TmapToGct Documentation

Description: Converts a TMAP file to GCT file format.

Contact: GenePattern Help, gp-help@broadinstitute.org

Summary

This module converts a file in TMAP format to GCT format.

The TMAP file format is a tab-delimited format produced by Cufflinks (in GenePattern, the module with this functionality is the Cufflinks.cuffcompare module). The TmapToGct module extracts expression data from a TMAP and writes it to a GCT file. Expression data in the TMAP file is in units of Fragments Per Kilobase of exon model per Million mapped fragments (FPKM). The text elements in the TMAP file used for the row names and descriptions in the GCT file can be selected from: Reference gene name, Reference transcript ID, Cufflinks gene ID, Cufflinks transcript ID, and Major isoform ID.

The GCT format is a tab-delimited file format that describes an expression dataset commonly used in GenePattern and other tools. More information can be found here: http://www.broadinstitute.org/cancer/software/genepattern/tutorial/gp_fileformats.html.

Links

Cufflinks documentation: http://cufflinks.cbcb.umd.edu/

Parameters

Name	Description
input.file	The TMAP file to be converted.
row.labels	The column of identifiers in the TMAP file that will be used for row names in the GCT file. Options include: - Reference gene name - Reference transcript ID - Cufflinks gene ID - Cufflinks transcript ID - Major isoform ID Default: Reference gene name

row.descriptions	The column of identifiers in the TMAP file that will be used for row descriptions in the GCT file. Options include: • Reference gene name • Reference transcript ID • Cufflinks gene ID • Cufflinks transcript ID • Major isoform ID Default: Reference gene name
filter.rows	If no, all rows in the input will appear in the output file. If yes, rows whose row label is specified as "-" in the input file will not appear in the output file. Default: no.
output.prefix	A label that will be used to name the GCT output file.

Output Files

1. <output.prefix>.gct

   This GCT file contains FPKM values, row names and row descriptions as selected from the TMAP file.

Platform Dependencies

Module type: RNA-seq

CPU type: any

OS: any

Language: Perl

2