

RemoveMissingValues Documentation

Description: Removes rows or columns containing missing values from a

dataset

Author: Michael Reich, michaelr@broadinstitute.org

Summary

RemoveMissingValues removes rows (e.g., genes) and/or columns (e.g., samples) from a dataset that contains missing values. Missing values are indicated by NA.

Parameters

Name	Description
Input file	File with missing values to be removed (gct format)
Remove opt	One of (rows,cols,both) – whether to remove rows, columns, or both rows and columns containing NA values
Output filename	Name of resulting file

Input Files

1. Input file

A .gct file with missing values indicated as "NA" (no quotes). If an input file does not have any missing values, then the output file will be identical to the input file.

Note that empty cells will also be treated as missing values, though NA is the preferred method of indicating a missing value in a .gct file.

Output Files

1. Output filename

The result is the input file with (rows, columns, or both) containing NA values removed. If the resulting dataset would contain no rows or columns, an error is generated with the message "The resulting dataset contains no values."



Example Data

ftp://ftp.broadinstitute.org/pub/genepattern/example_files/RemoveMissingValues/
missing_value_example.gct

Platform Dependencies

Module type: Missing Value Imputation

CPU type: Any **OS:** Any

Language: R (2.5 or greater)