



## ExpressionFileCreator Documentation

**Module name:** ExpressionFileCreator  
**Description:** Creates a res or gct file from a set of Affymetrix CEL files  
**Author:** Joshua Gould, gp-help@broad.mit.edu

**Summary:** Creates a res or gct file from a set of CEL files. The CEL files are given in one zip file. The CEL files contained in the zip file may also be zipped. Algorithms for converting from probe level data to expression level data are RMA, MAS5, and dChip. Note that the results obtained from running dChip and MAS5 differ slightly from the official implementations of these algorithms.

A clm file allows you to change the name of the samples in the expression matrix, reorder the sample names, and create a cls file. By default, sample names are taken from the CEL file names contained in the zip file. A clm file allows you to change the default sample names to your desired sample names. Additionally, the columns in the expression matrix are reordered so that they are in the same order as the sample names appear in the clm file. For example, the zip file my.cel.files.zip contains the files scan1.cel, scan2.cel, and scan3.cel. The clm file could contain the following text:

|           |         |        |
|-----------|---------|--------|
| scan3.cel | sample3 | tumor  |
| scan1.cel | sample1 | tumor  |
| scan2.cel | sample2 | normal |

The column names in the expression matrix would be: sample3, sample1, sample2.

### References:

1. Affymetrix. Affymetrix Microarray Suite User Guide. Affymetrix, Santa Clara, CA, version 5 edition, 2001.
2. Rafael A. Irizarry, Bridget Hobbs, Francois Collin, Yasmin D. Beazer-Barclay, Kristen J. Antonellis, Uwe Scherf, and Terence P. Speed. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics, 2003b. To appear.

### Parameters:

| Name                         | Description   |
|------------------------------|---|
| input filename               | A zip file containing CEL files or zipped CEL files   |
| method                       | The method to use   |
| quantile normalization       | (RMA only) Whether to normalize data using quantile normalization   |
| background correct           | (RMA only) Whether to background correct using RMA background correction  |
| compute present absent calls | (MAS5 only) Whether to compute Present/Absent calls   |
| normalization                | The normalization method to apply after expression values are computed  |
| reference.scan.name:         | (mean or median scaling only) The scan name to be used as a reference for normalization. Leave blank to use median scan as reference. |
| scale value                  | (MAS5 only) Value at which all arrays will be scaled to   |
| clm .filename                | tab-delimited text file containing one scan name, sample name,  |

# GenePattern

output file

and class name per line  
The base name of the output file

## Return Value:

1. gct or res file
2. cls file if clm.filename is supplied

## Platform dependencies:

|                   |                      |
|-------------------|----------------------|
| <b>Task type:</b> | Preprocess & Utility |
| <b>CPU type:</b>  | any                  |
| <b>OS:</b>        | any                  |
| <b>Language:</b>  | R                    |