← Go Back

Acgh2Tab, v4

Converts acgh files to a tab-delimited format usable by Genomica

Author: Kun Qu;Stanford   **Contact:** eby@broadinstitute.org   **Algorithm Version:**

Summary

There are cases when one would like to know a gene's copy number by offering a table with chromosome copy number data from aCGH microarray. For example, in Genomica, it is required to have an input file with genes and values attached to them in tab- separated format. This is the main reason why module AcghToTab is developed, to convert a table with chromosome location and copy number values to a table with genes and their copy number values.

Parameters

Name	Description
acgh input file *	The input file in aCGH format to be converted into TAB format for use in Genomica.
presentlist file *	A file containing the list of gene identifiers to be considered present.
absentlist file *	A file containing the list of gene identifiers to be considered absent.
genelocs file *	A file containing gene locations in UCSC whole BED format, as downloaded from UCSC Table Browser (select all fields from select table, see below for details).

* - required

Input Files

1. 1. Acgh input file
A tab-separated file with log2 ratio copy number data for each sample by gene. The file will contain the following columns (with the first row assumed to be a header):
    1. Clone sample name
    2. Target sample name
    3. Chromosome
    4. Position (in kb)
    5. A series of columns (one per gene) with the copy number alteration log2 ratio for the given samples.
2. Presentlist file
A single column of gene identifiers (no file headers).
3. Absentlist file
A single column of gene identifiers (no file headers).
4. Genelocs file
This annotation file comes from the USCS Genome Browser. See the USCS Genome Browser FAQ for general details of the bed file format. An example is included in the module as the default value for this parameter. This particular file was directly downloaded by performing the following steps:
a) Go to: http://genome.ucsc.edu/cgi-bin/hgTables?command=start
b) Chose "clade" -> Mammal,
"genome" -> human
"group" -> Genes and Gene Prediction Tracks, "track" -> RefSeq Genes
"output format" -> all fields from selected table
c) Fill in a file name and click "get output".

Output Files

1. Genomica Tab Format File
See the Genomica FAQ for details of their TAB file format

Example Data

A set of example breast cancer tumor data can be found at:
ftp://ftp.broadinstitute.org/pub/genepattern/example_files/Acgh2T ab/Acgh2T ab_testdata.zip

Platform Dependencies

Task Type:	CPU Type:	Operating System:	Language:
Data Format Conversion	any	any	Perl

Version Comments

Version	Release Date	Description
4	2014-06-02	Initial release on GParc