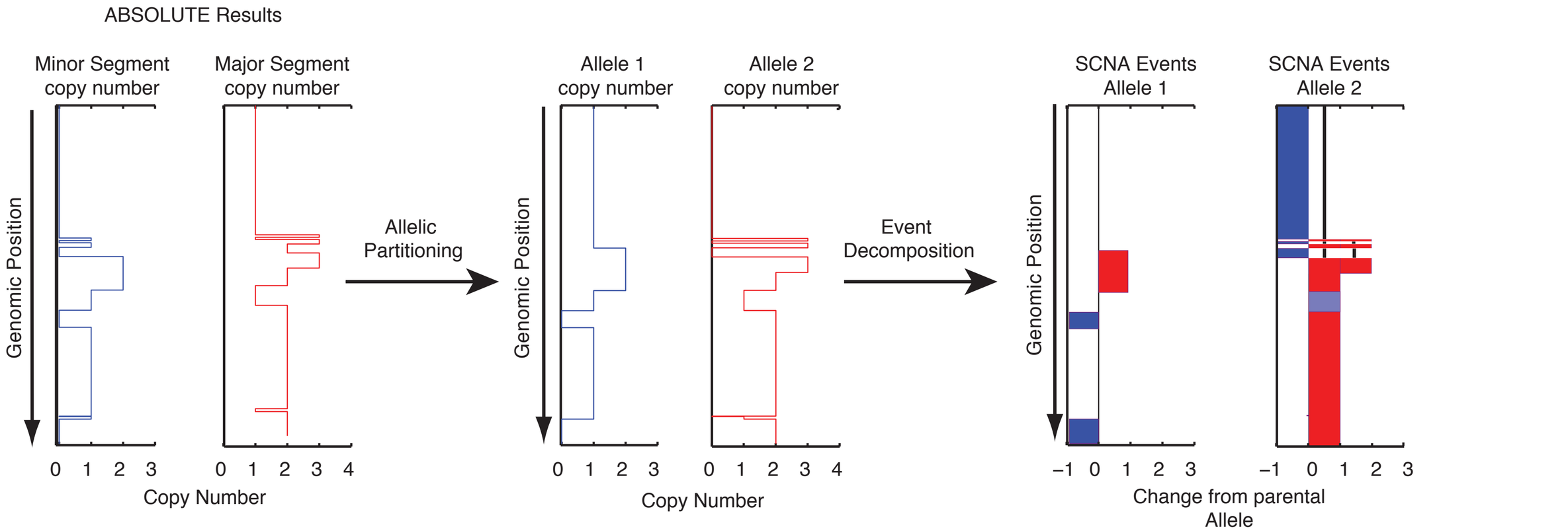
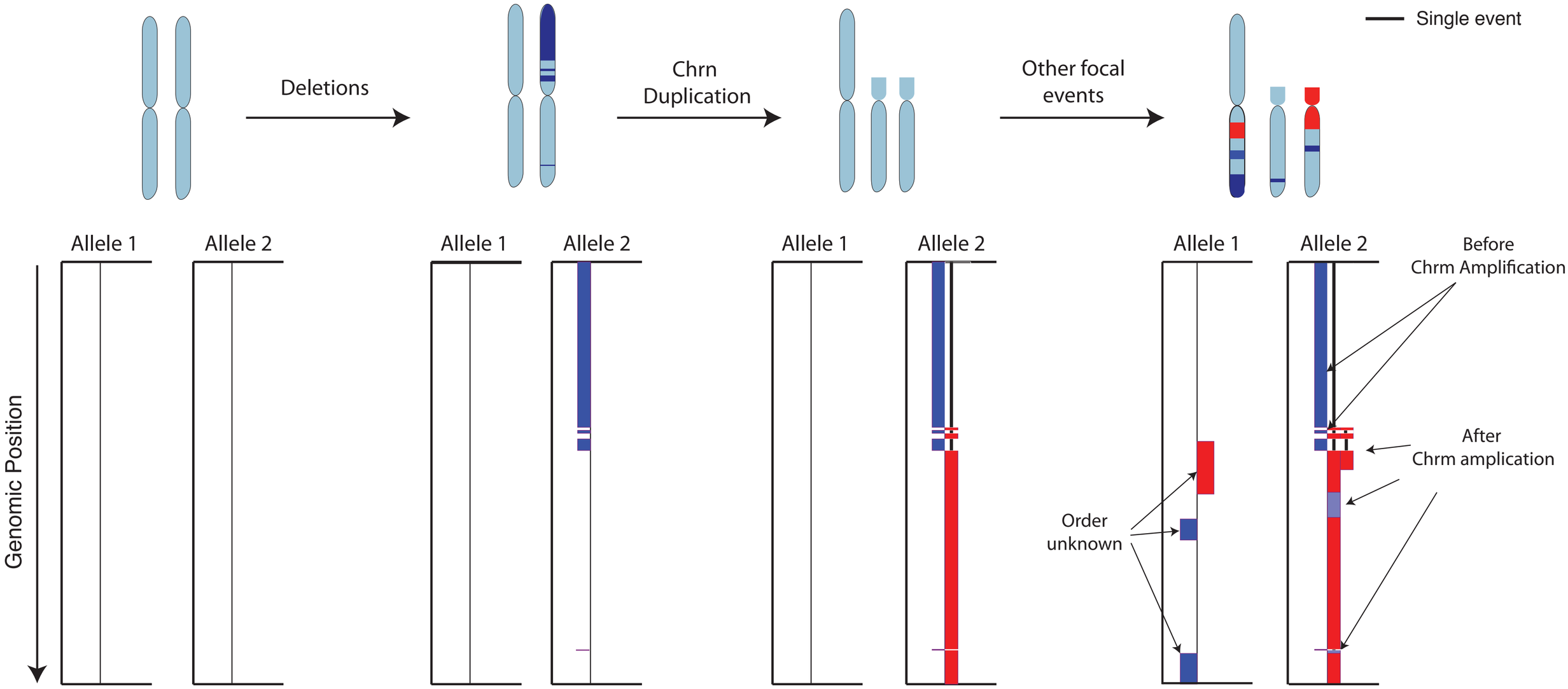
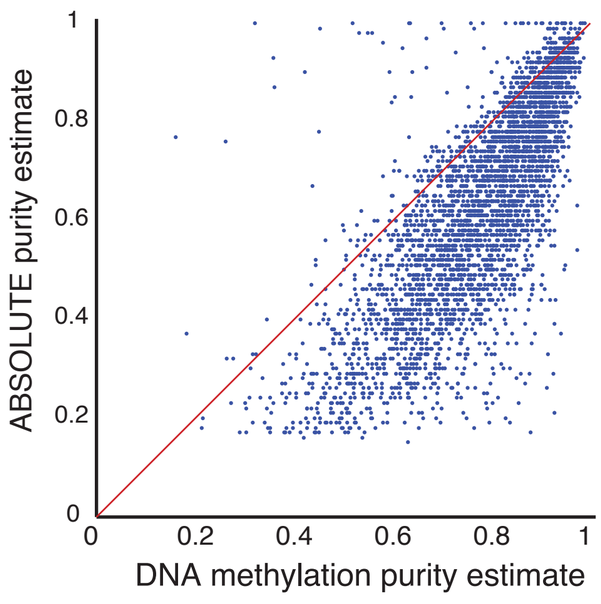
a


ABSOLUTE Results
Minor Segment copy number
Major Segment copy number
Genomic Position
0 1 2 3
0 1 2 3 4
Copy Number
Allelic Partitioning
Allele 1 copy number
Allele 2 copy number
Genomic Position
0 1 2 3
0 1 2 3 4
Copy Number
Event Decomposition
SCNA Events Allele 1
SCNA Events Allele 2
Genomic Position
−1 0 1 2 3
−1 0 1 2 3
Change from parental Allele
Amplification
Deletion
Single event

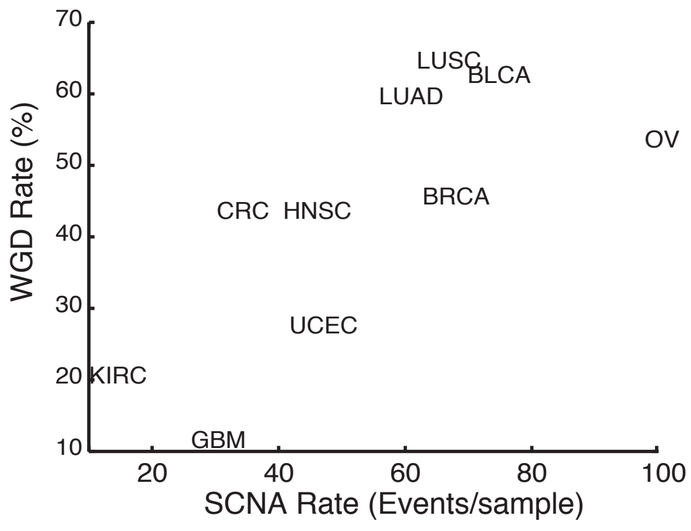
b


Deletions
Chrn Duplication
Other focal events
Allele 1
Allele 2
Genomic Position
Order unknown
Before Chrm Amplification
After Chrm amplication

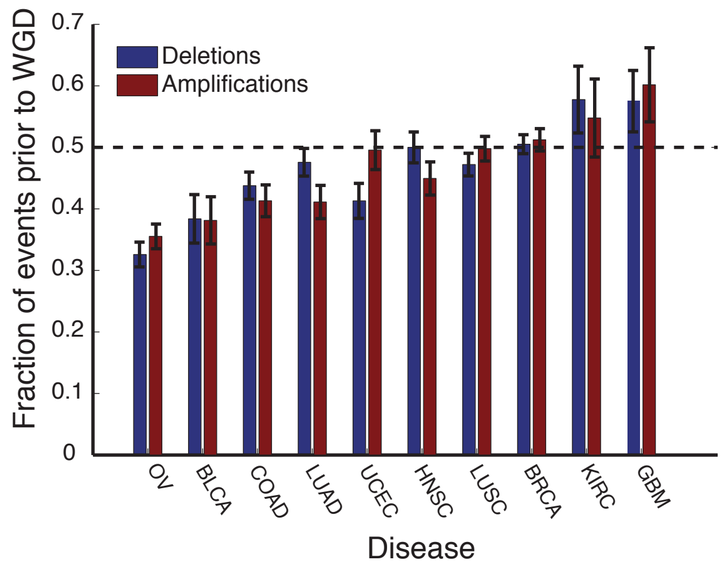
c


ABSOLUTE purity estimate
0 0.2 0.4 0.6 0.8 1
DNA methylation purity estimate

d

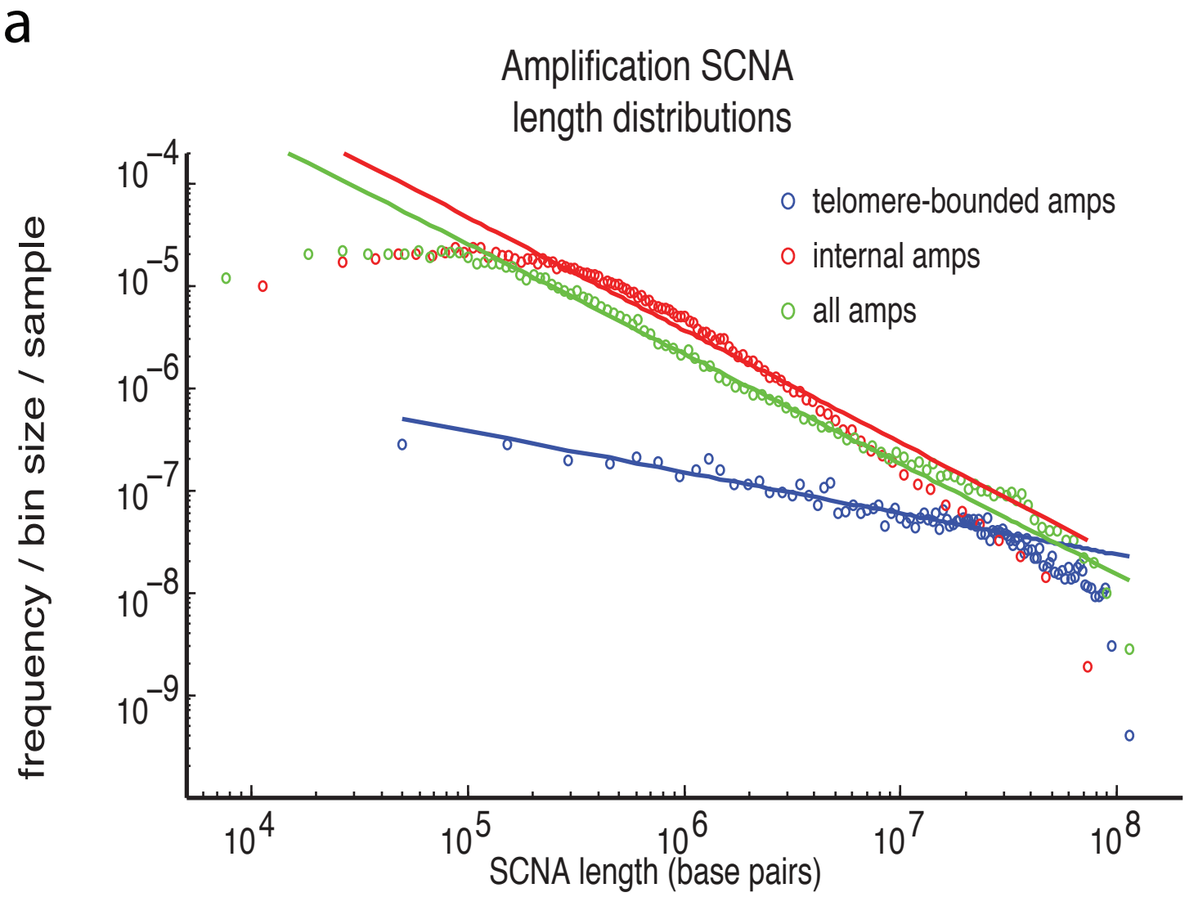

WGD Rate (%)
70 60 50 40 30 20 10
LUSC
BLCA
LUAD
OV
BRCA
CRC
HNSC
UCEC
KIRC
GBM
20 40 60 80 100
SCNA Rate (Events/sample)

e

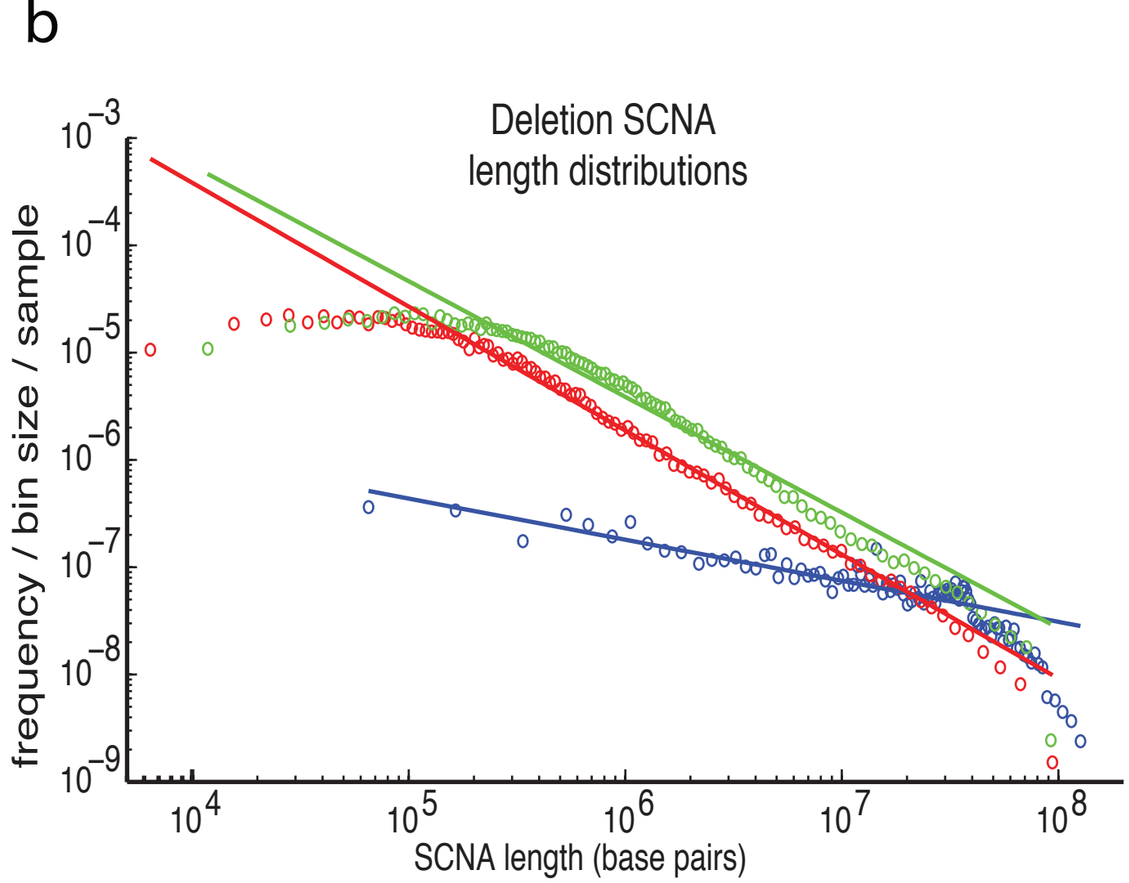

Fraction of events prior to WGD
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Deletions
Amplifications
OV
BLCA
COAD
LUAD
UCEC
HNSC
LUSC
BRCA
KIRC
GBM
Disease

Supplementary Figure 1: Determination of absolute allelic SCNAs
(**a**) Schematic indicating procedure to determine SCNAs. Absolute allelic copy-numbers generated by ABSOLUTE (left panel) are partitioned with the lowest copy-numbers on one allele (blue) and the higher copy-numbers on the other (red). We repartition the copy-numbers between alleles to test all possibilities if computationally efficient, or the most likely possibilities if not (middle panel). The most likely set of SCNAs generating each copy-number profile is then determined (right panel). Regions in red are amplifications; regions in blue are deletions. The black lines in allele 2 indicate deletions that preceded an amplification, so that segments that appear discontinuous on the reference genome were amplified in a single event. (**b**) Schematic indicating temporal order of SCNAs determined in (a). In this example, deletions followed by a chromosome level amplification can account for the copy-number profiles in (a). For the same profile to be generated by amplifications followed by deletions, either the deletions would have to remove two copy-levels, requiring sequential deletions with identical boundaries, or multiple neighboring, non-contiguous amplifications would be required in addition to multiple, non-contiguous deletions. We assume both of these possibilities are unlikely. (**c**) Estimates of purity according to ABSOLUTE (y-axis) against estimates of purity according to a lymphocyte/leukocyte DNA methylation signature (x-axis) across 3735 cancers. The estimate from the lymphocyte/leukocyte DNA methylation signature tended to provide higher purity estimates, suggesting other types of normal cells may contribute to impurities detected by ABSOLUTE. (**d**) Number of SCNA events in samples with WGD (y-axis) against number of SCNAs in near-diploid samples (x-axis) across diseases. Most diseases show a greater than 2:1 ratio (red line) between the average number of events observed in WGD samples versus their near-diploid counterparts. (**e**) Fraction of focal amplifications (red) and deletions (blue) occurring prior to WGD by disease. The fractions of amplifications and deletions that occur prior to WGD within each disease are highly correlated and vary across diseases. The amplifications estimate carries greater ambiguity because amplifications involve DNA being placed in disparate regions of the genome. SCNAs whose timing relative to WGD are ambiguous were omitted.

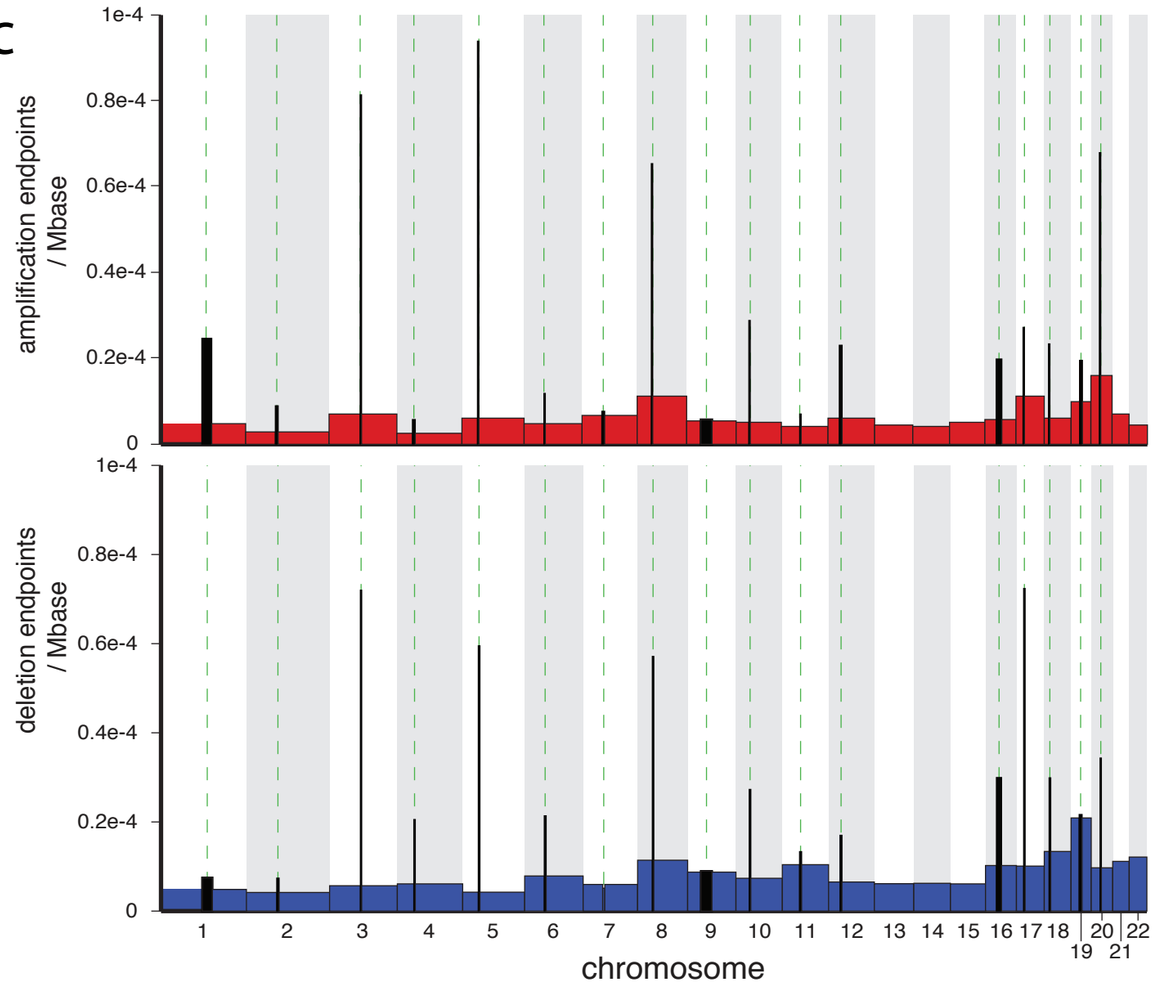
a

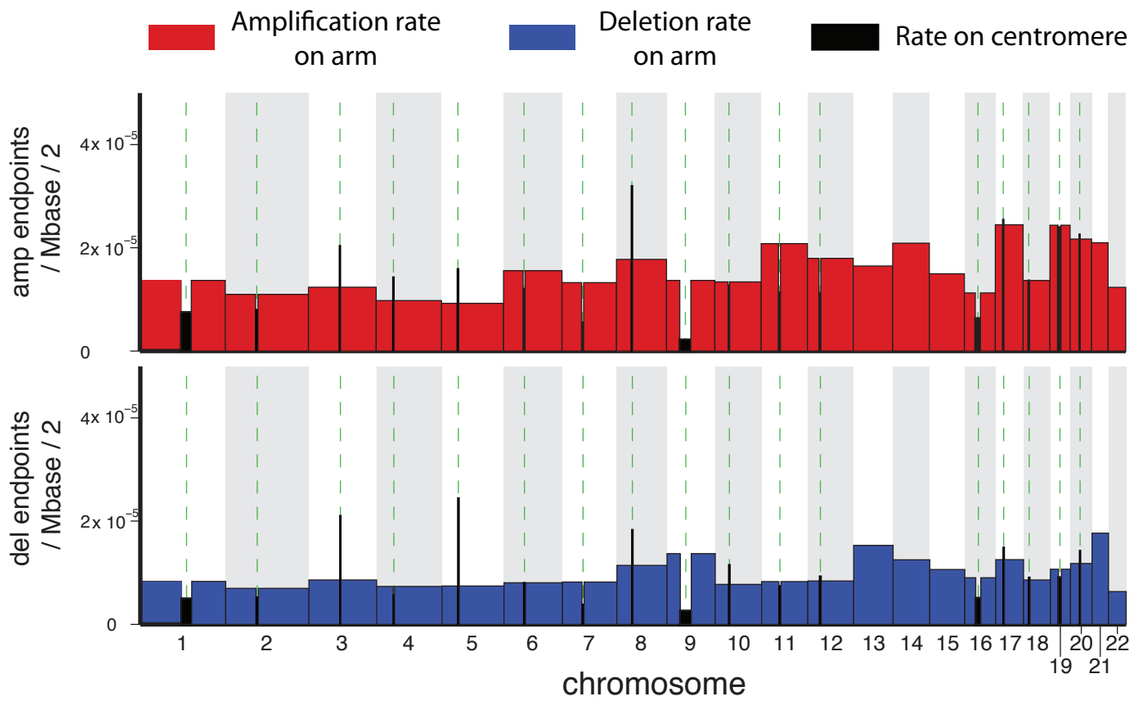

Amplification SCNA
length distributions
telomere-bounded amps
internal amps
all amps
frequency / bin size / sample
SCNA length (base pairs)

b

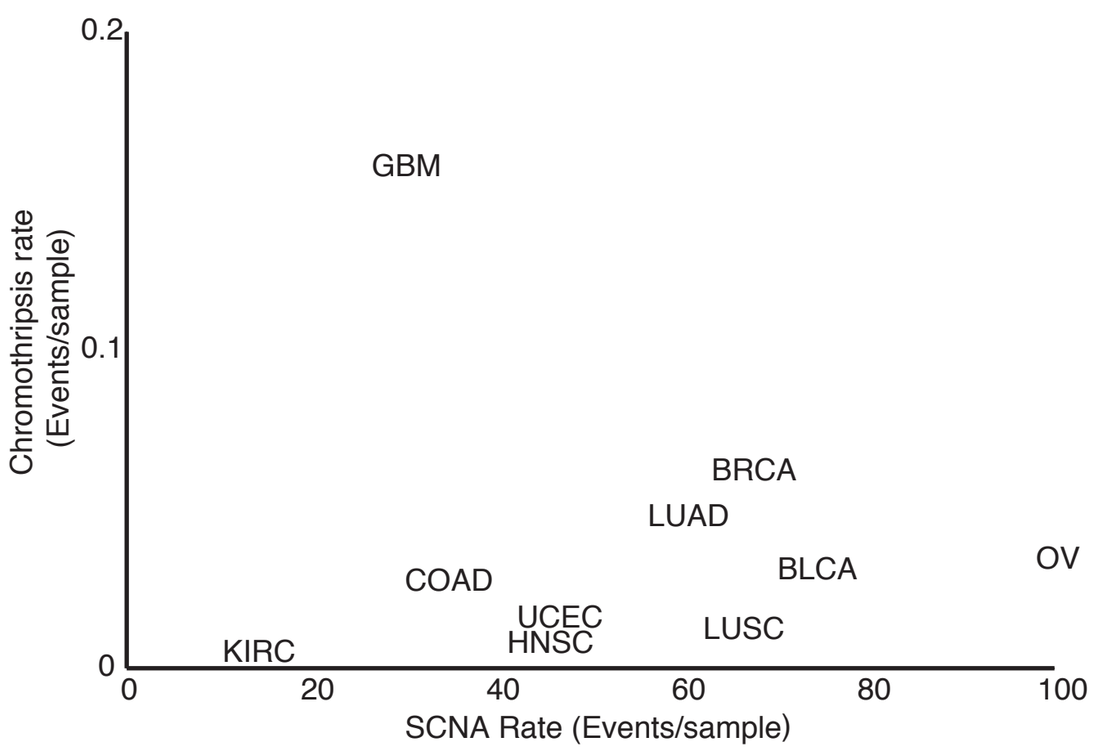

Deletion SCNA
length distributions
frequency / bin size / sample
SCNA length (base pairs)

c

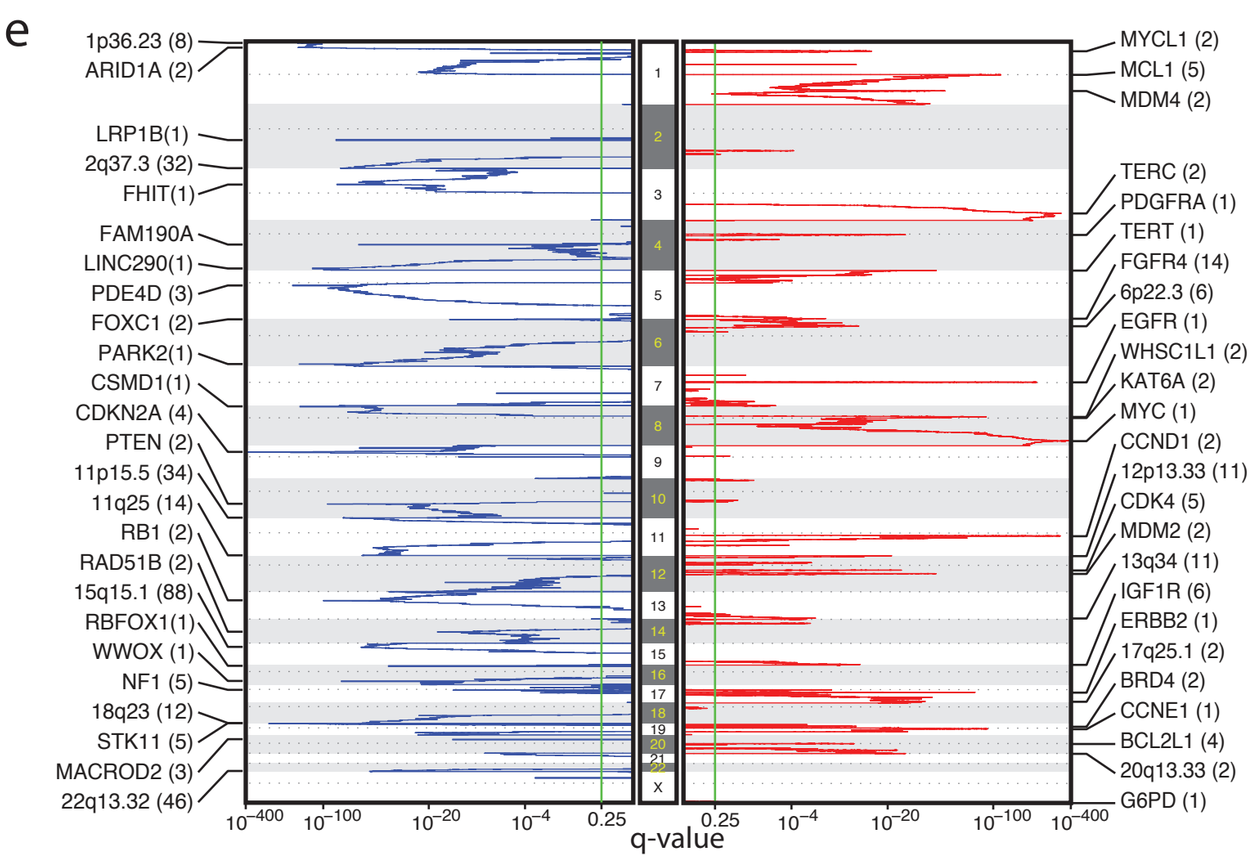

amplification endpoints / Mbase
deletion endpoints / Mbase
chromosome


Amplification rate on arm
Deletion rate on arm
Rate on centromere
amp endpoints / Mbase / 2
del endpoints / Mbase / 2
chromosome

d

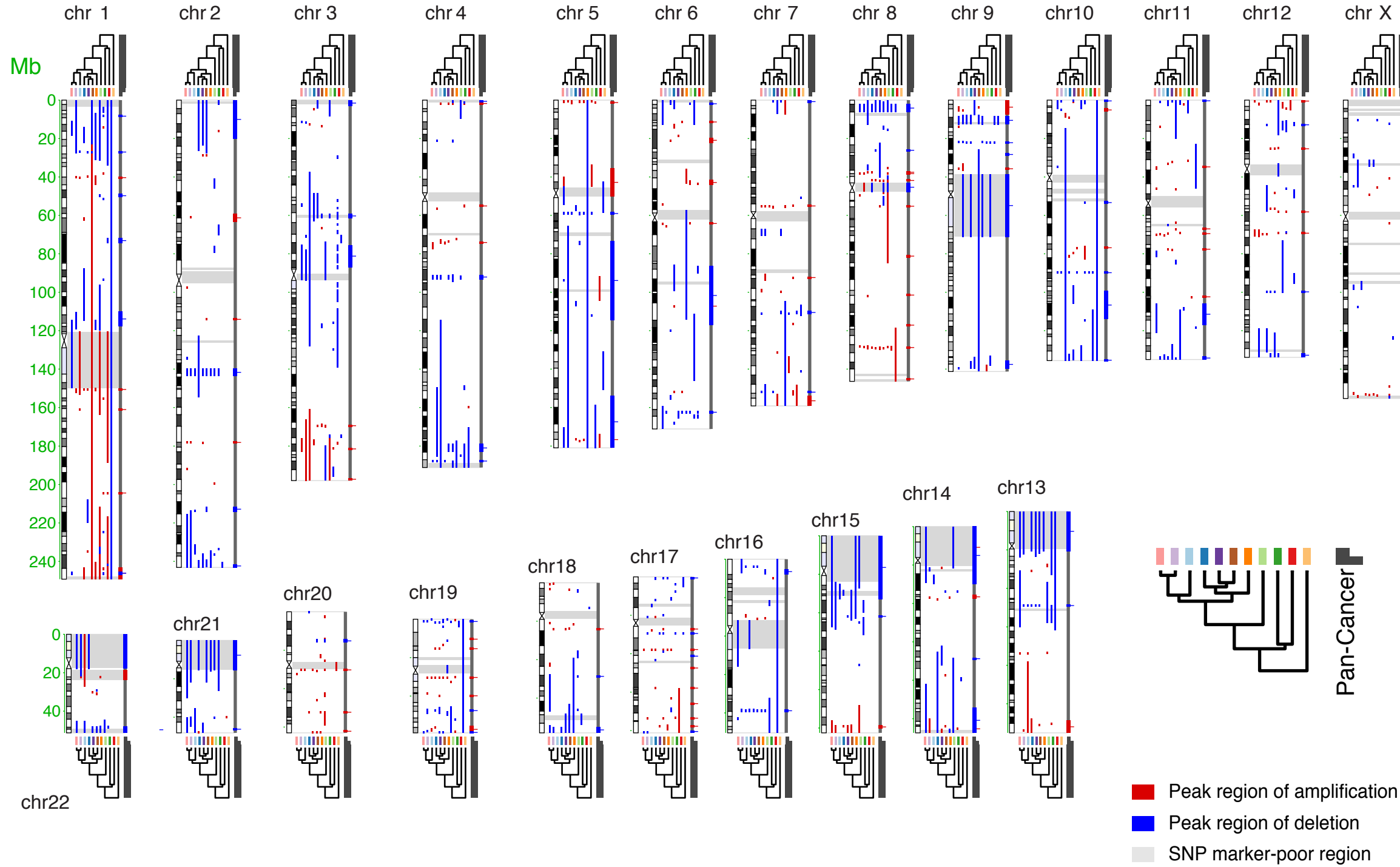

Chromothripsis rate (Events/sample)
GBM
BRCA
LUAD
BLCA
OV
COAD
UCEC
HNSC
LUSC
KIRC
SCNA Rate (Events/sample)

e


1p36.23 (8)
ARID1A (2)
LRP1B(1)
2q37.3 (32)
FHIT(1)
FAM190A
LINC290(1)
PDE4D (3)
FOXC1 (2)
PARK2(1)
CSMD1(1)
CDKN2A (4)
PTEN (2)
11p15.5 (34)
11q25 (14)
RB1 (2)
RAD51B (2)
15q15.1 (88)
RBFOX1(1)
WWOX (1)
NF1 (5)
18q23 (12)
STK11 (5)
MACROD2 (3)
22q13.32 (46)
MYCL1 (2)
MCL1 (5)
MDM4 (2)
TERC (2)
PDGFRA (1)
TERT (1)
FGFR4 (14)
6p22.3 (6)
EGFR (1)
WHSC1L1 (2)
KAT6A (2)
MYC (1)
CCND1 (2)
12p13.33 (11)
CDK4 (5)
MDM2 (2)
13q34 (11)
IGF1R (6)
ERBB2 (1)
17q25.1 (2)
BRD4 (2)
CCNE1 (1)
BCL2L1 (4)
20q13.33 (2)
G6PD (1)
q-value

Supplementary Figure 2: Features of SCNAs

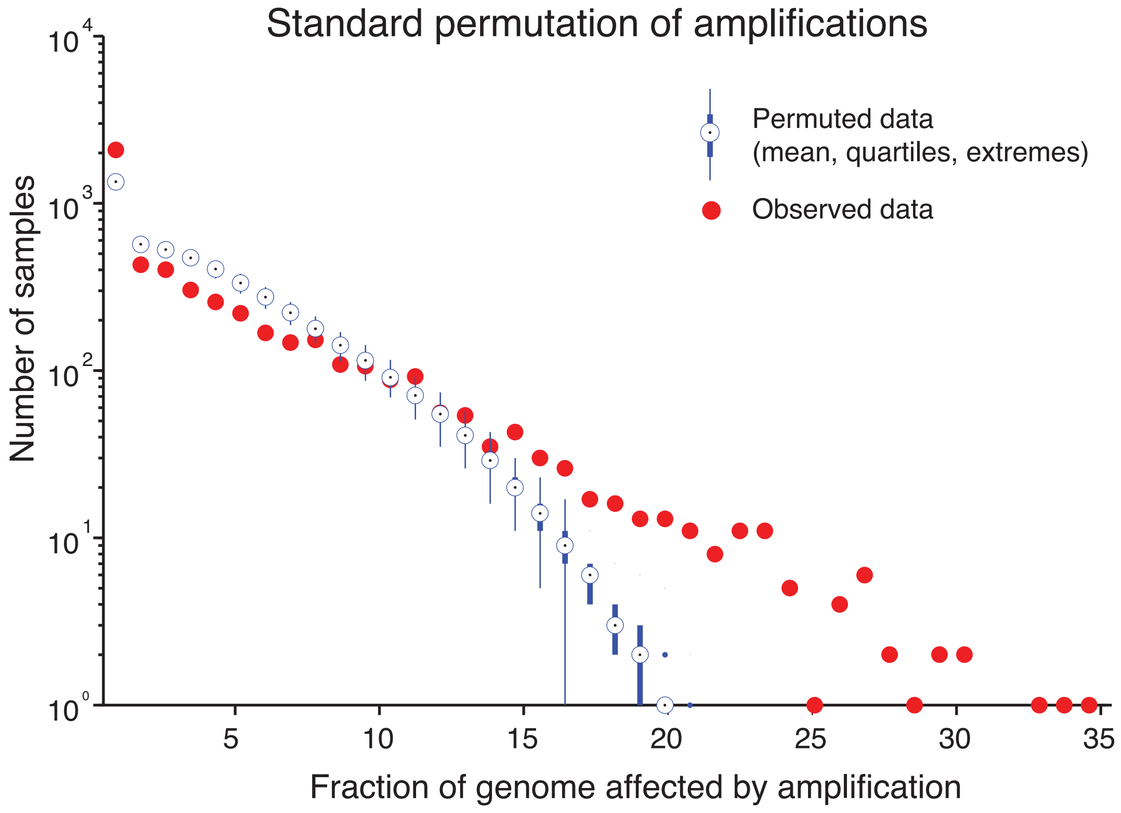
(**a**) Frequencies of all amplifications (green) and amplifications that begin at the telomere (blue) or are internal to the chromosome (red) against amplification length. We used a fixed number of events per bin as opposed to a fixed bin size to compute our least squared fit because there were many fewer long events. The frequencies of amplifications that are greater than 400 kb and less than a chromosome arm in length follow a power law $f(L) = 1/L^{\beta}$, where for all events, $\beta = 1.05$ and $r^2=0.99$; for telomeric events $\beta = 0.45$ and $r^2=0.90$; and for internal events, $\beta = 1.12$ and $r^2=0.97$ (**b**) Frequencies of all deletions (green) and deletions that begin at the telomere (blue) or are internal to the chromosome (red) against deletion length. The frequencies of deletions that are greater than 400 kb and less than a chromosome arm in length also follow a power law $f(L) = 1/L^{\beta}$, where for all events, $\beta = 1.05$ and $r^2=0.96$; for telomeric events, $\beta = 0.36$ and $r^2=0.79$; and for internal events, $\beta = 1.15$ and $r^2=0.99$. Note that the power law is a very poor fit for telomere-bounded events. **(c)** Number of amplification (red, top) and deletion (blue, bottom) endpoints in chromosome arms (colored regions) and centromeres in metacentric chromosomes (black regions) for telomere-bounded SCNAs (left panels) and internal SCNAs (right panels). SCNAs are more likely to end within the centromere than expected given the centromere's length for both telomeric and internal SCNAs ($p<0.0001$ in both cases). In the case of telomeric events, this tendency results in a propensity for SCNAs to be arm-level events involving precisely one chromosome arm and suggests that focal telomeric SCNAs are generated by similar mechanisms to such arm-level events. Note that centromeres often span large regions without SNP array markers, preventing detection of many SCNA endpoints in these regions. The width of each region reflects the size of its genomic locus. (**d**) Rates of cnLOH events that were internal to chromosomes (y-axis) against rates of cnLOH events that were telomere-bounded (x-axis) across diseases. Only 2% of focal SCNAs were cnLOH, and these events had more pronounced differences between telomeric and internal events than did amplifications and deletions. Most cnLOHs (58%) involved either whole chromosomes or exactly one chromosome arm, compared to an 18% rate of arm- and chromosome-level events for other SCNAs ($p < 0.0001$). Internal cnLOHs were typically much smaller than other focal internal SCNAs (median 0.2 Mb vs 0.8 Mb, Mann-Whitney (MW) $p < 0.0001$), whereas telomeric cnLOHs were much larger than other telomeric SCNAs (median 82Mb vs 27.2Mb, $p<0.0001$). Rates of telomeric and internal cnLOH show no correlation across diseases, suggesting the processes that lead to these events may be distinct. (**e**) Significance of focal SCNAs. GISTIC q-values (x-axis) for deletions (left, blue) and amplifications (right, red) are plotted across the genome (y-axis). Candidate gene targets within each peak are indicated for the 25 most significant peaks; in cases where no clear candidate was identified, the cytoband was indicated. Values in parentheses indicate the number of genes in each peak. Green lines indicate the significance threshold ($q=0.25$).


chr 1
chr 2
chr 3
chr 4
chr 5
chr 6
chr 7
chr 8
chr 9
chr10
chr11
chr12
chr X
Mb
0
20
40
60
80
100
120
140
160
180
200
220
240
chr22
chr21
chr20
chr19
chr18
chr17
chr16
chr15
chr14
chr13
Pan-Cancer
Peak region of amplification
Peak region of deletion
SNP marker-poor region

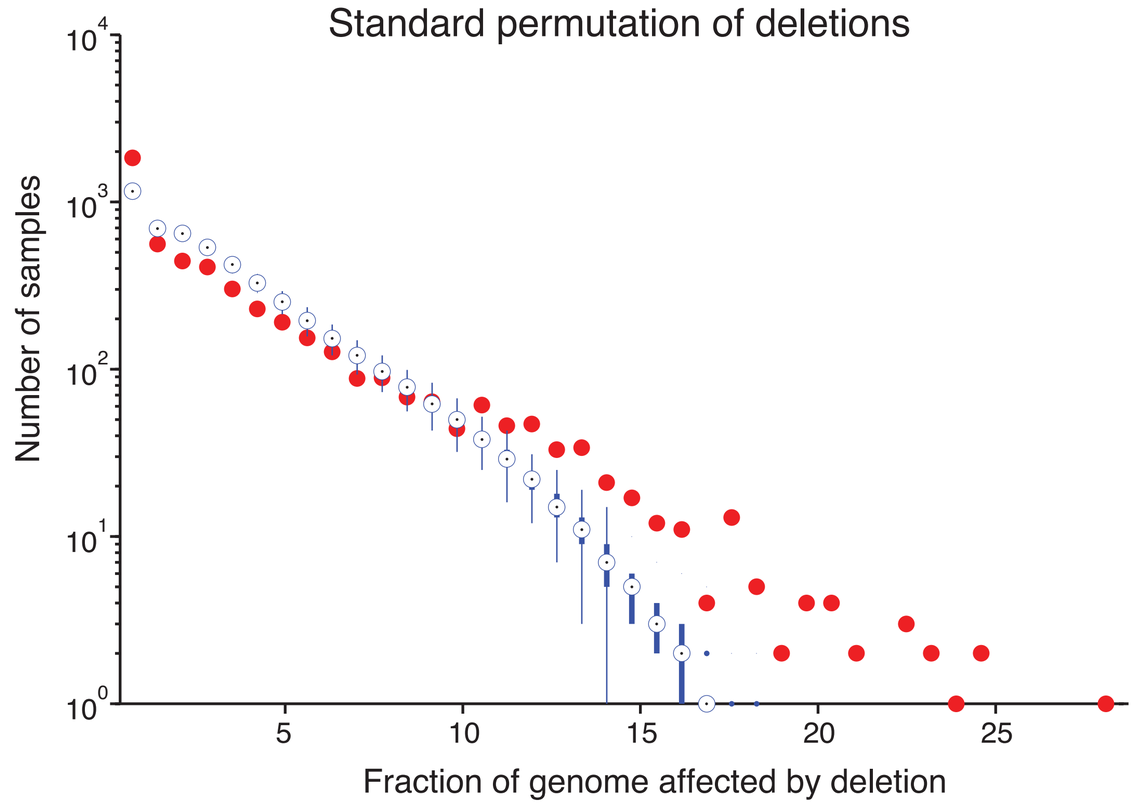
Supplementary Figure 3: Distribution and overlap of peak regions in disease-specific GISTIC analyses

Illustration of locations of peak regions within all chromosomes (indicated by green numbers) across cancer types (designated by boxes on top and bottom colored according to the scheme on the lower right and reflecting the clustering results in Figure 3a) and the Pan-Cancer analysis (right-most column, denoted by a black line). This is a complete, unlabeled version of the example illustrated by **Fig. 3a.** Detailed information about each peak can be found in Supplementary Table 3.

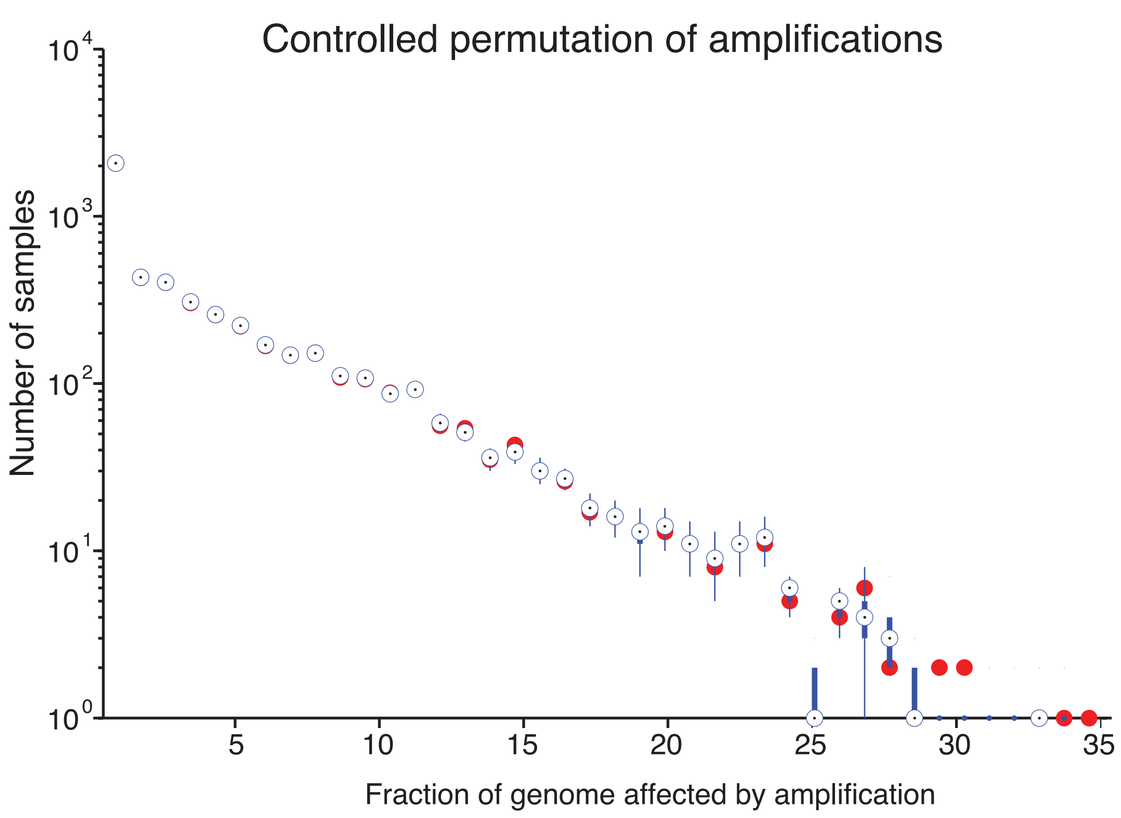
a


Standard permutation of amplifications
Permuted data
(mean, quartiles, extremes)
Observed data
Number of samples
Fraction of genome affected by amplification

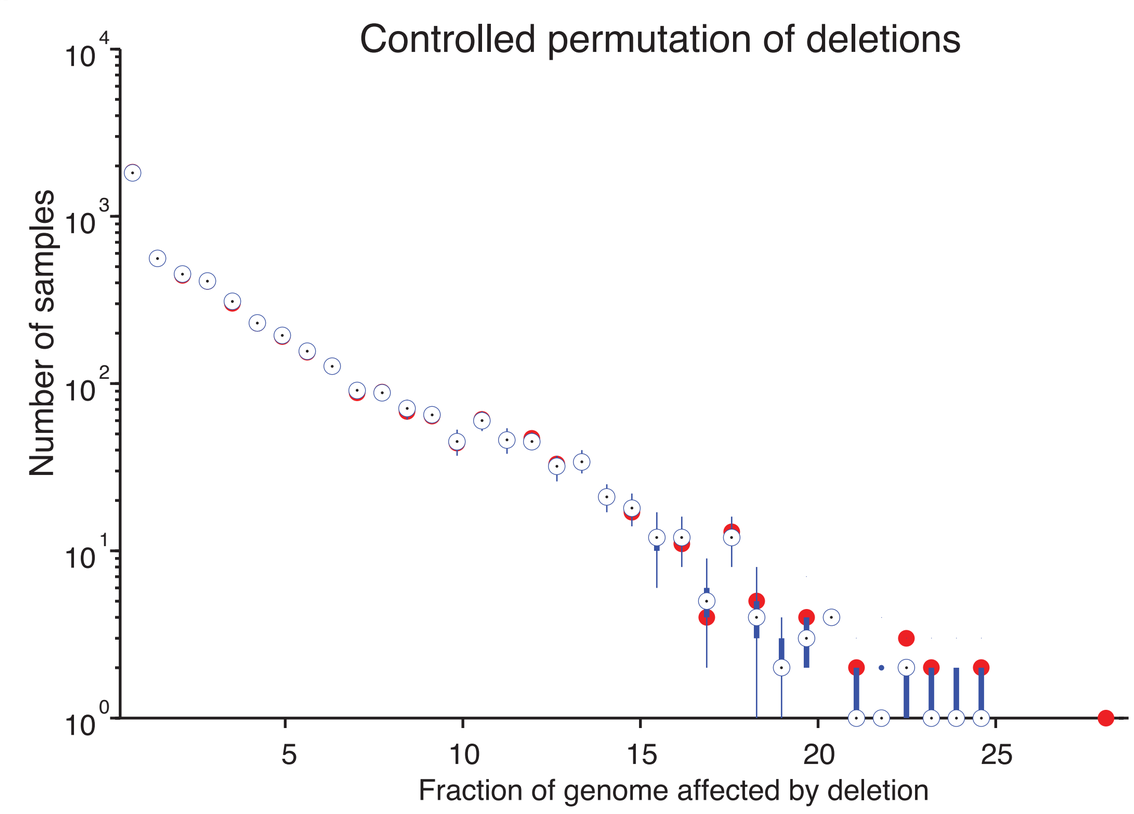
b


Standard permutation of deletions
Number of samples
Fraction of genome affected by deletion

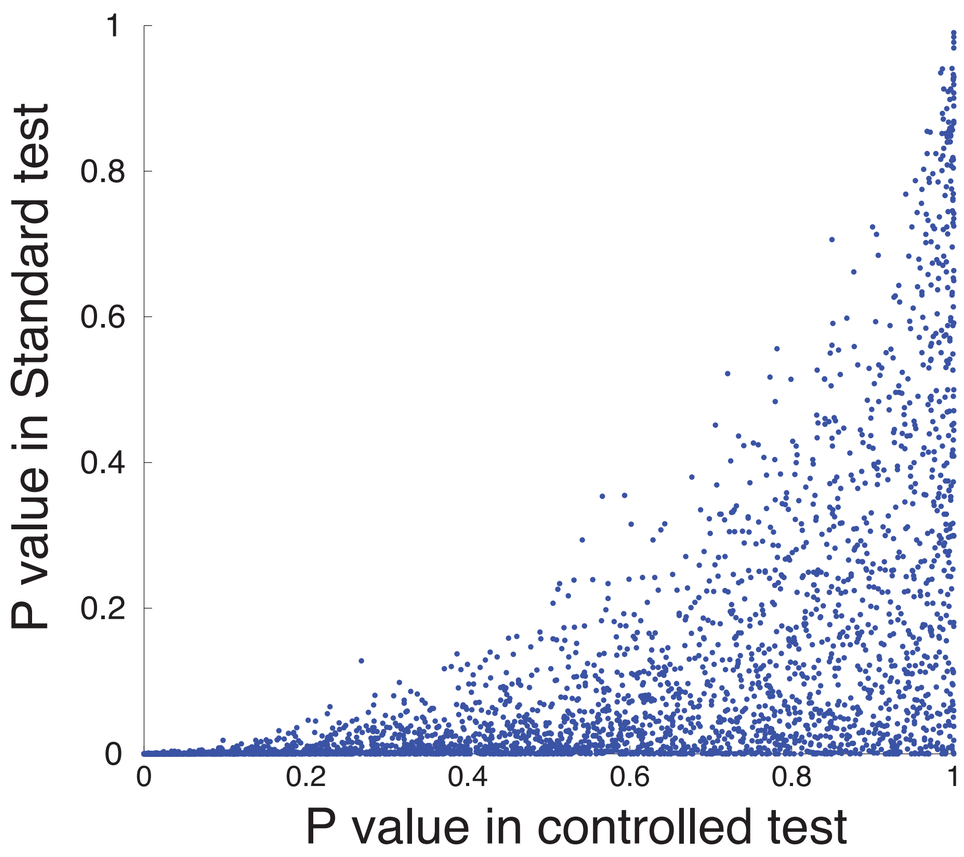
c


Controlled permutation of amplifications
Number of samples
Fraction of genome affected by amplification

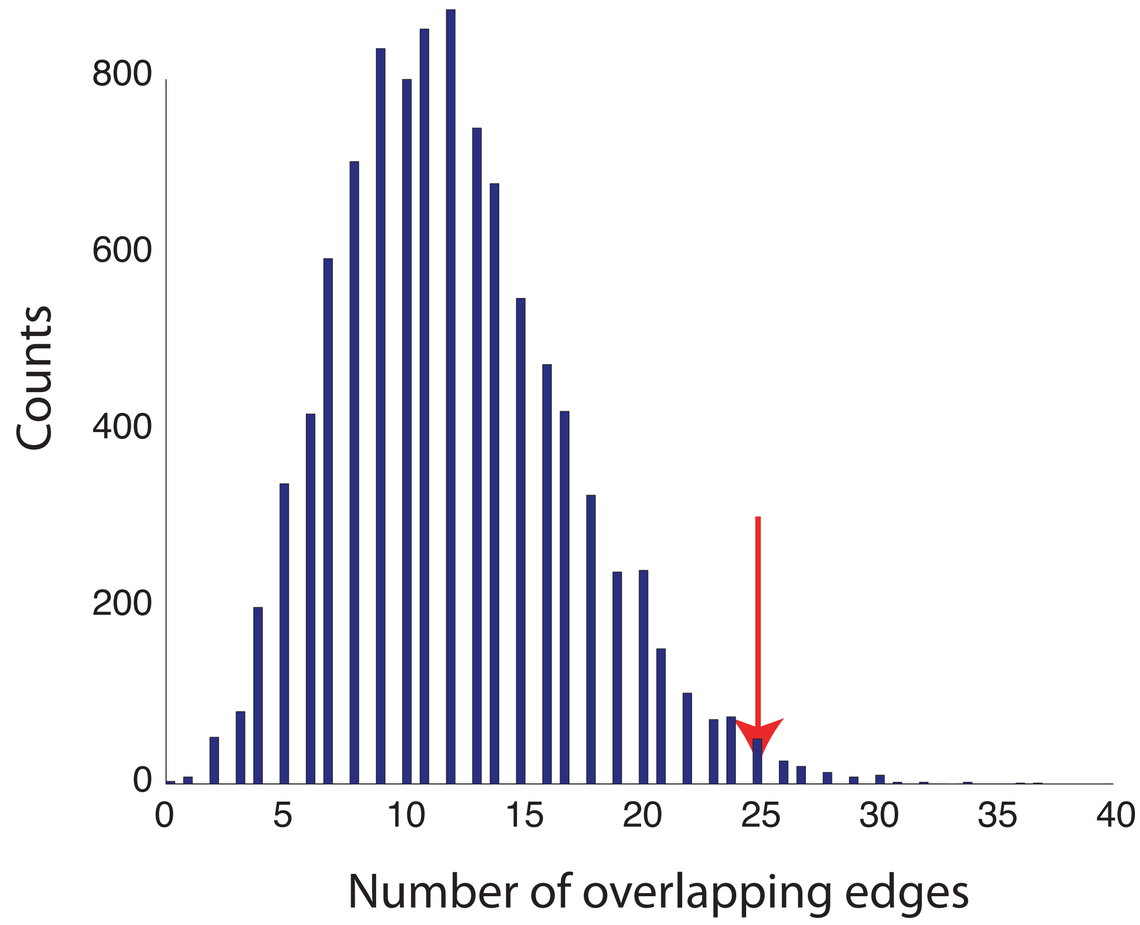
d


Controlled permutation of deletions
Number of samples
Fraction of genome affected by deletion

e


P value in Standard test
P value in controlled test

f


Counts
Number of overlapping edges

Supplementary Figure 4: Correlations analyses
Number of samples (y-axis) against fraction of genome amplified (**a**) and deleted (**b**) in each sample, in observed data (red dots) and in standard permutations in which SCNAs are randomly assigned to samples while maintaining the lineage assignments (blue box plots). (**c**) A comparison of p-values obtained from the lineage-controlled standard analysis of correlations (y-axis) and an analysis that controls for varying levels of genomic disruption (x-axis). Although the p-values tend to correlate, the ranking is not the same, and many interactions that appear significant in one test are not significant in the other. Most interactions appear significant in the standard analysis. (**d**) The number of significant anticorrelations that overlap known protein-protein interactions in the observed genetic interactome network (red arrow) and permuted networks (blue bars). These results are from the analysis of high-level SCNAs (amplifications > 4.4 copies and deletions to less than one copy); results from the analysis of all SCNAs are displayed in Figure 4e.