Up
<
>
Expand All
Collapse All
Set Auto Width
Print

SNP6 Copy number analysis (GISTIC2)

- Overview

- Introduction

GISTIC identifies genomic regions that are significantly gained or lost across a set of tumors. The pipeline first filters out normal samples from the segmented copy-number data by inspecting the TCGA barcodes and then executes GISTIC version 2.0.21 (Firehose task version: 127).

- Summary

There were 569 tumor samples used in this analysis: 32 significant arm-level results, 32 significant focal amplifications, and 36 significant focal deletions were found.

- Results

- Focal results

Figure 1. Genomic positions of amplified regions: the X-axis represents the normalized amplification signals (top) and significance by Q value (bottom). The green line represents the significance cutoff at Q value=0.25.

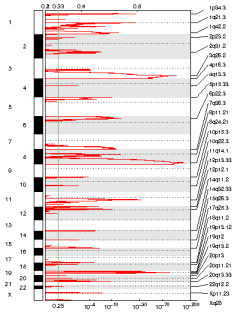


Table 1. Get Full Table Amplifications Table - 32 significant amplifications found. Click the link in the last column to view a comprehensive list of candidate genes. If no genes were identified within the peak, the nearest gene appears in brackets.

Cytoband	Q value	Residual Q value	Wide Peak Boundaries	# Genes in Wide Peak
8q24.21	6.8764e-153	6.8764e-153	chr8:128494482-129683450	11
3q26.2	6.1982e-119	6.1982e-119	chr3:168670994-168783199	0 [MECOM]
19q12	2.063e-93	3.53e-69	chr19:30306758-30331242	1
19p13.12	2.9546e-38	4.4985e-38	chr19:15333311-15422141	2
11q14.1	1.1081e-31	1.1081e-31	chr11:77620713-78035831	9
1p34.3	6.4195e-27	6.4195e-27	chr1:39887948-40168864	8
1q21.3	3.679e-26	4.0231e-18	chr1:150483517-150739128	9
6p22.3	9.9439e-16	9.9439e-16	chr6:18423406-18698180	3
7q36.3	6.2858e-13	6.2858e-13	chr7:145626981-159138663	110
5p15.33	6.3359e-11	6.3359e-11	chr5:1-1428010	26
20q13.33	2.1372e-10	9.7995e-10	chr20:62137482-63025520	41
1q42.2	1.2932e-13	3.8729e-09	chr1:234417530-235727932	17
Xp11.23	1.2878e-07	1.2878e-07	chrX:48532582-49592201	50
2q31.2	2.5197e-07	2.5197e-07	chr2:178462915-178609187	2
15q26.3	7.8057e-07	7.8057e-07	chr15:97998762-102531392	35
12p12.1	2.6134e-11	1.6879e-06	chr12:24880663-25881603	7
Xq28	1.6879e-06	1.6879e-06	chrX:152900017-154883511	77

Cytoband	Q value	Residual Q value	Wide Peak Boundaries	# Genes in Wide Peak
4p16.3	1.2475e-05	1.2475e-05	chr4:1653353-1991609	11
17q25.3	1.2757e-05	1.2757e-05	chr17:77641568-78054101	7
10p15.3	0.00018411	0.00018411	chr10:749239-1314547	8
14q11.2	6.4202e-06	0.00061977	chr14:21338460-21606746	13
8p11.21	0.00068231	0.00068231	chr8:41567872-42008185	2
12p13.33	1.9938e-06	0.00072368	chr12:1-3556349	35
19q13.2	7.2612e-27	0.0045813	chr19:39311157-40342966	41
10q22.3	0.0050822	0.0050822	chr10:79307097-79800033	5
18q11.2	0.0071339	0.0071339	chr18:23773718-24293308	3
4q13.3	0.0072495	0.0072495	chr4:73122319-74300609	4
2p23.2	0.011986	0.011986	chr2:28811078-28966789	1
20p13	0.082808	0.082808	chr20:1-4110252	89
22q12.2	0.12105	0.12105	chr22:30113489-30813423	15
20q11.21	0.027771	0.17383	chr20:30061713-30332464	10
14q32.33	0.0051711	0.18806	chr14:105216303-107349540	26

Figure 2. Genomic positions of deleted regions: the X-axis represents the normalized deletion signals (top) and significance by Q value (bottom). The green line represents the significance cutoff at Q value=0.25.

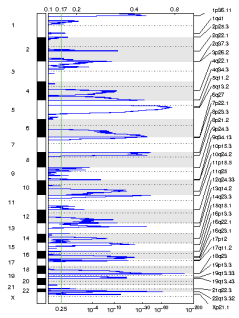


Table 2. Get Full Table Deletions Table - 36 significant deletions found. Click the link in the last column to view a comprehensive list of candidate genes. If no genes were identified within the peak, the nearest gene appears in brackets.

Cytoband	Q value	Residual Q value	Wide Peak Boundaries	# Genes in Wide Peak
19p13.3	2.9235e-140	2.9235e-140	chr19:1272041-1882761	24
22q13.32	8.8866e-64	8.8866e-64	chr22:49146971-51304566	42
11p15.5	1.3243e-41	1.3541e-41	chr11:502218-772981	17
6q27	2.5611e-36	2.5671e-36	chr6:156266097-171115067	92
13q14.2	2.4157e-35	2.4106e-35	chr13:48833767-49064807	2
1p36.11	8.5799e-35	8.5799e-35	chr1:26795113-27570286	16
5q13.2	4.7787e-89	4.026e-31	chr5:66492413-80256908	96
18q23	2.175e-29	2.1623e-29	chr18:67995567-78077248	36
5q11.2	2.4885e-75	1.4221e-26	chr5:58145167-59787985	3
7p22.1	3.069e-25	3.069e-25	chr7:1-6733205	81
4q34.3	5.6969e-33	8.3475e-25	chr4:178911874-183060693	1
15q15.1	6.945e-23	6.945e-23	chr15:41795901-42068054	5
8p23.3	1.4447e-41	2.3276e-19	chr8:1-1244373	6
17q11.2	3.7956e-18	5.6793e-18	chr17:29326736-29722618	5
10q24.2	5.0906e-17	5.0997e-17	chr10:88859357-111624768	239
8p21.2	5.3517e-36	9.5306e-15	chr8:25896447-26250295	1
Xp21.1	1.0638e-13	1.0638e-13	chrX:30865118-34644819	4
16q23.1	8.9636e-23	1.2204e-12	chr16:78016120-79627770	2

Cytoband	Q value	Residual Q value	Wide Peak Boundaries	# Genes in Wide Peak
12q24.33	1.0568e-11	1.0281e-11	chr12:131692956-133851895	27
2q22.1	4.0675e-12	1.4488e-09	chr2:139655617-143637838	1
16p13.3	4.2148e-09	4.1204e-09	chr16:3764793-4013466	1
14q23.3	2.6714e-08	2.6714e-08	chr14:59114448-73705982	105
3p26.2	6.1181e-08	6.1138e-08	chr3:1-4396118	8
9q34.13	1.9449e-06	1.9721e-06	chr9:131933735-140151408	181
4q22.1	8.6219e-12	4.7702e-06	chr4:90844993-93240505	1
19q13.43	4.2607e-17	3.4147e-05	chr19:58368882-59128983	35
2q37.3	8.1082e-09	0.0001243	chr2:234985548-243199373	82
19q13.33	6.0419e-17	0.00023297	chr19:46080683-49564123	151
9p24.3	0.00034178	0.00034501	chr9:1-12693402	48
16q22.1	5.695e-17	0.00087639	chr16:68383494-68879824	4
11q25	0.0019841	0.0019797	chr11:120006706-135006516	126
17p12	0.00397	0.0054218	chr17:11896630-12456081	3
21q22.3	0.018397	0.018561	chr21:42519390-48129895	103
10p15.3	0.076817	0.076711	chr10:1-1040137	5
2p25.3	0.098506	0.098506	chr2:1-25969186	126
1q41	0.14714	0.14644	chr1:207316102-220440260	83

- Arm-level results

Table 3. Get Full Table Arm-level significance table - 32 significant results found. The significance cutoff is at Q value=0.25.

Arm	# Genes	Amp Frequency	Amp Z score	Amp Q value	Del Frequency	Del Z score	Del Q value
1p	2121	0.32	2.38	0.0387	0.16	-5.04	1
1q	1955	0.40	5.63	8.88e-08	0.14	-6.13	1
2p	924	0.33	-1.74	1	0.16	-8.56	1
2q	1556	0.28	-1.78	1	0.16	-6.96	1
3p	1062	0.30	-2.44	1	0.26	-4.26	1
3q	1139	0.49	6.76	1.42e-10	0.18	-6.35	1
4p	489	0.20	-6.57	1	0.62	9.99	0
4q	1049	0.15	-6.44	1	0.66	14.8	0
5p	270	0.38	-1.5	1	0.42	-0.00973	0.806
5q	1427	0.15	-6.93	1	0.44	5.33	1.18e-07
6p	1173	0.39	1.59	0.185	0.43	3.55	0.000373
6q	839	0.27	-3.71	1	0.51	6.36	2.74e-10
7p	641	0.39	-0.106	1	0.32	-2.68	1
7q	1277	0.40	2.93	0.00836	0.22	-4.67	1
8p	580	0.31	-2.75	1	0.63	10.7	0
8q	859	0.48	4.66	1.26e-05	0.30	-2.7	1
9p	422	0.24	-5.74	1	0.55	6.64	4.34e-11
9q	1113	0.15	-7.26	1	0.54	8.9	0
10p	409	0.35	-2.26	1	0.29	-4.88	1
10q	1268	0.23	-4.73	1	0.27	-3.06	1
11p	862	0.18	-7.57	1	0.40	1.49	0.124
11q	1515	0.25	-2.97	1	0.33	0.434	0.554
12p	575	0.49	4.35	4.6e-05	0.28	-4.17	1
12q	1447	0.36	1.78	0.15	0.28	-2.12	1
13q	654	0.22	-5.61	1	0.60	9.7	0
14q	1341	0.16	-7.15	1	0.41	3.62	0.000295
15q	1355	0.13	-7.35	1	0.52	9.23	0

Arm	# Genes	Amp Frequency	Amp Z score	Amp Q value	Del Frequency	Del Z score	Del Q value
16p	872	0.21	-5.15	1	0.64	12.8	**0**
16q	702	0.18	-5.46	1	0.75	17.5	**0**
17p	683	0.18	-3.96	1	0.86	23.1	**0**
17q	1592	0.23	-2.37	1	0.73	20.3	**0**
18p	143	0.31	-4.23	1	0.54	4.79	**1.71e-06**
18q	446	0.24	-5.51	1	0.59	8.34	**0**
19p	995	0.40	1.67	**0.174**	0.47	4.81	**1.68e-06**
19q	1709	0.38	3.01	**0.00739**	0.44	6.21	**6.67e-10**
20p	355	0.54	5.79	**4.61e-08**	0.19	-7.78	1
20q	753	0.59	9.94	**0**	0.15	-7.63	1
21q	509	0.29	-4.39	1	0.43	1.2	**0.202**
22q	921	0.17	-5	1	0.77	19.7	**0**
Xq	1312	0.30	-1.3	1	0.61	12.6	**0**

- Methods & Data

- Input

Description

- Segmentation File: The segmentation file contains the segmented data for all the samples identified by GLAD, CBS, or some other segmentation algorithm. (See GLAD file format in the Genepattern file formats documentation.) It is a six column, tab-delimited file with an optional first line identifying the columns. Positions are in base pair units.The column headers are: (1) Sample (sample name), (2) Chromosome (chromosome number), (3) Start Position (segment start position, in bases), (4) End Position (segment end position, in bases), (5) Num markers (number of markers in segment), (6) Seg.CN (log2() -1 of copy number).
- Markers File: The markers file identifies the marker names and positions of the markers in the original dataset (before segmentation). It is a three column, tab-delimited file with an optional header. The column headers are: (1) Marker Name, (2) Chromosome, (3) Marker Position (in bases).
- Reference Genome: The reference genome file contains information about the location of genes and cytobands on a given build of the genome. Reference genome files are created in Matlab and are not viewable with a text editor.
- CNV Files: There are two options for the cnv file. The first option allows CNVs to be identified by marker name. The second option allows the CNVs to be identified by genomic location. Option #1: A two column, tab-delimited file with an optional header row. The marker names given in this file must match the marker names given in the markers file. The CNV identifiers are for user use and can be arbitrary. The column headers are: (1) Marker Name, (2) CNV Identifier. Option #2: A 6 column, tab-delimited file with an optional header row. The 'CNV Identifier' is for user use and can be arbitrary. 'Narrow Region Start' and 'Narrow Region End' are also not used. The column headers are: (1) CNV Identifier, (2) Chromosome, (3) Narrow Region Start, (4) Narrow Region End, (5) Wide Region Start, (6) Wide Region End
- Amplification Threshold: Threshold for copy number amplifications. Regions with a log2 ratio above this value are considered amplified.
- Deletion Threshold: Threshold for copy number deletions. Regions with a log2 ratio below the negative of this value are considered deletions.
- Cap Values: Minimum and maximum cap values on analyzed data. Regions with a log2 ratio greater than the cap are set to the cap value; regions with a log2 ratio less than -cap value are set to -cap. Values must be positive.
- Broad Length Cutoff: Threshold used to distinguish broad from focal events, given in units of fraction of chromosome arm.
- Remove X-Chromosome: Flag indicating whether to remove data from the X-chromosome before analysis. Allowed values= {1,0} (1: Remove X-Chromosome, 0: Do not remove X-Chromosome.
- Confidence Level: Confidence level used to calculate the region containing a driver.
- Join Segment Size: Smallest number of markers to allow in segments from the segmented data. Segments that contain fewer than this number of markers are joined to the neighboring segment that is closest in copy number.
- Arm Level Peel Off: Flag set to enable arm-level peel-off of events during peak definition. The arm-level

peel-off enhancement to the arbitrated peel-off method assigns all events in the same chromosome arm of the same sample to a single peak. It is useful when peaks are split by noise or chromothripsis. Allowed values= {1,0} (1: Use arm level peel off, 0: Use normal arbitrated peel-off).

- Maximum Sample Segments: Maximum number of segments allowed for a sample in the input data. Samples with more segments than this threshold are excluded from the analysis.
- Gene GISTIC: When enabled (value = 1), this option causes GISTIC to analyze deletions using genes instead of array markers to locate the lesion. In this mode, the copy number assigned to a gene is the lowest copy number among the markers that represent the gene.

Values

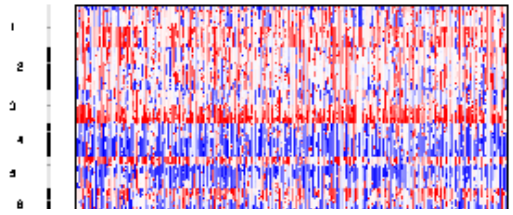
List of inputs used for this run of GISTIC2. All files listed should be included in the archived results.

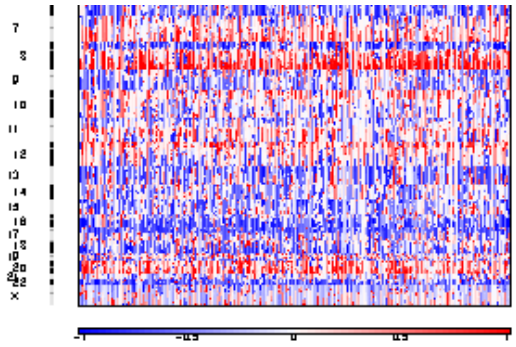
- *Segmentation File* = `/xchip/cga/gdac-prod/tcga-gdac/jobResults/GDAC_MergeDataFilesPipeline/OV-TP/9826235/GDAC_MergeDataFiles_3126532/OV-TP.snp__genome_wide_snp_6__broad_mit_edu__Level_3__segmented_scna_minus_germline_cnv_hg19__seg.seg.txt`
- *Markers File* = `/xchip/cga/reference/gistic2/genome.info.6.0_hg19.na31_minus_frequent_nan_probes_sorted_2.1.txt`
- *Reference Genome* = `/xchip/cga/reference/gistic2/hg19_with_miR_20120227.mat`
- *CNV Files* = `/xchip/cga/reference/gistic2/CNV.hg19.bypos.111213.txt`
- *Amplification Threshold = 0.1*
- *Deletion Threshold = 0.1*
- *Cap Values = 1.5*
- *Broad Length Cutoff = 0.7*
- *Remove X-Chromosome = 0*
- *Confidence Level = 0.99*
- *Join Segment Size = 4*
- *Arm Level Peel Off = 1*
- *Maximum Sample Segments = 2000*
- *Gene GISTIC = 1*

Table 4. Get Full Table First 10 out of 569 Input Tumor Samples.

Tumor Sample Names
TCGA-04-1331-01A-01D-0428-01
TCGA-04-1332-01A-01D-0428-01
TCGA-04-1335-01A-01D-0428-01
TCGA-04-1336-01A-01D-0428-01
TCGA-04-1337-01A-01D-0428-01
TCGA-04-1338-01A-01D-0428-01
TCGA-04-1341-01A-01D-0428-01
TCGA-04-1342-01A-01D-0428-01
TCGA-04-1343-01A-01D-0428-01
TCGA-04-1346-01A-01D-0428-01

Figure 3. Segmented copy number profiles in the input data





- Output

All Lesions File (all_lesions.conf_##.txt, where ## is the confidence level)

The all lesions file summarizes the results from the GISTIC run. It contains data about the significant regions of amplification and deletion as well as which samples are amplified or deleted in each of these regions. The identified regions are listed down the first column, and the samples are listed across the first row, starting in column 10.

Region Data

Columns 1-9 present the data about the significant regions as follows:

1. Unique Name: A name assigned to identify the region.
2. Descriptor: The genomic descriptor of that region.
3. Wide Peak Limits: The 'wide peak' boundaries most likely to contain the targeted genes. These are listed in genomic coordinates and marker (or probe) indices.
4. Peak Limits: The boundaries of the region of maximal amplification or deletion.
5. Region Limits: The boundaries of the entire significant region of amplification or deletion.
6. Q values: The Q value of the peak region.
7. Residual Q values: The Q value of the peak region after removing ('peeling off') amplifications or deletions that overlap other, more significant peak regions in the same chromosome.
8. Broad or Focal: Identifies whether the region reaches significance due primarily to broad events (called 'broad'), focal events (called 'focal'), or independently significant broad and focal events (called 'both').
9. Amplitude Threshold: Key giving the meaning of values in the subsequent columns associated with each sample.

Sample Data

Each of the analyzed samples is represented in one of the columns following the lesion data (columns 10 through end). The data contained in these columns varies slightly by section of the file. The first section can be identified by the key given in column 9 - it starts in row 2 and continues until the row that reads 'Actual Copy Change Given.' This section contains summarized data for each sample. A '0' indicates that the copy number of the sample was not amplified or deleted beyond the threshold amount in that peak region. A '1' indicates that the sample had low-level copy number aberrations (exceeding the low threshold indicated in column 9), and a '2' indicates that the sample had high-level copy number aberrations (exceeding the high threshold indicated in column 9).The second section can be identified the rows in which column 9 reads 'Actual Copy Change Given.' The second section exactly reproduces the first section, except that here the actual changes in copy number are provided rather than zeroes, ones, and twos.The final section is similar to the first section, except that here only broad events are included. A 1 in the samples columns (columns 10+) indicates that the median copy number of the sample across the entire significant region exceeded the threshold given in column 9. That is, it indicates whether the sample had a geographically extended event, rather than a focal amplification or deletion covering little more than the peak region.

Amplification Genes File (amp_genes.conf_##.txt, where ## is the confidence level)

The amp genes file contains one column for each amplification peak identified in the GISTIC analysis. The first four rows are:

1. Cytoband
2. Q value
3. Residual Q value

4. Wide Peak Boundaries

These rows identify the lesion in the same way as the all lesions file.The remaining rows list the genes contained in each wide peak. For peaks that contain no genes, the nearest gene is listed in brackets.

Deletion Genes File (del_genes.conf_##.txt, where ## is the confidence level)

The del genes file contains one column for each deletion peak identified in the GISTIC analysis. The file format for the del genes file is identical to the format for the amp genes file.

Gistic Scores File (scores.gistic)

The scores file lists the Q values [presented as -log10(q)], G scores, average amplitudes among aberrant samples, and frequency of aberration, across the genome for both amplifications and deletions. The scores file is viewable with the Genepattern SNPViewer module and may be imported into the Integrated Genomics Viewer (IGV).

Segmented Copy Number (raw_copy_number.{fig|pdf|png})

The segmented copy number is a pdf file containing a colormap image of the segmented copy number profiles in the input data.

Amplification Score GISTIC plot (amp_qplot.{fig|pdf|png|v2.pdf})

The amplification pdf is a plot of the G scores (top) and Q values (bottom) with respect to amplifications for all markers over the entire region analyzed.

Deletion Score GISTIC plot (del_qplot.{fig|pdf|png|v2.pdf})

The deletion pdf is a plot of the G scores (top) and Q values (bottom) with respect to deletions for all markers over the entire region analyzed.

Tables (table_{amp|del}.conf_##.txt, where ## is the confidence level)

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.

Broad Significance Results (broad_significance_results.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.

Broad Values By Arm (broad_values_by_arm.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.

All Data By Genes (all_data_by_genes.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.

Broad Data By Genes (broad_data_by_genes.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`.

Focal Data By Genes (focal_data_by_genes.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`.

All Thresholded By Genes (all_thresholded.by_genes.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.

Sample Cutoffs (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.txt` 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.

Focal Input To Gistic (focal_input.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.

Gene Counts vs. Copy Number Alteration Frequency (freqarms_vs_ngenes.{fig|pdf})

An image showing the correlation between gene counts and frequency of copy number alterations.

Confidence Intervals (regions_track.conf_##.bed, where ## is the confidence level)

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.

- GISTIC

GISTIC identifies genomic regions that are significantly gained or lost across a set of tumors. It takes segmented copy number ratios as input, separates arm-level events from focal events, and then performs two tests: (i) identifies significantly amplified/deleted chromosome arms; and (ii) identifies regions that are significantly focally amplified or deleted. For the focal analysis, the significance levels (Q values) are calculated by comparing the observed gains/losses at each locus to those obtained by randomly permuting the events along the genome to reflect the null hypothesis that they are all 'passengers' and could have occurred anywhere. The locus-specific significance levels are then corrected for multiple hypothesis testing. The arm-level significance is calculated by comparing the frequency of gains/losses of each arm to the expected rate given its size. The method outputs genomic views of significantly amplified and deleted regions, as well as a table of genes with gain or loss scores. A more in depth discussion of the GISTIC algorithm and its utility is given in [1], [3], and [5].

CNV Description

Regions of the genome that are prone to germ line variations in copy number are excluded from the GISTIC analysis using a list of germ line copy number variations (CNVs). A CNV is a DNA sequence that may be found at different copy numbers in the germ line of two different individuals. Such germ line variations can confound a GISTIC analysis, which finds significant somatic copy number variations in cancer. A more in depth discussion is provided in [6]. GISTIC currently uses two CNV exclusion lists. One is based on the literature describing copy number variation, and a second one comes from an analysis of significant variations among the blood normals in the TCGA data set.

- References

[1] Beroukhim et al, Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma, *Proc Natl Acad Sci U S A*. **Vol. 104**:50 (2007)

[2] GISTIC version 1

[3] Mermel et al, GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers, *Genome Biology* **Vol. 12**:4 (2011)

[4] GISTIC version 2

[5] Beroukhim et al., The landscape of somatic copy-number alteration across human cancers, *Nature* **Vol. 463**:7283 (2010)

[6] McCarroll, S. A. et al., Integrated detection and population-genetic analysis of SNPs and copy number variation, *Nat Genet* **Vol. 40**(10):1166-1174 (2008)

[7] The Sanger Institute: Cancer Gene Census

MADE WITH NOZZLE