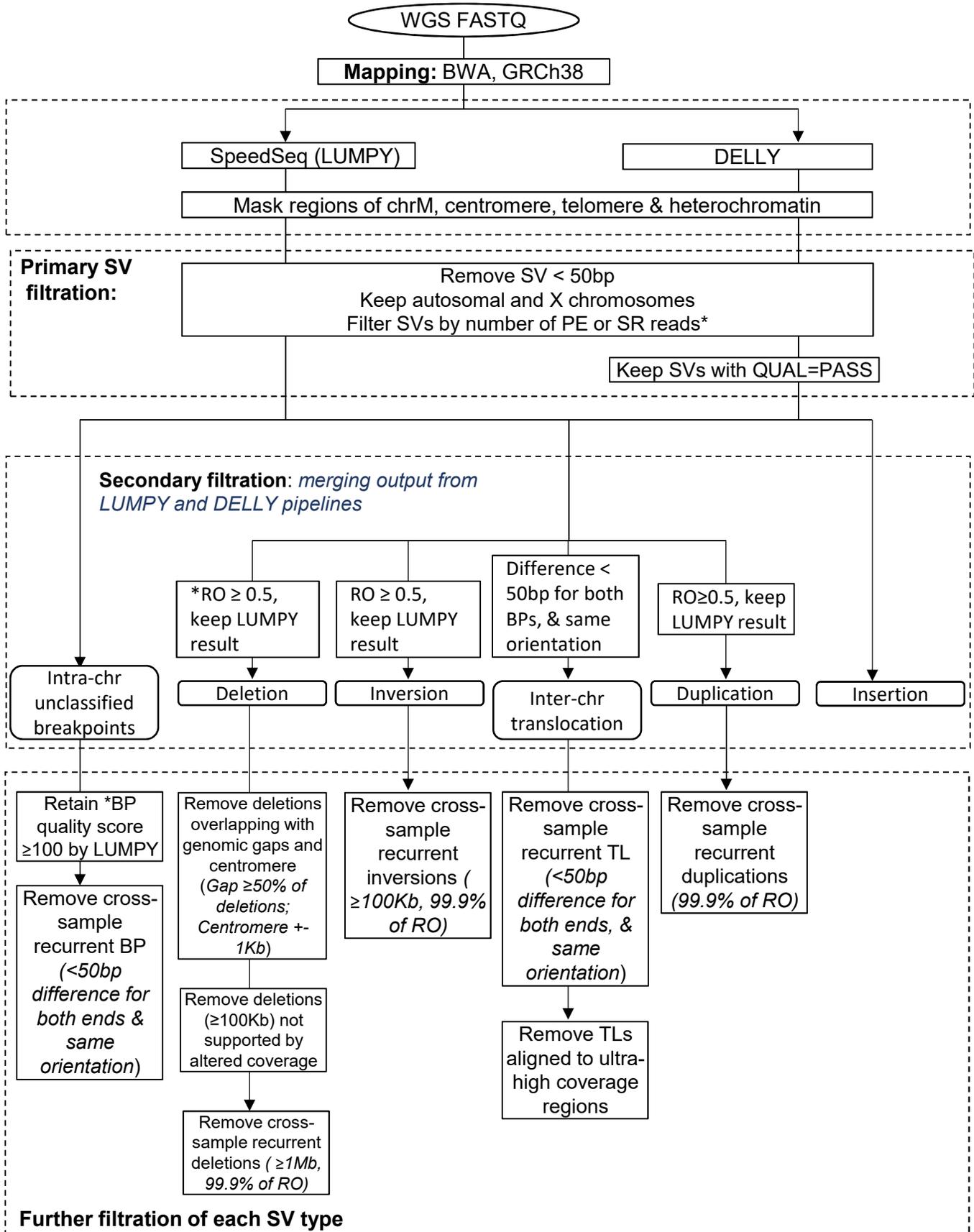


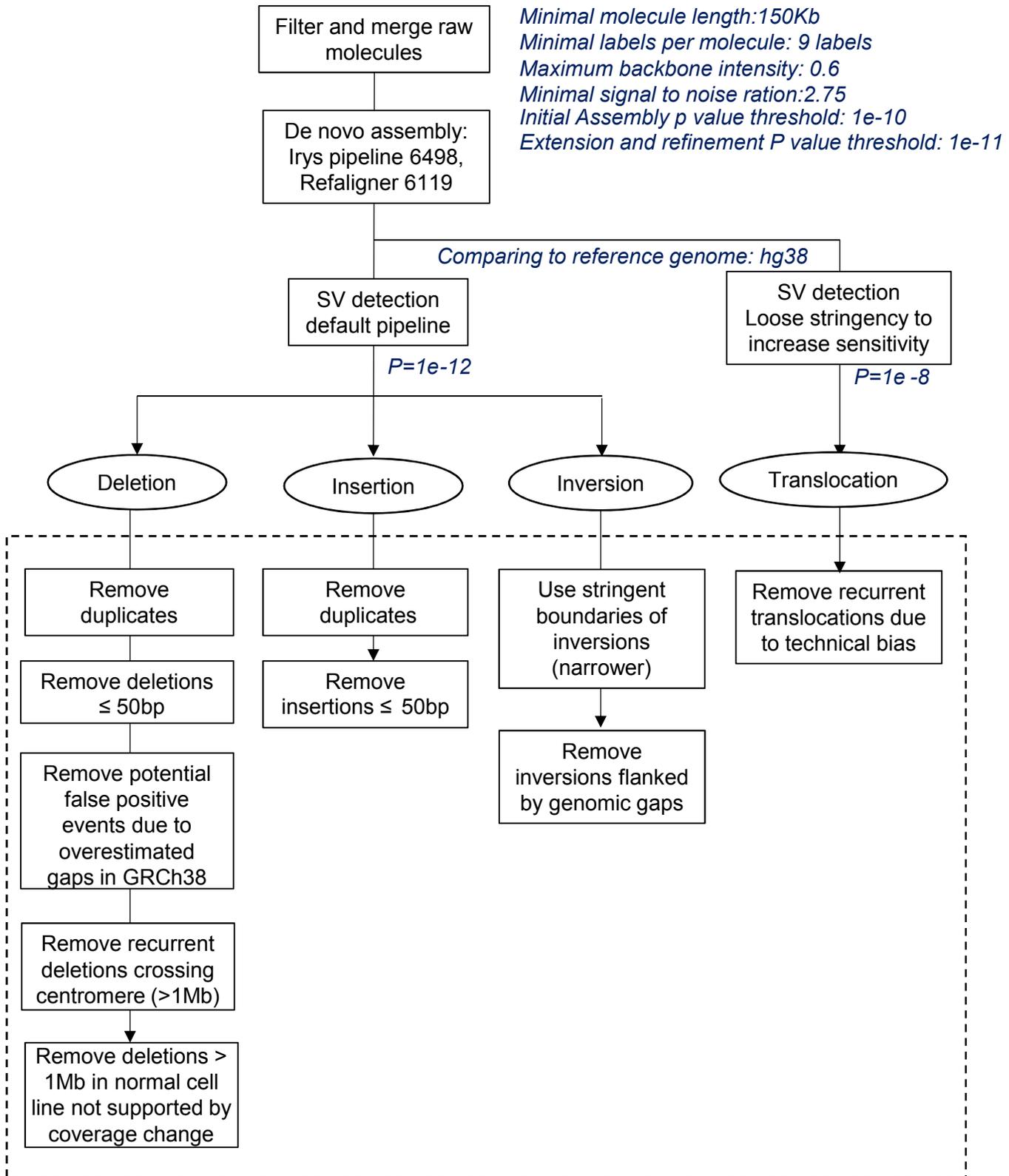
## Supplementary Figures

- Supplementary Figure 1: Pipeline of structural variants detection and filtration by WGS.
- Supplementary Figure 2: Pipeline of structural variants detection and filtration by optical mapping.
- Supplementary Figure 3: Overview of methodology for finding translocations in cancer genomes using Hi-C data.
- Supplementary Figure 4: Hi-C demarcates inversions, deletions and tandem duplication.
- Supplementary Figure 5: Cancer genomes possess extensive CNVs and translocations.
- Supplementary Figure 6: Identification of re-arrangement breakpoints at kilobase resolution.
- Supplementary Figure 7: FISH validation of translocations detected by Hi-C.
- Supplementary Figure 8. Assessment of sensitivity and internal reproducibility of intra-chromosomal re-arrangements identified by Hi-C.
- Supplementary Figure 9. Validation of breakpoints using replication timing.
- Supplementary Figure 10: Comparative integration of inter-chromosomal translocations and large intra-chromosomal SVs ( $\geq 1\text{Mb}$ ).
- Supplementary Figure 11: Contribution to overall profile of large structural variants in T47D and K562 cells by different methods.
- Supplementary Figure 12: Cross-platform validation of structural variants detected by Hi-C using other five methods.
- Supplementary Figure 13: Deletions predicted by Irys overlap with multiple smaller WGS predicted deletions.
- Supplementary Figure 14: Examples of large deletions in normal and cancer genomes.
- Supplementary Figure 15: Overlap of large SVs detected by Hi-C, optical mapping, and WGS.
- Supplementary Figure 16: Hi-C and optical mapping detect translocations with un-alignable junctions.
- Supplementary Figure 17: Examples of using Hi-C and optical mapping to reconstruct the overall structure of complex translocations.
- Supplementary Figure 18: Impact of exon deletion and copy loss on gene expression.
- Supplementary Figure 19: Copy number alterations of COSMIC tumor-related genes, which are computed based on its surrounding 50 kb regions by optical mapping.
- Supplementary Figure 20: List of non-COSMIC tumor-related genes that have significant copy number changes.
- Supplementary Figure 21: Comparison of the frequency of enhancer disruption versus expectation.
- Supplementary Figure 22: Characterization of known polymorphic deletions and novel deletions.
- Supplementary Figure 23: Genome-wide CNVs predicted by optical mapping and WGS are consistent.
- Supplementary Figure 24: Annotation of known recurrent structural variants in cancer cell lines.

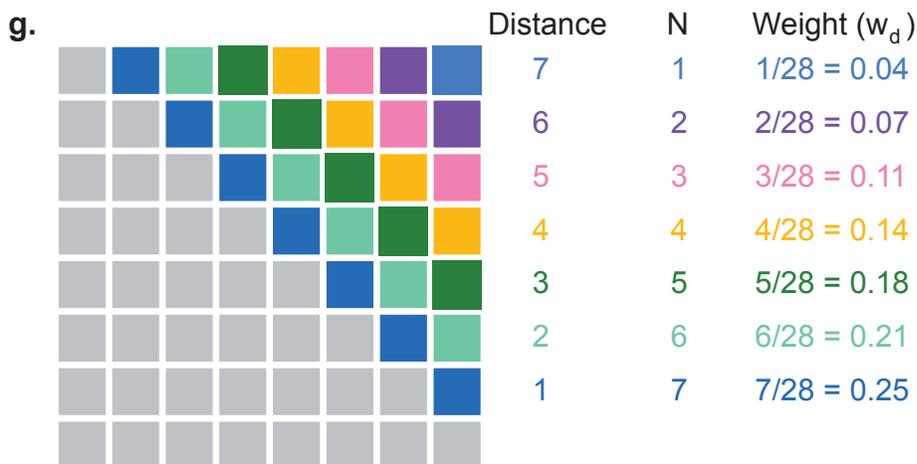
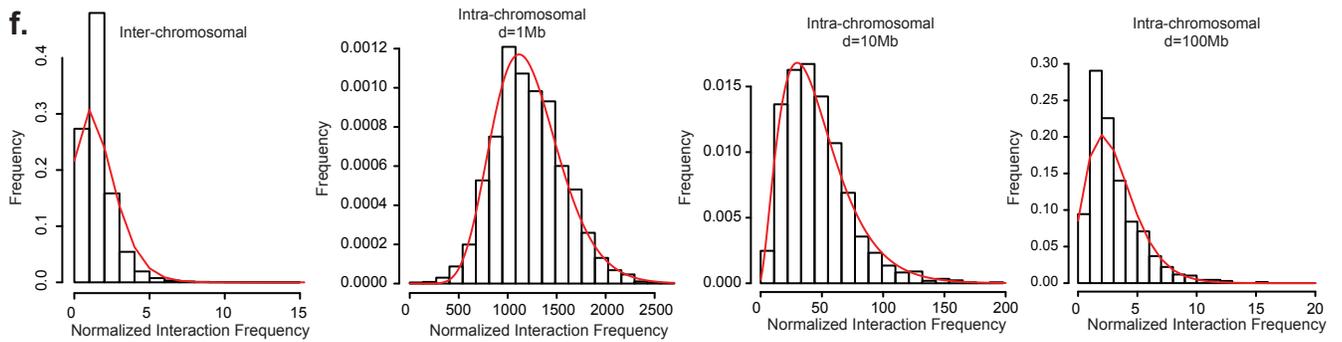
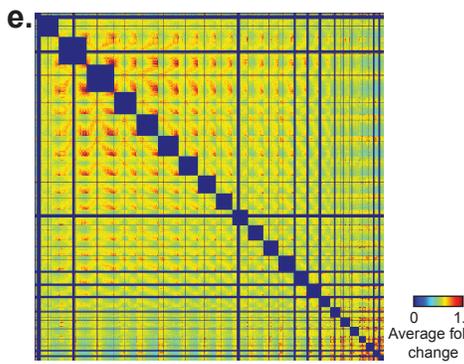
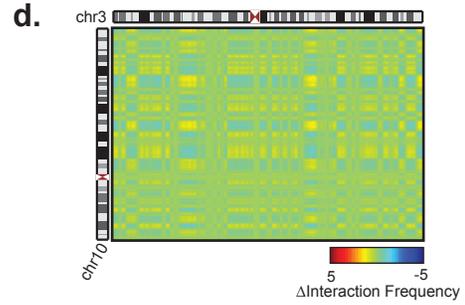
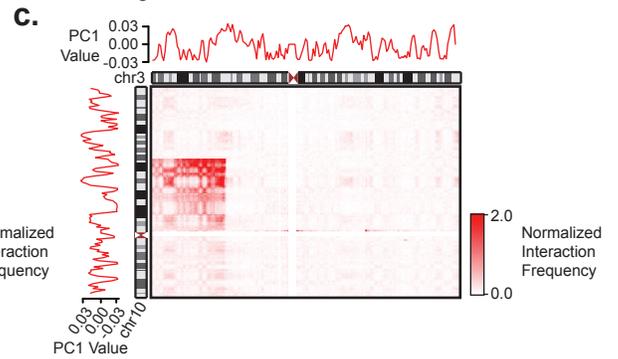
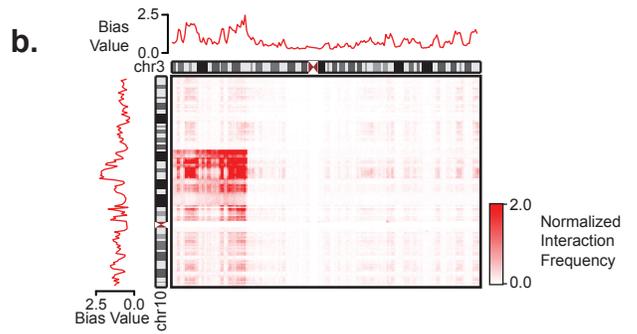
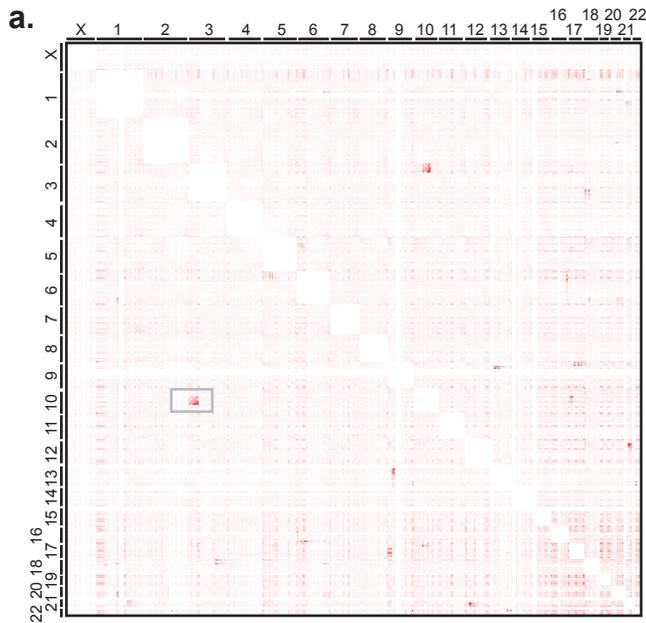


\* Number of PE or SR reads are adjusted to the WGS coverage and the ploidy of cell lines  
The procedure of Control FREEC is described in the methods section  
\* RO: ratio of reciprocal overlap. \*BP: breakpoint

Supplementary Figure 1 | Pipeline of structural variants detection and filtration by WGS.

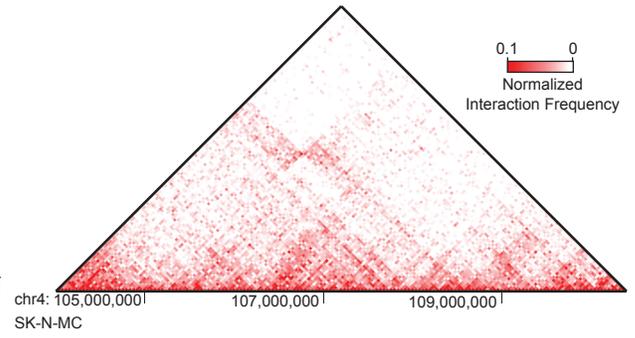
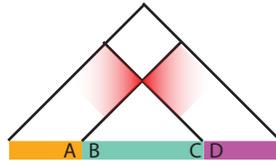
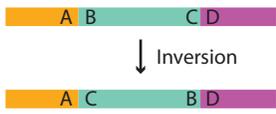


**Supplementary Figure 2 | Pipeline of structural variants detection and filtration by optical mapping.**

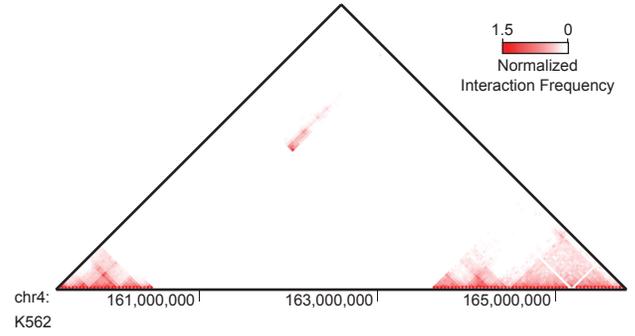
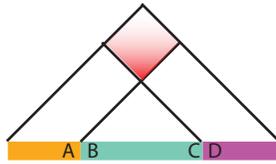
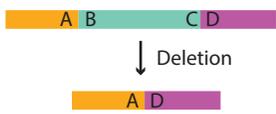


**Supplementary Figure 3 | Overview of methodology for finding translocations in cancer genomes using Hi-C data.** **a.** Whole genome Hi-C interaction heat map from K562 cells. Strong local increases in Hi-C signal between heterologous chromosomes likely represent translocations. The grey box is the outline of the interaction between chromosome 3 and 10, where there is a known translocation in K562 cells. **b.** Raw heat map of interaction frequencies between chromosomes 3 (x-axis) and 10 (y-axis). The red lines represent the intrinsic vector of biases calculated for each bin. **c.** Normalized heat map between chromosomes 3 and 10. Red lines represent the first principal component values for each bin. **d.** Matrix showing the additive increase or decrease in interaction frequency due to the A/B compartment signal. **e.** Heat map of the average inter-chromosomal interaction frequency across 9 normal cell lines after subtracted the A/B compartment signals. **f.** Histogram of interaction frequencies in K562 cells for interactions separated by 1Mb, 10Mb, and 100Mb. The red line shows the fitted probability density function. **g.** Example diagram of how the weights are calculated for the negative binomial mixture model. There are 7 different pixels in the matrix separated by a distance of 1 bin (dark blue boxes near the diagonal), 6 pixels in the matrix separated by 2 bins, and so on. In total, there are 28 bins, so the weight of each distance is calculated by taking the number of pixels at that distance and dividing it by the total number of pixels.

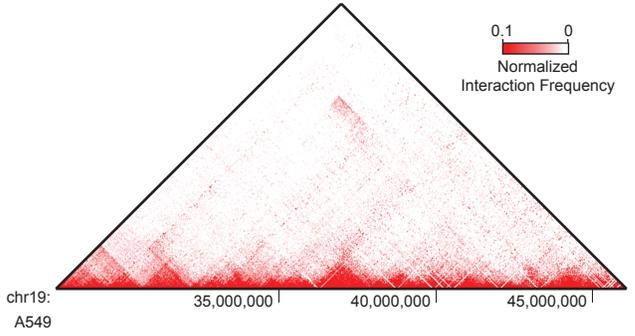
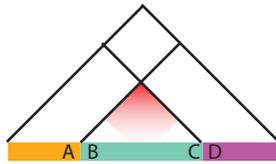
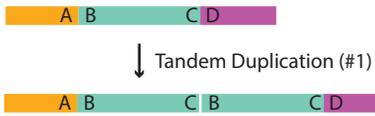
a.



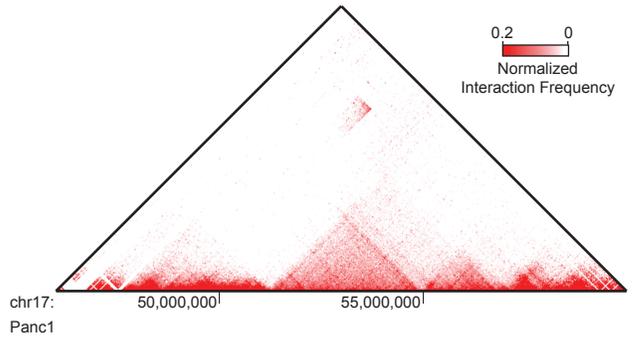
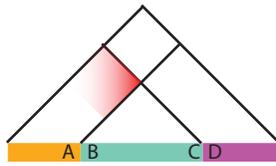
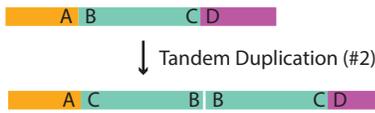
b.



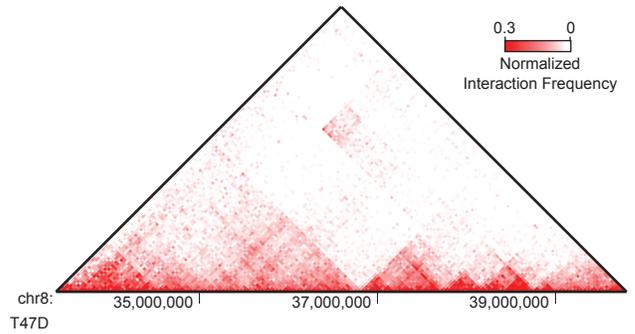
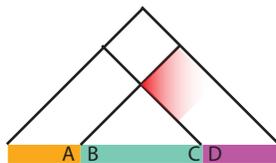
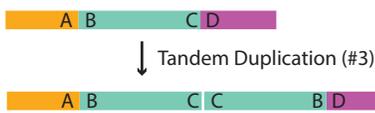
c.



d.

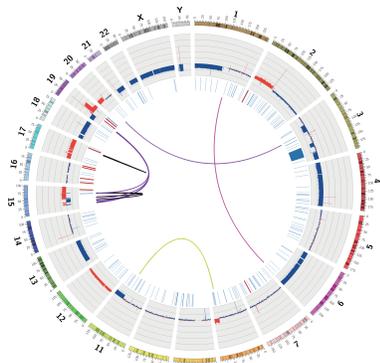


e.



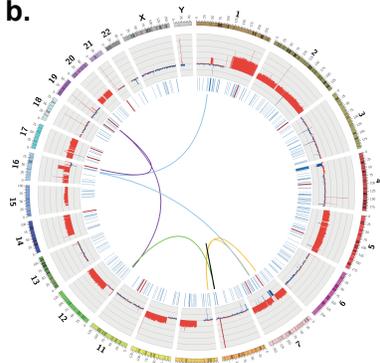
**Supplementary Figure 4 | Hi-C identifies inversions, deletions and tandem duplications.** **a.** Example of an inversion identified by Hi-C. The left hand cartoon shows the example of the genetic structure of an inversion, juxtaposing regions A and C as well as B and D. The cartoon in the middle depicts the expected alteration to chromatin interaction frequencies by such an event, showing increased interaction frequencies between regions A and C or between B and D as a result of the altered linear proximity of these regions (a “butterfly pattern”). The right-hand panel shows an example of an inversion identified in SK-N-MC cells by Hi-C, optical mapping, and WGS. **b.** Example of a deletion and its effects on Hi-C data. The deleted region (left panel), removes the B-to-C region in the diagram, and results in the juxtaposition of the A and D regions, which would result in an increase in the interactions between regions flanking the deleted region (middle panel). The right panel shows a deletion in K562 identified by Hi-C, optical mapping, and WGS. **c-e.** Examples of tandem duplications in Hi-C data, with different orientations of the duplicated region (left-hand diagram), and their expected changes in interaction frequencies (middle panel). The right-hand examples in panel c shows tandem duplications identified in A549 cells by Hi-C, optical mapping, and WGS. The right-hand panel in d shows a tandem duplication in Panc1 cells identified by Hi-C and WGS. Panel e shows a tandem duplication identified in T47D cells by Hi-C and WGS.

**a.**



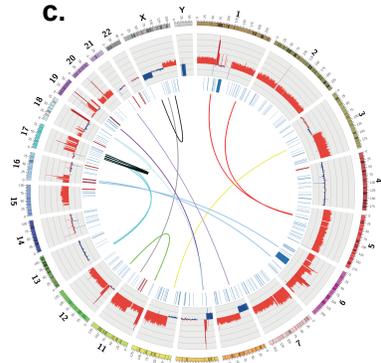
**A549**

**b.**



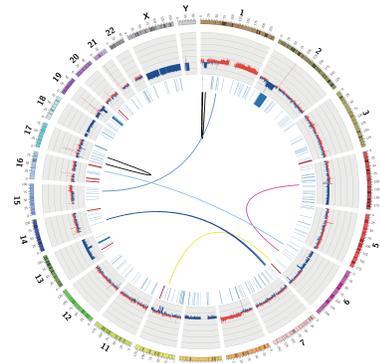
**NCI-H460**

**c.**



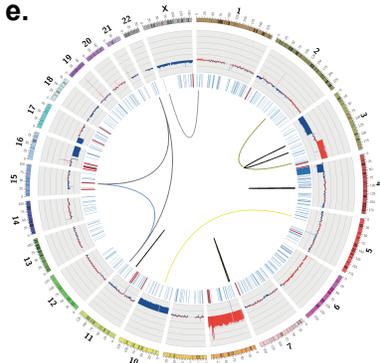
**PANC-1**

**d.**



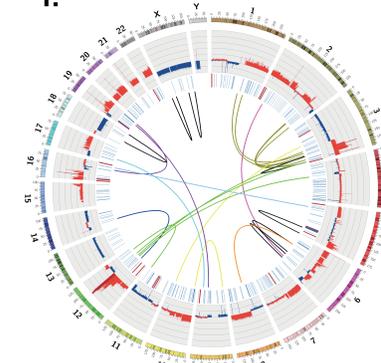
**LNCaP**

**e.**



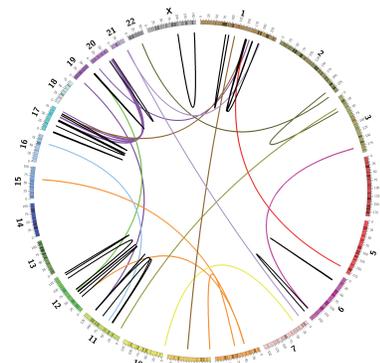
**SK-N-MC**

**f.**



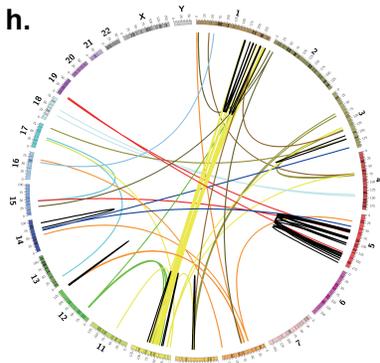
**CAKI2**

**g.**



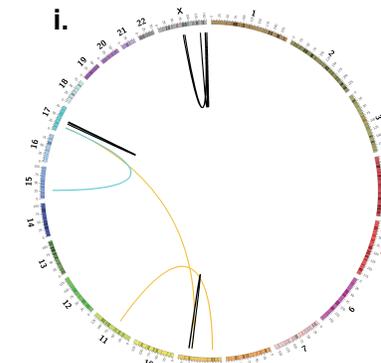
**MCF7**

**h.**



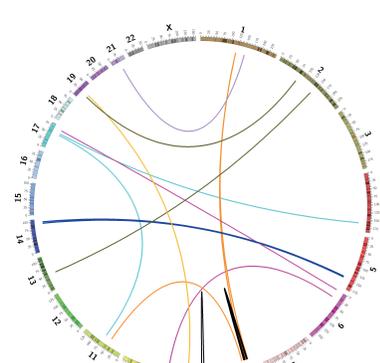
**PC3**

**i.**



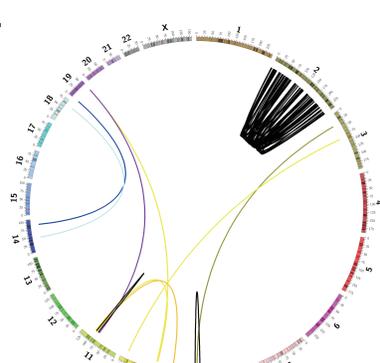
**RPMI-7951**

**j.**



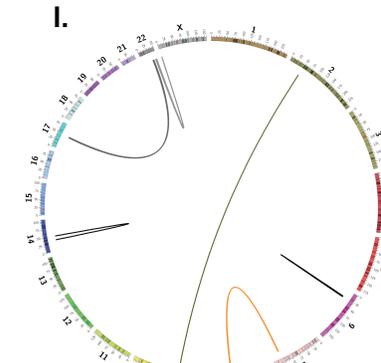
**SK-MEL-5**

**k.**



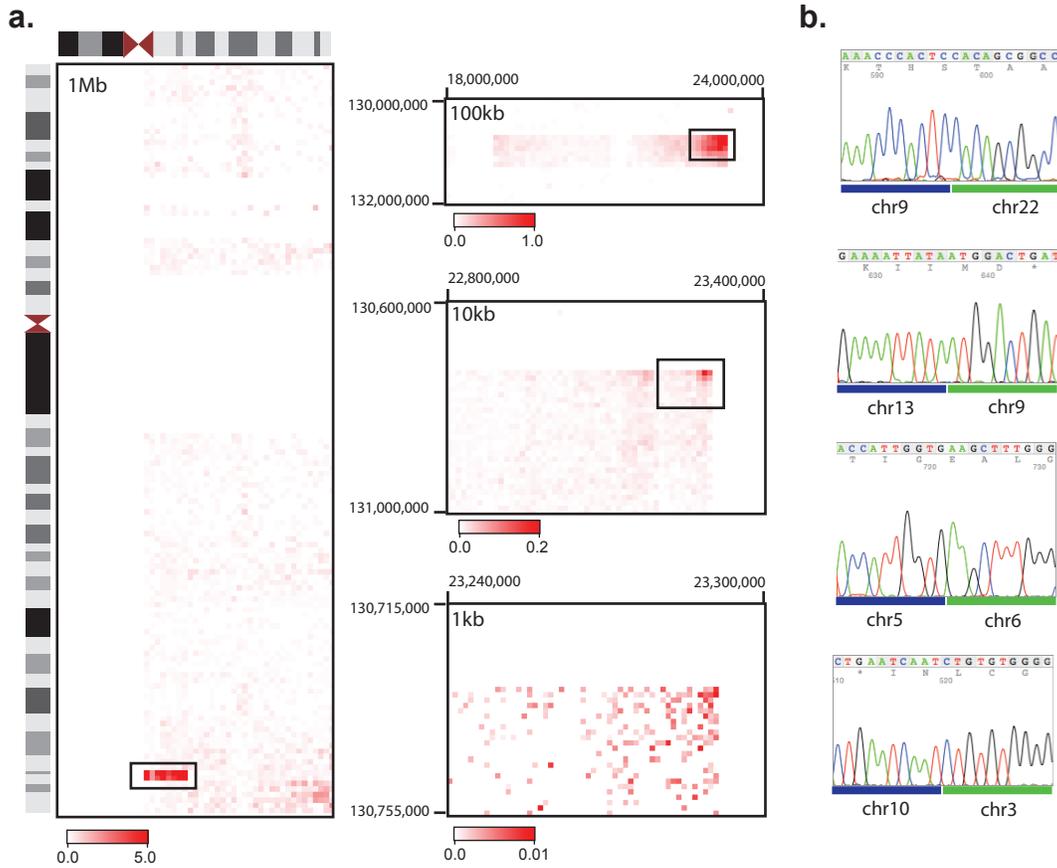
**SK-N-DZ**

**l.**

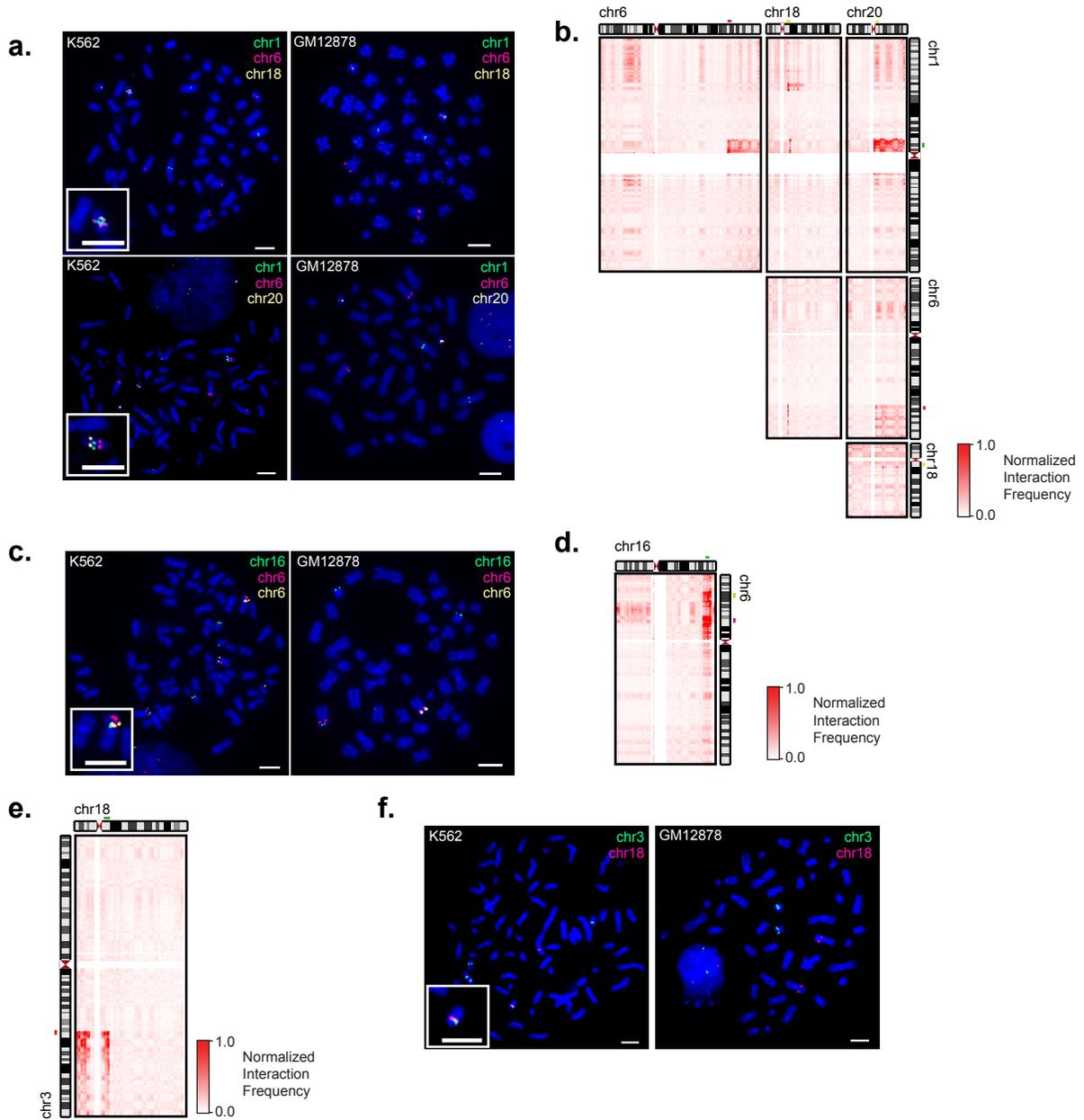


**SK-N-SH**

**Supplementary Figure 5 | Cancer genomes possess extensive CNVs and translocations. a-f.** Genome profiles of 6 cancer cell lines. All SVs are detected by at least two out the three methods (Hi-C, optical mapping and WGS). Tracks from outer to inner circles are chromosome coordinates, CNVs, deletions (blue) and duplications (red), and positional rearrangements including inversions, unclassified rearrangements and inter-chromosomal translocations. Outward red bars in CNV track indicate gain of copies ( $>2$ ), and inward blue loss of copies ( $<2$ ). CNVs are profiled based on WGS data binned at 50-kb resolution. **g-i.** Large intra-chromosomal rearrangements and inter-chromosomal translocations detected by Hi-C in 6 cancer cell lines.



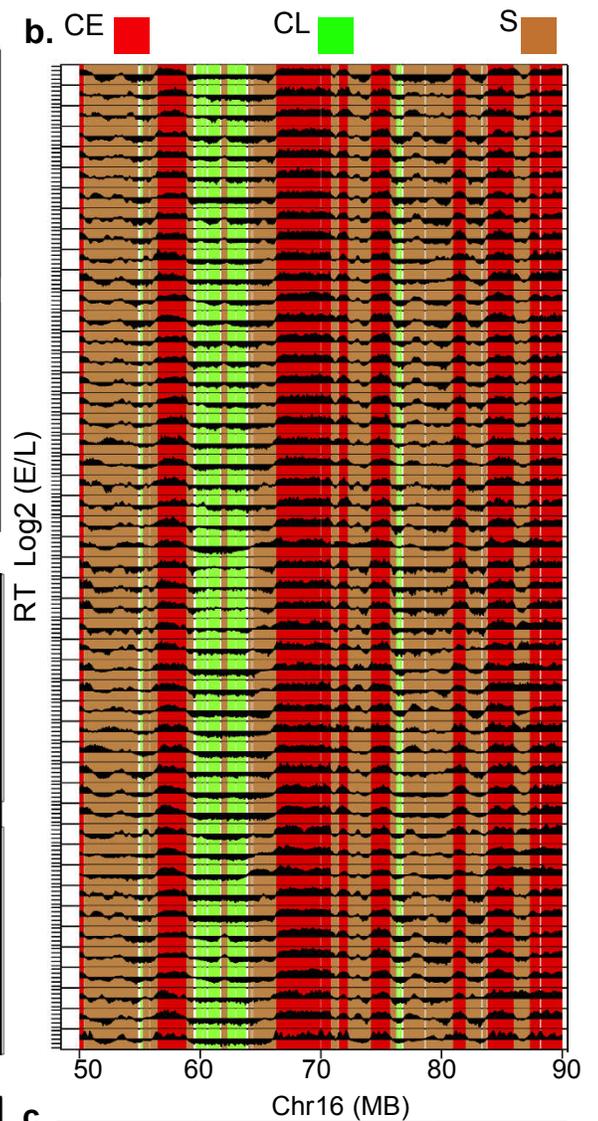
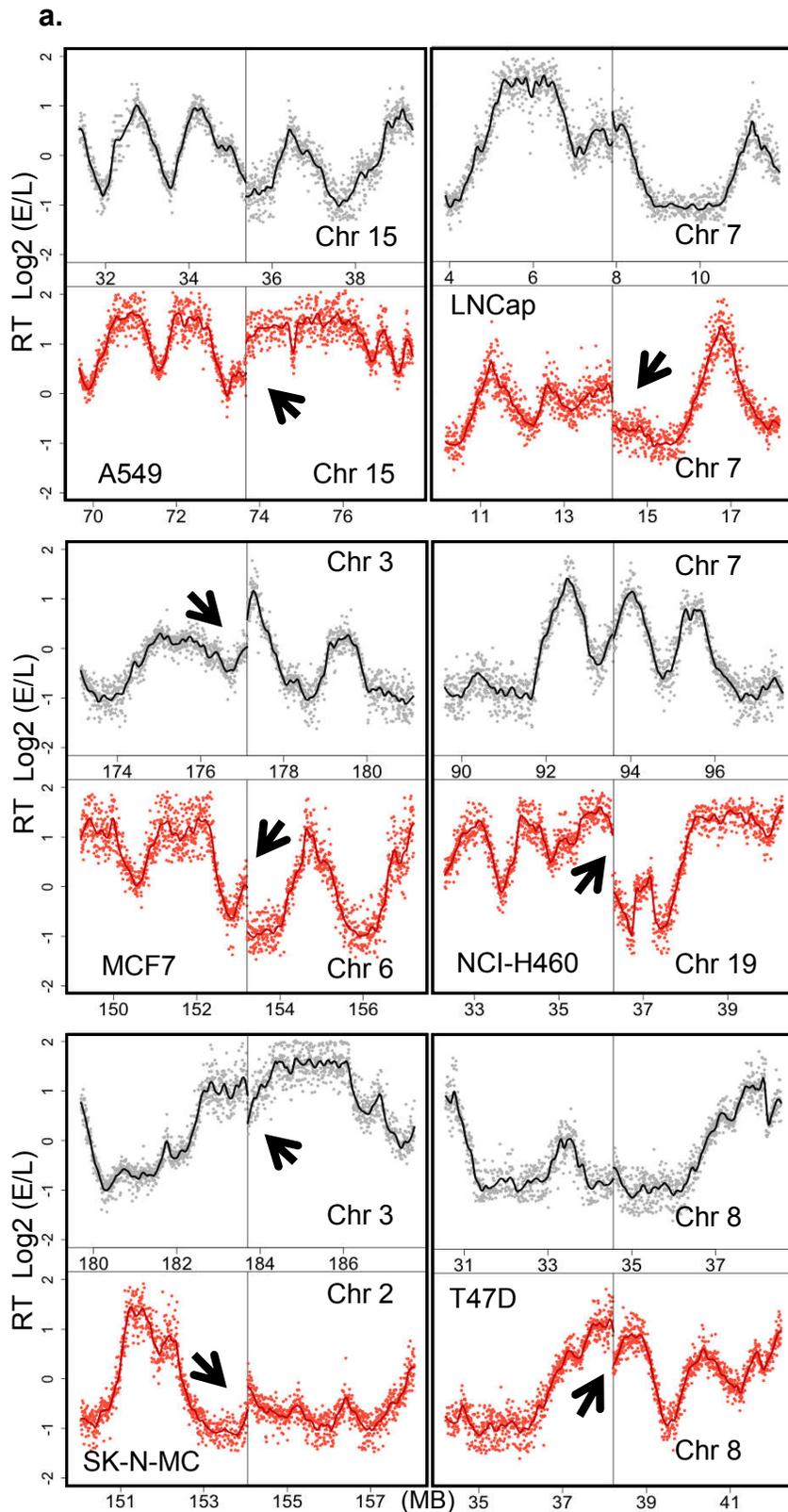
**Supplementary Figure 6 | Identification of rearrangement breakpoints at kilobase resolution by Hi-C. a.** Hi-C signal between chromosomes 9 and 22 in K562 cells. The left hand matrix is shown at a resolution of 1Mb, and the smaller matrices use progressively smaller bin sizes to narrow in on the breakpoint. **b.** Sanger sequencing results confirmed 4 translocations predicted by Hi-C at 1 kb resolution in K562 cells.



**Supplementary Figure 7 | FISH validation of translocations detected by Hi-C.** **a.** Fluorescence in situ hybridization (FISH) data for K562 and GM12878 cells of chromosome 1, 6, 18, and 20. FISH probes targeting chromosomes 1, 6, and 18 are shown in the top panels, and probes targeting chromosomes 1, 6, and 20 are shown in the bottom panels. Similar results for FISH validation experiments were performed using 20 independent metaphase nuclei. Scale bars (white) represent 5  $\mu$ M. **b.** Hi-C signal between chromosomes 1, 6, 18, and 20, showing increased interacting regions which are called as translocations, as well as the locations of the probes used in FISH in panel a. **c.** FISH in K562 and GM12878 cells using probes derived from translocated regions on chromosomes 6 and 16. Similar results for FISH validation experiments were performed using 20 independent metaphase nuclei. Scale bars (white) represent 5  $\mu$ M. **d.** Hi-C signal from between chromosomes 6 and 16 showing the rearrangement and the location of the FISH probes. **e.** Hi-C data showing a translocation between chromosomes 3 and 18. Also shown is the location of FISH probes used for validation. **f.** FISH in K562 (left panel) and negative control GM12878 (right panel) showing rearrangement between chromosomes 3 and 18 in K562 cells. Similar results for FISH validation experiments were performed using 20 independent metaphase nuclei. Scale bars (white) represent 5  $\mu$ M.



**Supplementary Figure 8 | Sensitivity and internal reproducibility of rearrangements identified by Hi-C.** **a.** Hi-C data from human chromosome 21 Tc1 cells. Dotted lines indicate regions shown in panels e and f. **b.** Sensitivity of Hi-C to detect gold standard SV calls at different sequencing depths. SVs were considered as matched if within 10kb with identical strandedness (purple line), within 10kb (pink), within 50kb (green), and within 50kb or internal to the rearrangement region (light blue). **c, d.** Internal consistency of sub-sampled calls at a resolution of 100kb (c) and 10kb (d). The number of reads sub-sampled reads is shown on the axes. **e.** Example of an SV where the breakpoint site (\*) matches but the strandedness does not. Hi-C strandedness is "+/-", while gold standard is "-/-" (red arrows). **f.** Example of a region where Hi-C merged multiple rearrangements together. **g.** Example of an SV with strand discrepancy between Hi-C and WGS in Panc-1 cells (breakpoint marked with an asterisk). Hi-C indicates strandedness as +/+ (red arrows), while WGS indicates -/+ (black arrows). **h.** Diagram of the breakpoint shown in panel g. WGS identifies a small inversion (8kb) on chromosome 11 near the translocation breakpoint, such that the breakpoint lies within the inverted region. As a result, the global structure of the translocation is "+/+" (consistent with Hi-C), while the exact fusion is "-/+". **i.** Sensitivity to detect SVs using K562 (tumor) and GM12878 (normal) Hi-C data mixed at various fractions. **j.** Histogram of SV sizes detected by Hi-C.

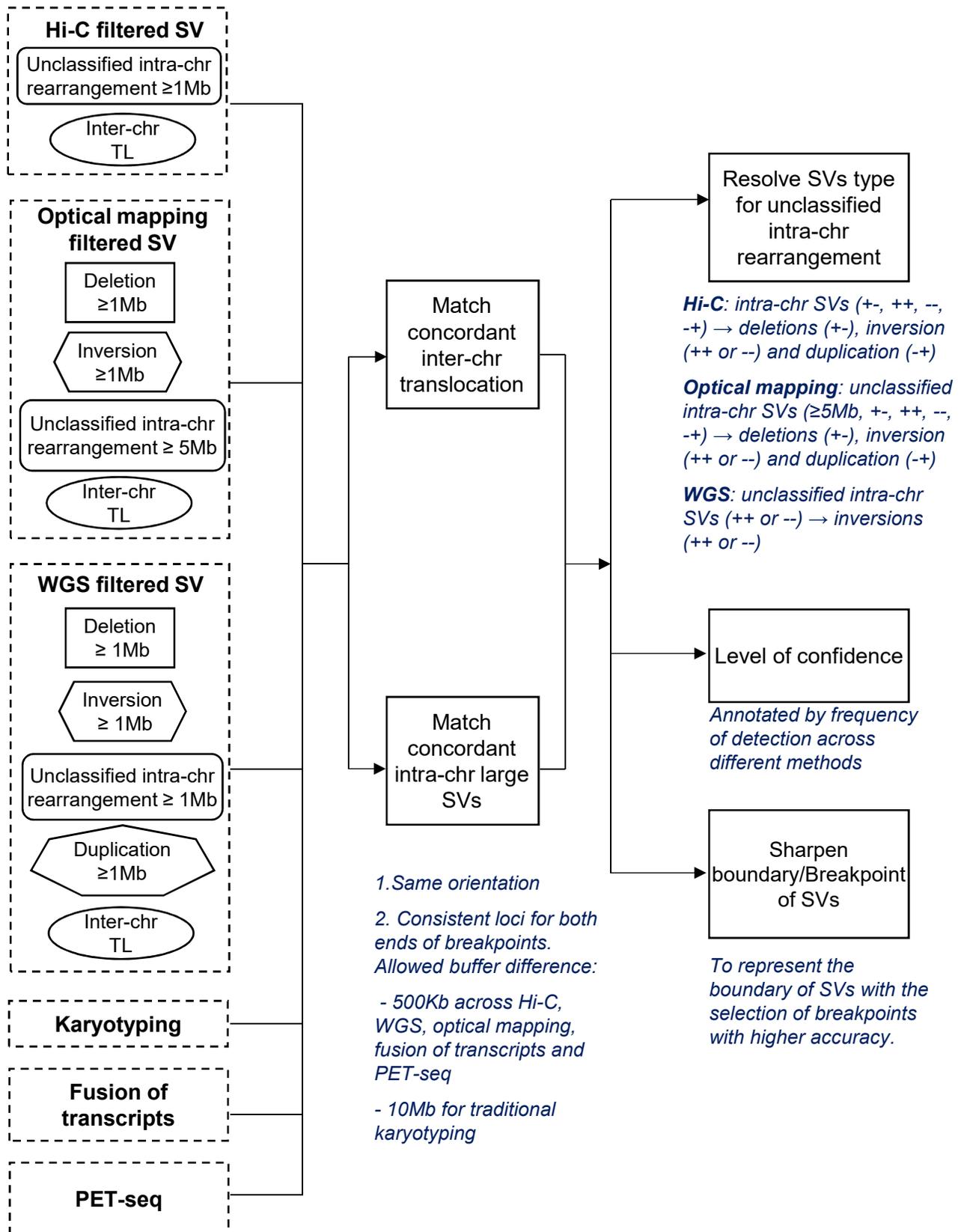


**c.**

Class	Breakpoints		Abrupt shift
	Hi-C	Expected	
CE to CE	27	9.42	3
CL to CL	5	4.50	1
CE to CL	9	13.02	6
S to S	80	61.50	20
S to CE/CL	89	81.41	34
NA	39	79.15	11
<b>Total</b>	<b>249</b>	<b>249</b>	<b>75</b>

**Supplementary Figure 9 | Validation of breakpoints using replication timing.** **a.** Examples of abrupt shift in replication timing for translocations detected by Hi-C when mapped to the reference genome configuration. Abrupt shifts are marked by arrows. Solid black and red lines indicate loess smoothed RT data for the pair of locations involved in each translocation. As RT experiments were designed for validation purposes, one replicate was performed for RT experiments. **b.** Classification of the genome into constitutive early (CE, red), constitutive late (CL, green) or switching (S, brown) segments (50 Kb bins) based on the developmental regulation of replication timing. **c.** Table listing translocations classified into the above categories and the number of abrupt shifts observed in each category.

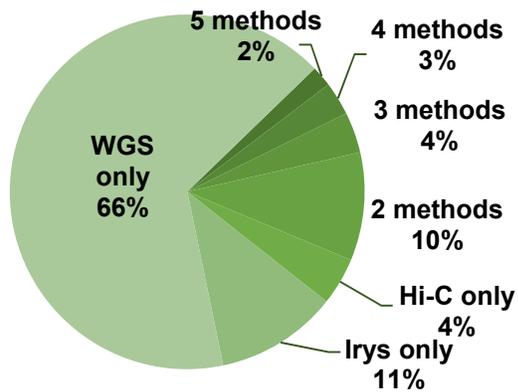
# Comparative integration of inter-chr translocation and large intra-chr SVs (≥1Mb)



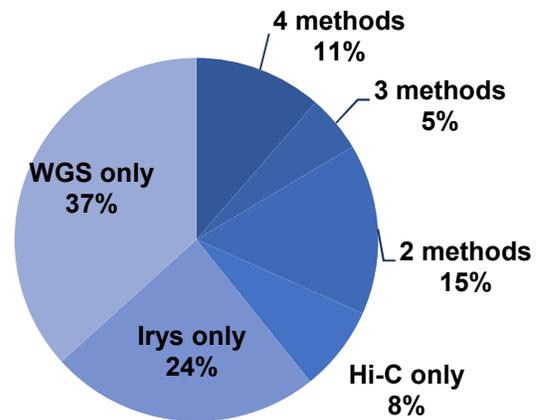
**Supplementary Figure 10 | Comparison and integration of inter-chromosomal translocations and large intra-chromosomal SVs ( $\geq 1\text{Mb}$ ).** We compared the SV calls by Hi-C, optical mapping and WGS, and we also included SV calls from additional methods, including karyotyping, fusion transcripts, and paired-end tag sequencing (PET-seq) when available from the same cell line. For the comparison, we first converted the strand orientation for SVs detected from different methods to a unified system, in which “+” indicates the breakpoint locates at the 3’ end of the joined arm, and “-” indicates the breakpoint at the 5’ end of the joined arm. For WGS data, this dictates that SV originally classified as deletions are given the strand orientation of “+”, inversions as “++ and - -”, duplications as “-+” and unclassified intra-chromosomal rearrangement as “++” or “- -”. Optical mapping originally reports deletions, which are assigned a strand orientation of “+”, inversions as “++” or “- -”, and also intra-chromosomal rearrangements  $>5\text{Mb}$  as “unclassified intra-chromosomal rearrangements” for which the software reports the strand orientation. The same SV from distinct methods is considered a match when they have the same orientation and loci for both ends of breakpoint. The confidence level for each SV is represented by the times that the SV is independently reported by different methods. Further, the breakpoint/boundary of each SV is sharpened by choosing loci determined by the highest-resolution method. Finally, unclassified intra-chromosomal variants from WGS or optical mapping can be re-classified if resolved by an alternative method.

a. T47D:

205 Inter-chr translocation

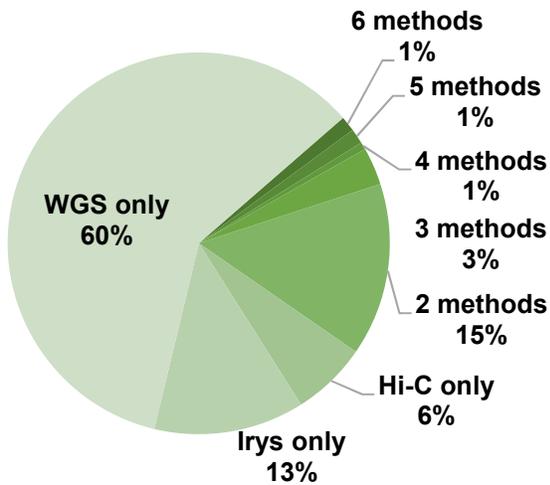


80 Intra-chr large SVs ( $\geq 1\text{Mb}$ )

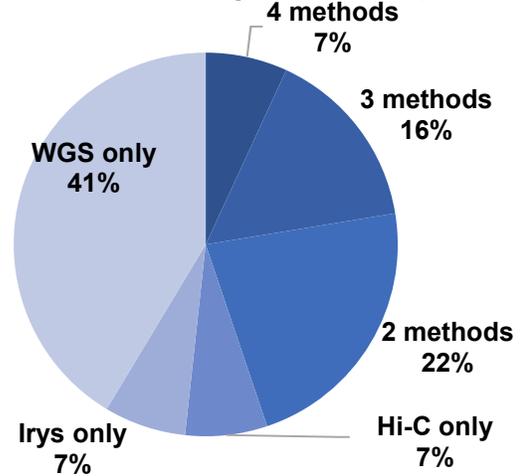


b. K562:

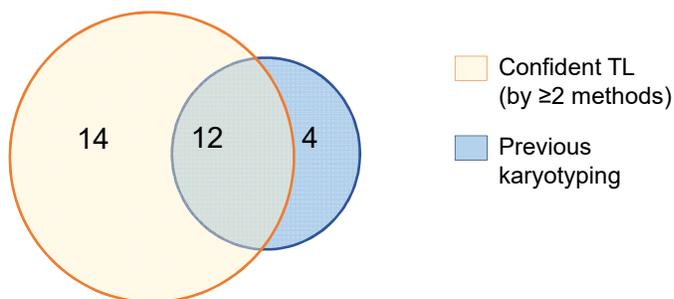
178 Inter-chr translocation



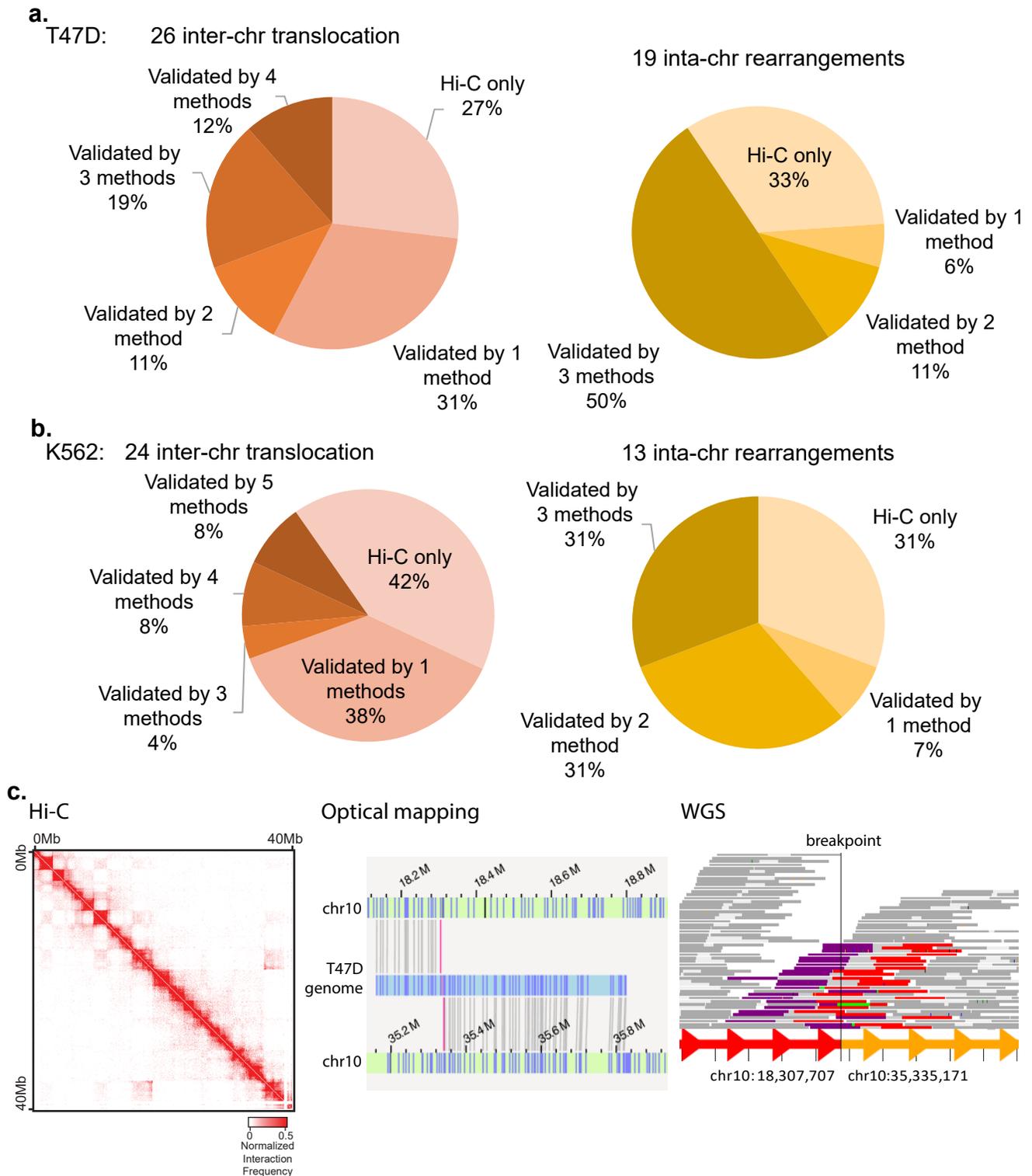
58 Intra-chr large SVs ( $\geq 1\text{Mb}$ )



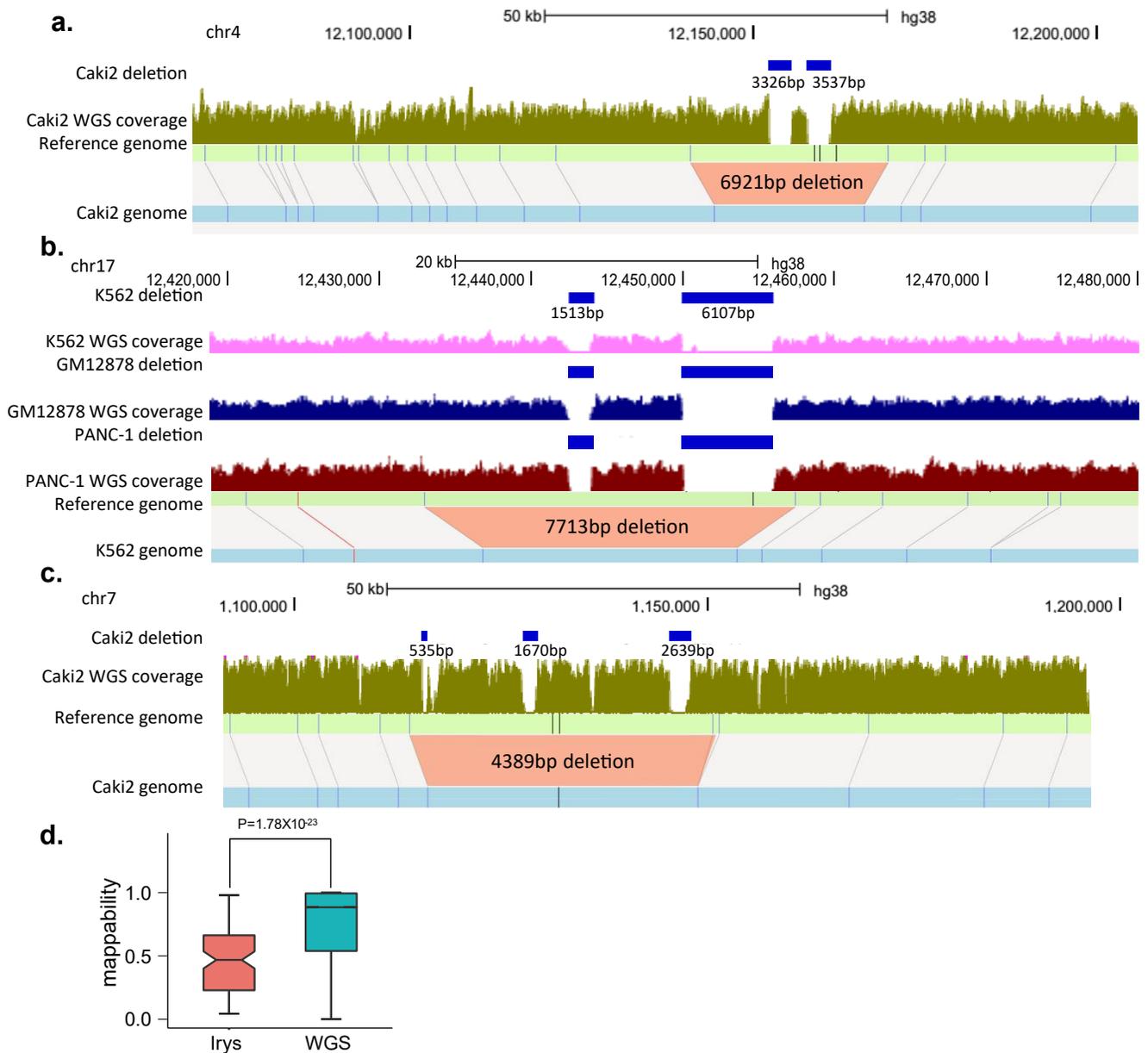
c. T47D



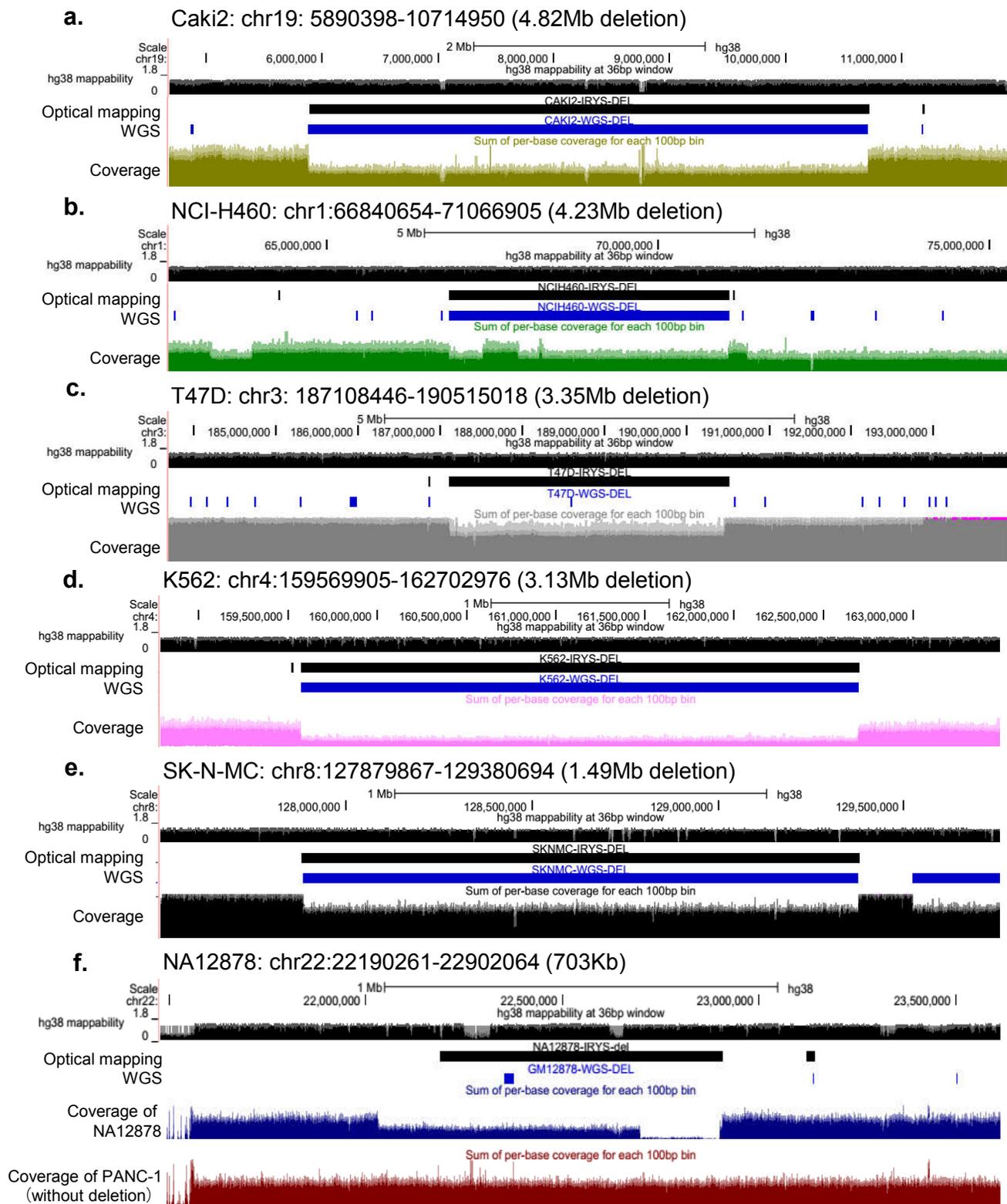
**Supplementary Figure 11 | Contribution to the overall profile of large structural variants in T47D and K562 cells by different methods.** a,b. We first compared inter-chromosomal translocations and large intra-chromosomal rearrangements ( $\geq 1\text{Mb}$ ) detected by Hi-C, optical mapping and WGS and then merged them into a non-redundant union set in T47D (a) and K562 (b). Then we compared this list with karyotyping, fusion of transcripts, and PET-seq and reported how many times each SV was reported by these six methods. c. Comparison of high-confidence TLs detected in this study with previously known karyotypes in T47D cells.



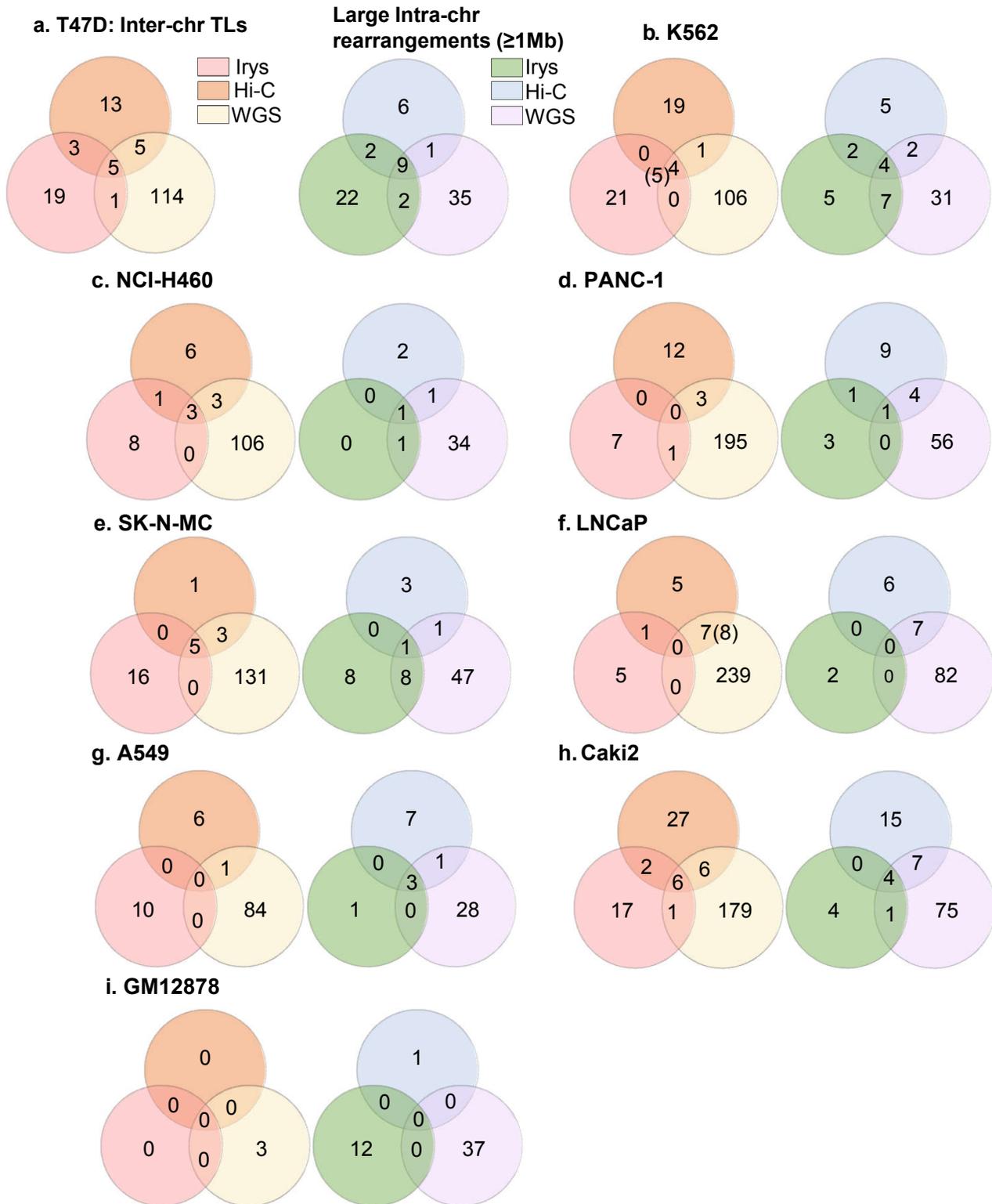
**Supplementary Figure 12 | Cross-platform validation of structural variants detected by Hi-C using other five methods.** Inter-chromosomal translocations and large intra-chromosomal rearrangements ( $\geq 1\text{Mb}$ ) detected by Hi-C are validated by WGS, optical mapping, karyotyping, fusion of transcripts and PET-seq. **a.** The validation rate in T47D cells. **b.** The validation rate in K562 cells. **c.** The same SV detected by all three methods in T47D cells. We report the coordinates of the highest resolution method (in this case WGS) when SVs are identified by two methods. Each line in the WGS panel represents a read pair. Reads that support the breakpoint site are marked as purple (forward strand) and red (reverse strand).



**Supplementary Figure 13 | Deletions predicted by Irys overlap with multiple smaller WGS predicted deletions.** **a.** Optical mapping detects a 6,921 bp deletion within chr4: 12,140,782-12,169,591 in Caki2 cells. In the same region, there are two deletions reported by WGS (Del1: 12152,224-12,155,550, Del2: 12,157,718-12,161,255). The sum of their sizes is 6,863 bp, which is similar to that of the Irys predicted deletion. **b.** Similar as in a, optical mapping detected a shared polymorphic deletion of 7,713bp within chr17:12,432,762-12,457,176 in K562, GM12878 and PANC-1 cells. Again, this deletion can be supported by two smaller deletions detected by WGS (Del1: 12,442,344-12,443,887, Del2:12,449,829-12,455,936), whose summed size is 7,650bp. **c.** An Irys-detected 4,389bp deletion within chr7:1113898-1151045 in Caki2 cells overlaps with three WGS-detected deletions (Del1:1,115,577-1,116,112, Del2:1,127,730-1,129,400, Del3:1,145,442-1,148,018), whose summed size is 4,781 bp. **d.** Deletions detected by Irys have overall lower mappability compared to deletions detected by WGS (by two sided Wilcoxon rank-sum test). For WGS deletions, we computed the average of mappability scores for the 500bp regions upstream and downstream of the deletions (immediately outside the two breakpoints, n=26,255). For Irys-detected deletions, we computed the average mappability score between the two nicking enzymes (labels). We also require the size of deletions to count for at least 80% of the genomic distance between the two labels (n=103). For boxplots, the box represents the interquartile range (IQR), and the whiskers extend to 1.5 times the IQR or to the maximum/minimum if less than 1.5x IQR.

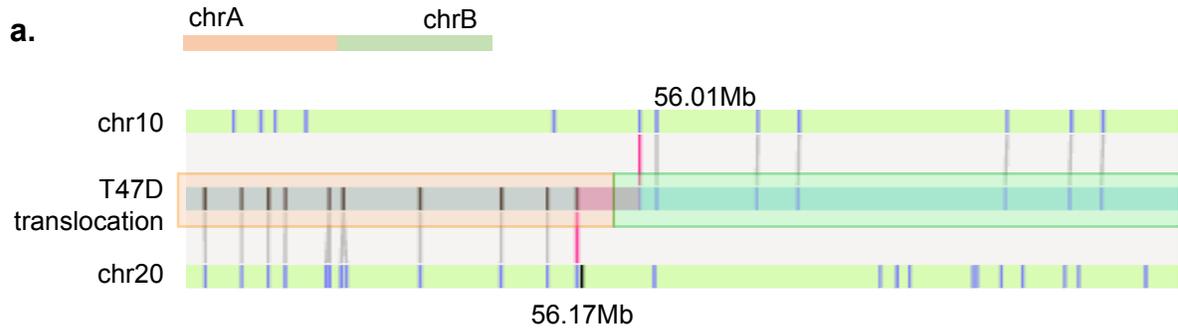


**Supplementary Figure 14 | Examples of large deletions in normal and cancer genomes. a-e.** Large deletions in cancer genomes, detected by both optical mapping and WGS coverage. **f.** A 703kb deletion in GM12878 cell line, which overlaps with IGL locus and represents potential V(D)J recombination.

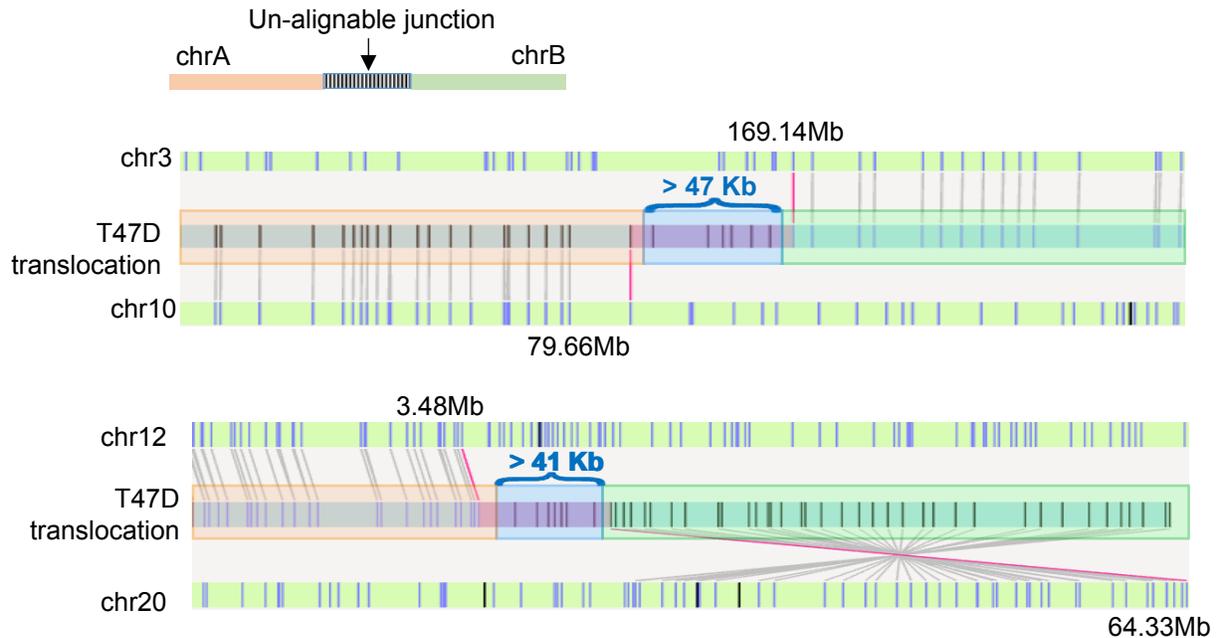


**Supplementary Figure 15 | Overlap of large SVs detected by Hi-C, optical mapping, and WGS.** Number of inter-chromosomal translocations (left panel) and large intra-chromosomal rearrangements ( $\geq 1\text{Mb}$ , right panel) detected by optical mapping, Hi-C, and WGS in T47D (a), K562 (b), NCI-H460 (c), PANC-1 (d), SK-N-MC (e), LNCaP (f), A549 (g), Caki2 (h) and GM12878 (i).

## A simple translocation detected by WGS, Hi-C and optical mapping

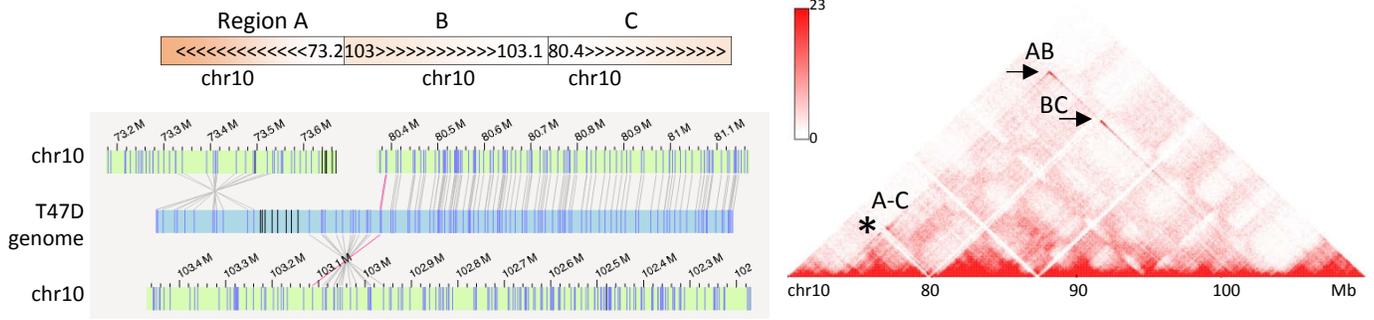


**b. Complex translocations detected by Hi-C and optical mapping with un-alignable junctions**

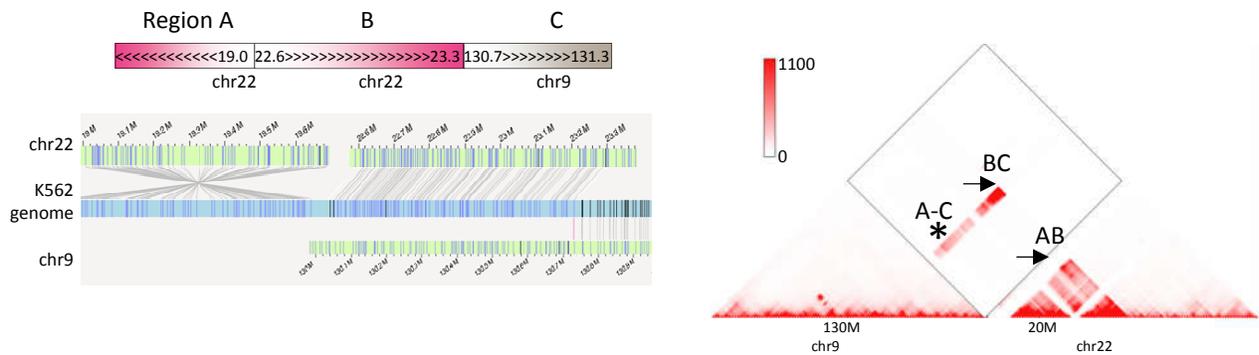


**Supplementary Figure 16 | Hi-C and optical mapping detect translocations with unalignable junctions.** **a.** An example of a simple translocation detected by WGS, Hi-C, and optical mapping. The predicted breakpoint is located between the two labels (nicking enzymes) and there is no unalignable region between them. **b.** Two examples of complex translocations with unalignable junctions detected by Hi-C and Irys but missed by WGS. In both scenarios, the large DNA fragments (> 40kb) between the two translocated arms were not mapped to human reference genome.

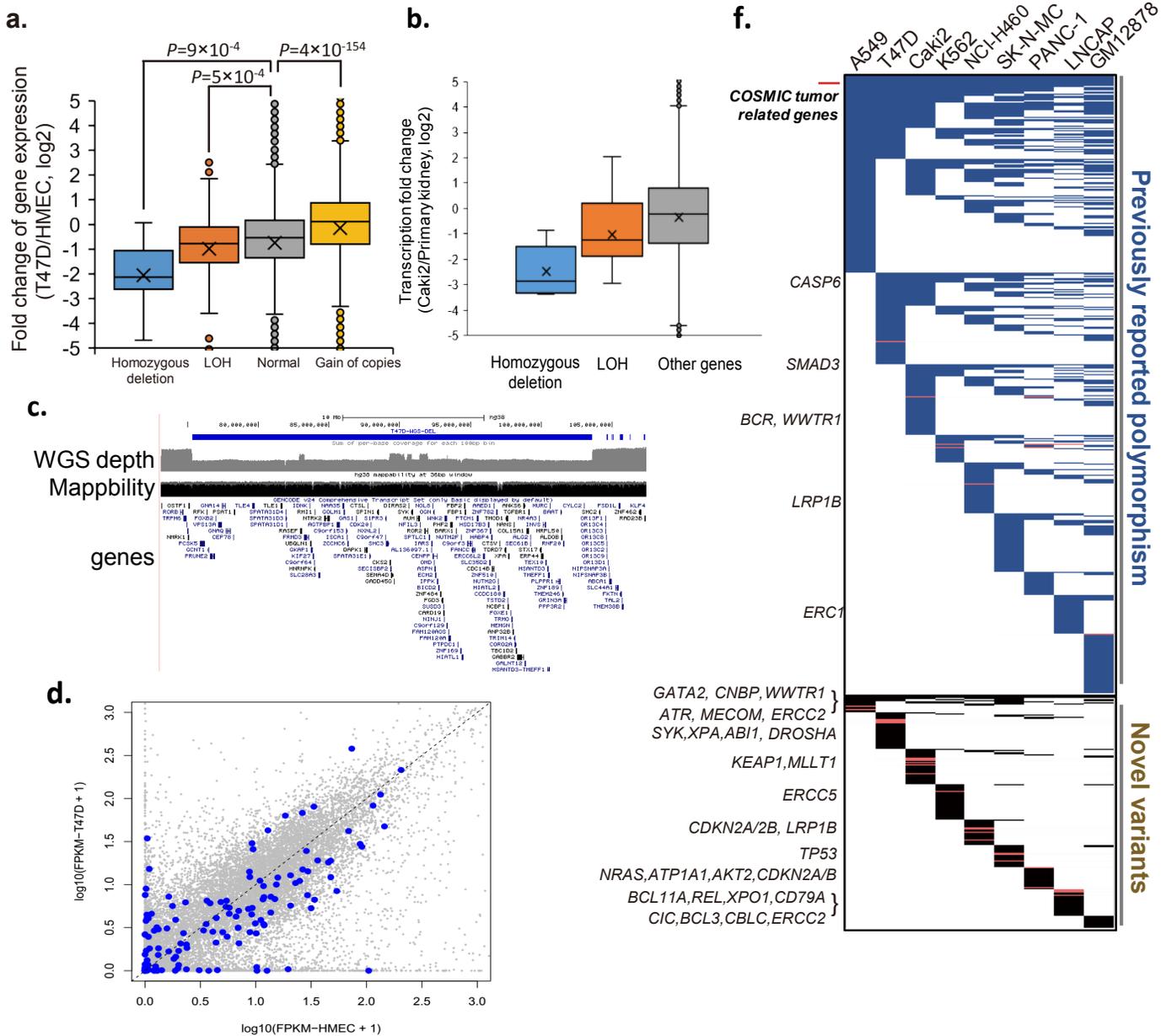
**a. T47D**



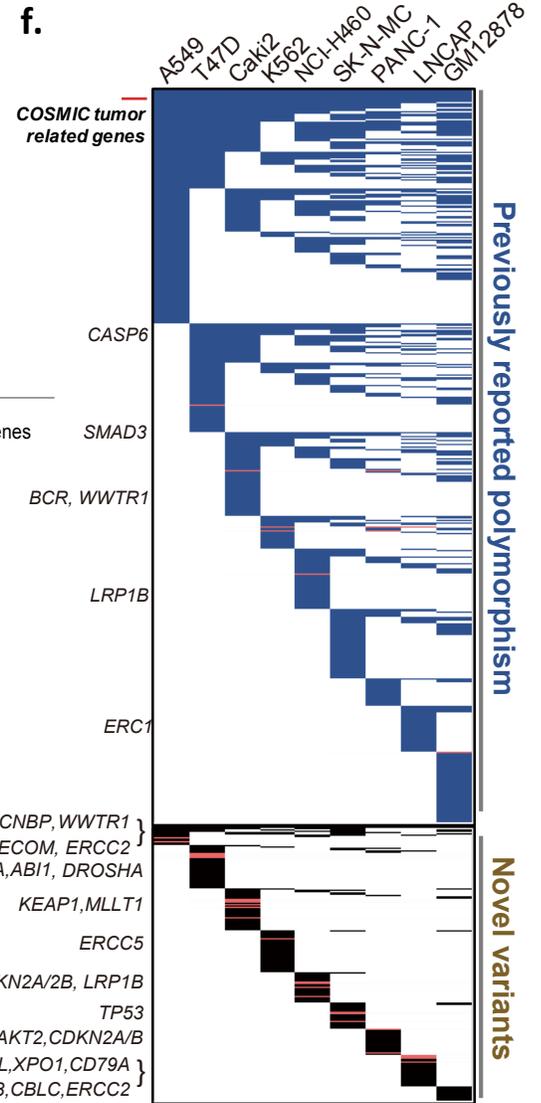
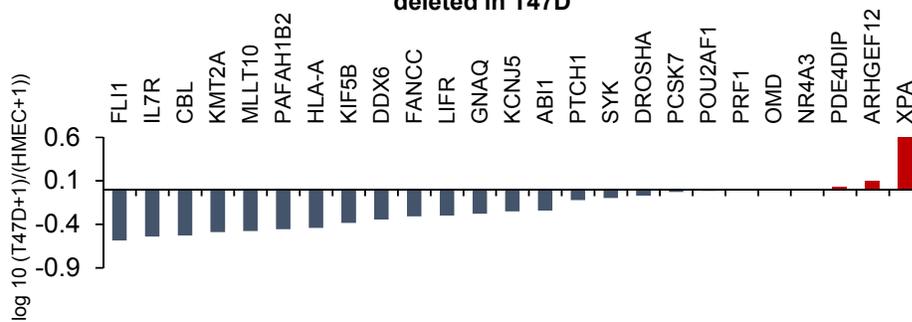
**b. K562**



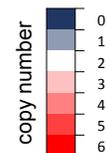
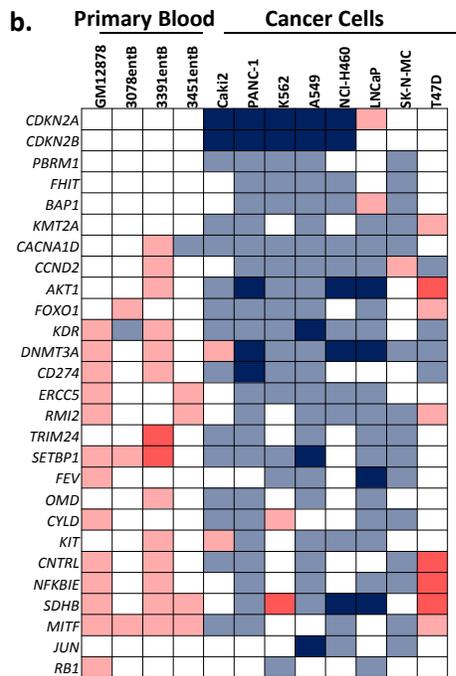
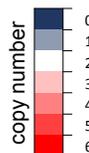
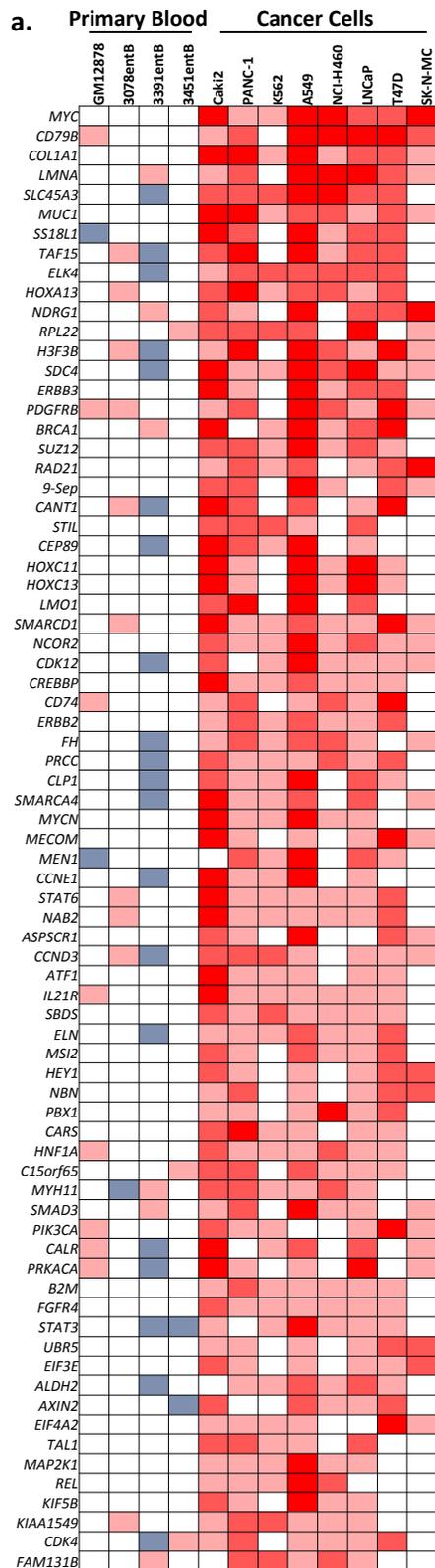
**Supplementary Figure 17 | Examples of using Hi-C and optical mapping to reconstruct the overall structure of complex translocations.** Similar to Fig. 3d, Arrow (->) indicates directly jointed translocation and asterisk (\*) marks the linked adjacent SV. **a.** Schematic of the local chromosome structure in T47D cells, which consist of 3 translocated regions: A (chr10:73.5-73.5M), B (chr10:80.4M-81.1M), and C (103-103.1M). **b.** Another example of locally resolved SV in LNCaP cell line. A ~8mb region on chr 7 (A) is inversely inserted between regions B and C on chr14.



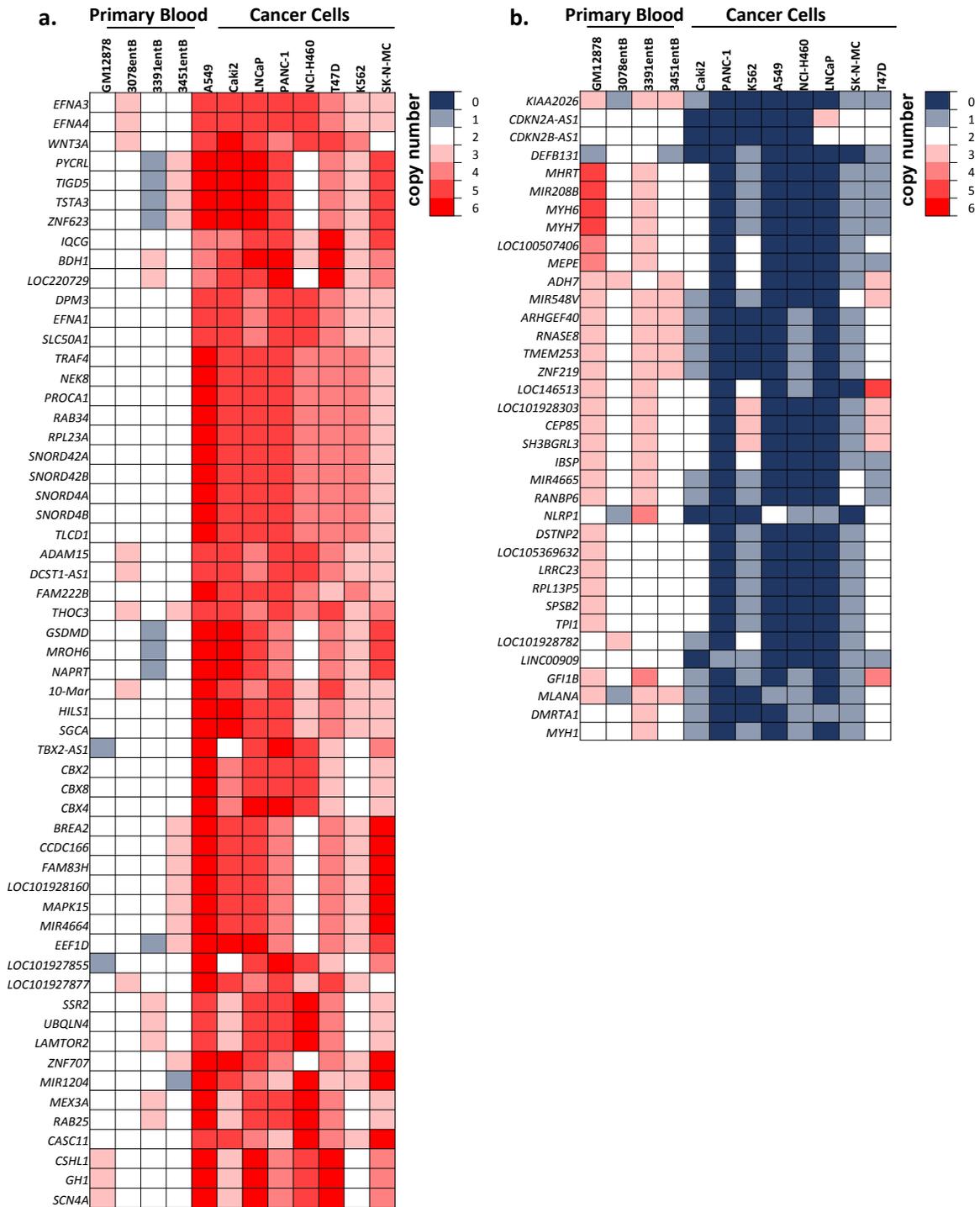
**e.** Transcriptional fold change of 25 Cosmic cancer censor genes deleted in T47D



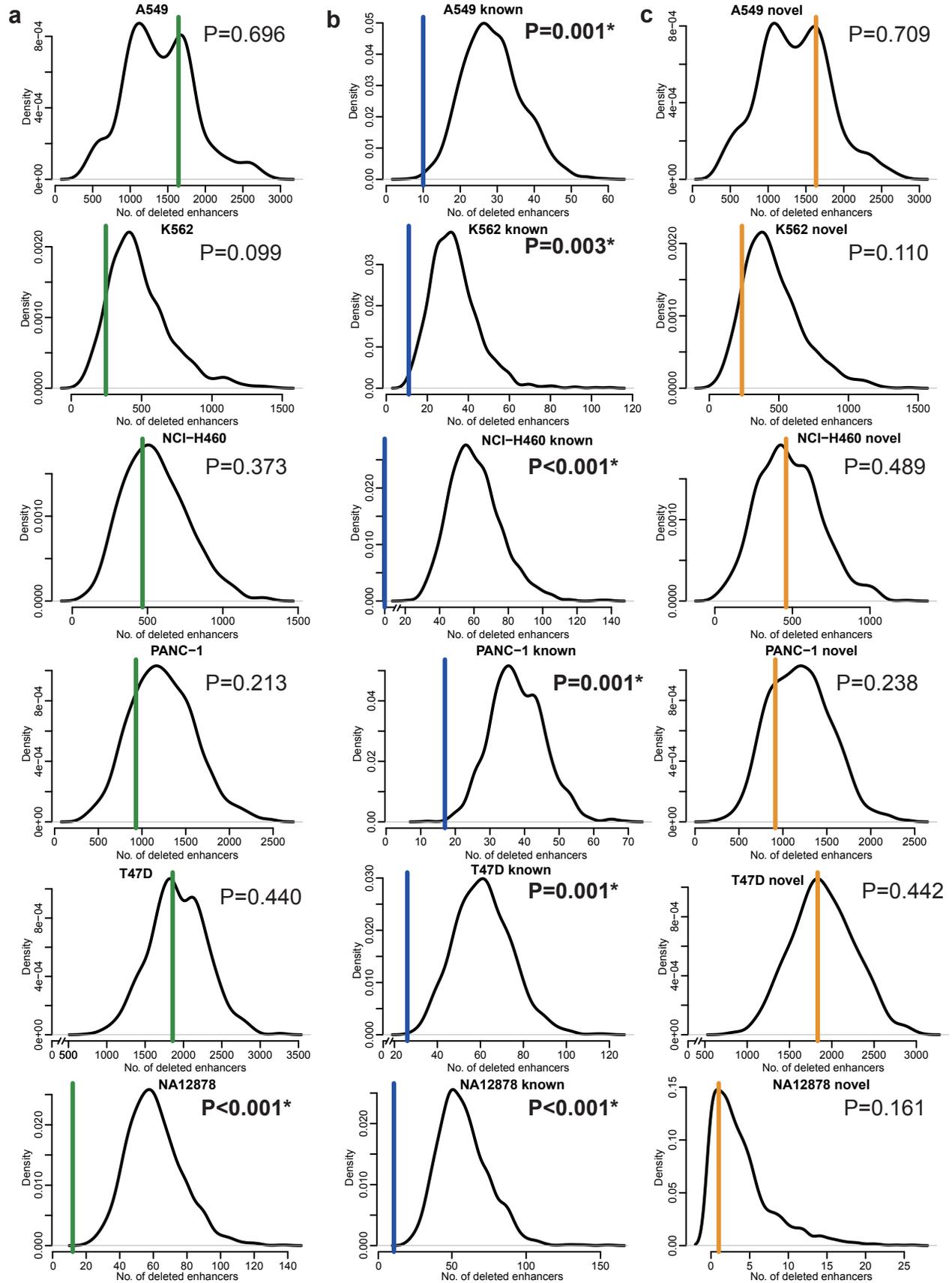
**Supplementary Figure 18 | Impact of exon deletion and copy loss on gene expression.** **a.** Compared with HMEC cells, expressed genes (FPKM>1 in HMEC cells) with homozygous deletions (n=10) and LOH (n=325) in T47D cells show reduced expression compared to copy-neutral genes (n=5113, P = 0.009 and 0.003 respectively, two-sided Wilcoxon rank sum test), and compared to gain of copy genes (n=6413, p=4×10<sup>-79</sup>, two sided Wilcoxon rank sum test). For all boxplots in the figure, the box represents the interquartile range (IQR), and the whiskers extend to 1.5 times the IQR or to the maximum/minimum if less than 1.5x IQR. **b.** Expressed genes (FPKM>1 in primary kidney epithelium cells) with homozygous deletions (n=5) or LOH (n=28) in Caki2 show reduced expression relative to non-copy number reduced genes (n=13859). **c.** A 28Mb deletion (chr9:75,335,996-103,526,867) in T47D cells causing LOH of over 400 genes. **d.** Deleted genes in T47D show reduced transcription. **e.** 25 COSMIC tumor-related genes have deletions overlapping with exons and the majority show reduced transcription. **f.** Cancer-specific novel deletions are enriched in COSMIC cancer-related genes. High-confidence deletions are classified as either known polymorphisms (from DGV database) or novel variants. In karyotypically normal cells (GM12878), 95% of deletions are polymorphic and 5% are novel, while in cancer genomes, over 10% of the deletions are novel. Novel deletions in cancer genomes are enriched for tumor related genes annotated by COSMIC database (red bars in the heat map).



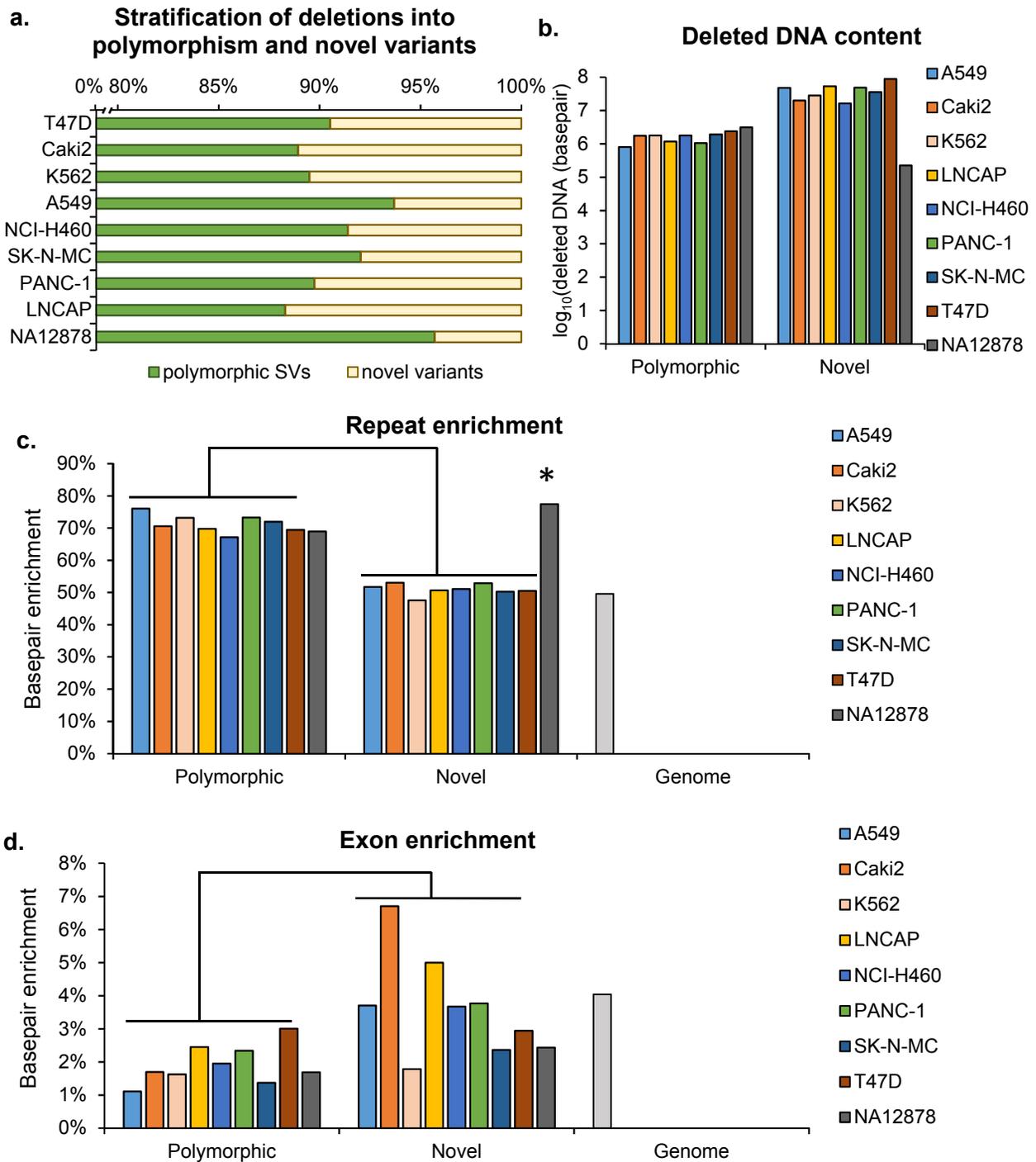
**Supplementary Figure 19 | Copy number alterations of COSMIC tumor-related genes, which are computed based on its surrounding 50 kb regions by optical mapping. a. COSMIC tumor-related genes with extensive gain of copies in cancer cell lines. b. COSMIC tumor-related genes with significant loss of copies in cancer cell lines.**



**Supplementary Figure 20 | List of non-COSMIC tumor-related genes that have significant copy number changes. Copy number is computed based on the surrounding 50Kb regions by optical mapping. a. 58 Genes with most significant amplifications. b. 37 genes with most significant loss of copies.**



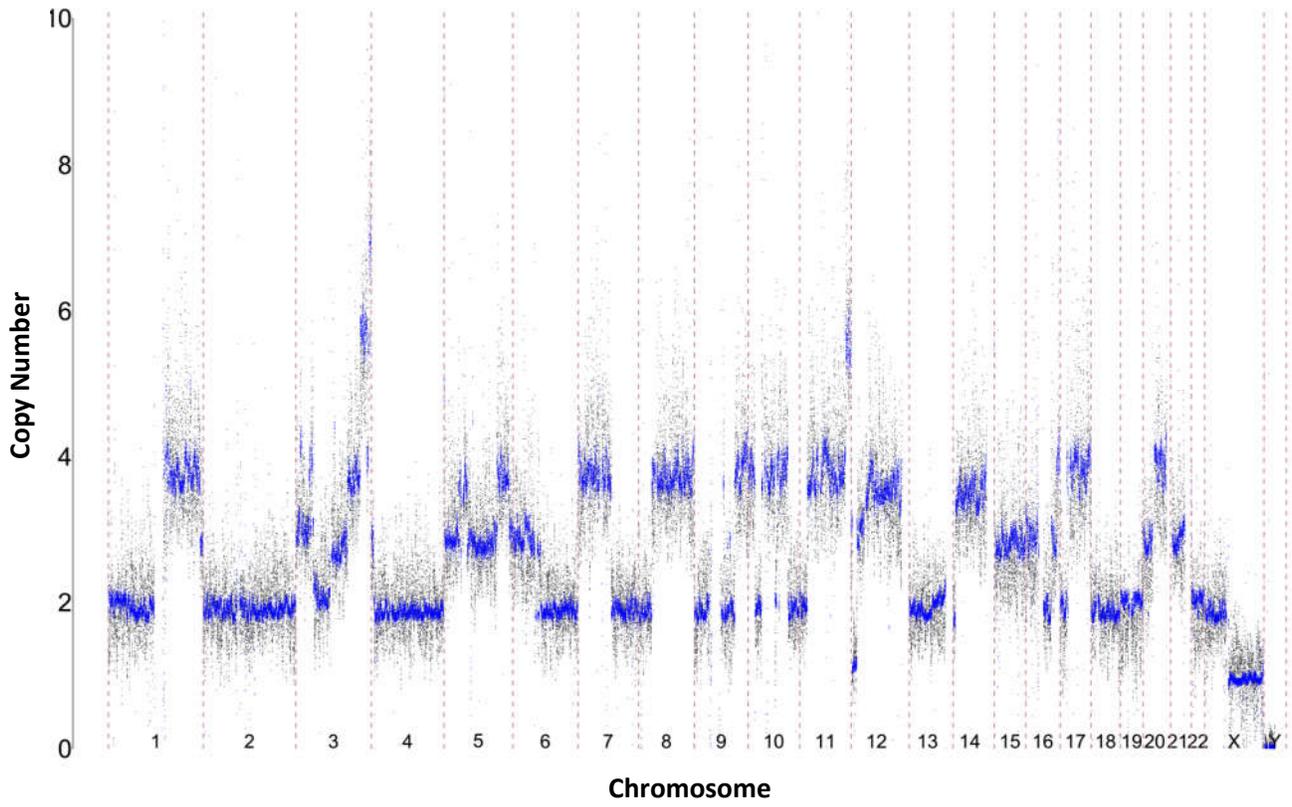
**Supplementary Figure 21 | Comparison of the frequency of enhancer disruptions versus expectation. a.** Overall, we found that deletions in normal cell types (GM12878) are less likely to delete enhancer that would be expected at random, while the enrichment level of deletions in enhancers in cancer cells are close to the values expected at random. For this analysis, we matched each cancer cell line with a control normal cell type that is developmentally from the same/similar tissue type: T47D vs. HMEC, K562 vs. mononuclear cells, PANC-1 vs. primary pancreatic tissues, A549 and NCI-H460 vs. NHLF cells. We used the H3K27ac peaks in the normal cell/tissue type as enhancer set. Then, we randomly shuffled the deletions in the cancer genomes 1,000 times and overlapped them with the enhancer set to compute the expected value (number of deletions: A549=237, K562=435, NCI-H460=405, PANC-1=320, T47D=454, NA12878=535). The curve shows the distribution of simulated results and the vertical line shows the observed value. The empirical P value is then calculated based on how many times the simulated number is smaller than the observed value ( $P < 0.001$  means no such incidence was observed in the 1000 simulations). **b,c.** We stratified the deletions into two categories by comparing them with DGV database: polymorphic deletions (A549=223, K562=392, NCI-H460=372, PANC-1=289, T47D=411, NA12878=513) and novel deletions (A549=14, K562=43, NCI-dH460=33, PANC-1=31, T47D=43, NA12878=22). We found that polymorphic deletions are less likely to delete enhancer, while novel deletions are reflect the genome wide distribution of enhancers.



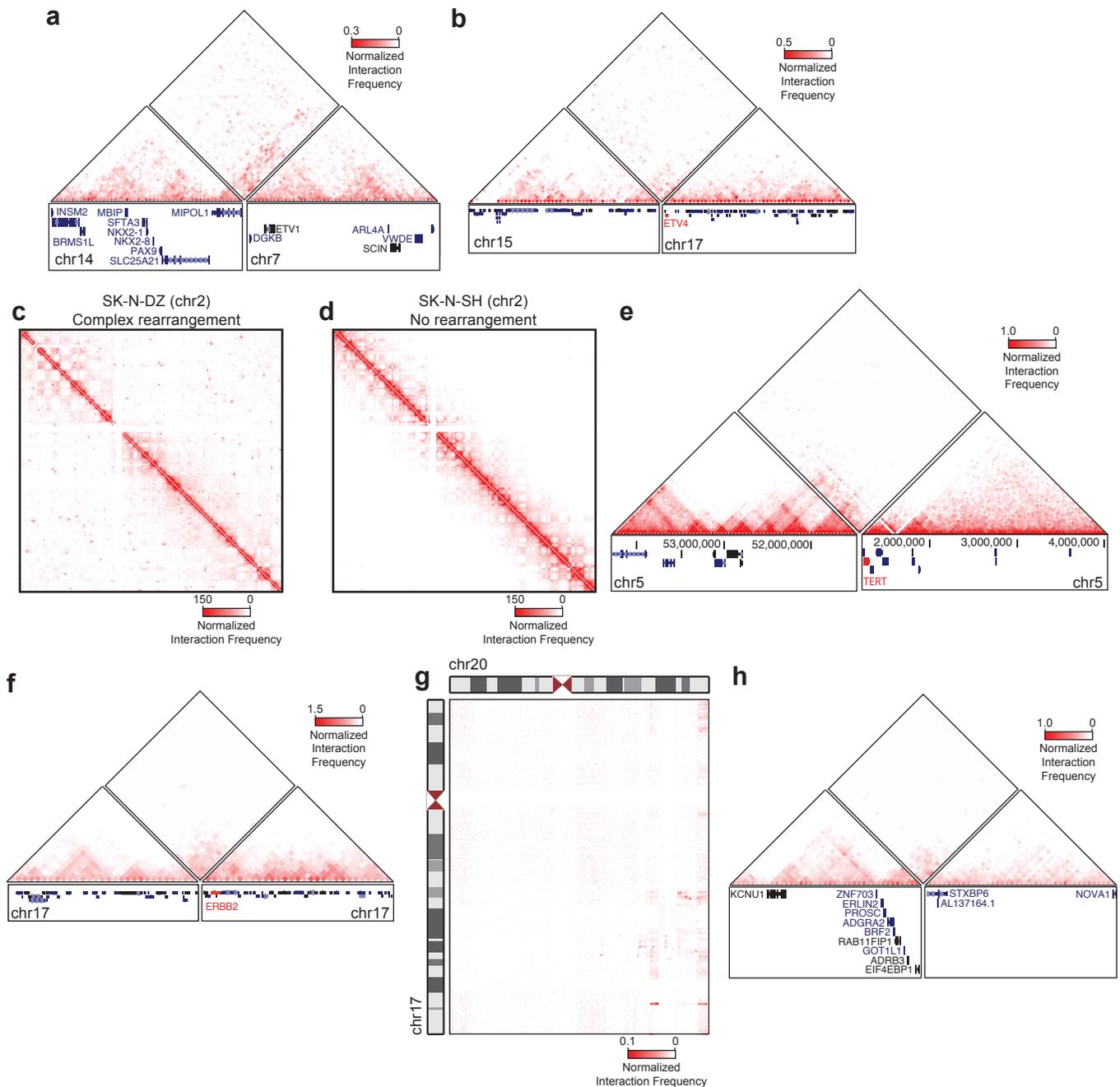
**Supplementary Figure 22 | Characterization of known polymorphic deletions and novel deletions.** We stratified the deletions into two categories by comparing with DGV database: polymorphic deletions and novel deletions. **a.** Only 5% of deletions in GM12878 cells are novel variants, whereas on average 10% of deletions found in cancer cells are novel variants. **b.** In cancer genomes, the loss of DNA content due to novel deletions is more than ten times of those induced by polymorphic deletions. **c.** Polymorphic deletions show higher enrichment of repetitive elements (70%) than genome background (50%), whereas novel deletions in cancer cells are not enriched for repeats. **d.** In general, polymorphic deletions are resistant to exon deletions, compared with novel deletions and genomic background.

**T47D cells:**

- ..... *Copy number predicted by WGS*
- ..... *Copy number predicted by optical mapping*



**Supplementary Figure 23 | Genome-wide CNVs predicted by optical mapping and WGS are consistent.**



**Supplementary Figure 24 | Annotation of known recurrent structural variants in cancer cell lines.** **a.** Effects of a chr7-chr14 translocation in LNCaP cells. The fusion generates a neo-TAD that encompasses the ETV1 gene. **b.** Effects of a chr15-chr17 translocation in PC-3 cells. The translocation generates a TAD fusion event that encompasses the ETV4 gene. **c.** Interaction frequency heat map of chromosome 2 in SK-N-DZ cells. We identified 46 rearrangements in chromosome 2 of SK-N-DZ cells, indicative of a complex chromosomal rearrangement. **d.** For comparison with panel c, the interaction frequency heat map of chromosome 2 SK-N-SH cells. We identify no rearrangements on chromosome 2 in SK-N-SH cells. **e.** Effects of an inversion on chromosome 5 in SK-N-AS cells. We identified an inversion in the proximity of the TERT gene (shown in red) that creates a neo-TAD as a result of the inversion. **f.** Effects of a ~3.5Mb deletion on chromosome 17 in Panc1 cells. The deletion results in a neo-TAD between the regions immediately flanking the deleted region. This neo-TAD includes the ERBB2 oncogene. **g.** Interaction frequency heat map of interactions between chromosome 17 and chromosome 20 in MCF7 cells. There is a complex rearrangement involving both chromosomes in MCF7 breast cancer cells. Of note, the rearranged region on chromosome 20 is recurrently amplified in breast cancer. **h.** Effects of a chr8-chr14 translocation in T47D breast cancer cells. This rearrangement appears to create a neo-TAD in the vicinity of the ZNF703 gene, a known oncogene in breast cancer.

## Supplementary Tables

Supplementary Table 1: List of cell/tissue types with performed experiments and analysis.

Supplementary Table 2: Number of SVs detected by WGS, Hi-C, and Optical mapping in eight cancer cell lines and NA12878.

Supplementary Table 3: SVs detected by WGS in 8 cancer cell lines and NA12878 ([Excel](#)).

Supplementary Table 4: SVs detected by optical mapping in 8 cancer cell lines and NA12878 ([Excel](#)).

Supplementary Table 5: SVs detected by Hi-C in 36 cell lines ([Excel](#)).

Supplementary Table 6: High-confidence SV calls from integration([Excel](#)).

Supplementary Table 7: Validated translocations and deletions in K562, Caki and T47D cells.

Supplementary Table 8: Cross comparison of large intra-chromosomal rearrangements ( $\geq 1\text{Mb}$ ) and inter-chromosomal translocations ([Excel](#)).

Supplementary Table 9: Contribution by each method and their overlapping percentage with high-confidence SVs.

Supplementary Table 10: Integration of intra-chromosomal rearrangements ( $< 1\text{Mb}$ ) ([Excel](#)).

Supplementary Table 11: Irys-detected deletions encompass multiple smaller WGS-detected deletions with the same total deleted sizes ([Excel](#)).

Supplementary Table 12: Optical mapping predicts the size of unresolved genome gap in hg19.

Supplementary Table 13: Optical mapping provides estimation of gap size in hg38 and comparison to previous gap assessment of hg38 ([Excel](#)).

Supplementary Table 14: SV-induced fused genes detected by RNA-seq.

Supplementary Table 15. Summary of genes, repetitive elements and insulators overlapping with high-confidence deletions.

Supplementary Table 16. Frequency of enhancer deletions versus simulated expectation in cancer cells and normal cells.

Supplementary Table 17: Deleted potential enhancers and insulators in T47D, Caki2, K562, NCIH460 ([Excel](#)).

**Supplementary Table 1. List of cell/tissue types with performed experiments and analysis**

Tissue type	Cell line	Hi-C	Optical mapping	WGS	replication timing	Karyotyping	RNA-seq	PET-seq
CML	K562	● ●	●	●	●	●	●	●
Kidney cancer	Caki2	●	●	●	●	●	●	
Breast cancer	T47D	●	●	●	●	● ●	●	●
Lung caner	NCI-H460	●	●	●	●	●	●	
Askin's tumor	SK-N-MC	●	●	●	●	● ●	●	
Prostate cancer	LNCaP	●	●	●	●	●	●	
Lung cancer	A549	●	●	●	●	●	●	
Pancreatic cancer	Panc1	●	●	●		●	●	
Wilms' tumor	G401	●			●		●	
<b>Cancer</b>	Melonoma	●					●	
<b>cel lines</b>	Neuroblastoma	●					●	
	Neuroblastoma	●						
	Rhabdomyosarcoma	●						
	Melonoma	●						
	Breast cancer	●			●		●	● ●
	Prostate cancer	●					●	
	Neuroblastoma	●			●		●	
	CML	●						
	ALL	●						
	Lymphoma	●						
	Glioblastomas	●						
	Glioblastomas	●						
<b>Primary</b>	Glioblastomas	●						
<b>tumor</b>	Glioblastomas	●						
<b>samples</b>	Glioblastomas	●						
	Glioblastomas	●						
	Glioblastomas	●						
	Leukemia	●						
	hESC	●						
	Breast epithelial	●					●	
	Endothelial	●						
<b>Normal</b>	Lung	●						
<b>cell</b>	mESC	●						
<b>lines</b>	Mesenchymal stem cell	●						
<b>and</b>	Neural progenitor	●						
<b>primary</b>	Trophectoderm	●						
<b>samples</b>	Lymphoblastoid		●					
	Lymphoblastoid		●					
	Lymphoblastoid		●					
	Lymphoblastoid	●	●	●				
	Primary kidney						●	

● We performed the experiments and analysis

● We downloaded the raw data from public resources for analysis

**Supplementary Table 2. Number of SVs detected by Hi-C, optical mapping and WGS in 8 cancer cell lines and NA12878 cells**

**Hi-C**

	<b>Intra-chr SVs (&gt;1Mb)</b>	<b>Inter-chr translocations</b>
T47D	18	26
CAKI2	26	41
K562	13	24
A549	11	7
NCI-H460	4	13
PANC-1	15	15
LNCAP	13	13
SK-N-MC	5	9
NA12878	1	0

**Optical mapping**

	<b>Deletion</b>	<b>Insertion</b>	<b>Inversion</b>	<b>Unclassified intra-chr SVs (≥5Mb)</b>	<b>Inter-chr translocation</b>
T47D	1128	2233	13	23	28
CAKI2	1483	2456	7	6	26
K562	1447	2531	14	10	26
A549	1112	1949	4	3	10
NCI-H460	1411	2483	9	1	12
PANC-1	1266	2083	6	3	8
LNCAP	1021	1891	3	1	6
SK-N-MC	1463	2513	18	6	21
NA12878	1151	2425	19	1	0

**WGS**

	<b>Coverage (×)</b>	<b>Deletion</b>	<b>Duplication</b>	<b>Inversion</b>	<b>Unclassified breakpoints</b>	<b>Inter-chr translocation</b>
T47D	37	2945	736	27	234	125
CAKI2	46	3202	863	29	290	192
K562	35	3223	702	31	240	111
A549	47	3654	715	27	210	85
NCI-H460	35	2951	699	37	228	112
PANC-1	52	3165	887	48	313	199
LNCAP	81	4151	840	62	225	247
SK-N-MC	41	3036	712	49	278	139
NA12878	50	3473	760	66	278	3

**Non-redundant SVs from integration of three methods**

	<b>Deletion</b>	<b>Duplication</b>	<b>Insertion</b>	<b>Inter-chr translocation</b>	<b>Inversion</b>	<b>Unclassified intra-chr breakpoints</b>
T47D	3624	736	2437	160	49	252
CAKI2	4282	865	2652	238	38	309
K562	4240	703	2720	151	47	248
A549	4529	717	2131	101	31	218
NCI-H460	3956	699	2691	127	48	227
PANC-1	4118	887	2310	218	53	326
LNCAP	4906	840	1878	257	67	233
SK-N-MC	4015	714	2721	156	72	284
NA12878	4094	761	2439	3	92	279

**Supplementary Table 7. Validated translocations and deletions in K562, CAK12 and T47D cells**

Hi-C source		Region I		Region II		Known/novel	Validation	Result			
<b>Validating translocations detected by Hi-C in K562 cells</b>											
Rao et al.	Dixon et al.	chr13	+	107854000	108009000	chr9	+	131176000 131280000	Known*		
Rao et al.	Dixon et al.	chr13	-	19000000	47000000	chr9	-	27000000 39000000	Known		
Rao et al.	Dixon et al.	chr13	+	107800000	108000000	chr22	+	22000000 23300000	Known		
	Dixon et al.	chr17	-	27000000	29000000	chr9	+	0 21000000	Known		
Rao et al.	Dixon et al.	chr17	-	51000000	57000000	chr9	+	0 21000000	Known		
Rao et al.	Dixon et al.	chr17	+	19000000	23000000	chr10	-	43000000 51000000	Known		
Rao et al.	Dixon et al.	chr9	-	130731000	131000000	chr22	+	22958000 23291000	Known		
Rao et al.	Dixon et al.	chr3	+	48147000	48186000	chr10	+	86065000 86089000	Known		
Rao et al.	Dixon et al.	chr5	-	51084000	51094000	chr6	+	37789000 37856000	Known		
Rao et al.	Dixon et al.	chr22	-	22500000	22700000	chr2	+	150400000 150900000	Known		
Rao et al.	Dixon et al.	chr12	-	22621000	22633000	chr21	-	24258000 24281000	Known		
Rao et al.	Dixon et al.	chr3	-	138000000	162000000	chr18	-	26000000 27000000	Novel	FISH	Confirmed
Rao et al.	Dixon et al.	chr3	+	138000000	150000000	chr18	+	4000000 8000000	Novel	FISH	Confirmed
Rao et al.	Dixon et al.	chr1	-	107000000	112000000	chr20	-	30000000 35000000	Novel	FISH	Confirmed
Rao et al.	Dixon et al.	chr1	+	54500000	54800000	chr18	+	24400000 25900000	Novel	FISH	Not Confirmed
Rao et al.	Dixon et al.	chr1	+	106780000	106820000	chr18	-	27260000 27450000	Novel	FISH	Confirmed
	Dixon et al.	chr1	+	115000000	120000000	chr6	-	135000000 140000000	Novel	FISH	Confirmed
Rao et al.	Dixon et al.	chr16	-	85528000	85548000	chr6	-	16766000 16770000	Novel	FISH	Confirmed
Rao et al.	Dixon et al.	chr18	-	27000000	27300000	chr6	-	135400000 136200000	Novel	FISH	Confirmed
<b>Validating translocations in T47D</b>											
		chr3	-	136170000	137100000	chr5	+	171830000 171430000	Known		
		chr3	-	169130000	170110000	chr10	+	73240000 73280000	Known		
		chr3	-	193000000	193620000	chr12	-	15100000 15580000	Known		
		chr6	+	46000000	58000000	chrX	+	36000000 64000000	Known		
		chr7	+	86890000	87700000	chr15	-	29660000 30080000	Known		
		chr8	+	36770000	38090000	chr14	-	24870000 25550000	Known		
		chr9	-	68000000	101000000	chr17	-	19000000 36000000	Known		
		chr10	+	53650000	56020000	chr20	-	56170000 56580000	Known		
		chr10	-	56020000	58280000	chr20	+	54080000 56170000	Known		
		chr12	-	21300000	22100000	chr13	-	78800000 79300000	Known		
		chr12	-	15150000	15860000	chr16	-	67060000 67360000	Known		
Hillmer et al.		chr4	+	6590000	6800000	chr5	+	900000 1380000	Reported	PCR	Confirmed
Hillmer et al.		chr6	+	71240000	72130000	chr22	+	16920000 17230000	Reported	PCR	Confirmed
Hillmer et al.		chr9	+	15500000	17340000	chr15	-	27200000 28140000	Reported	PCR	Confirmed
Hillmer et al.		chr5	+	1640000	1750000	chr5	-	40600000 40870000	Reported	PCR	Confirmed
Hillmer et al.		chr9	+	75040000	75340000	chr9	-	103600000 104790000	Reported	PCR	Confirmed
		chr3	+	45740000	46390000	chr9	+	89250000 89420000	novel	PCR	Confirmed
		chr3	-	169130000	170110000	chr10	+	79230000 79650000	novel	PCR	Confirmed
		chr10	+	18080000	18280000	chr10	+	36210000 36880000	novel	PCR	Confirmed
<b>Validated translocations in Caki2</b>											
		chr12		66571831		chr4		64330748		PCR	Confirmed
		chr9		85978709		chr19		45733773		PCR	Confirmed
		chr6		56750050		chr8		58550779		PCR	Confirmed
<b>Validating deletions detected by optical mapping in T47D cells</b>											
		chrX		42652746		chrX		42656304		PCR	Not confirmed
		chr2		212590110		chr2		212720073		PCR	Confirmed
		chr2		97188517		chr2		97190465		PCR	Confirmed
		chr14		104948976		chr14		104951429		PCR	Confirmed
		chr3		58586154		chr3		58586217		PCR	Confirmed
		chr4		165081464		chr4		165083902		PCR	Confirmed
		chr2		28466613		chr2		28469693		PCR	Confirmed
		chr7		6861596		chr7		6887316		PCR	Confirmed
		chr1		207523594		chr1		207546536		PCR	Confirmed
		chr12		58325913		chr12		58339245		PCR	Confirmed
		chr11		107361838		chr11		107374676		PCR	Confirmed
		chr7		97762466		chr7		97773481		PCR	Confirmed
		chr7		70969523		chr7		70979773		PCR	Confirmed
		chr6		85998091		chr6		86007304		PCR	Confirmed
		chr1		53126296		chr1		53129986		PCR	Not confirmed
		chr13		69400712		chr13		69404714		PCR	Confirmed

\*Known translocations are validated by previous karyotyping

**Supplementary Table 9. Contribution by each method and their overlapping percentage with high-confidence SVs.**

**All large SVs (inter-chromosomal TL and intra-chromosomal SVs  $\geq 1\text{Mb}$ )**

SV detection methods	Average contribution	Average overlap with high confidence SVs
Hi-C	48%	66%
Irys	40%	43%
WGS	64%	22%
<b>3 Methods</b>	90%	23%
Karyotype	23%	88%
Transcript fusion	18%	NA
PET-seq	73%	12%

**Inter-chromosomal translocations**

SV detection methods	Average contribution	Average overlap with high confidence SVs
Hi-C	53%	66%
Irys	24%	28%
WGS	56%	15%
<b>3 Methods</b>	88%	18%
Karyotype	56%	88%
Transcript fusion	24%	NA
PET-seq	61%	7%

**Intra-chromosomal large SVs ( $\geq 1\text{Mb}$ )**

SV detection methods	Average contribution	Average overlap with high confidence SVs
Hi-C	43%	71%
Irys	59%	62%
WGS	74%	42%
<b>3 Methods</b>	92%	36%
Karyotype	2%	50%
Transcript fusion	12%	NA
PET-seq	88%	23%

**Supplementary Table 12. Optical mapping predicts the size of unresolved genome gap in hg19**

	GAP location in hg19		Gap type	hg19 size (Kb)	Optical mapping prediction (Kb)	hg38 size (Kb)
<b>Prediction consistent with GRCH38</b>						
chr1	3845268	3995268	contig	150.00	6.68	6.51
chr1	29878082	30028082	contig	150.00	3.43	3.67
chr1	103863906	103913906	clone	50.00	-27.22*	-27.07*
chr1	144710724	144810724	clone	100.00	-101.17	-128.94
chr1	223747846	223797846	clone	50.00	37.76	35.64
chr1	235192211	235242211	clone	50.00	23.99	22.43
chr1	248908210	249058210	contig	150.00	-33.48	-48.49
chr10	47792476	47892476	contig	100.00	74.46	72.30
chr10	128616069	128766069	contig	150.00	47.60	40.31
chr10	133381404	133431404	clone	50.00	11.34	15.93
chr11	69089801	69139801	clone	50.00	2.65	2.70
chr11	69724695	69774695	clone	50.00	19.02	18.63
chr11	96287584	96437584	contig	150.00	12.73	12.16
chr12	7189876	7239876	contig	50.00	4.15	4.71
chr12	109373470	109423470	contig	50.00	5.99	5.97
chr12	122530623	122580623	contig	50.00	3.43	3.36
chr12	132706992	132806992	contig	100.00	14.96	13.70
chr13	114639948	114739948	contig	100.00	33.98	33.56
chr15	29159443	29209443	contig	50.00	3.59	2.94
chr16	8636921	8686921	clone	50.00	7.07	6.15
chr16	88389383	88439383	contig	50.00	18.27	17.20
chr18	52059136	52209136	contig	150.00	5.36	9.14
chr18	72283353	72333353	clone	50.00	6.45	5.28
chr18	75721820	75771820	clone	50.00	2.76	1.95
chr19	7346004	7396004	contig	50.00	0.74	0.00
chr19	8687198	8737198	contig	50.00	12.60	-0.10
chr19	20523415	20573415	clone	50.00	-20.00	-22.00
chr2	3529312	3579312	contig	50.00	6.73	6.18
chr2	5018788	5118788	contig	100.00	7.94	7.46
chr2	16279724	16329724	contig	50.00	8.37	8.85
chr2	21153113	21178113	contig	25.00	1.51	1.89
chr2	110109337	110251337	contig	142.00	0.85	0.88
chr2	149690582	149790582	contig	100.00	1.14	1.06
chr2	239801978	239831978	contig	30.00	19.72	16.95
chr2	240784132	240809132	contig	25.00	8.59	7.28
chr20	34897085	34947085	clone	50.00	11.02	9.52
chr20	61091437	61141437	clone	50.00	29.94	27.85
chr20	61213369	61263369	contig	50.00	16.36	15.86
chr21	42955559	43005559	contig	50.00	1.68	1.77
chr22	50364777	50414777	contig	50.00	6.15	5.22
chr3	66170270	66270270	contig	100.00	34.86	35.25
chr4	1423146	1478646	contig	55.50	50.73	47.56
chr5	91636128	91686128	contig	50.00	9.78	10.11
chr5	138787073	138837073	contig	50.00	5.28	6.10
chr5	155138727	155188727	contig	50.00	1.52	2.55
chr7	232484	282484	clone	50.00	12.84	10.03
chr7	50370631	50410631	contig	40.00	12.01	11.90
chr7	74715724	74765724	clone	50.00	-164.41	-165.17

\* A negative value indicates that the size of the gap is around "0" and the gap-surrounding regions in hg19 is also removed or shortened in hg38

	GAP location in hg19		Type	hg19 size (Kb)	Optical mapping predict (Kb)	hg38 size (Kb)
chr7	130154523	130254523	clone	100.00	55.32	55.57
chr7	139379377	139404377	contig	25.00	9.55	9.95
chr7	154270634	154370634	contig	100.00	5.44	5.38
chr8	142766515	142816515	clone	50.00	-15.38	-21.26
chr8	145332588	145432588	contig	100.00	-57.90	-68.78
chr9	133073060	133223060	contig	150.00	37.47	36.89
chr9	137041193	137091193	contig	50.00	22.58	23.25
chr9	139166997	139216997	contig	50.00	47.92	47.39
chrX	7623882	7673882	clone	50.00	0.20	0.00
chrX	10738674	10788674	clone	50.00	-0.22	-0.08
chrX	76653692	76703692	contig	50.00	15.08	14.97
chrX	148906424	148956424	clone	50.00	3.00	2.85
chrX	149032062	149082062	contig	50.00	12.05	10.56
chrX	152277099	152327099	clone	50.00	-45.33	-59.07
chrY	20143885	20193885	clone	50.00	-1.15	-0.01
<b>Inconsistent regions</b>						
chr1	205922707	206072707	contig	150.00	-47.82	445.77
chr1	206332221	206482221	contig	150.00	-47.82	445.78
chr11	87688378	87738378	clone	50.00	4.28	27.59
chr13	86760324	86910324	contig	150.00	24.45	-0.10
chr13	114325993	114425993	contig	100.00	1.86	51.34
chr15	22212114	22262114	contig	50.00	-32.44	402.90
chr17	34675848	34725848	contig	50.00	9.02	21.59
chr17	79709049	79759049	contig	50.00	9.50	59.15
chr4	8799203	8818203	contig	19.00	0.52	19.00
chr4	9274642	9324642	clone	50.00	2.03	50.00
chr4	31820917	31837417	contig	16.50	3.96	16.50
chr4	59739333	59789333	contig	50.00	8.90	50.49
chr4	75427379	75452279	contig	24.90	-32.77	-0.10
chr5	17530657	17580657	clone	50.00	29.50	50.00
chr6	157559467	157609467	clone	50.00	-18.18	-49.90
chr6	157641300	157691300	clone	50.00	-18.18	-49.90
chr6	167942073	168042073	clone	100.00	65.45	111.51
chr7	100556043	100606043	clone	50.00	37.13	-0.12
chr7	143347897	143397897	clone	50.00	-25.73	50.00
chr8	86576451	86726451	contig	150.00	137.60	50.00
chr9	92343416	92443416	clone	100.00	-44.61	13.47
chr9	92528796	92678796	clone	150.00	5.39	89.17
chrX	37098256	37148256	contig	50.00	16.56	208.82
chrX	49242997	49292997	contig	50.00	-42.53	141.86
chrX	49974173	50024173	contig	50.00	24.49	71.65
chrX	115682290	115732290	contig	50.00	13.89	47.19
chrX	120013235	120063235	clone	50.00	-26.06	50.00
chrX	143507324	143557324	contig	50.00	4.98	51.63
chrY	8914955	8964955	contig	50.00	30.15	80.43
chrY	9241322	9291322	contig	50.00	22.29	50.00

\* A negative value indicates that the size of the gap is around "0" and the gap-surrounding regions in hg19 is also removed or shortened in hg38

**Supplementary Table 14. SV-induced fused genes detected by RNA-seq**

Gene Name1	Gene Name2	chrA	Breakpoint1	chrB	Breakpoint2	Crossing reads	Spanning reads	SV Detection methods
<b>T47D</b>								
MECOM	CFAP70	3	169143698	10	73275598	7	3	Hi-C
ZDHHC3	RPS14	3	44926161	5	150446962	4	50	Hi-C, Karyotype
MB21D2	RERG	3	192917629	12	15151819	16	3	Hi-C, WGS
VPS26A	FAM149B1	2	69133047	2	73208619	2	3	WGS, PET-seq
<b>Caki2</b>								
DST	SDCBP	6	56843055	8	58565019	149	3	Hi-C, IRYS, WGS
PCMT1	PDSS2	6	149749956	6	107212282	14	2	Hi-C
<b>K562</b>								
BCR	ABL1	22	23290413	9	130854064	41	45	Hi-C, IRYS, Karyotype
GSE1	RP1-151F17.1	16	85556363	6	16762352	10	0	Hi-C
NUP214	XKR3	9	131199015	22	16808083	18	9	Hi-C, IRYS, WGS, Karyotype
<b>MCF7</b>								
BCAS4	BCAS3	20	50795172	17	61368326	83	116	Hi-C, PET-seq
B3GNTL1	SLC9A8	17	82956695	20	49877981	5	1	Hi-C
SGPP2	ULK4	2	222474725	3	41249575	3	7	Hi-C
TBL1XR1	RGS17	3	177197121	6	153044043	11	0	Hi-C, PET-seq
LEMD3	RP1-74B13.2	12	65218619	12	110845599	3	3	Hi-C
<b>PC3</b>								
KDM5B	LINC01351	1	202745857	1	190480344	13	3	Hi-C
PLOD2	IQCJ	3	146160789	3	159245843	13	7	Hi-C
<b>NCIH460</b>								
ANKRD11	ZNF585B	16	89418284	19	37207254	122	47	Hi-C
ZRANB2	WDR78	1	71067336	1	66840952	34	5	WGS
<b>SKNMC</b>								
KIF1B	CNKSR2	1	10261970	X	21497787	36	40	Hi-C, IRYS, WGS
EWSR1	FLI1	22	29287134	11	128805366	99	130	Hi-C, IRYS, Karyotype
YEATS2	GALNT13	3	183697993	2	154140337	45	75	Hi-C, IRYS, WGS, Karyotype
<b>A549</b>								
SERAC1	SNX27	6	158110261	1	151695563	3	7	Hi-C, Karyotype
NCEH1	MUC13	3	172710847	3	124939520	2	8	Hi-C
SCAMP2	WDR72	15	74845473	15	53706074s	79	32	Hi-C

**Supplementary Table 15. Summary of genes, repetitive elements and insulators overlapping with high-confidence deletions.**

**Deletions of genes and repetitive elements**

<b>Cell lines</b>	<b>No. of confident deletion</b>	<b>No. of genes disrupted by confident deletions</b>	<b>Percentage of repetitive elements (basepair enrichment)</b>
CAKI2	404	586	54.38%
T47D	454	1097	50.98%
K562	435	398	48.99%
A549	237	624	52.16%
NCI-H460	405	414	52.58%
PANC-1	320	558	53.26%
LNCAP	281	658	51.07%
SK-N-MC	487	965	51.34%
NA12878	535	273	69.02%

**Deletions of insulators**

<b>Cell lines</b>	<b>Tissue for enhancer annotation (H3K27ac)</b>	<b>No. of all confident deletion</b>	<b>No. of deleted insulator annotated by CTCF binding sites (Tissue for annotation)</b>
T47D	HMEC	454	1019 (HMEC)
K562	Primary blood mononuclear cells	435	228 (NA12878)
A549	NHLF	237	2125 (NHLF)
NCI-H460	NHLF	405	663 (NHLF)
PANC-1	Primary pancreatic tissue	320	457 (Primary pancreatic tissue)

**Supplementary table 16. Frequency of enhancer deletions versus simulated expectation in cancer cells and normal cells**

Cancer cell lines	Number of deletions	Total deleted base pairs	Control tissue	Number of enhancers in control tissue (H3K27ac)	Number of enhancers deleted in cancer cells VS expectation (p value)	Number of deleted enhancers per 100 Kb deletion	Number of deleted enhancers per 100 Kb deletion normalized to 100,000 total enhancers
T47D (breast cancer)	454	91,072,914	HMEC (human mammary epithelium cells)	66,066	1859 : 1928 (p = 0.440)	2.04	3.09
K562 (chronic leukemia)	435	29,756,292	Primary blood mononuclear cells	51,862	246 : 484 (p = 0.099)	0.83	1.60
A549 (lung carcinoma)	237	48,725,609	NHLF (lung fibroblast)	91,440	1643 : 1398 (p = 0.696)	3.37	3.68
NCI-H460 (lung carcinoma)	405	18,082,136	NHLF (lung fibroblast)	91,440	467 : 556 (p = 0.373)	2.58	2.82
PANC-1 (pancreatic cancer)	320	49,875,753	Primary pancreatic tissue	78,896	931 : 1237 (p = 0.213)	1.86	2.35
<b>NA12878</b> (lymphoblastoid)	<b>535</b>	<b>3,359,296</b>	Primary blood mononuclear cells	51862	12 : 60 (p < 0.001)*	<b>0.36</b>	<b>0.69</b>

\*From left to right, the second and third columns show the total number of deletion incidence and the sum of deleted DNA content (basepair). The fourth column shows the control tissue/cell lines that are in close developmental relationship to the tested cell lines, and we use the H3K27ac marks in the control cell line to annotate the enhancers for that tissue type. The fifth column indicates the total number of annotated enhancers from each control tissue. The sixth column shows the number of deleted enhancer in each cell line, and we also include the result of simulation to approximate how many deletions of enhancers are likely to occur if the deletions are stochastically distributed in the genome. In the seventh and eighth columns we calculate the number of enhancer deletion per 100Kb deletion, and that value normalized to the total number of enhancers from that tissue type.