



## ReorderGCT Documentation

- Description:** Reorder a GCT by sorting on a property such as the gene symbol, average expression level, sum and variance
- Author:** Warren Kaplan (Garvan Institute of Medical Research, Sydney Australia), [w.kaplan@garvan.org.au](mailto:w.kaplan@garvan.org.au)  
Kevin Ying (Garvan Institute of Medical Research, Sydney Australia), [k.ying@garvan.org.au](mailto:k.ying@garvan.org.au)
- Date:** September, 2011
- Release:** 6.4

### Summary

This GenePattern module will sort a GCT file according to:

- **Gene Symbol**

Assumes the Gene symbol is at the start of the description field either by itself or followed by a semicolon (:).

TIP: The Reannotate module will annotate GCT's to this format

For example:

```
1.2
```

```
3      3
```

Name	Description	Sample1	Sample2	Sample3
122_z	TP	4.13	2.123	2.3
123_a	AGRN: agrin	15.13	13.123	13.4
124_b	ESPN:: espin	160.13	124.123	124.5

Output:

```
1.2
```

```
3      3
```

Name	Description	Sample1	Sample2	Sample3
123_a	AGRN: agrin	15.13	13.123	13.4
124_b	ESPN:: espin	160.13	124.123	124.5
122_z	TP	4.13	2.123	2.3

- **Average Expression Level**  
This is calculated as the mean of the expression values.
- **Variance** of the expression values
- **Sum** of the expression values

### Missing data

The follow values (without quotes:”) will be read as missing data and ignored:

” - Empty value

”na”

”NA”

### References & Links

Official Broad documentation on GCT file formats:

[http://www.broadinstitute.org/cancer/software/genepattern/tutorial/gp\\_fileformats.html#gct](http://www.broadinstitute.org/cancer/software/genepattern/tutorial/gp_fileformats.html#gct)

### Parameters (\* = required)

Name	Description
gct file*	The GCT file to be reordered.
sort feature*	Drop down menu with the choice of: <ul style="list-style-type: none"> <li>• Gene Symbol</li> <li>• Average Expression Level</li> <li>• Variance</li> <li>• Sum</li> </ul>
order as*	Ascending/Descending order
output file*	The name of the reordered GCT file. Default: <gct.file_basename>_reord.gct

## Input Files

### 1. **gct file**

A file to be reordered.

## Output Files

### **output file.gct**

The reordered GCT file

### **stdout.txt**

Details of the run

## Warning/Error Messages

### **stderr.txt**

*Error: There is a value the module cannot read as a number: at row name: \_\_\_\_*

*[ '...' '...' '...' ]*

This error occurs when the expression value is not a number or any of the missing value formats listed above. The error will output the name of the row and the values for that row. One or more of these values is likely not numeric or a missing value.

*Warning: header value: \_\_\_\_ does not match actual no. of rows \_\_\_\_, , check the help file for more details.*

The module will check the data with the information in the GCT header. If the number of actual rows in the files does not match the number in the header, this message will be given. Correct the file and try again.

*Warning: Column names (row 3 of the GCT) may be malformed, check the help file for more details*

Using the number of samples specified in the GCT header, the module will check for the correct number of columns in row 3 of the GCT.

## Example Data

[ftp://ftp.broadinstitute.org/pub/genepattern/datasets/all\\_aml/all\\_aml\\_test.gct](ftp://ftp.broadinstitute.org/pub/genepattern/datasets/all_aml/all_aml_test.gct)

## Citing this module

Kaplan, W., Ying, K., *ReorderGCT* – a GenePattern module for sorting GCT files (not published).

## Platform Dependencies

<b>Module type:</b>	Preprocess & Utilities
<b>CPU type:</b>	any
<b>OS:</b>	any Tested on Ubuntu 10.10
<b>software</b>	<i>numpy (Python module)</i>
<b>Language:</b>	Python