Fpkm_trackingToGct Documentation

Description: Converts an FPKM_TRACKING file to GCT format.

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

Summary

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

The FPKM_TRACKING file format is a tab-delimited format produced by Cufflinks (in GenePattern, the module with this functionality is the Cufflinks.cuffdiff module). The Fpkm_trackingToGct module extracts expression data from a FPKM_TRACKING file and writes it to a GCT file. In the FPKM_TRACKING file, expression values are in units of Fragments Per Kilobase of exon model per Million mapped fragments (FPKM). The text elements in the FPKM_tracking file used for the row names and descriptions in the GCT file can be selected from: Tracking ID, Nearest Reference Transcript ID, Gene Name, TSS ID and Locus.

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/gp_guides/file-formats/sections/gct.

Links

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

Parameters

Name	Description
input.file	The FPKM_TRACKING 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: • Tracking ID • Nearest Reference Transcript ID • Gene Name • TSS ID • Locus Default: Tracking ID

row.descriptions	The column of identifiers in the TMAP file that will be used for row descriptions in the GCT file. Options include: • Locus • Tracking ID • Nearest Reference Transcript ID • Gene Name • TSS ID • Class Code Default: Locus
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.
data.columns	The column(s) to include in the output file. If All Columns is selected then all columns excluding the columns used for row names and descriptions are included as data columns in the output file. Default: FPKM.
output.prefix	A label that will be used to name the GCT output file.

Output Files

1. <output.prefix>.gct

   This GCT file contains the selected output column values, row names and row descriptions as selected from the FPKM_TRACKING file.

Platform Dependencies

Module type: RNA-seq

CPU type: any

OS: any

Language: Perl

GenePattern Module Version Notes

Date	Version	Description
09/28/12	2	Added data columns parameter and fixed error with specifying non default values for row labels and row descriptions