# Revision History

|  |  |  |
| --- | --- | --- |
| **Date** | **Author** | **Description** |
| July 8, 2011 | Steve Schumacher | Initial draft for comments. |
|  |  |  |

# Summary

## Confidence-level multiplicity

At the request of a reviewer of the GISTIC 2.0 methods paper, a feature was added to GISTIC 2.0.0 to call amplification and deletion peaks at multiple confidence levels. (Essentially, the higher the confidence level required of a gene, the wider the boundaries of a peak must be to ensure the inclusion of a driver gene.) As a result of this change, several of the output files gained a new ".conf\_level.*XX*" subextension to their names, where *.XX* refers to the numeric confidence level, e.g. .99. Additionally, when multiple confidence levels be specified, there would be as many multiples of each of these files.

* Table\_Amp\_.txt Table\_Amp\_.conf\_level.99.txt
* Table\_Del\_.txt Table\_Del\_.conf\_level.99.txt
* all\_lesions.txt all\_lesions\_.conf\_level.99.txt
* amplification.{fig|pdf|png} amplification.conf\_level.99.{fig|pdf|png}
* deletion.{fig|pdf|png} deletion.conf\_level.99.{fig|pdf|png}
* Raw\_copy\_number.{fig|pdf|png} Raw\_copy\_number.conf\_level.99.{fig|pdf|png}

The figures do not change with the confidence level, and should not have been multiplexed along with the tables. This will change with this release of GISTIC 2.0.*TBA*.

## New naming convention

GISTIC 2.0 allowed an optional user extension delimited by a "." to be added to most of the output files. The extension was a penultimate one, the final extension still designates the file type. The extension was specified by setting *params.ext* to the desired string (including the leading period).

The new version of GISTIC 2.0.*TBA*, applies the extension consistently across all output files, plus it adds an optional base name at the beginning of the file name, consistent with most GenePattern modules.Table of name changes

|  |  |  |
| --- | --- | --- |
| **new name** | **old name** | **notes** |
| [*fname.*]**all\_lesions.conf\_XX**[*.ext*]**.txt** | all\_lesions\_[.*ext*].conf\_level.XX.txt | 1 |
| [*fname****.***]**amp\_genes.**conf\_XX[.*ext*]**.txt** | Amp\_genes\_[.*ext*].conf\_level.XX.txt | 1 |
| [*fname.*]**del\_genes.**conf\_XX[.*ext*]**.txt** | Del\_genes\_[.*ext*].conf\_level.XX.txt | 1 |
| [*fname*.]**amp\_qplot**[.*ext***].{fig|pdf|png|v2.pdf}** | amplification.conf\_level.XX[.*ext*].{fig|pdf|png|v2.pdf} | 1,2,4 |
| [*fname*.]**del\_qplot**[.*ext***].{fig|pdf|png|v2.pdf}** | deletion.conf\_level. [.*ext*].{fig|pdf|png|v2.pdf} | 1,2,4 |
| [*fname*.]**raw\_copy\_number**[.*ext***].{fig|pdf|png}** | Raw\_copy\_number[.*ext*].conf\_level.xx.{fig|pdf|png} | 1,2,3,4 |
| [*fname*.]**all\_data\_by\_genes**[.*ext*]**.txt** | all\_data\_by\_genes[.*ext*].txt | 8 |
| [*fname*.]**broad\_data\_by\_genes**[.*ext*]**.txt** | broad\_data\_by\_genes[.*ext*].txt | 8 |
| [*fname*.]**focal\_data\_by\_genes**[.*ext*]**.txt** | focal\_data\_by\_genes}[.*ext*].txt | 8 |
| [*fname*.]**all\_thresholded.by\_genes**[.*ext*]**.txt** | all\_thresholded.by\_genes[.*ext*].txt | 8 |
| [*fname*.]**table\_{amp|del}.**conf\_XX[.*ext*]**.txt** | Table\_{Amp|Del}\_[.*ext*].conf\_level.XX.txt | 1,C |
| [*fname*.]**broad\_significance\_results**[.*ext*]**.txt** | broad\_significance\_results[.*ext*].txt | 7 |
| [*fname*.]**broad\_values\_by\_arm**[.*ext*]**.txt** | broad\_values\_by\_arm[.*ext*].txt | 7 |
| [*fname*.]**focal\_input**[.*ext***].seg.txt** | focal\_input\_to\_gistic.seg.txt | C |
| [*fname*.]**freqarms\_vs\_ngenes**[.*ext*]**.{fig|pdf}** | frequency\_arm\_events\_vs\_num\_genes[.*ext*].{fig|pdf} | 4,7 |
| [*fname*.]**regions\_track**.conf\_XX[.*ext*]**.bed** | regions\_track[.*ext*].conf\_level.XX.bed | 1 |
| [*fname*.]**scores**[.*ext*]**.gistic** | scores\_[.*ext*].conf\_level.XX.gistic.txt | 1,2 |
| [*fname*.]**sample\_cutoffs**[.*ext*]**.txt** | sample\_cutoffs.txt | 8 |
| [*fname*.]a**rraylistfile**[.*ext*]**.txt** | arraylistfile.txt | 5 |
| [*fname*.]**broad\_gistic\_plot**[.*ext*]**.{fig|pdf}** | broad\_gistic\_plot[.*ext*].{fig|pdf} | 4,6,7 |
| [*fname*.]**D.capC.C**[.*ext*]**.mat** | D.C.C[.*ext*].mat | A,B |
| [*fname*.]**broad\_results**[.*ext*]**.mat** | broad\_results[.*ext*].mat | C |
| [*fname*.]**focal\_dat.*L.L***[.*ext*]**.mat** | focal\_dat.L.L[.*ext*].mat | A,B |
| [*fname*.]**scores.*L.L***[.*ext*]**.mat** | scores.L.L[.*ext*].mat | A,B |
| [*fname*.]**gene\_stats**[.*ext*]**.mat** | gene\_stats[.*ext*].mat | A,C |
| [*fname*.]**orig\_stats**[.*ext*]**.mat** | orig\_stats[.*ext*].mat | A,C |
| [*fname*.]**all\_thresholded.by\_genes**[.*ext*]**.mat** | all\_thresholded.by\_genes[.*ext*].mat | C |
| [*fname*.]**peak\_regs**[.*ext*]**.mat** | peak\_regs[.*ext*].mat | A |
| [*fname*.]**perm\_ads**[.*ext*]**.mat** | perm\_ads[.*ext*].mat | A |
| [*fname*.]**wide\_peak\_regs**[.*ext*]**.mat** | wide\_peak\_regs[.*ext*].mat | A |
| [*fname*.]**segmented\_data**[.*ext*]**.mat** | segmented\_data[.*ext*].mat | C,9 |

|  |
| --- |
| **Notes** |
| 1 - this file was multiplexed one per confidence level (specified by conf\_XX) for GISTIC 2.0 |
| 2 - unnecessary multiplexing wil be removed for this release |
| 3 - documented for GISTIC 1.0 as segmented\_copy\_number[.ext].pdf |
| 4 - if genepattern, the .fig version of the image is not exported |
| 5 - only output if params.array\_list\_file is set to an input array list |
| 6 - only output if params.use\_two\_sided = 1 |
| 7 - only output if params.run\_broad\_analysis = 1 |
| 8 - only output if params.write\_gene\_files = 1 |
| 9 - output may be supressed by setting params.save\_seg\_data = 0 |
| A - output may be suppressed by setting params.save\_data\_files = 0 |
| B - output may be suppressed by setting params.conserve\_disk\_space = 1 |
| C - output may be supressed by setting params.genepattern = 1 |

# File Descriptions

## Already documented output files

### [*fname.*]all\_lesions.conf\_level.*XX*[.*ext*].txt

(was all\_lesions\_.conf\_level.*XX*.txt)

### [*fname.*]{amp|del}\_genes.conf\_level.XX[.ext].txt

(was {Amp|Del}\_genes\_[.*ext*].conf\_level.*XX*.txt)

### [*fname.*]amp\_qplot[.*ext*].{fig|pdf|png|v2.pdf})

(was amplification.conf\_level. [.*ext*].{fig|pdf|png|v2.pdf})

### [*fname.*]del\_qplot[.*ext*].{fig|pdf|png|v2.pdf}

(was deletion.conf\_level. [.*ext*].{fig|pdf|png|v2.pdf})

### [*fname.*]raw\_copy\_number[.*ext*].{fig|pdf|png}

(was Raw\_copy\_number[.*ext*].conf\_level.xx.{fig|pdf|png}, also appears to be documented as "segmented\_copy\_number.[*.ext*].pdf" )

## Previously undocumented output files

### [*fname.*]table\_{amp|del}.conf\_ XX[.ext].txt

Tables of basic information about the genomic regions (peaks) that GISTIC determined to be significantly amplified or deleted. These describe three kinds of peak boundaries, and list the genes contained in two of them. The *region start* and *region end* columns (along with the *chromosome* column) delimit the entire area containing the peak that is above the significance level. The region may be the same for multiple peaks. The *peak* start and end delimit the maximum value of the peak. The extended peak is the peak determined by robust, and is contained within the wide peak reported in {*amp|del}\_genes.txt* by one marker.

### [*fname.*]broad\_significance\_results[.*ext*].txt

A table of per-arm statistical results for the data set. Each arm is a row in the table. The first column specifies the arm and the second column counts the number of genes known to be on the arm. For both amplification and deletion, the table has columns for the frequency of amplification or deletion of the arm, and a z-score and q-value.

### [*fname.*]broad\_values\_by\_arm[.*ext*].txt

A table of chromosome arm amplification levels for each sample. Each row is a chromosome arm, and each column a sample. The data are in units of absolute copy number - 2.

### [*fname.*]all\_data\_by\_genes[.*ext*].txt

A gene-level table of copy number values for all samples. Each row is the data for a gene. The first three columns name the gene, its NIH locus ID, and its cytoband - the remaining columns are the samples. The copy number values in the table are in units of (copy number - 2), so that no amplification or deletion is 0, genes with amplifications have positive values, and genes with deletions are negative values. The data are converted from marker level to gene level using the extreme method: a gene is assigned the greatest amplification or the least deletion value among the markers it covers.

### [*fname.*]broad\_data\_by\_genes[.*ext*].txt

A gene-level table of copy number data similar to the *all\_data\_by\_genes.txt* output, but using only broad events with lengths greater than the broad length cutoff. The structure of the file and the methods and units used for the data analysis are otherwise identical to *all\_data\_by\_genes.txt*.

### [*fname.*]focal\_data\_by\_genes[.*ext*].txt

A gene-level table of copy number data similar to the *all\_data\_by\_genes.txt* output, but using only focal events with lengths greater than the focal length cutoff. The structure of the file and the methods and units used for the data analysis are otherwise identical to *all\_data\_by\_genes.txt*.

### [*fname.*]all\_thresholded.by\_genes[.*ext*].txt

A gene-level table of discrete amplification and deletion indicators at for all samples. There is a row for each gene. The first three columns name the gene, its NIH locus ID, and its cytoband - the remaining columns are the samples. A table value of 0 means no amplification or deletion above the threshold. Amplifications are positive numbers: 1 means amplification above the amplification threshold; 2 means amplifications larger to the arm level amplifications observed for the sample. Deletions are represented by negative table values: -1 represents deletion beyond the threshold; -2 means deletions greater than the minimum arm-level deletion observed for the sample.

### [*fname.*]sample\_cutoffs[.*ext*].txt (was sample\_cutoffs.txt)

A table of the per-sample threshold cutoffs (in units of absolute copy number - 2) used to distinguish the high level amplifications (±2) from ordinary amplifications (±1) in the all\_thresholded.by\_genes output file. The table contains three columns: the sample identifier followed by the low (deletion) and high (amplification) cutoff values. The cutoffs are calculated as the minimum arm-level amplification level less the deletion threshold for deletions and the maximum arm-level amplification plus the amplification threshold for amplifications.

### [*fname.*]focal\_input\_to\_gistic[.*ext*].seg.txt

A list of copy number segments describing just the focal events present in the data. The segment amplification/deletion levels are in units of (copy number - 2), with amplifications positive and deletions negative numbers. This file may be viewed with IGV.

*Changes since 2.0*: the column header no longer incorrectly indicates log2 ratio units for copy number. Optional extension added.

VERIFY: This file should now have an extension. Copy number units in header should not indicate log

Created by *run\_focal\_gistic()* calling *write\_seg\_file()*.

### [*fname.*]freqarms\_vs\_ngenes[.*ext*].{fig|pdf}

An image showing the correlation between gene counts and frequency of copy number alterations, similar to Supplementary Figure S6 in (Beroukhim et al., 2010). Output was renamed from frequency\_arm\_events\_vs\_num\_genes for brevity.

### [*fname.*]regions\_track.conf\_level.XX[.ext].bed

A file indicating the position of the confidence intervals around GISTIC peaks that can be loaded as a track in a compatible viewer browser such as IGV or the UCSC genome browser.

### [*fname.*]scores.gistic[.*ext*].txt (was scores\_.conf\_level.*XX.*gistic.txt)

* [*fname.*]arraylistfile.txt Only created if params.arraylist is non-empty.
* [*fname.*]broad\_gistic\_plot[.*ext*].{fig|pdf} Only created if params.use\_two\_sided is set

## Matlab Data File Outputs

These are completely suppressed if params.genepattern=1 and params.save\_data\_files=0 for the GISTIC run.

Descriptions TBA...

### [*fname.*]D.cap*C.C*.mat (was D.*C.C*.mat)

### [*fname.*]broad\_results[.*ext*].mat

### [*fname.*]focal\_dat.*L.L*[.*ext*].mat

### [*fname.*]scores.*L.L*[.*ext*].mat

### [*fname.*]gene\_stats[.*ext*].mat

### [*fname.*]orig\_stats[.*ext*].mat

### [*fname.*]all\_thresholded.by\_genes}[.*ext*].mat

### [*fname.*]peak\_regs[.*ext*].mat

### [*fname.*]perm\_ads[.*ext*].mat

### [*fname.*]wide\_peak\_regs[.*ext*].mat

### [*fname.*]segmented\_data[.*ext*].mat

### [*fname.*]ts[.*ext*].mat