

COPA Documentation

Description: Performs Cancer Outlier Profile Analysis (COPA)

Author: Alper Uzun, Joshua Gould, Marc-Danie Nazaire (Broad Institute),
gp-help@broad.mit.edu

Summary

Cancer Profile Outlier Analysis (COPA) is used for finding recurrent chromosomal translocations in cancer. COPA uses microarray data to find pairs of genes with mutually exclusive outlier tumor samples and either few or no outlier normal samples. This implementation is an adaptation of the Bioconductor (<http://www.bioconductor.org/>) package `copa` by MacDonald (2006).

References:

- Tomlins, S.A. et al. (2005) Recurrent fusion of Tmprss2 and Ets transcription factor genes in prostate cancer. *Science*, 310, 644–648.
- MacDonald JW, Ghosh D. COPA--cancer outlier profile analysis. *Bioinformatics*. 2006 Dec 1;22(23):2950-1.

Parameters:

Name	Description
dataset.file	Input dataset file - .gct or .res
sample.info.file	Sample information file (for example, a dChip format sample information file). This is a tab-delimited file where <ul style="list-style-type: none"> ▪ The first row contains labels that identify the sample information provided in each column. ▪ Each remaining row provides information for one sample. COPA requires the following columns (column labels are case-sensitive and spacing must be exact), all other columns are ignored: <ul style="list-style-type: none"> ▪ Sample: contains the sample name ▪ Ploidy(numeric): samples with ploidy=2 are normal samples, all others are target samples; there are no spaces in the column label
outlier cutoff	The cutoff to determine 'outlier' status. Default is 5.
pre-filter percentile	The percentile to use for prefiltering the data. All genes with a number of outlier samples less than the top (default 95th) percentile will be removed from further consideration. This value must be between 0 and 1 exclusive. Default is 0.95.
compute large	Whether to continue analysis when the number of genes remaining after pre-filtering is greater than 1000. The default is no.
max overlap	The maximum number of samples that can be considered 'outliers' when comparing two genes. The default is 0, indicating that there can be no overlap.
normal outliers	The number of normal samples that can be considered 'outliers'. The default is 0, meaning that no normals may be outliers.
gene pairs to plot	A range of gene pairs to plot. (e.g., 2-5) based on the gene pair rank. The number of outliers determines the rank. The more outliers a gene pair has

GenePattern

	the greater the ranking. Specify none if no gene pair plots should be generated. Default: 1-5.
output file	The base name of the output file. The default is the <i>dataset file</i> name.

Output Files:

1. <output.file>.zip containing the following files:
 - a. <output.file>_copa_table.txt, a table of the number of gene pairs with their number of outliers
 - b. <output.file>_copa_summary.txt, a detailed table showing each gene pair along with their number of outliers. The gene pairs are ranked by their number of outliers.
 - c. <output.file>_copa_filtered.gct, a gct file containing the pre-filtered dataset
 - d. If gene pairs to plot is not none:
 - a. <output.file>.#.plot.png, a plot for each specified gene pair

Platform dependencies:

Module type:	Gene List Selection
CPU type:	any
OS:	any
Language:	R