is specified it also indicates for each retrieved gene whether it belongs to the CGC list and its molecular genetics information. Created when `create_summary_table` is TRUE.

Processed maf files Depending on the parameter settings the additional maf files are returned. It is advised to use these maf files, rather than the original one, for additional analysis to enhance the processing speed.

maf_filename_CDS is the original maf file with a column added containing the CDS (numeric) position of the mutation. Created when `calculate_CDS` is TRUE.

- *maf_filename_HGNC* is the original maf file with the genenames converted to HGNC genenames. Created when `convert_genenames_to_HGNC` is TRUE.
- *maf_filename_converted* is the original Cosmic maf file in TCGA format. Created when `convert_cosmic_to_tcga` is TRUE.

Examples

```
# To run SomInaClust_det on the example file with default settings
som_ex<-SomInaClust_det(maf = maf_brca, calculate_CDS = TRUE,convert_genenames_to_HGNC=FALSE)
som_ex[1:10,] # to visualize the results of the 10 top ranked_genes
# A log file, results file and the maf file containing information on the mutations CDS positions are

# For additional analysis the maf file containing CDS position information can be used
# When an additional pyramid plot and summary table containing information regarding CGC genes is req
filename<-paste(getwd(), '/', 'SomInaClust_results_CDS.maf', sep="")
maf_brca_converted<-read.table(filename,header=TRUE,sep='\t',row.names=NULL,
  colClasses = 'character',quote=NULL,fill=TRUE)
som_ex2<-SomInaClust_det(
  maf = maf_brca_converted,
  convert_genenames_to_HGNC=FALSE,
  calculate_CDS=FALSE,
  create_pyramidplot=TRUE,
  create_summary_table=TRUE,
  filename='SomInaClust_results2.txt',
  CGC=CGC,
)
# The pyramid plot and the summary table is now present in the working directory as well.
```