BMC
Bioinformatics

SOFTWARE　　　　　　　　　　　　　　　　　　　　　**Open Access**

JISTIC: Identification of Significant Targets in Cancer

Felix Sanchez-Garcia*[1,2], Uri David Akavia[2,3], Eyal Mozes[2] and Dana Pe'er[2,3]

Abstract

Background: Cancer is caused through a multistep process, in which a succession of genetic changes, each conferring a competitive advantage for growth and proliferation, leads to the progressive conversion of normal human cells into malignant cancer cells. Interrogation of cancer genomes holds the promise of understanding this process, thus revolutionizing cancer research and treatment. As datasets measuring copy number aberrations in tumors accumulate, a major challenge has become to distinguish between those mutations that drive the cancer versus those passenger mutations that have no effect.

Results: We present JISTIC, a tool for analyzing datasets of genome-wide copy number variation to identify driver aberrations in cancer. JISTIC is an improvement over the widely used GISTIC algorithm. We compared the performance of JISTIC versus GISTIC on a dataset of glioblastoma copy number variation, JISTIC finds 173 significant regions, whereas GISTIC only finds 103 significant regions. Importantly, the additional regions detected by JISTIC are enriched for oncogenes and genes involved in cell-cycle and proliferation.

Conclusions: JISTIC is an easy-to-install platform independent implementation of GISTIC that outperforms the original algorithm detecting more relevant candidate genes and regions. The software and documentation are freely available and can be found at: http://www.c2b2.columbia.edu/danapeerlab/html/software.html

Background

A comprehensive study of the genomic alterations that occur in cancer is vital for understanding this disease. Technological advances have made it possible to detect chromosomal regions of amplifications and deletions genome-wide and at high resolution. Datasets measuring such aberrations in patient tumors are accumulating at a staggering rate for multiple types of cancer [1-4]

However, tumors harbor a great number of copy number alterations and it is difficult to distinguish between driver aberrations (functional changes vital for cancer progression) and passenger aberrations (random and with no selective advantage). Thus, the distinction between driver and passenger mutations has become one of the key challenges in cancer genomics. A very successful algorithm to address this is "Genomic Identification of Significant Targets in Cancer (GISTIC)" [1], that identifies aberrant regions more likely to drive cancer pathogenesis. GISTIC calculates the background rate of random chromosomal aberrations and identifies those regions that are aberrant more often than would be expected by chance, with greater weight given to high amplitude events that are less likely to represent random aberrations. There are other algorithms that tackle this task such as GLAD [5], RAE [6] and STAC [6]. However, GISTIC is unique in its ability to combine magnitude and frequency of the alteration into a statistical score. This algorithm has been successfully applied to various datasets [2,4,7,8] and the approach is becoming widespread.

GISTIC identifies those regions of the genome that are aberrant more often than would be expected by chance. While successful in most scenarios, GISTIC has trouble identifying the relevant sub-region when a very large region is amplified or deleted. Such large chromosomal aberrations frequently occur in cancer and this leaves the user with two less than optimal options - consider only a single peak within the region, or consider an entire chromosomal arm. However, we have observed that in many cases there are other small sub-regions for which the aberration is significantly stronger than in the rest of the

* Correspondence: fs2282@columbia.edu

[1] Department of Computer Science, Columbia University, New York, NY, USA
Full list of author information is available at the end of the article

large region. Moreover, these regions often contain known oncogenes.

To address this issue, we developed JISTIC, a tool that implements all of GISTIC's capabilities, with an additional new variant of the algorithm capable of detecting multiple significant sub-regions within large aberrant regions.

Implementation

JISTIC is based on the GISTIC algorithm [1]. JISTIC implements the previously published variants of GISTIC (standard, focal and arm-peel-off) and can also deal with LOH in the same way that the original algorithm does. More detailed information on GISTIC can be found in the Supporting Information of [1].

In brief, GISTIC calculates a statistic (G-score) which represents the strength of the aberration for each marker. The G-score for a marker m is the summation of the copy number across all samples. For this summation, the samples in which the copy number is less significant than an empirical aberration threshold (Θ^{AMP} for amplification, Θ^{DEL} for deletion) is set to zero. Therefore, the G-score in the case of amplifications is:

$$G^{AMP}(m) = \sum_i CN(m,i) \times I(CN(m,i) > \Theta^{AMP})$$

Where I(x) is the indicator function and CN(m, i) the copy number for marker m and sample i. The score is defined similarly for deletions, taking into account the change in sign. While standard GISTIC uses a fixed aberration threshold for each type of aberration, focal GISTIC uses sample-specific high-level thresholds. This threshold is set for each sample by adding the standard threshold to the maximum (minimum for deletions) of medians observed for each chromosome arm.

GISTIC assesses the significance of the G-score for each marker by comparing this score to results expected by chance using genome-wide permutation testing. This significance is then corrected using False Discovery Rate (FDR) using Benjamini and Hochberg method [9], and a q-value is obtained. All regions with a q-value below a threshold (0.25 in previously published articles) are deemed significant. For large aberrations, the sub-region with a minimal q-value is identified as a peak driver region.

To identify independent peaks within a region and discard spurious peaks, GISTIC uses a peel-off algorithm. The algorithm iteratively finds the most significant peak and then, for each sample, if the peak is included in the aberration, it sets to zero all markers consecutive with the peak, thus removing all aberrations overlapping the peak. Once a peak has been detected, the peak region is extended by leaving each sample out in turn, and recalculating the peak boundaries.
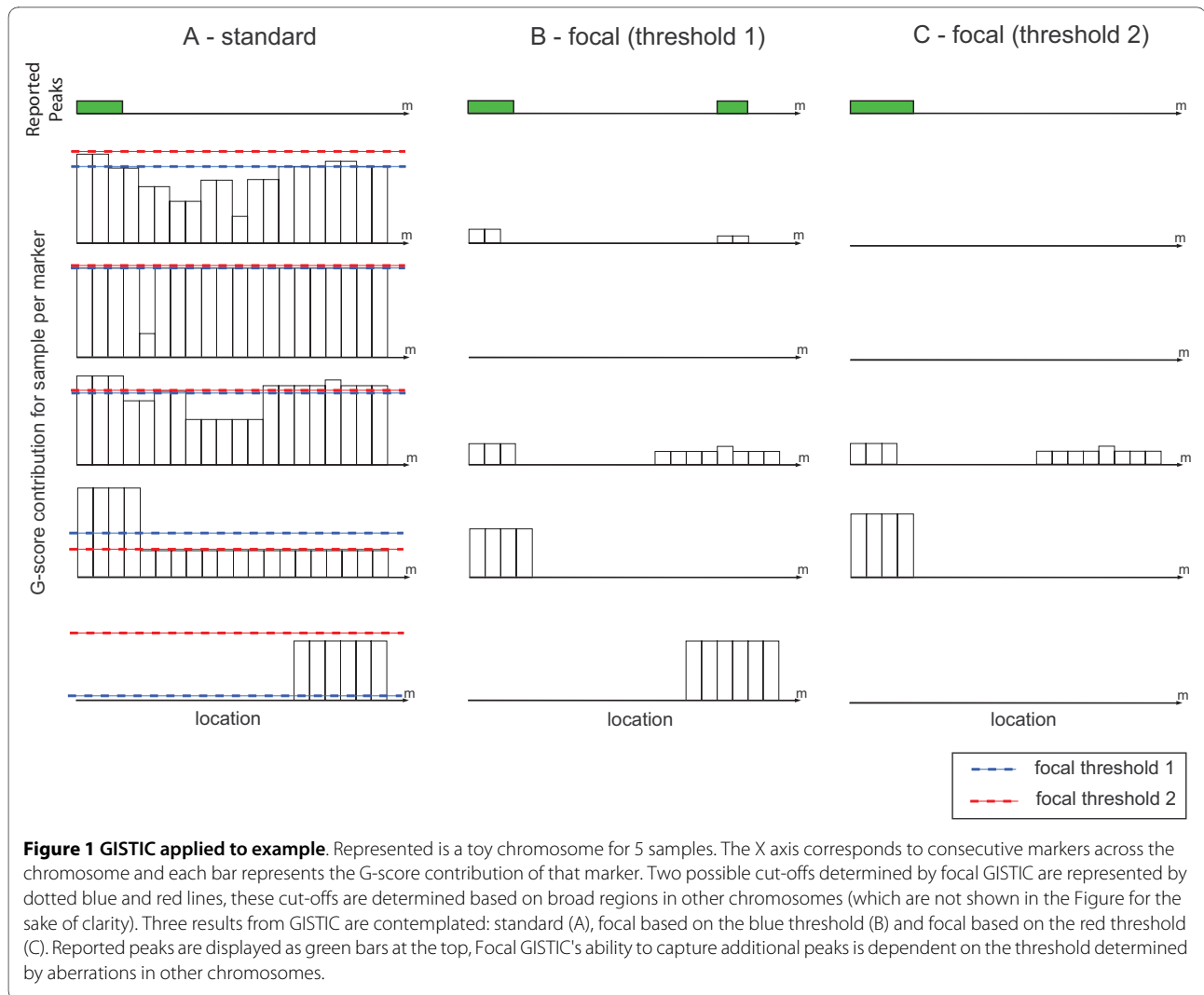
Typical to cancer genomes is the presence of broad copy number aberrations, defined as aberrant regions at least as large as one half of a chromosomal arm [1]. Peel-off on broad regions using standard GISTIC usually results in identifying only a single peak. GISTIC also has a focal variant of the algorithm which can potentially capture more aberrations, but misses many peaks in practice.

The crux of the matter is to distinguish between genuine multiple peaks and a single peak within noisy fluctuations of microarray intensities. Focal peel-off is designed to deal with this issue, but since the focal threshold is defined according to the highest broad aberration genome-wide, this variant not only requires the aberration to have a strong focal behavior, but also depends on the strength of other broad aberrations across the genome. In Figure 1 we demonstrate how different thresholds (determined by broad aberrations in other chromosomes not shown in the figure) can lead to either detection or failure to detect a second peak. In this example, standard GISTIC only captures a single peak and misses a second peak that seems equally significant, with just a slight difference in G-score. In this type of situations standard GISTIC is doomed to fail. The detection of the second peak with focal GISTIC highly depends on the focal threshold. This threshold is set for each sample by the highest broad aberration genome-wide and there are cases, such as C in Figure 1, in which the focal events that define the aberration are masked by broad aberrations in other chromosome (not shown in Figure). In the next section we will see that this problem is prevalent in real datasets.

For JISTIC, we developed a new variant called limited peel-off, designed to tackle the problem posed in Figure 1. While GISTIC sets to zero any aberration containing the peak, our algorithm "peels off" only part of the aberration. Limited peel-off uses the global G-score to determine the extent of "peel-off. Given a single aberration, we expect G(m) would decrease as we get further from it until it reaches the noise level. The idea behind limited peel-off is to decompose the G-score for a marker in two parts, one that represents what remains from the peak that we are peeling off ($G_r(m)$) and another that depicts contribution independent of the peak ($G_n(m)$). We use a threshold value (G_{thres}) to estimate whether $G_n(m)$ contains only noise or the aberration is likely due to an independent peak.

Denote **G(m, i)** as the G-score contribution of sample i in marker m, which can be defined as:

$$G(m,i) = CN(m,i) \times I(CN(m,i) > \Theta^{AMP})$$

Figure 1 GISTIC applied to example. Represented is a toy chromosome for 5 samples. The X axis corresponds to consecutive markers across the chromosome and each bar represents the G-score contribution of that marker. Two possible cut-offs determined by focal GISTIC are represented by dotted blue and red lines, these cut-offs are determined based on broad regions in other chromosomes (which are not shown in the Figure for the sake of clarity). Three results from GISTIC are contemplated: standard (A), focal based on the blue threshold (B) and focal based on the red threshold (C). Reported peaks are displayed as green bars at the top, Focal GISTIC's ability to capture additional peaks is dependent on the threshold determined by aberrations in other chromosomes.

For each sample i, the algorithm iteratively calculates on both sides of a newly discovered peak the magnitude **$G_r(m, i)$**. For the right side of the peak:

$$G_r(m, i) = \min(G_r(m-1, i), \max_{k \in (0, s)}(G(m+k, i)))$$

the left side of the peak is the symmetric equation:

$$G_r(m, i) = \min(G_r(m+1, i), \max_{k \in (0, s)}(G(m-k, i)))$$

G_r(m, i) represents the amount of aberration that remains from the primary peak in marker m. JISTIC considers any reduction in the aberration that is consistent for at least a minimal number of markers, s. This parameter was introduced after observing that transient fluctuations in copy number result in a greater number of false positives. For s = 0 a total of 147 peaks are detected with limited peel-off. However, many of those peaks are located next to each other in the genome and the biologi-cal assessment suggested that those were spurious peaks. The number of reported peaks decreases when the parameter s increases, eliminating most of those spurious peaks.

In order to determine whether a new peak exists, we need the complementary component of G_r(m, i), G_n(m, i), measuring the component of the aberration that is independent from the primary peak, for each sample:

$$G_n(m, i) = \max(G(m, i) - (G_r(m, i), 0)$$

The condition that JISTIC checks in order to abort the peel-off at a marker m is

$$\sum_i G_n(m, i) = G_n(m) > G_{thres}$$

The results critically depend on the threshold G_{thres}. If G_{thres} is too high, the algorithm will obtain exactly the same result than the standard variant of GISTIC. If the

value is too low, the algorithm will obtain many spurious peaks. For consistency, we use the G-score value that corresponds to the threshold q-value when running the focal variant, thus using the same threshold both for a peak on baseline copy number and a sub-peak within a broader region.

Figure 2 shows how limited peel-off successfully detects the second peak in the example from Figure 1. Contrary to the focal variant, this success is independent of broad regions in other chromosomes. It is important to note that limited peel-off fully replaces the standard variant, as any peak captured by the latter, will also be captured by the former. On the other hand, focal and limited peel-off can be considered complementary. While limited peel-off looks at the average of the aberration across samples, focal concentrates on the most acute aberrations for each sample.
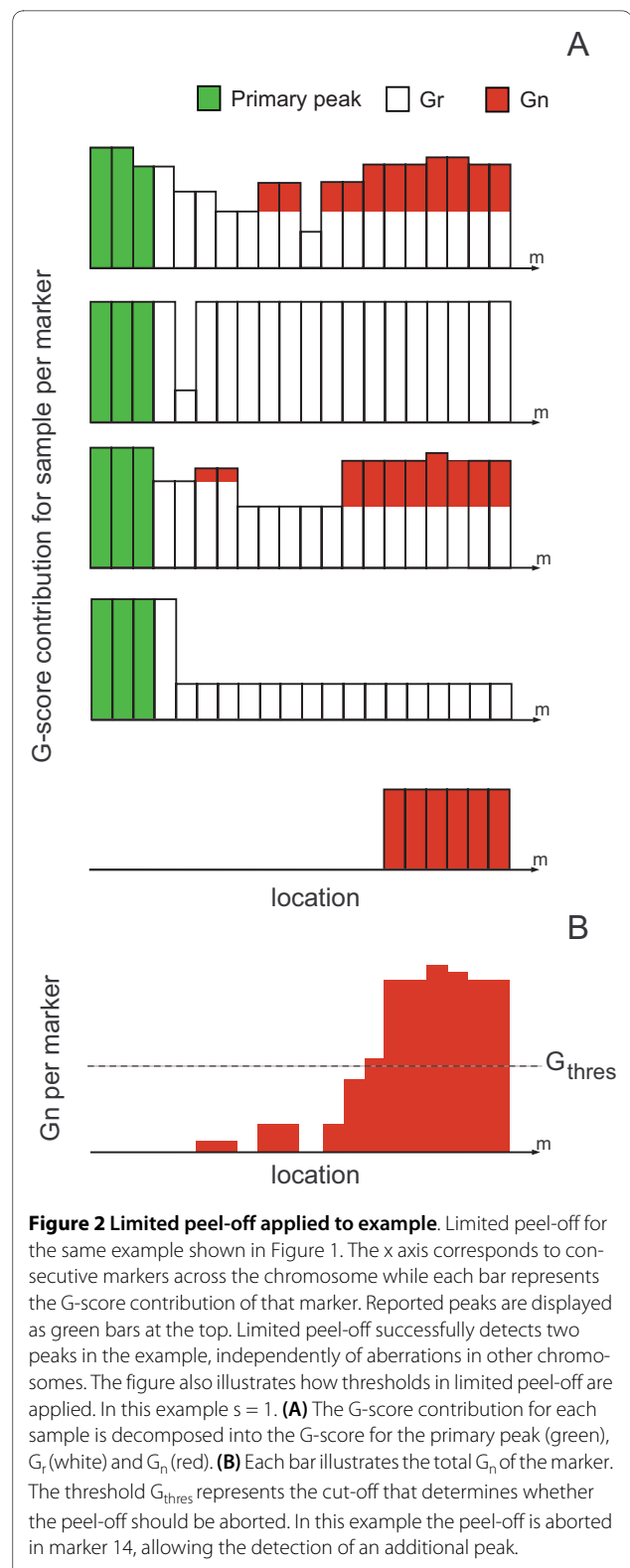
Results

We tested JISTIC on 178 glioblastoma samples that had been previously used to test the latest version of GISTIC [10]. This dataset was generated by Harvard Medical School using Agilent 244 K Arrays.

The parameter s was tuned using the histogram of segments sizes to obtain candidate values and the optimal value was chosen based on the distribution of minimum distance between peaks. In the test dataset, the histogram of segment size showed that the number of segments decreases with the segment size, as it was expected. We observed a stronger decrease in the number of segments for 3 different segment sizes (3, 8 and 12); those sizes were considered as candidate values for s. The distribution of minimum distance between peaks was used to estimate the number of spurious peaks for each of those values. This distribution shows a considerable change when s was increased to 8, as the number of adjacent peaks decreased. The distribution did not show any significant change when s was increased to 12 and we adopted a conservative approach by setting s = 10, which corresponds to 126.87 kb on average. Limited peel-off detected 124 peaks using s = 10.

Table 1 shows the peaks and genes obtained with JISTIC in the three variants: standard, focal and limited peel-off. Note that standard and focal are equivalent to implementations of the GISTIC algorithm [10]. When looking at peaks by chromosome (Table 2), we observe that most of the novel peaks appear in chromosomes with a broad region for which standard detects only a single peak and focal detects one or no peaks.

Figure 3 demonstrates an example of such a broad peak in chromosome 19. Standard (A) captures only a single peak and focal (B) fails to capture any aberration. On the other hand, limited peel-off (C) successfully captures multiple relevant regions. The ultimate assessment is

Figure 2 Limited peel-off applied to example. Limited peel-off for the same example shown in Figure 1. The x axis corresponds to consecutive markers across the chromosome while each bar represents the G-score contribution of that marker. Reported peaks are displayed as green bars at the top. Limited peel-off successfully detects two peaks in the example, independently of aberrations in other chromosomes. The figure also illustrates how thresholds in limited peel-off are applied. In this example s = 1. **(A)** The G-score contribution for each sample is decomposed into the G-score for the primary peak (green), G_r (white) and G_n (red). **(B)** Each bar illustrates the total G_n of the marker. The threshold G_{thres} represents the cut-off that determines whether the peel-off should be aborted. In this example the peel-off is aborted in marker 14, allowing the detection of an additional peak.

whether the additional peaks identified are indeed drivers. To evaluate this we assessed whether the new peaks in chromosome 19 contained solid candidates for oncogenes. Standard GISTIC identified a single peak contain-

Table 1: Summary of peaks and genes for the three variants.

	Amplification		Deletion	
	Peaks	Genes	Peaks	Genes
Standard	14	35	12	59
Focal	74	202	21	53
Standard+ Focal	75	204	28	99
Limited peel-off	64	272	60	205
Novel by limited peel-off	29	186	41	136

For each variant and type of aberration, the number of peak regions and genes contained within them are illustrated. The table also summarizes results for all GISTIC variants (non-overlapping peaks obtained by standard and focal GISTIC) and novel peak regions reported by limited peel-off. Limited peel-off detects 70 novel peaks and 322 genes.

ing the gene (ZNF626), whose role is unknown. In comparison, limited peel-off detected 7 additional regions, where each includes at least one oncogene and multiple genes related to cell replication. Examples of oncogenes include AP1 M2 [11], Cyclin E1 [12], LYPD3 [13], see Table 3 for list of potential oncogenes in each peak.

Figure 4 shows another example, a deletion of the entire chromosome 22. Standard finds a single peak and focal finds no peaks, whereas limited peel-off finds 3 additional peaks. The peak identified by standard includes FBLN1, a gene that has been previously selected as candidate tumor suppressor [14]. Each additional peak identified by limited peek-off has at least one tumor suppressor, such

as SMARCB1 [15], HIRA [16] and LARGE [17], see Table 4 for list of potential tumor suppressors in each peak.

For a systematic evaluation we designed an automated statistical test to study the limited peel-off's specificity and establish that the increased number of peaks detected by limited peel-off does not increase the false positive rate. Our permutation based test estimates the number of candidate cancer driver genes expected in random regions matching in size with the regions detected by JISTIC. We compiled a list of 1880 likely driver genes based on two different sources:

- Genes listed in the Sanger Cancer Gene Census [18].
- Genes annotated in GO [19] for at least one of the following processes: cell cycle, cell proliferation, cell death and neurogenesis.

Table 2: Peaks by chromosome.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
AMP	**Standard**	2	0	1	4	0	0	**3**	0	0	0	0	2	0	0	0	0	0	0	**1**	**1**	0	0
	Focal	9	6	5	6	2	3	**17**	1	0	0	3	16	1	1	1	0	1	0	**0**	**1**	0	1
	Limited	6	0	2	4	0	0	**28**	0	0	0	0	13	0	0	0	0	0	0	**7**	**4**	0	0
DEL	**Standard**	1	0	0	0	1	**1**	0	0	**1**	**2**	1	0	**1**	**1**	1	0	0	0	1	0	0	**1**
	Focal	4	1	0	1	2	**1**	0	1	**7**	**1**	0	0	**1**	**1**	0	0	0	0	1	0	0	**0**
	Limited	5	0	0	0	1	**6**	0	0	**16**	**12**	1	0	**7**	**4**	3	0	0	0	1	0	0	**4**

Table 3: Novel amplification peaks in chrom. 19 reported by limited peel-off.

Location	Candidate driver genes
265456-781585 (19p13:3)	PTBP1 [21], BSG [22]
4105670-4353981 (19p13:3)	CHAF1A [23], SH3GL1 [24]
10539157-10628316 (19p13:2)	AP1 M2 [11], ILF3 [25]
34895742-35531840 (19q12)	Cyclin E1 [12]
48658163-48753325 (19q13:31)	LYPD3 [13]
61586503-63564951 (19q13:43)	BMZF3 [26]

For each novel peak region of amplification detected by limited peel-off in chromosome 19 its location in the genome and suspected driver genes are shown. Each driver gene is accompanied by a reference that supports its oncogenic nature. All novel peaks contain at least one candidate driver gene.

Table 4: Novel deletion peaks in chrom. 22 reported by limited peel-off.

Location	Candidate driver genes
17482013-17787269 (22q11:21)	HIRA [16]
22370186-22476593 (22q11:23)	SMARCB1 [15]
32114340-33392384 (22q12:3)	LARGE [17]

For each novel peak region of deletion detected by limited peel-off in chromosome 22 its location in the genome and suspected driver genes are shown. Each driver gene is accompanied by a reference that supports its behavior as tumor suppressor. All novel peaks contain at least one candidate driver gene.

We generated 10,000 random outputs with similar characteristics to the real output. A random output is generated by iterating over all detected peaks and shifting the peak's location to a random position in the genome. For each output, we count the number of regions containing at least one gene in the list of driver candidates above. This provides us with a distribution for the number of potential driver genes expected in the null model.

To compare focal and limited peel-off, we tested three sets of peaks: the peaks that are exclusive to each of the two methods and the peaks that are common to both. Note that all the peaks obtained by standard GISTIC are also obtained by limited peel-off. The results of the 10,000 iterations for each of the three sets are shown in Table 5.

As expected, the p-value obtained for the overlapping peaks is more significant than the p-value for the exclusive peaks for each method.

For the 78 peaks obtained exclusively by limited peel-off, 25 contained likely driver genes, only 23/10,000 permutations reached the value obtained by the true output, giving a p-value of 0.0023. In comparison, for exclusive peaks for the focal variant, 9 of the 49 peaks contained likely driver genes. In 73/10,000 permutations the test obtained as many peaks with likely driver genes, giving a p-value of 0.0073.
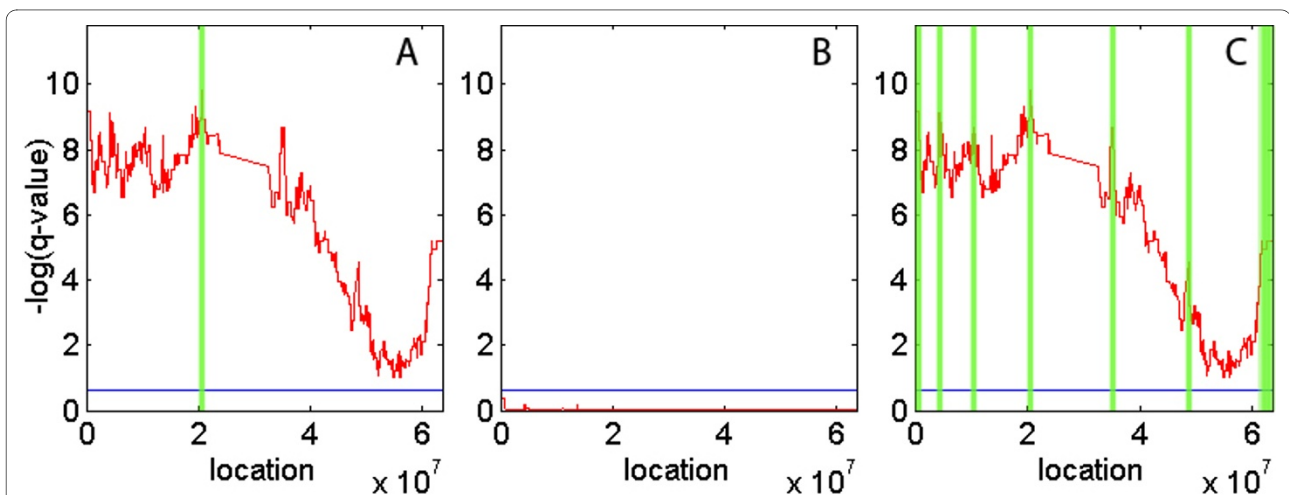

Figure 3 Amplification aberrations for chromosome 19. Represented are standard (A), focal (B) and limited peel-off (C) results for amplifications in the whole chromosome 19. The x axis corresponds to consecutive markers across the chromosome and the y axis corresponds to the q-value in logarithmic scale. The significance threshold for aberrations (q-value = 0.25) is represented by a blue line. Reported peak regions are illustrated in green. Limited peel-off detects 6 novel peaks.
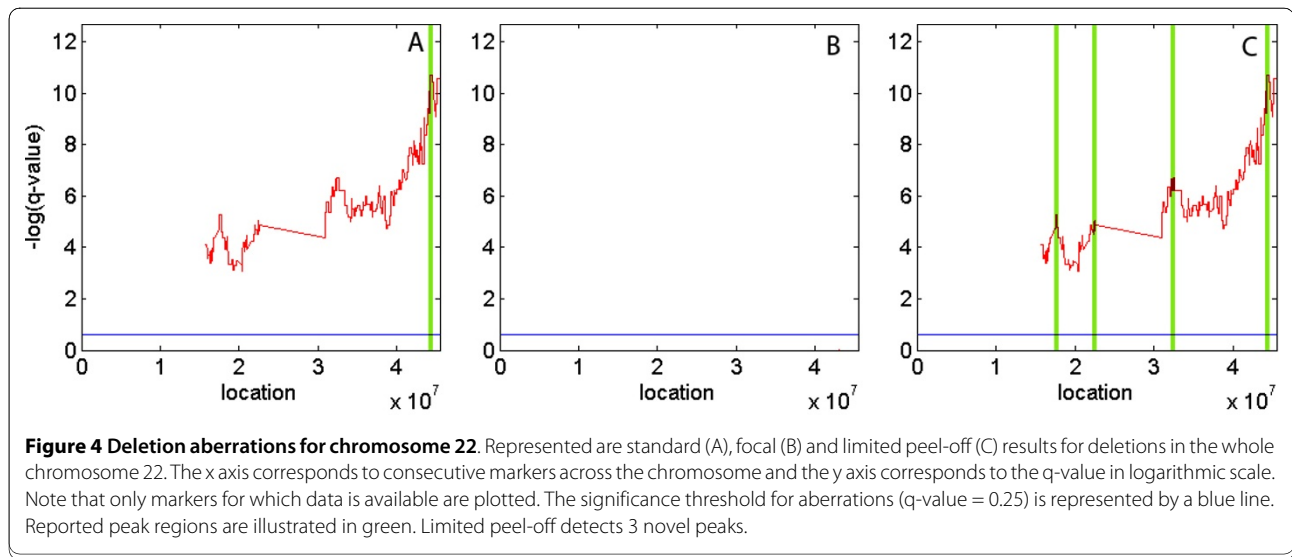
Figure 4 Deletion aberrations for chromosome 22. Represented are standard (A), focal (B) and limited peel-off (C) results for deletions in the whole chromosome 22. The x axis corresponds to consecutive markers across the chromosome and the y axis corresponds to the q-value in logarithmic scale. Note that only markers for which data is available are plotted. The significance threshold for aberrations (q-value = 0.25) is represented by a blue line. Reported peak regions are illustrated in green. Limited peel-off detects 3 novel peaks.

Table 5: Results for permutation test for focal and limited peel-off peaks.

	Peaks	Driver Peaks	P-value
Focal exclusive	49	9	7.3E-3
Limited peel-off exclusive	78	25	2.3E-3
Overlapping between both	46	14	3.5E-6

The p-value for the permutation test estimates the probability of detecting as many peaks containing candidate driver genes if random regions of the same size across genome are selected. The table shows results for mutually exclusive peaks between focal GISTIC and limited peel-off and for overlapping peaks between both variants.

Conclusions

The analysis performed on limited peel-off against standard and focal GISTIC demonstrates the superiority of limited peel-off to achieve both better specificity and dramatically increase recall by obtaining a large number of novel peaks.

In conclusion, JISTIC is a significantly improved algorithm to distinguish between driver and passenger copy number aberrations in cancer genomes. Importantly, it detects a significant number of additional driver regions while maintaining a similar false positive rate. We conclude that both limited peel-off and focal GISTIC should be used together as they provide complementary and significant results.

The tool is implemented in Java, has been tested on Linux and Windows. It can be downloaded from http://www.c2b2.columbia.edu/danapeerlab/html/software.html. It does not have dependencies to external libraries and can be downloaded as a single Java JAR file. The execution time for the glioblastoma dataset on a standard desktop computer (Intel Xeon W3505 @ 2.53 GHz, 3GB of RAM) was 8 minutes for all variants.
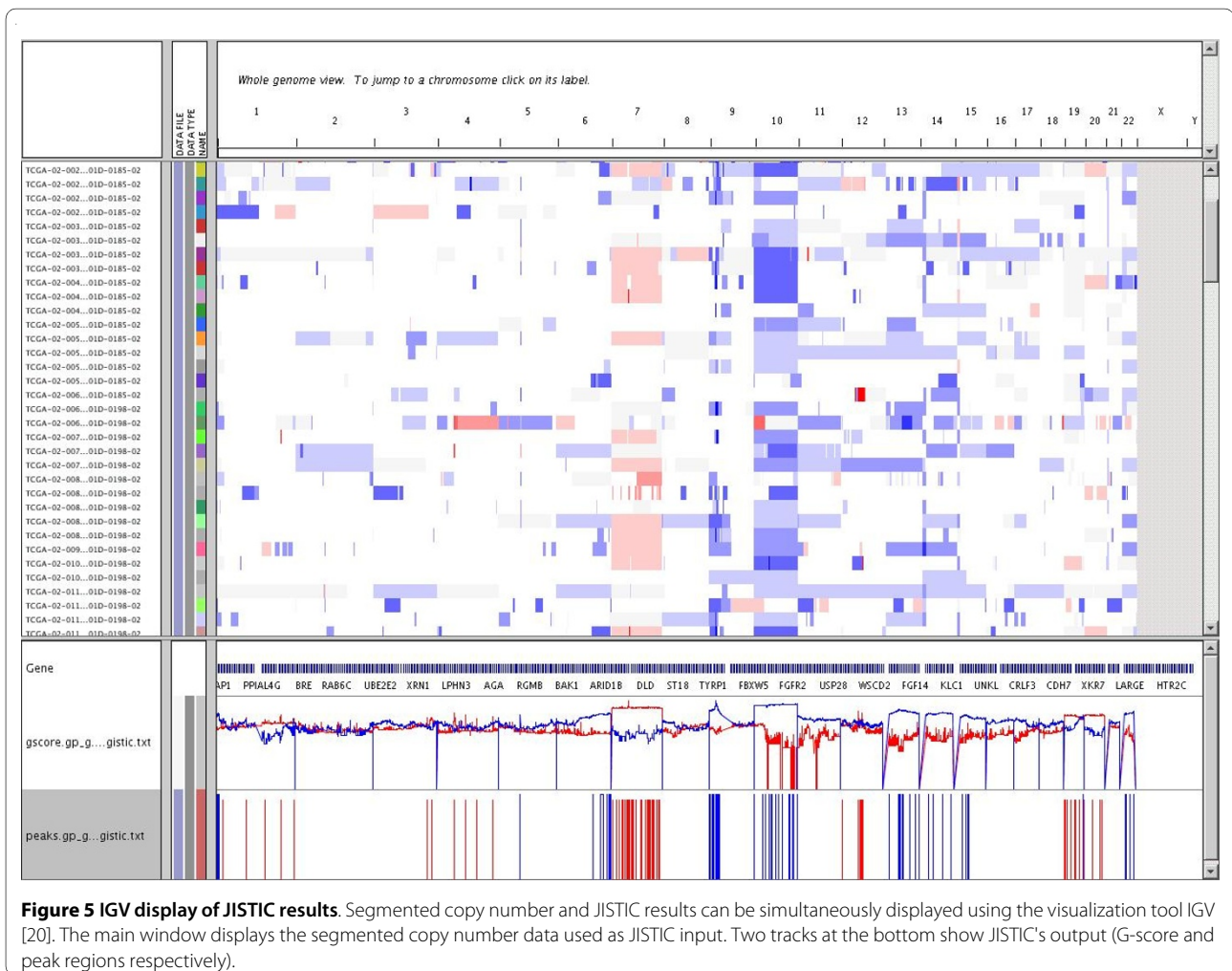
Matlab scripts are provided in order to visualize the output and obtain different statistics. The tool can also convert to the format used by the open-source visualization tool IGV [20], used to display cancer genomic data using a user-friendly interface, demonstrated in Figure 5.

Availability and Requirements

- **Project name**: JISTIC
- **Project home page**: http://www.c2b2.columbia.edu/danapeerlab/html/software.html
- **Operating system(s):** Platform independent
- **Programming language**: Java
- **Other requirements**: Java 1.5 or higher
- **License**: LGPL
- **Any restrictions to use by non-academics**: None

Conflict of interests

The authors declare that they have no competing interests.

Figure 5 IGV display of JISTIC results. Segmented copy number and JISTIC results can be simultaneously displayed using the visualization tool IGV [20]. The main window displays the segmented copy number data used as JISTIC input. Two tracks at the bottom show JISTIC's output (G-score and peak regions respectively).

Authors' contributions
DP, FSG and UDA designed research. FSG and UDA developed the limited-peel off algorithm EM and FSG implemented the JAVA code. FSG and UDA analyzed and evaluated the algorithm. FSG and DP wrote the manuscript. All authors read and approved the final manuscript.

Author Details
[1]Department of Computer Science, Columbia University, New York, NY, USA, [2]Department of Biological Sciences, Columbia University, New York, NY, USA and [3]Center for Computational Biology and Bioinformatics, Columbia University, New York, NY, USA

References
1. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, Vivanco I, Lee JC, Huang JH, Alexander S, *et al.*: **Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma.** *Proc Natl Acad Sci USA* 2007, **104(50):**20007-20012.
2. Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, Lin WM, Province MA, Kraja A, Johnson LA, *et al.*: **Characterizing the cancer genome in lung adenocarcinoma.** *Nature* 2007, **450:**893-898.
3. Pollack JR, Sørlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Børresen-Dale A-L, Brown PO: **Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99(20):**12963-12968.
4. Northcott PA, Nakahara Y, Wu X, Feuk L, Ellison DW, Croul S, Mack S, Kongkham PN, Peacock J, Dubuc A, *et al.*: **Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma.** *Nat Genet* 2009, **41(4):**465-472.
5. Hupe P, Stransky N, Thiery J-P, Radvanyi F, Barillot E: **Analysis of array CGH data: from signal ratio to gain and loss of DNA regions.** *Bioinformatics* 2004, **20(18):**3413-3422.
6. Taylor BS, Barretina J, Socci ND, DeCarolis P, Ladanyi M, Meyerson M, Singer S, Sander C: **Functional Copy-Number Alterations in Cancer.** *PLoS ONE* 2008, **3(9):**e3179.
7. Firestein R, Bass AJ, Kim SY, Dunn IF, Silver SJ, Guney I, Freed E, Ligon AH, Vena N, Ogino S, *et al.*: **CDK8 is a colorectal cancer oncogene that regulates [bgr]-catenin activity.** *Nature* 2008, **455(7212):**547-551.
8. Salvesen HB, Carter SL, Mannelqvist M, Dutt A, Getz G, Stefansson IM, Raeder MB, Sos ML, Engelsen IB, Trovik J, *et al.*: **Integrated genomic profiling of endometrial carcinoma associates aggressive tumors with indicators of PI3 kinase activation.** *Proceedings of the National Academy of Sciences* 2009, **106(12):**4834-4839.
9. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.** *Journal of the Royal Statistical Society Series B (Methodological)* 1995, **57(1):**289-300.

10. Cancer Genome Atlas Research Network: **Comprehensive genomic characterization defines human glioblastoma genes and core pathways.** *Nature* 2008, **455(7216):**1061-1068.

11. Rohan S, Tu JJ, Kao J, Mukherjee P, Campagne F, Zhou XK, Hyjek E, Alonso MA, Chen YT: **Gene expression profiling separates chromophobe renal cell carcinoma from oncocytoma and identifies vesicular transport and cell junction proteins as differentially expressed genes.** *Clin Cancer Res* 2006, **12(23):**6937-6945.

12. Jeon HM, Jin X, Lee JS, Oh SY, Sohn YW, Park HJ, Joo KM, Park WY, Nam DH, DePinho RA, *et al.*: **Inhibitor of differentiation 4 drives brain tumor-initiating cell genesis through cyclin E and notch signaling.** *Genes Dev* 2008, **22(15):**2028-2033.

13. Jacobsen B, Ploug M: **The urokinase receptor and its structural homologue C4.4A in human cancer: expression, prognosis and pharmacological inhibition.** *Curr Med Chem* 2008, **15(25):**2559-2573.

14. Cheng YY, Jin H, Liu X, Siu JMT, Wong YP, Ng EKO, Yu J, Leung Wk, Sung JJY, Chan FKL: **Fibulin 1 is downregulated through promoter hypermethylation in gastric cancer.** *Br J Cancer* 2008, **99(12):**2083-2087.

15. Bug S, Durig J, Oyen F, Klein-Hitpass L, Martin-Subero JI, Harder L, Baudis M, Arnold N, Kordes U, Duhrsen U, *et al.*: **Recurrent loss, but lack of mutations, of the SMARCB1 tumor suppressor gene in T-cell prolymphocytic leukemia with TCL1A-TCRAD juxtaposition.** *Cancer Genet Cytogenet* 2009, **192(1):**44-47.

16. Hall C, Nelson DM, Ye X, Baker K, DeCaprio JA, Seeholzer S, Lipinski M, Adams PD: **HIRA, the Human Homologue of Yeast Hir1p and Hir2p, Is a Novel Cyclin-cdk2 Substrate Whose Expression Blocks S-Phase Progression.** *Mol Cell Biol* 2001, **21(5):**1854-1865.

17. de Bernabe DB-V, Inamori K-i, Yoshida-Moriguchi T, Weydert CJ, Harper HA, Willer T, Henry MD, Campbell KP: **Loss of {alpha}-Dystroglycan Laminin Binding in Epithelium-derived Cancers Is Caused by Silencing of LARGE.** *J Biol Chem* 2009, **284(17):**11279-11284.

18. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR: **A census of human cancer genes.** *Nat Rev Cancer* 2004, **4(3):**177-183.

19. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al.*: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25(1):**25-29.

20. Gey S, Nedelec E: **Model selection for CART regression trees.** *Information Theory, IEEE Transactions on* 2005, **51(2):**658-670.

21. Bushell M, Stoneley M, Kong YW, Hamilton TL, Spriggs KA, Dobbyn HC, Qin X, Sarnow P, Willis AE: **Polypyrimidine tract binding protein regulates IRES-mediated gene expression during apoptosis.** *Mol Cell* 2006, **23(3):**401-412.

22. Su J, Chen X, Kanekura T: **A CD147-targeting siRNA inhibits the proliferation, invasiveness, and VEGF production of human malignant melanoma cells by down-regulating glycolysis.** *Cancer Lett* 2009, **273(1):**140-147.

23. Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL: **Gene expression profiling predicts clinical outcome of prostate cancer.** *J Clin Invest* 2004, **113(6):**913-923.

24. Yam JW, Jin DY, So CW, Chan LC: **Identification and characterization of EBP, a novel EEN binding protein that inhibits Ras signaling and is recruited into the nucleus by the MLL-EEN fusion protein.** *Blood* 2004, **103(4):**1445-1453.

25. Guan D, Altan-Bonnet N, Parrott AM, Arrigo CJ, Li Q, Khaleduzzaman M, Li H, Lee C-G, Pe'ery T, Mathews MB: **Nuclear Factor 45 (NF45) Is a Regulatory Subunit of Complexes with NF90/110 Involved in Mitotic Control.** *Mol Cell Biol* 2008, **28(14):**4629-4641.

26. Suzuki C, Murakumo Y, Kawase Y, Sato T, Morinaga T, Fukuda N, Enomoto A, Ichihara M, Takahashi M: **A novel GDNF-inducible gene, BMZF3, encodes a transcriptional repressor associated with KAP-1.** *Biochemical and Biophysical Research Communications* 2008, **366(1):**226-232.