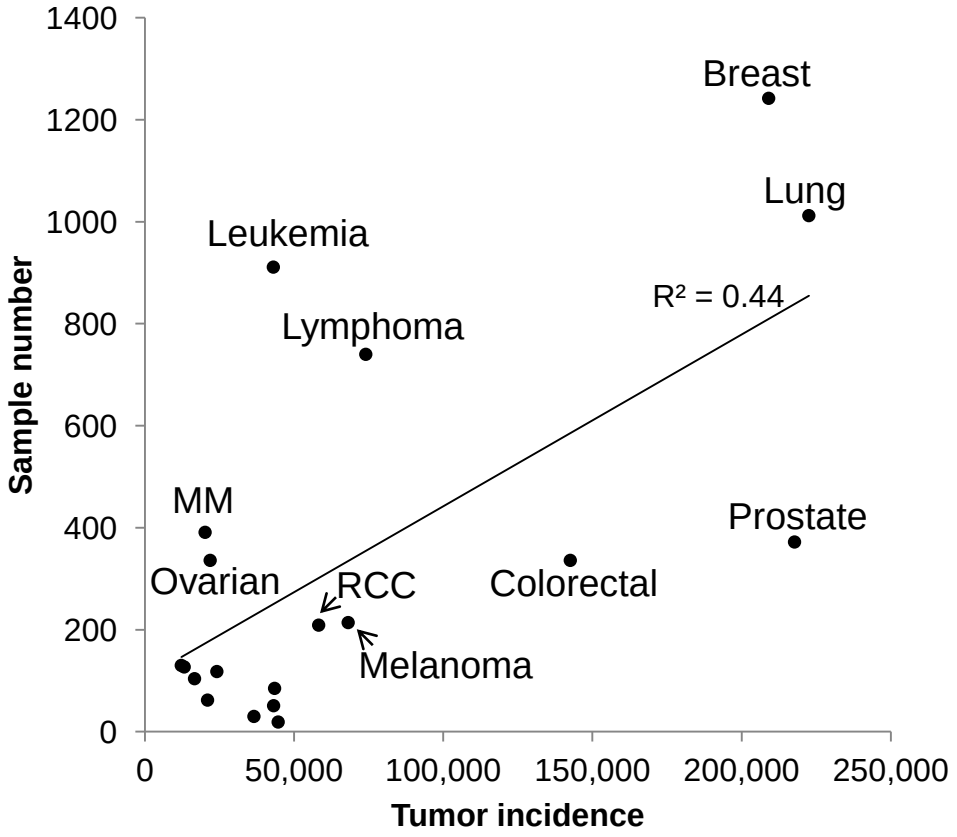
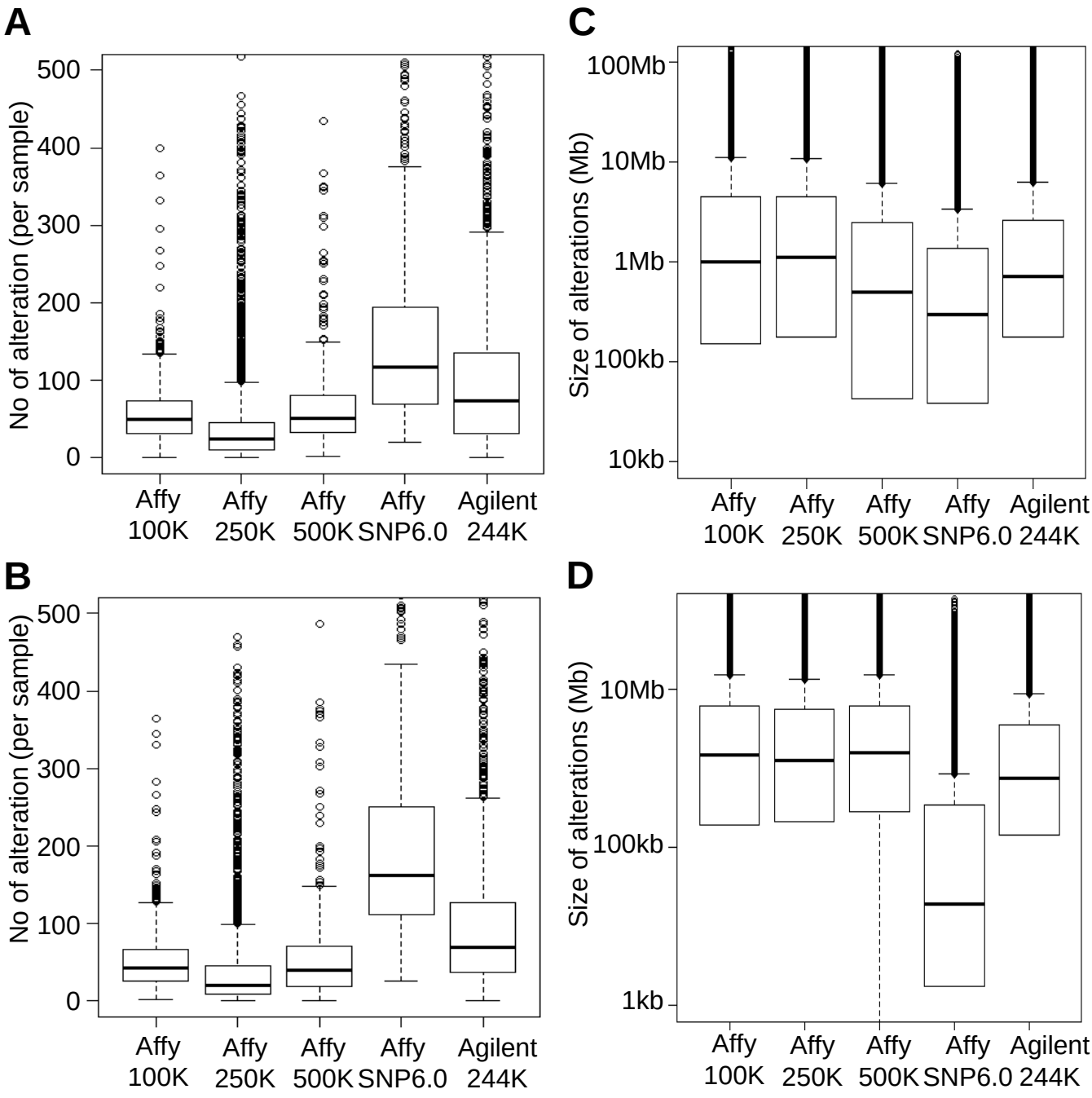
Supplementary Fig. 1



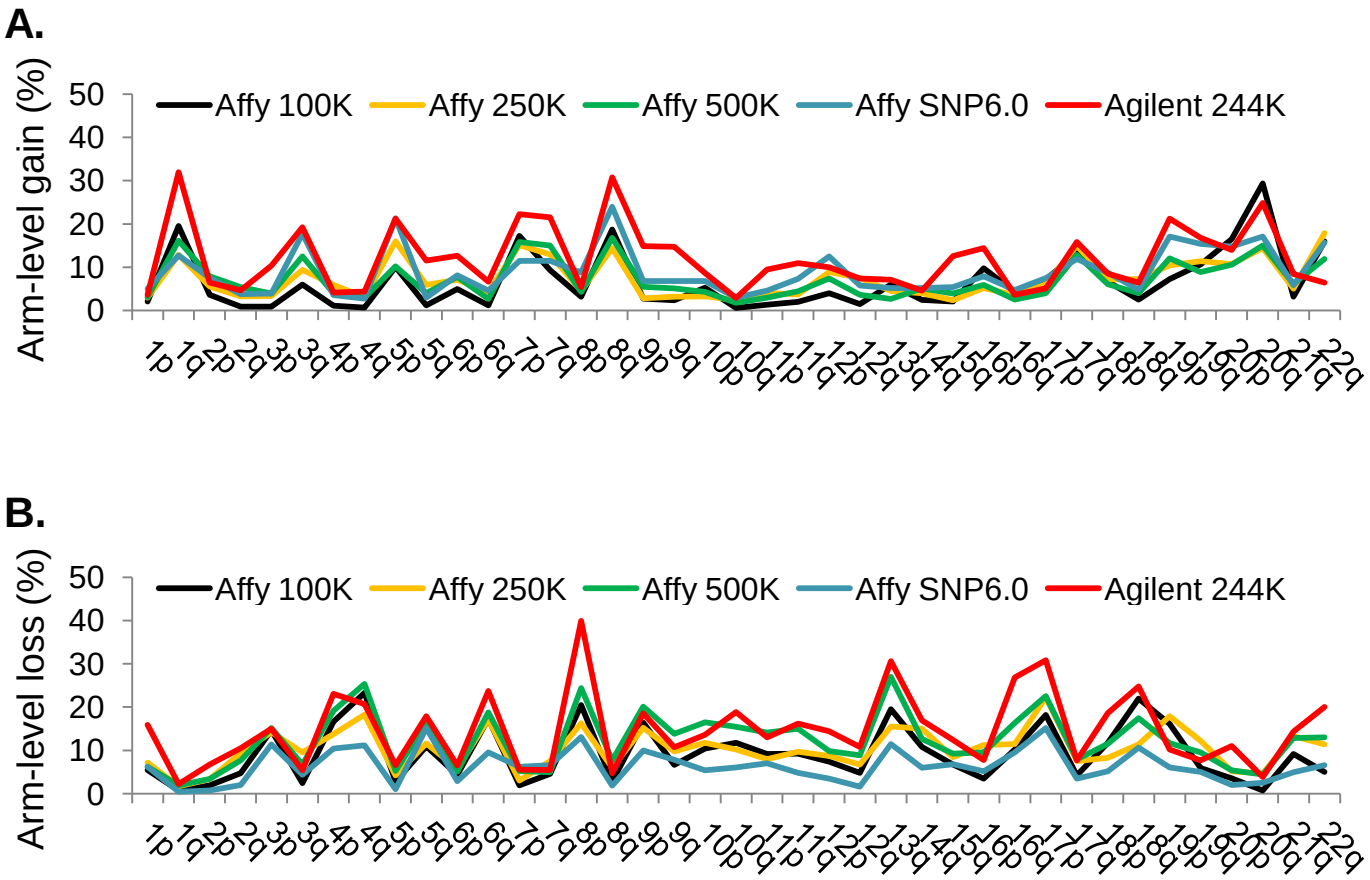
Supplementary Figure 1. The frequency of major tumor types in the study and their actual tumor incidence rates. The sample numbers in our dataset (y-axis) are plotted against the incidence rates for tumor types (x-axis; obtained from American Cancer Society, Cancer Facts & Figures 2010).

Supplementary Fig. 2



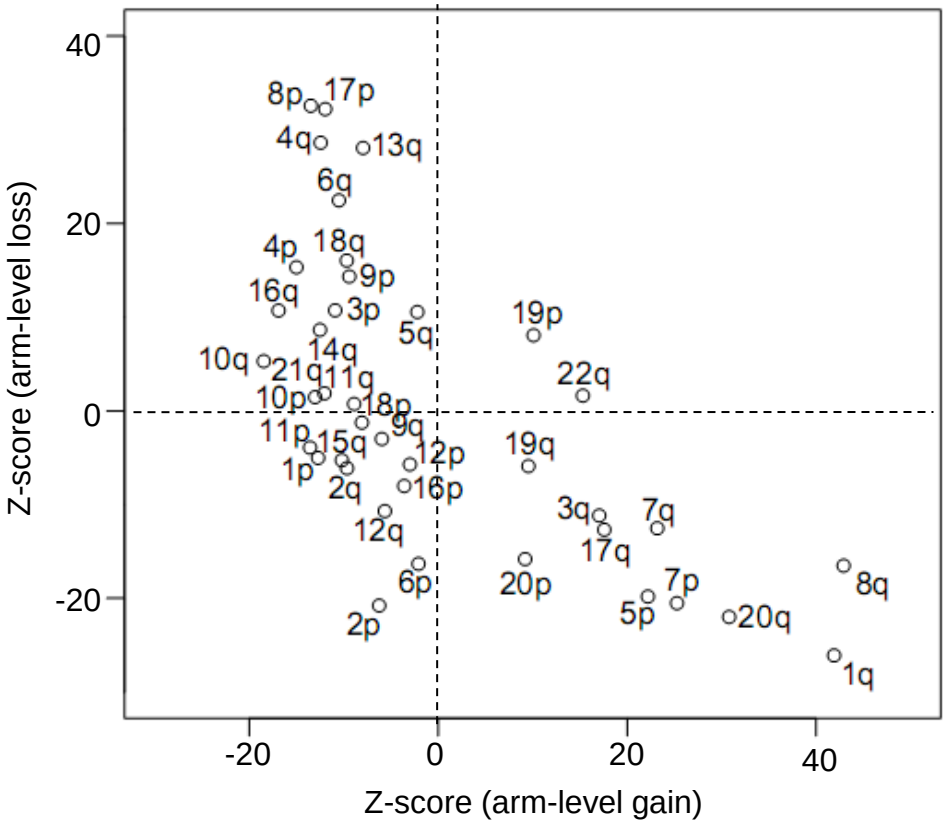
Supplementary Figure 2. The distributions of alteration numbers and sizes across array-CGH platforms. The average numbers of alterations across the five platforms are shown for chromosomal gains (A) and losses (B). The distributions of alteration sizes are shown for chromosomal gains (C) and losses (D).

Supplementary Fig. 3



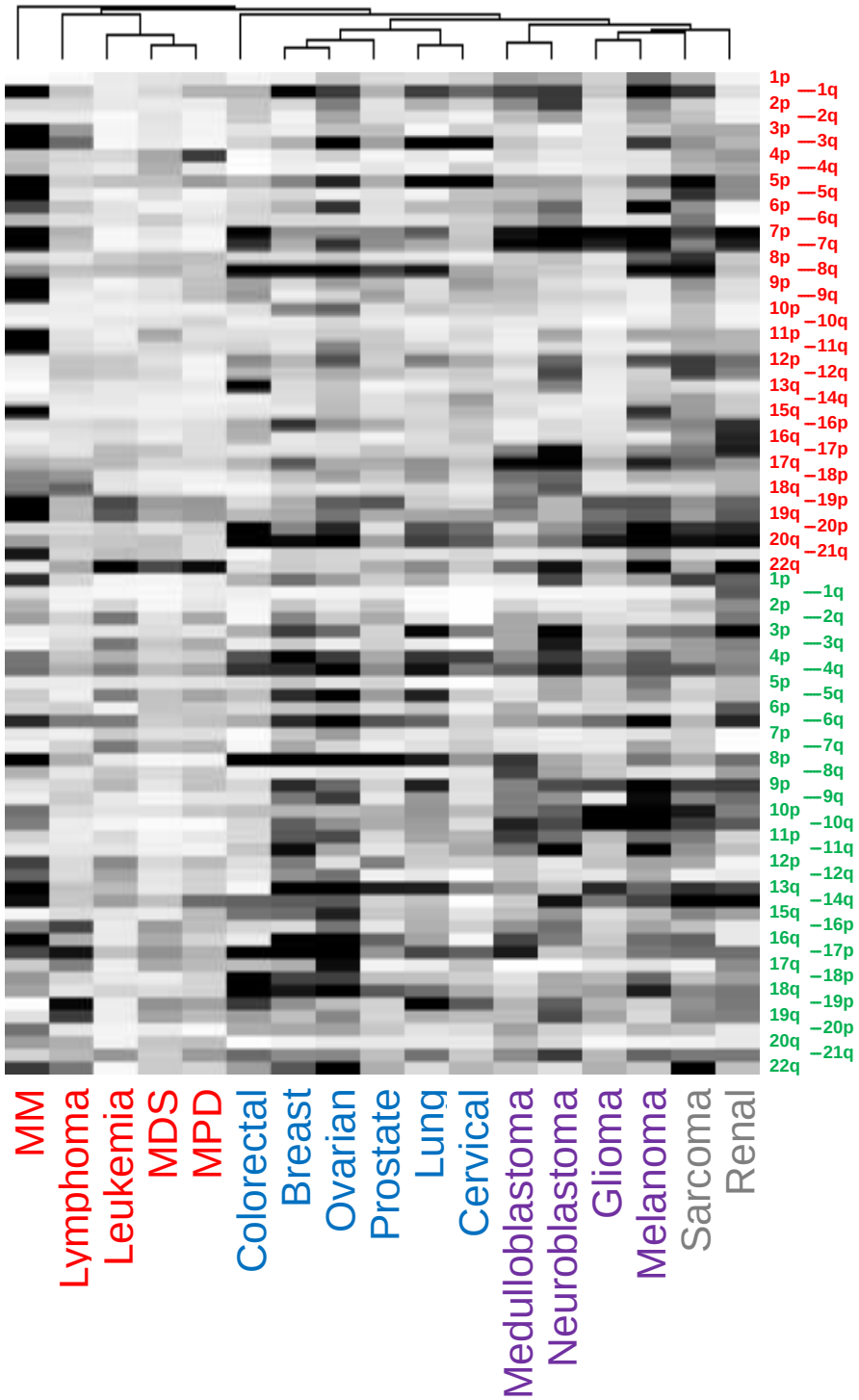
Supplementary Figure 3. The chromosomal arm-level alteration frequencies across array-CGH platforms. (A) gains; (B) losses.

Supplementary Fig. 4



Supplementary Figure 4. The size-adjusted frequencies of arm-level alterations. Chromosomal size-adjusted *Z* scores were computed for the arm-level alteration frequencies. The *Z* scores represent the extent of deviation of the corresponding background frequencies.

Supplementary Fig. 5



Supplementary Figure 5. The hierarchical clustering of chromosomal arm-level alteration frequencies. The clustering was performed for 5966 samples after removing the samples from Affymetrix 250K (StyI) and 500K (StyI and NspI) platforms. The clustering pattern is similar to that obtained with the entire dataset (*n* = 8337; **Fig. 2B**).

Supplementary Fig. 6

A.

Amplification (before adjustment)

	Affy 100K	Affy 250K	Affy 500K	Affy SNP6	Agilent 244K
Affy100K		46.9	44.9	40.8	59.2
Affy250K	42.5		50.7	50.7	49.3
Affy500K	50.0	63.0		48.1	50.0
AffySNP6	43.6	63.6	56.4		65.5
Agilent244K	40.8	50.0	36.8	35.5	

Deletion (before adjustment)

	Affy 100K	Affy 250K	Affy 500K	Affy SNP6	Agilent 244K
Affy100K		17.1	31.7	34.1	48.8
Affy250K	50.0		80.0	40.0	60.0
Affy500K	42.9	47.6		57.1	66.7
AffySNP6	27.6	22.4	27.6		36.8
Agilent244K	48.2	37.5	39.3	53.6	

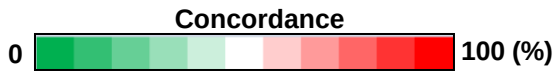
B

Amplification (after adjustment)

	Affy 100K	Affy 250K	Affy 500K	Affy SNP6	Agilent 244K
Affy100K		58.0	54.0	58.0	74.0
Affy250K	52.6		55.3	52.6	57.9
Affy500K	57.1	67.9		58.9	69.6
AffySNP6	61.8	67.3	63.6		69.1
Agilent244K	55.4	63.5	54.1	50.0	

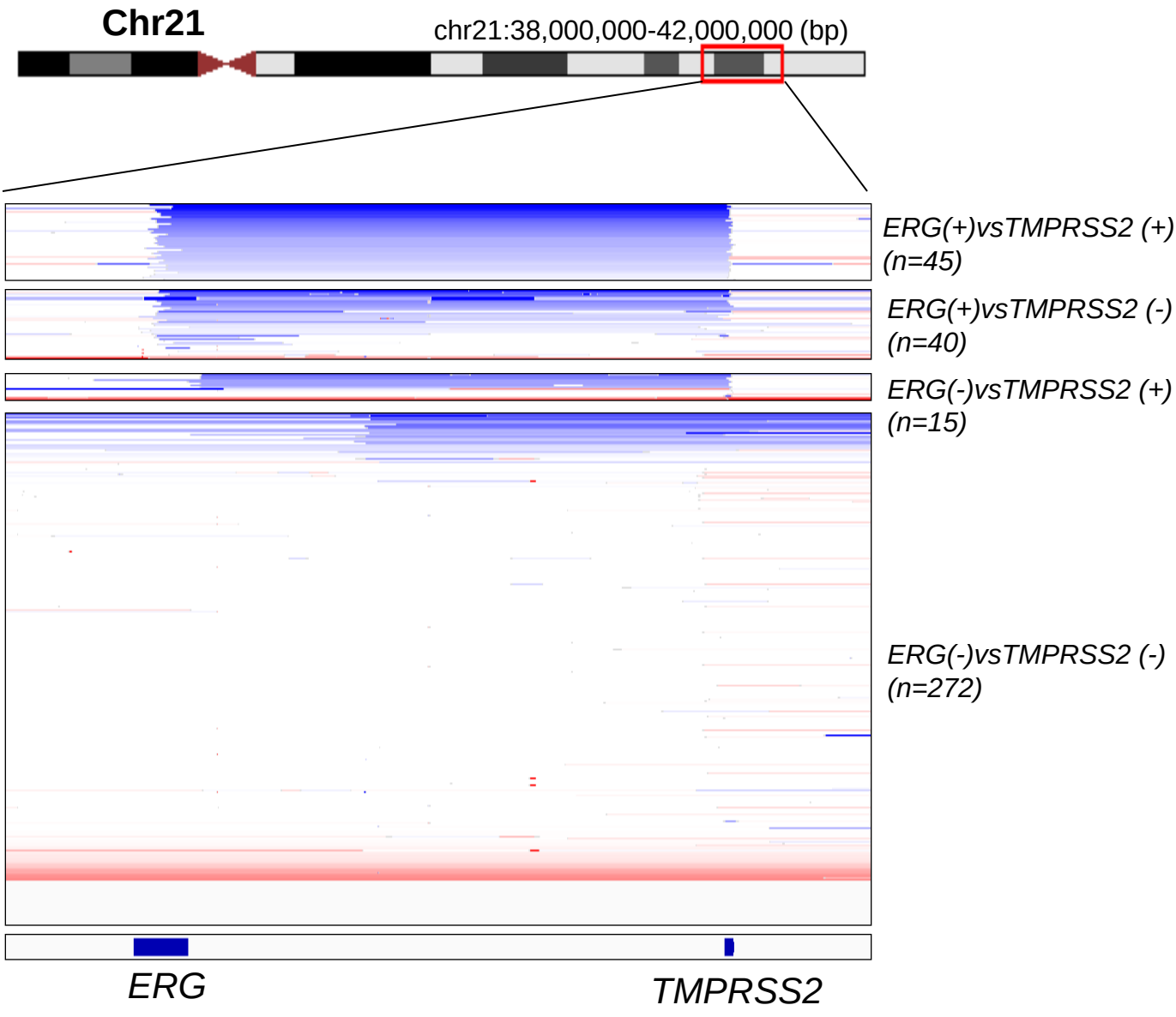
Deletion (after adjustment)

	Affy 100K	Affy 250K	Affy 500K	Affy SNP6	Agilent 244K
Affy100K		26.7	42.2	40.0	48.9
Affy250K	50.0		85.7	57.1	71.4
Affy500K	55.6	63.0		66.7	74.1
AffySNP6	36.8	27.6	38.2		51.3
Agilent244K	61.8	41.8	50.9	63.6	



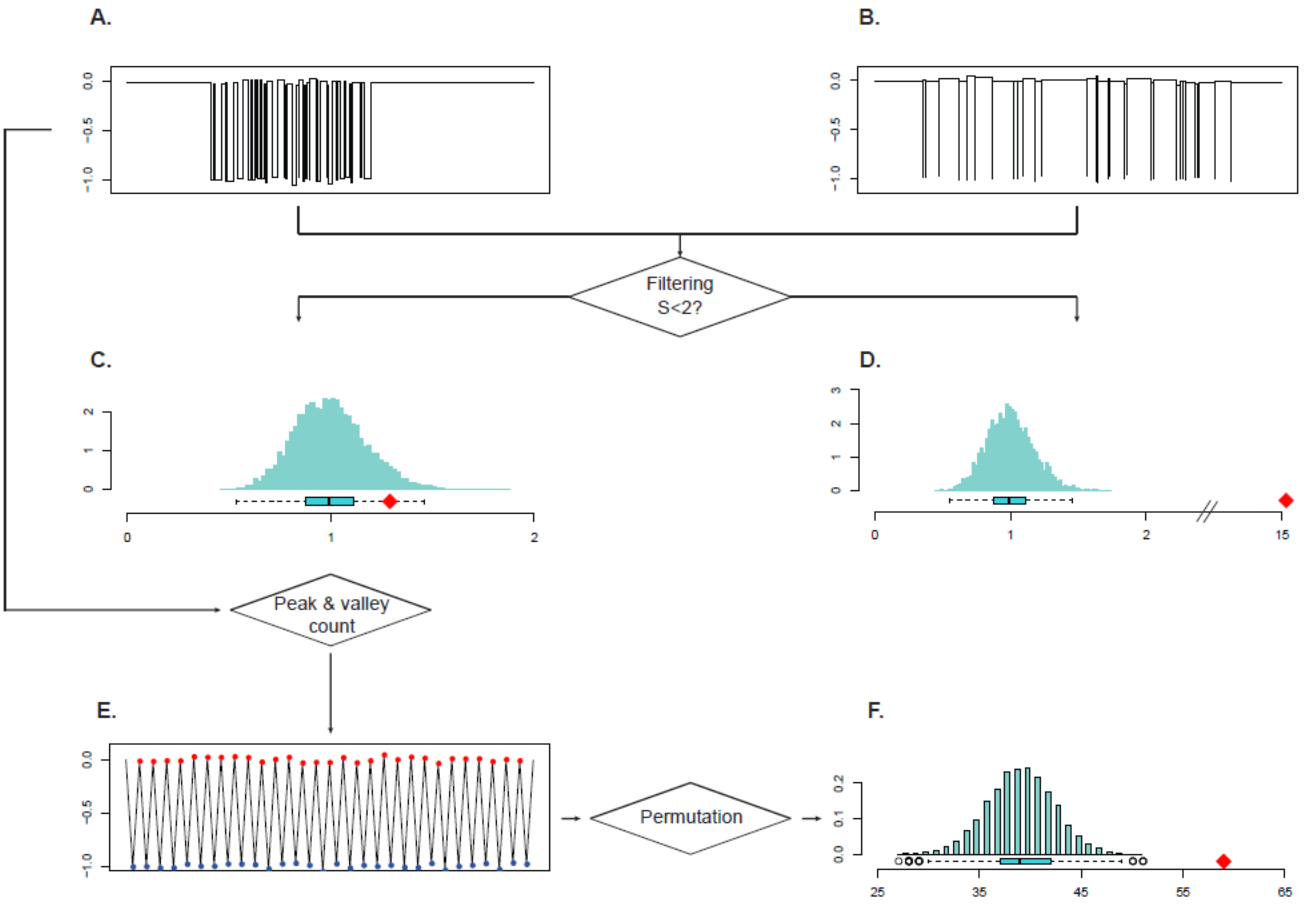
Supplementary Figure 6. The removal of potential platform biases and the impact on the level of concordance between platform-specific peaks. (A) The concordance levels (i.e., the extent of overlap between GISTIC peaks) were measured across the different platforms for amplifications (above) and deletions (below). The first row shows that 46.9% and 44.9% of peaks from Affymetrix 100K platform were rediscovered in the analysis using only Affymetrix 250K and 500K platforms, respectively. (B) To remove the potential platform biases, we employed a linear mixed model. The adjusted segmentation profiles show an overall increase in the extent of overlap between those identified from each of the platforms. Red and green are for amplification and deletion peaks, respectively.

Supplementary Fig. 7



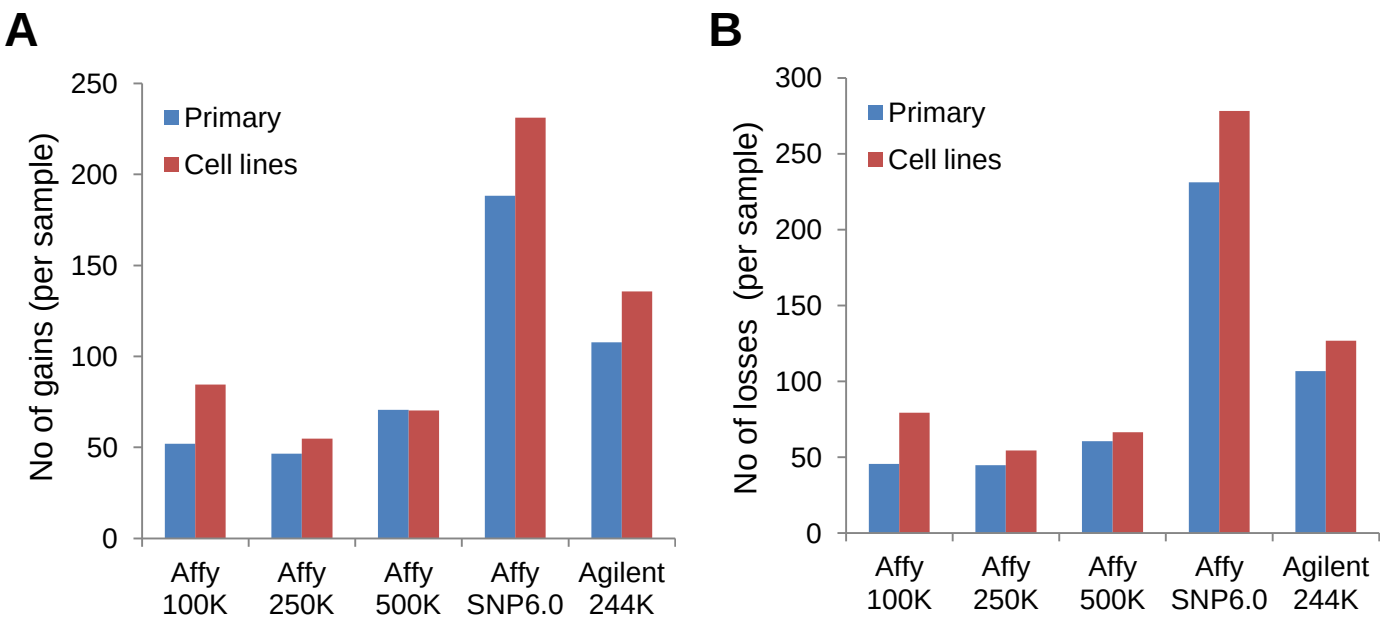
Supplementary Figure 7. Copy number profile of the *TMPRSS2-ERG* loci in prostate cancer. The 372 prostate cancers were classified according to the presence of intragenic deletion breakpoints in the *ERG* and *TMPRSS2* loci (chromosome 21). ERG (+/-) and TMPRSS2 (+/-) cases are those with/without intragenic deletion breakpoints within the *ERG* and *TMPRSS2* loci, respectively. The significance for co-occurrence of intragenic breakpoints in both loci was calculated using Fisher's exact test ($P = 9.1 \times 10^{-25}$). The 45 deletions encompassing *ERG* (3') and *TMPRSS2* (5') follow the previously proposed 'fusion breakpoint principle.'

Supplementary Fig. 8



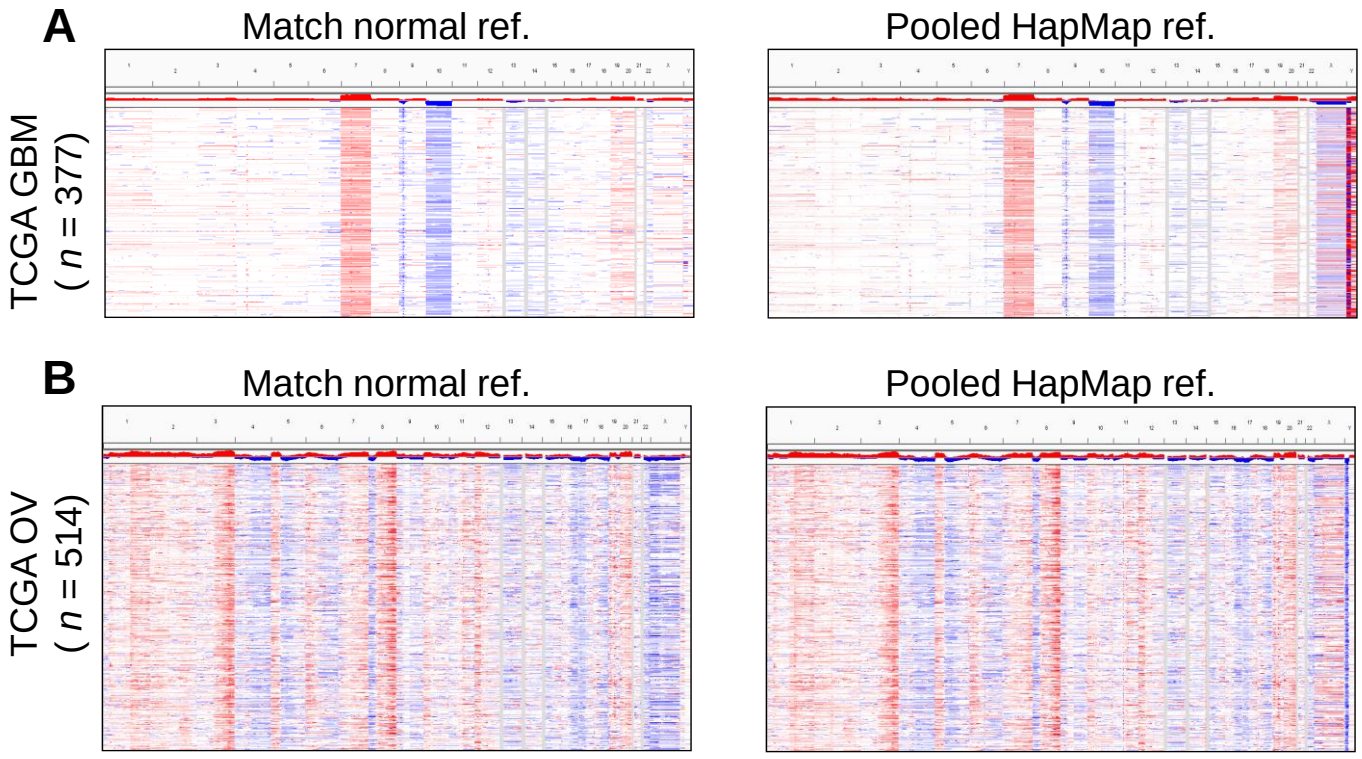
Supplementary Figure 8. A schematic of the algorithm to detect chromothripsis. Given a set of candidate profiles (A and B), we first filter out the profiles that are unlikely to be chromothripsis based on the *S* statistic as described in **Methods**. For the profile A, the *S* statistic is 1.3 and hence A will be kept for further analysis. However, the *S* statistic for B is 15, which is far larger than the cutoff value 2, and thus the profile B is filtered out. The expected distribution of the *S* statistic under the hypothesis that the breakpoints are randomly distributed is shown in C and D for A and B, respectively. The red diamonds in C and D are the observed *S* statistic correspondingly. After the filtering, we count the number of peaks and valleys in the profile. In E, peaks are shown in red and valleys are shown in blue. Of note, peaks (valleys) are defined as segments whose log copy ratios are larger (less) than those of their direct neighboring segments. Lastly, a p-value is assigned to each remaining profile by permuting the order of the segments and comparing the peak/valley count of the observed profile with the permuted profile (F). The red diamond in (F) is the observed peak/valley count and the bars represent the distribution of the peak/valley counts of the permuted profile.

Supplementary Fig. 9



Supplementary Figure 9. The number of alterations between primary cases and cell lines. Cell lines show a substantially higher number of gains (A) and losses (B) compared to primary cases across the different array-CGH platforms.

Supplementary Fig. 10

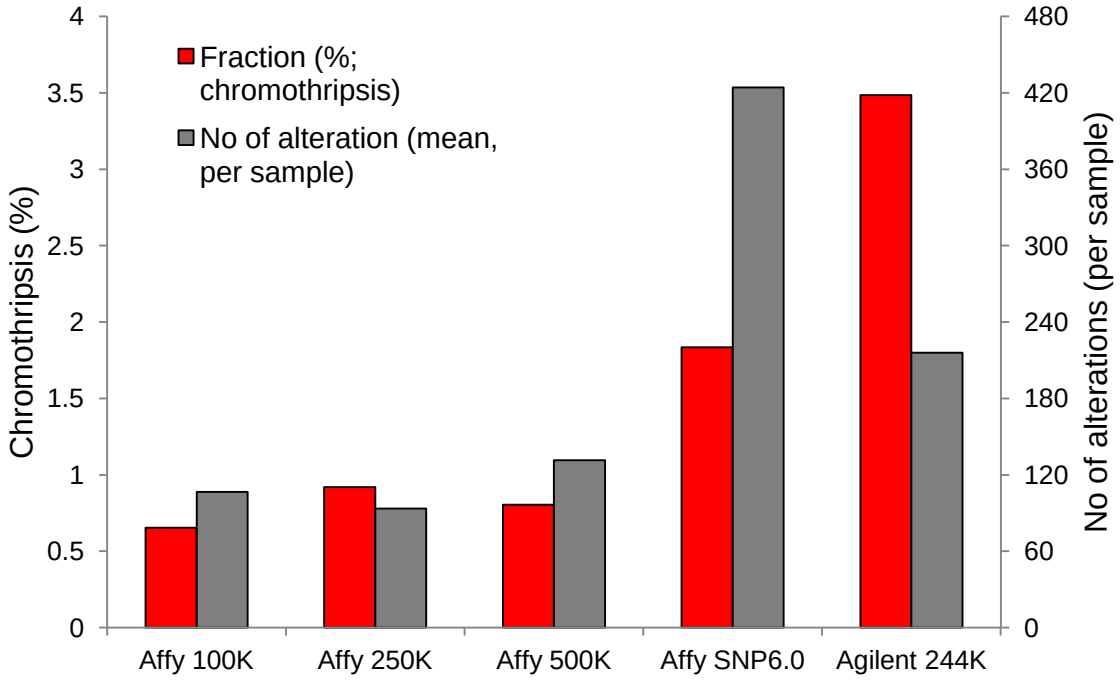


C

	Concordance		Concordance for pan-lineage MCRs		Pan-lineage MCRs in non-overlapping peaks	
	Amplification	Deletion	Amplification	Deletion	Amplification	Deletion
TCGA_GBM (matched Ref)	69% (50/72)	44% (44/100)	54% (39/72)	76% (76/100)	9% (2/22)	8% (5/56)
TCGA_GBM (pooled Ref)	67% (49/73)	56% (46/82)	56% (41/73)	83% (68/82)	8% (2/24)	8% (3/36)
TCGA_OV (matched Ref)	74% (35/47)	45% (45/99)	81% (38/47)	77% (76/99)	8% (1/12)	9% (5/54)
TCGA_OV (pooled Ref)	73% (36/49)	60% (35/58)	73% (36/49)	83% (48/58)	15% (2/13)	13% (3/23)

Supplementary Figure 10. Comparison of using pooled HapMap reference vs matched normal controls for TCGA GBM and Ovarian cancers. (A) The IGV snapshots show the copy number profiles (red and blue for chromosomal gains and losses, respectively) of 377 TCGA GBM samples using matched normal controls (left) and using the pooled HapMap reference (right). (B) Similarly for 514 TCGA ovarian cancers. (C) The overlap in alteration peaks between matched controls and pooled HapMap references ('Concordance') or pan-lineage MCRs ('Concordance for pan-lineage MCRs') was counted. The first row, for example, shows that 72 and 73 amplification peaks were observed in the matched and pooled references and that 50 out of 72 and 49 out of 73 are shared, corresponding to 69% and 67%. We also counted the number of pan-lineage MCRs in non-overlapping peaks (i.e., peaks exclusive to matched or pooled references). For example, among the 22 and 56 amplification and deletion peaks observed for GBM using matched references, 9% and 8% showed overlap with pan-lineage MCRs. The percentages are similar between the matched and pooled cases, but the number of peaks is too small to draw a conclusion.

Supplementary Fig. 11



Supplementary Figure 11. The frequency of chromothripsis in different array-CGH platforms. The frequencies of chromothripsis are shown for the five array-CGH platforms (red). The average number of alterations per sample is also shown (grey).