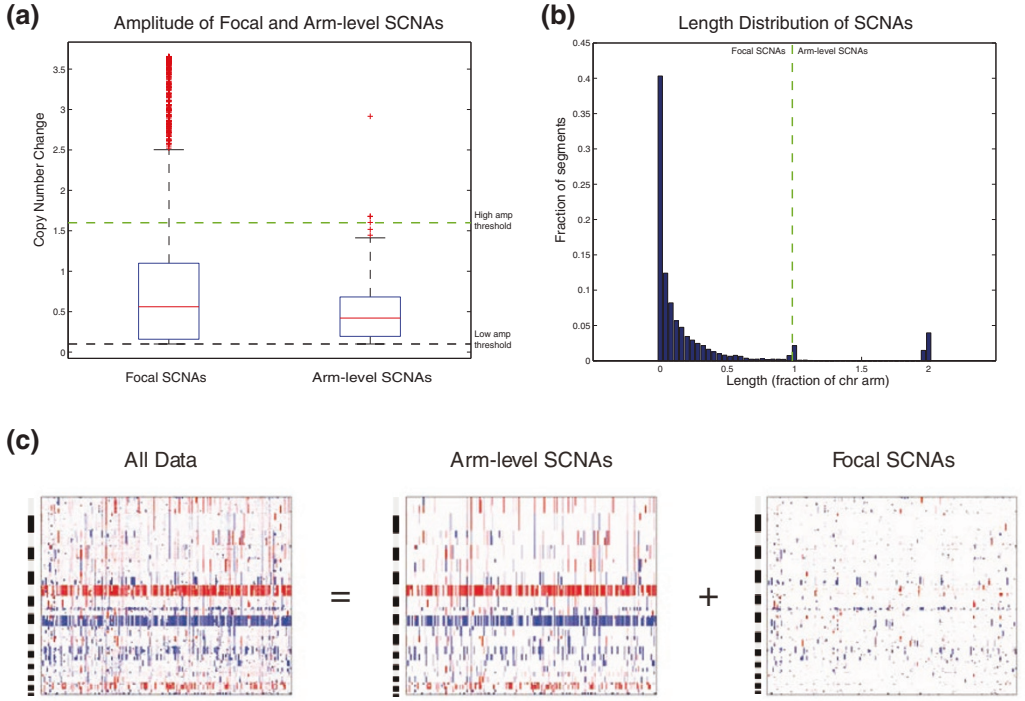

Figure 2 Computational separation of arm-level and focal SCNAs. **(a)** Boxplot showing the distribution of copy-number changes for amplified focal (length < 98% of a chromosome arm) and arm-level (length > 98% of a chromosome arm) SCNAs across 178 GBM profiles from TCGA. The black dotted line denotes a typical low-level amplitude threshold used to eliminate artifactual SCNAs, while the green dotted line denotes a typical high-level amplitude threshold used in previous version of GISTIC to eliminate arm-level SCNAs. **(b)** Histogram showing the frequency of observing SCNAs of a given length across 178 GBM samples. The high frequency of events occupying exactly one chromosome arm led us to distinguish between focal and arm-level SCNAs. **(c)** Heatmaps showing the total segmented copy-number profile of the TCGA GBM set (leftmost panel), and the results of computationally separating these samples into arm-level profiles (middle panel) and focal profiles (rightmost panel) by summing arm-level and focal SCNAs. In each heatmap, the chromosomes are arranged vertically from top to bottom and samples are arranged from left to right. Red and blue represent gain and loss, respectively.

alterations were detected using either the high amplitude (Figure 3b) or the focal length filters (Figure 3c).

The benefits of length-based filtering result from the inclusion of low- to moderate-amplitude focal events. Amplification of *PIK3CA* and *AKT1* and deletion of *WWOX* are detected using length-based filtering, but are not significant under the high amplitude filter (compare Figure 3b and 3c). Moreover, the length-based analysis identified significant SCNAs detected in neither of the amplitude-based analyses, including amplifications of *MLLT10* and deletions of *CDKN1B* and *NF1*.

No known GBM target gene was detected in either of the amplitude-based analyses that was not also detected by the length-based analysis. These results suggest that length-based filtering of arm-level events greatly improves the sensitivity of GISTIC to identify relevant regions of focal SCNA.

Probabilistic scoring of SCNAs

We set out to define a scoring framework for SCNAs that more accurately reflects the background rates of alteration. Ideally, we aim to score each region of the genome according to the probability with which the observed set of SCNAs would occur by chance alone. Scores using this framework have a clear interpretation: the higher the score assigned to a region, the less likely that the SCNAs in that region are observed entirely by chance, and the more likely that they underwent positive selection.

The probability of observing a single SCNA of given length and amplitude can be approximated by the frequency of occurrence of events of similar length and amplitude across the entire dataset (as in Supplementary Figure S2 in Additional file 3). However, since cancer genomes do contain drivers, this procedure is likely to overestimate the probability of observing SCNAs under the null model. Specifically, driver events tend to be shorter in length and of higher amplitude than passengers and therefore constitute the majority of events in their length/amplitude neighborhood (Supplementary Figure S3 in Additional file 5).