

In the format provided by the authors and unedited.

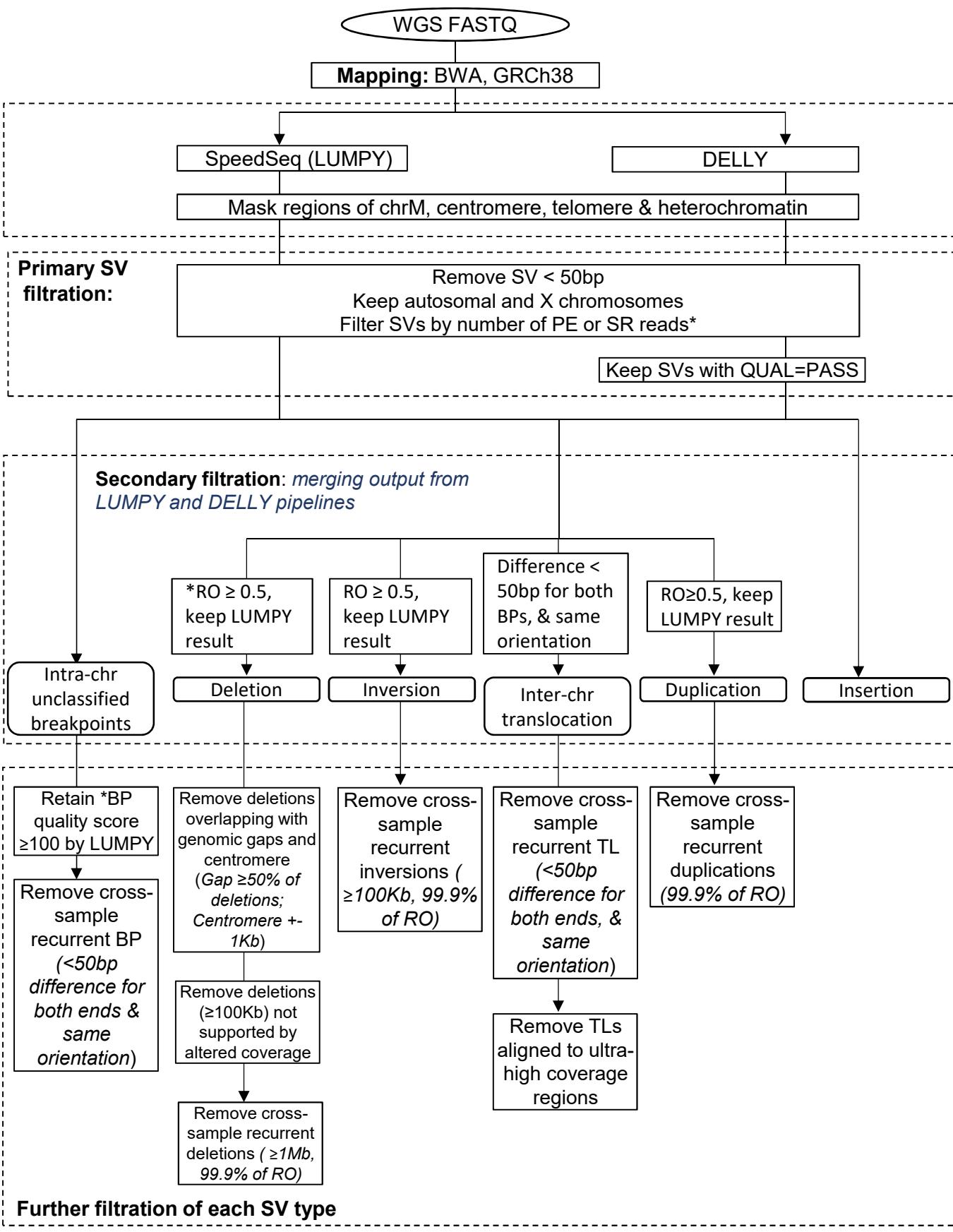
Integrative detection and analysis of structural variation in cancer genomes

Jesse R. Dixon^{ID 1,17*}, Jie Xu^{2,17}, Vishnu Dileep^{3,17}, Ye Zhan^{ID 4,17}, Fan Song^{5,17}, Victoria T. Le¹, Galip Gürkan Yardımcı⁶, Abhijit Chakraborty⁷, Darrin V. Bann⁸, Yanli Wang⁵, Royden Clark⁹, Lijun Zhang², Hongbo Yang², Tingting Liu², Sriranga Iyyanki², Lin An⁵, Christopher Pool⁸, Takayo Sasaki³, Juan Carlos Rivera-Mulia³, Hakan Ozadam⁴, Bryan R. Lajoie⁴, Rajinder Kaul¹⁰, Michael Buckley^{ID 10}, Kristen Lee¹⁰, Morgan Diegel¹⁰, Dubravka Pezic¹¹, Christina Ernst^{ID 12}, Suzana Hadjur¹¹, Duncan T. Odom^{ID 12,13}, John A. Stamatoyannopoulos¹⁰, James R. Broach², Ross C. Hardison¹⁴, Ferhat Ay^{ID 7,15*}, William Stafford Noble^{ID 6*}, Job Dekker^{ID 4,16*}, David M. Gilbert^{ID 3*} and Feng Yue^{ID 2,5*}

¹Salk Institute for Biological Studies, La Jolla, CA, USA. ²Department of Biochemistry and Molecular Biology, College of Medicine, The Pennsylvania State University, Hershey, PA, USA. ³Department of Biological Science, Florida State University, Tallahassee, FL, USA. ⁴Program in Systems Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA. ⁵Bioinformatics and Genomics Program, The Pennsylvania State University, University Park, PA, USA. ⁶Department of Genome Sciences, University of Washington, Seattle, WA, USA. ⁷La Jolla Institute for Allergy and Immunology, La Jolla, CA, USA. ⁸Division of Otolaryngology, Head & Neck Surgery, Milton S. Hershey Medical Center, Hershey, PA, USA. ⁹Penn State College of Medicine, Informatics and Technology, Hershey, PA, USA. ¹⁰Altius institute for Biomedical Sciences, Seattle, WA, USA. ¹¹Research Department of Cancer Biology, Cancer Institute, University College London, London, UK. ¹²Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK. ¹³German Cancer Research Center (DKFZ), Division Signaling and Functional Genomics, Heidelberg, Germany. ¹⁴Center for Comparative Genomics and Bioinformatics, Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA, USA. ¹⁵School of Medicine, University of California San Diego, La Jolla, CA, USA. ¹⁶Howard Hughes Medical Institute, Chevy Chase, MD, USA. ¹⁷These authors contributed equally to this work: Jesse R. Dixon, Jie Xu, Vishnu Dileep, Ye Zhan, and Fan Song. *e-mail: jedixon@salk.edu; ferhatay@lji.org; william-noble@uw.edu; job.dekker@umassmed.edu; gilbert@bio.fsu.edu; fyüe@hmc.psu.edu

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 lrys 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.

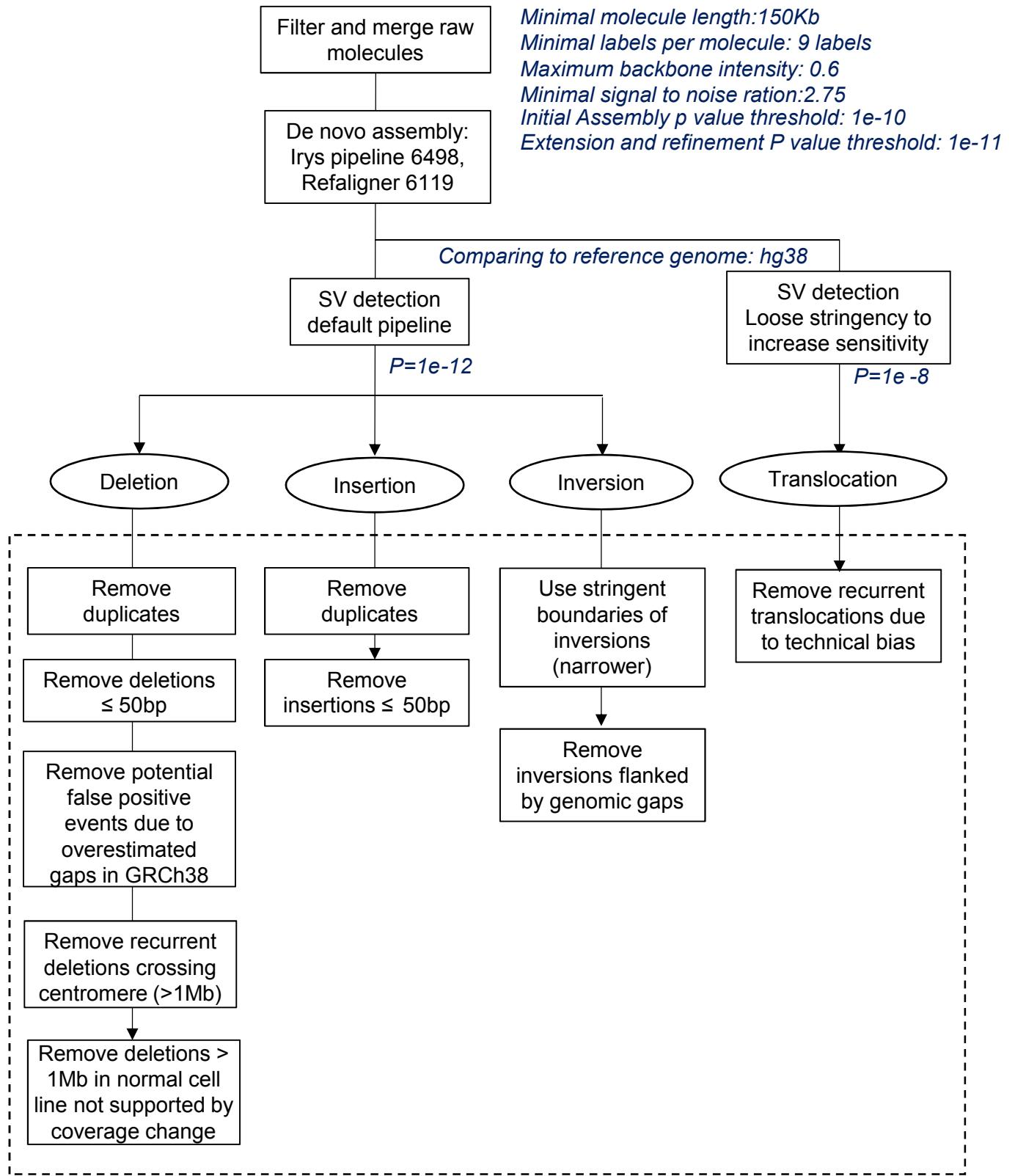


* Number of PE or SR reads are adjusted to the WGS coverage and the ploidy of cell lines

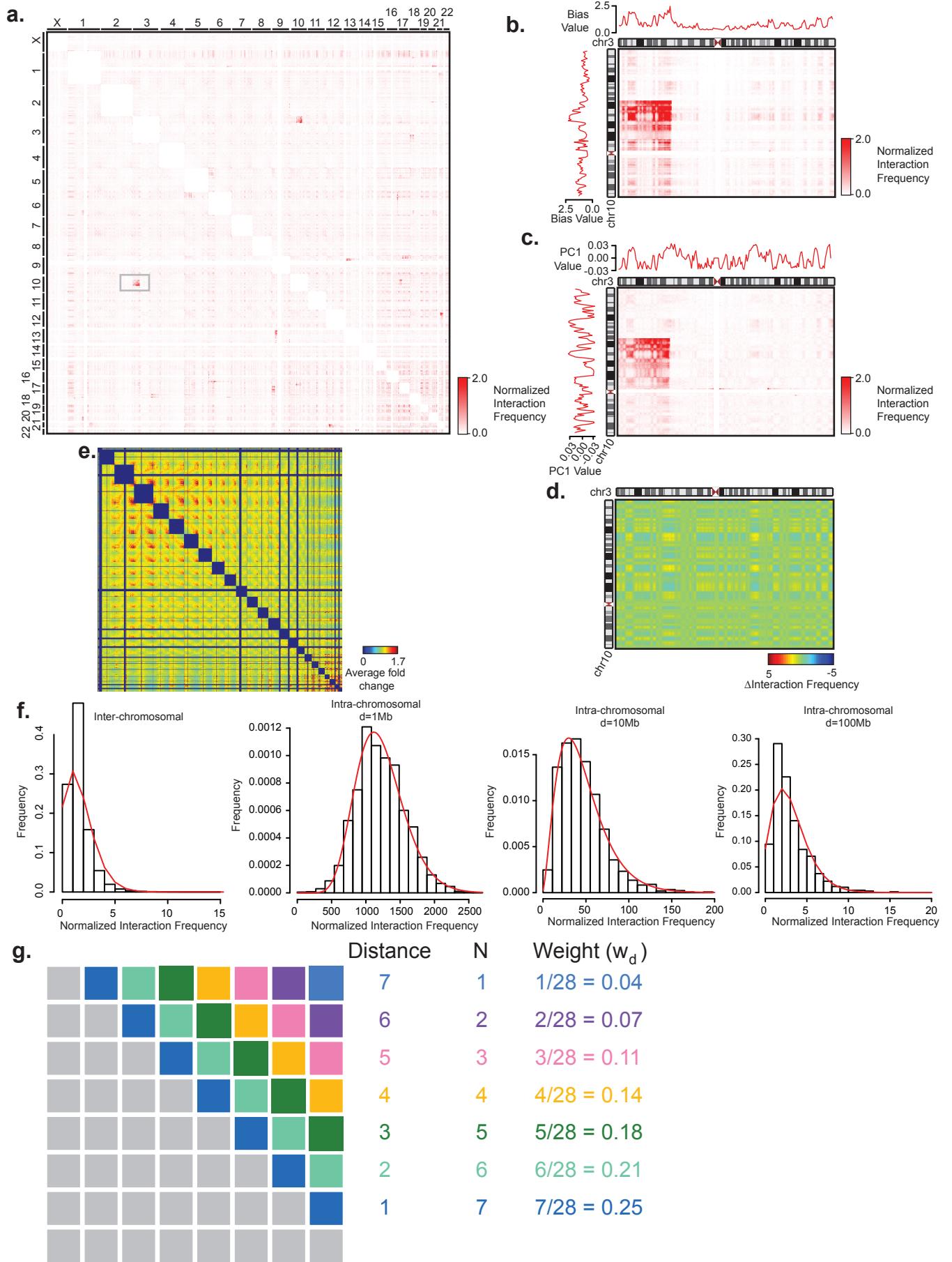
The procedure of Control FREEC is described in the methods section

* RO: ratio of reciprocal overlap. *BP: breakpoint

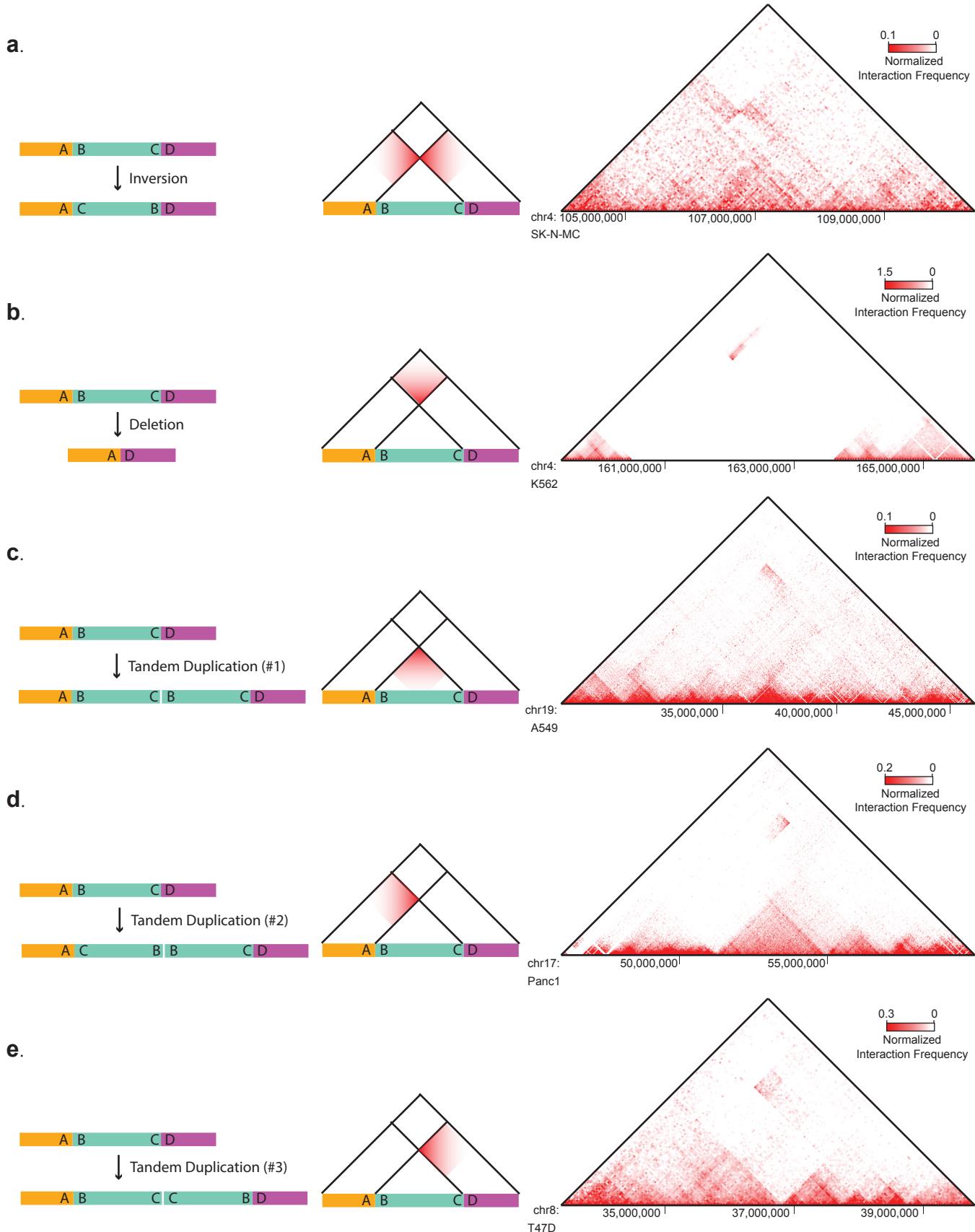
Supplementary Figure 1 | Pipeline of structural variants detection and filtration by WGS.



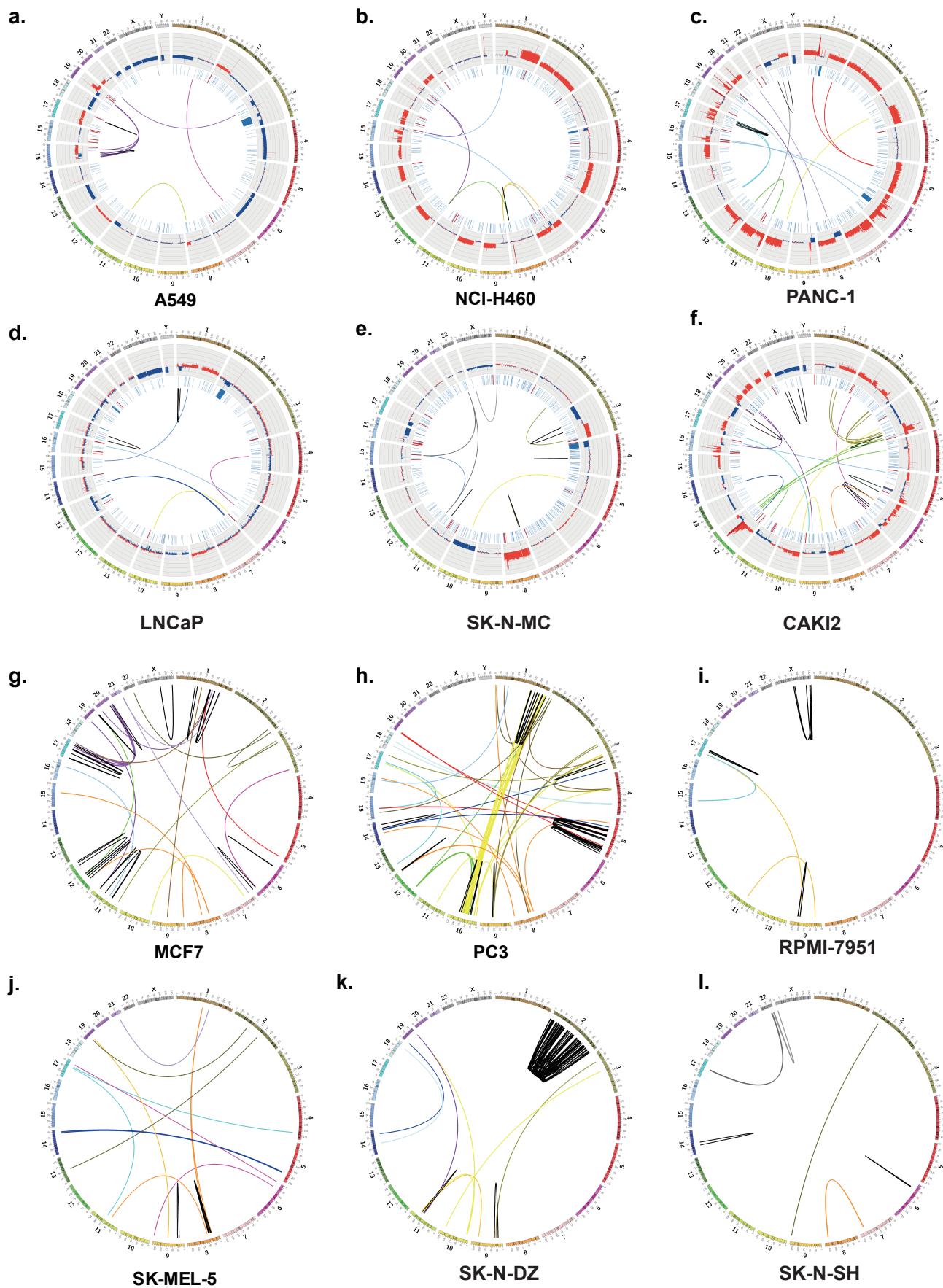
Supplementary Figure 2 | Pipeline of structural variants detection and filtration by optical mapping.



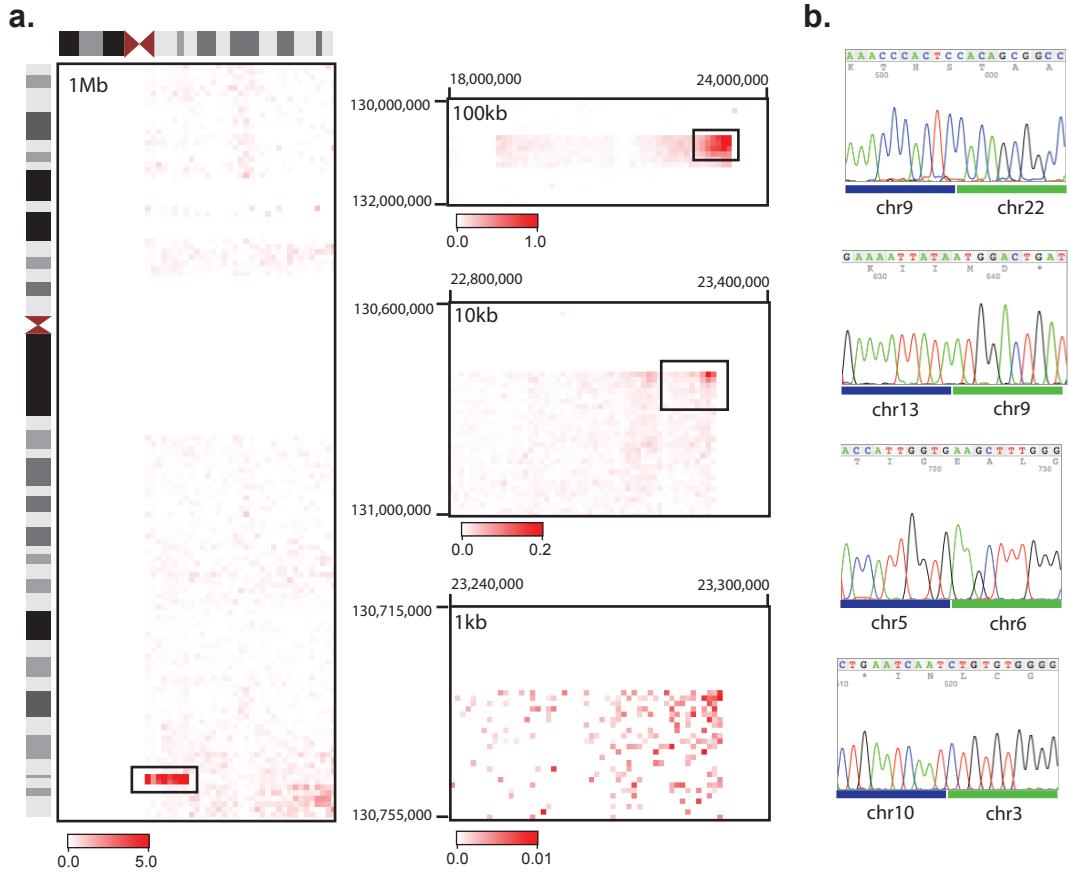
Supplementary Figure 3 | Overview of methodology for finding translocations in cancer genomes using Hi-C data. **a.** Whole genome Hi-C interaction heat map from K562 cells. Strong local increases in Hi-C signal between heterologous chromosomes likely represent translocations. The grey box is the outline 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.



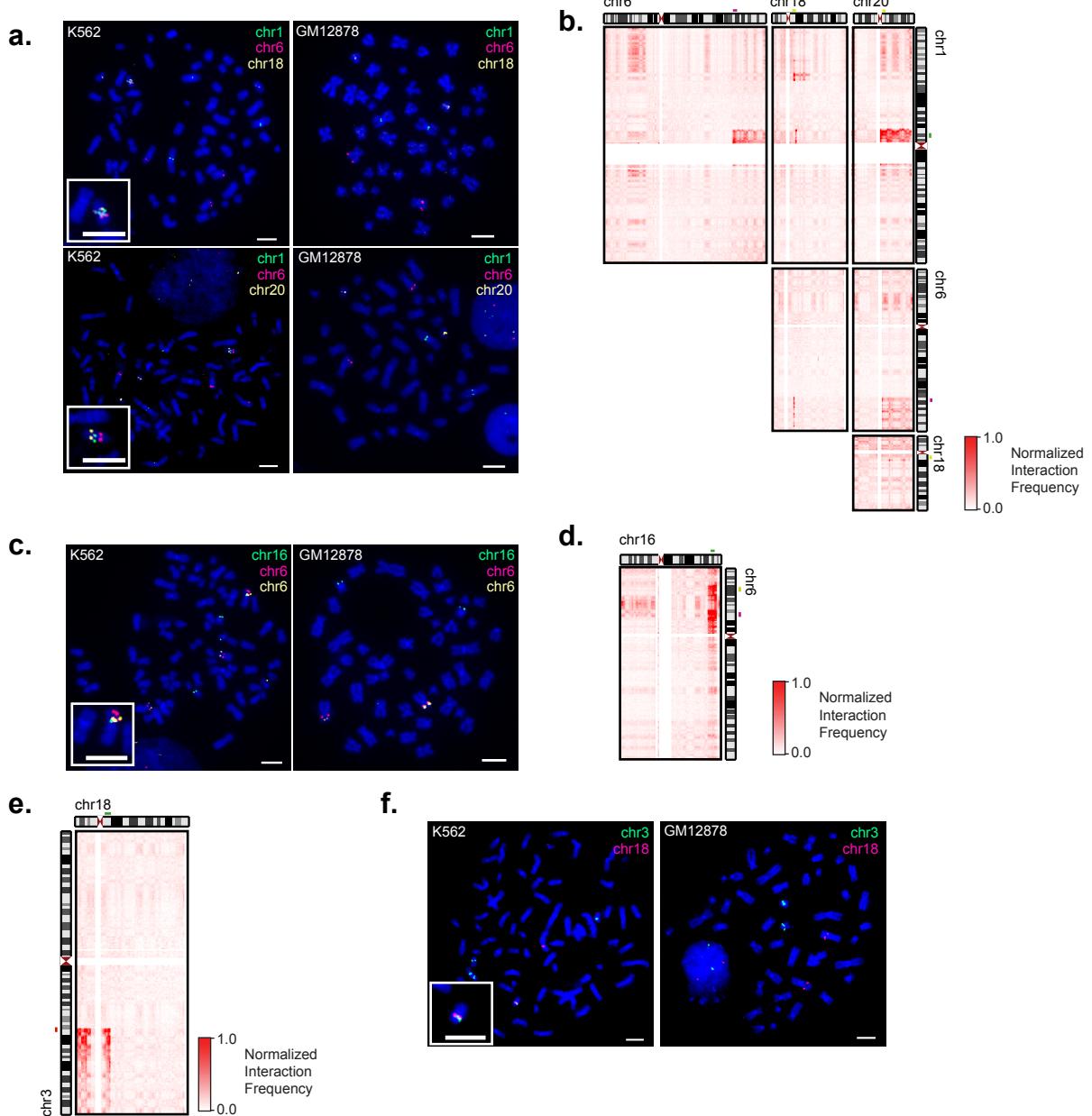
Supplementary Figure 4 | Hi-C identifies inversions, deletions and tandem duplications. **a.** Example of an inversion identified by Hi-C. The left hand cartoon shows the example of the genetic structure of an inversion, juxtaposing regions A and C as well as B and D. The cartoon in the middle depicts 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.



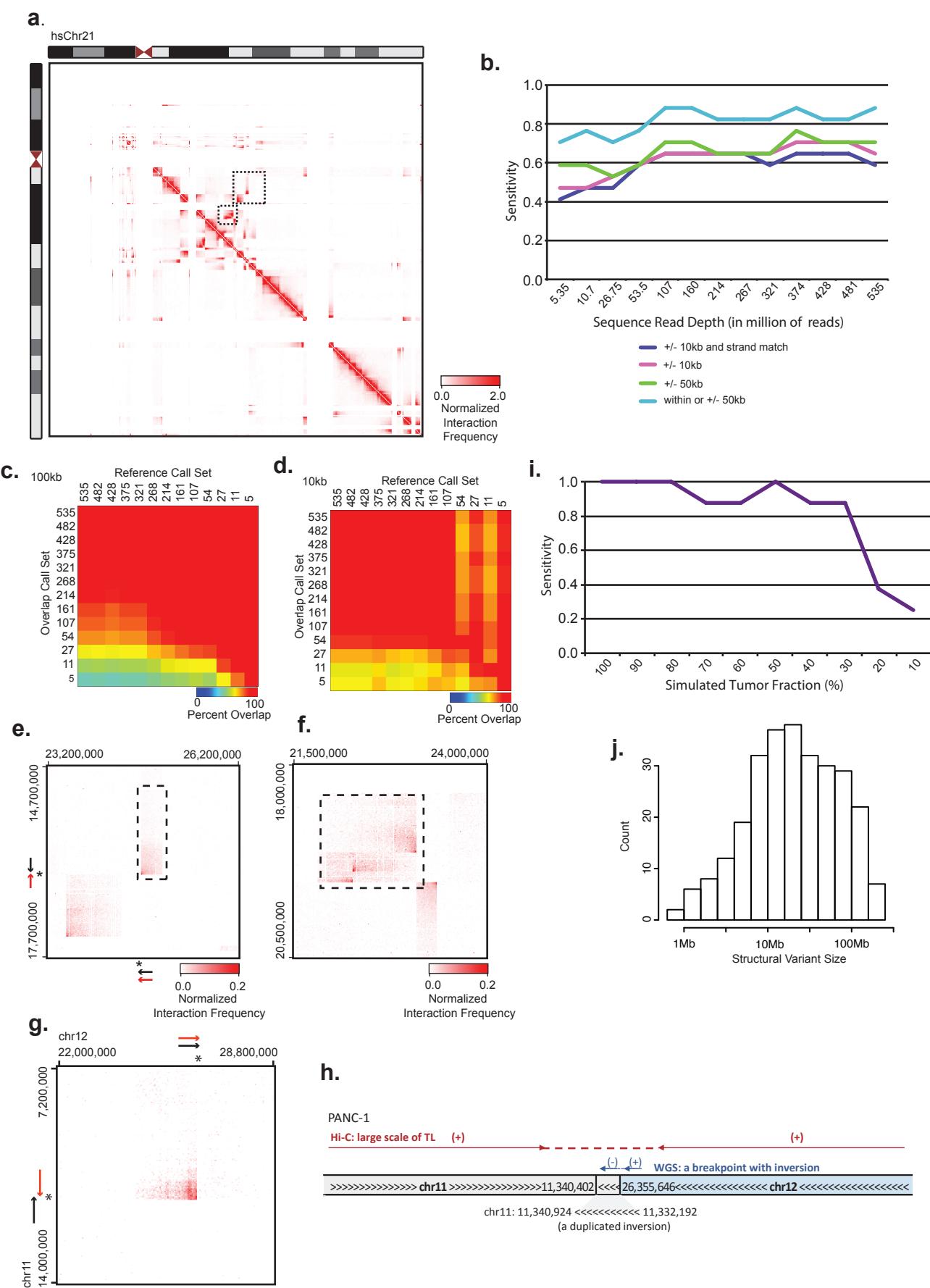
Supplementary Figure 5 | Cancer genomes possess extensive CNVs and translocations. **a-f.** Genome profiles of 6 cancer cell lines. All SVs are detected by at least two out of the three methods (Hi-C, optical mapping and WGS). Tracks from outer to inner circles are chromosome coordinates, CNVs, deletions (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-l.** Large intra-chromosomal rearrangements and inter-chromosomal translocations detected by Hi-C in 6 cancer cell lines.



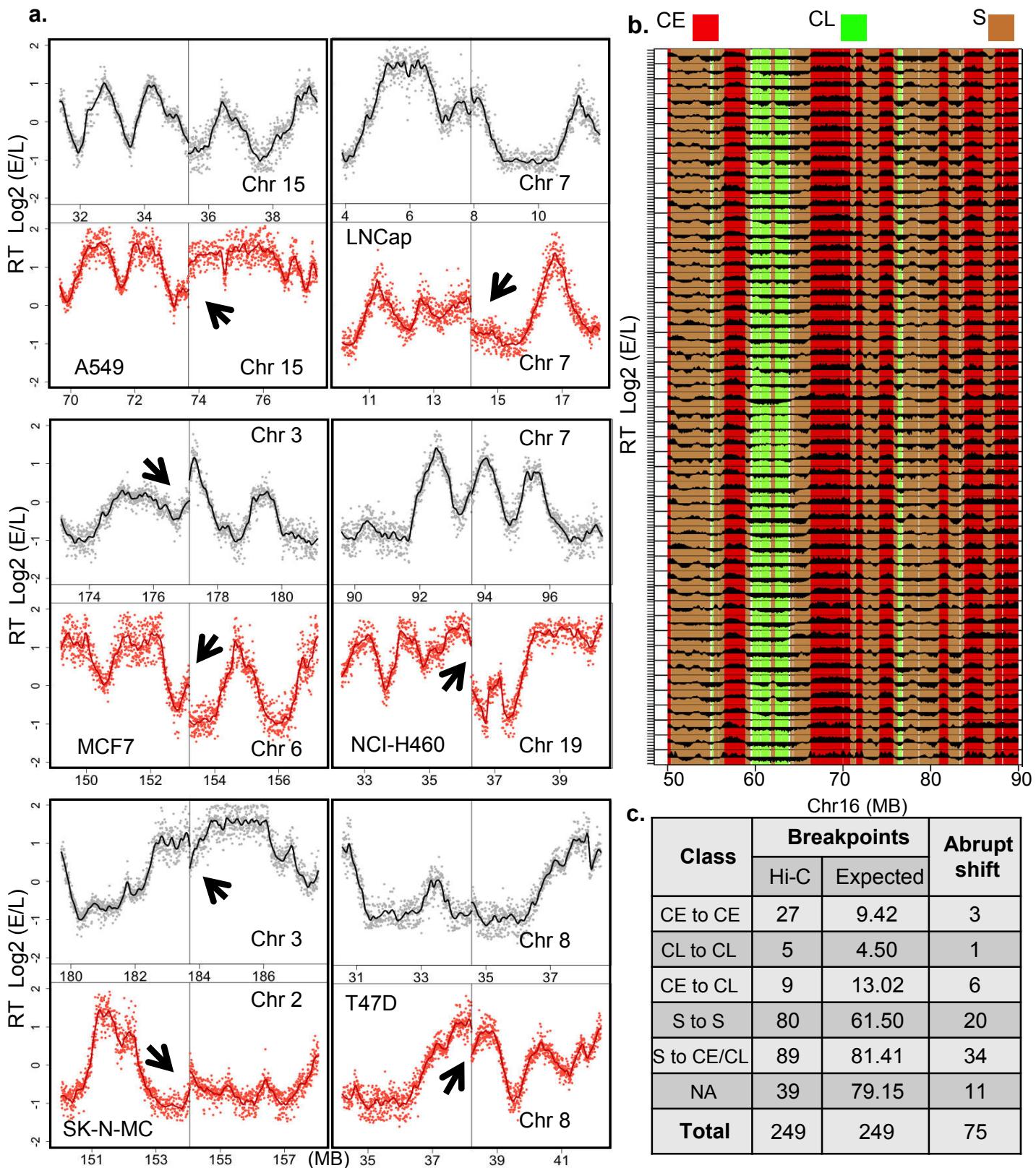
Supplementary Figure 6 | Identification of rearrangement breakpoints at kilobase resolution by Hi-C. a. Hi-C signal between chromosomes 9 and 22 in K562 cells. The left hand matrix is shown at a resolution of 1Mb, and the smaller matrices use progressively smaller bin sizes to narrow in on the breakpoint.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

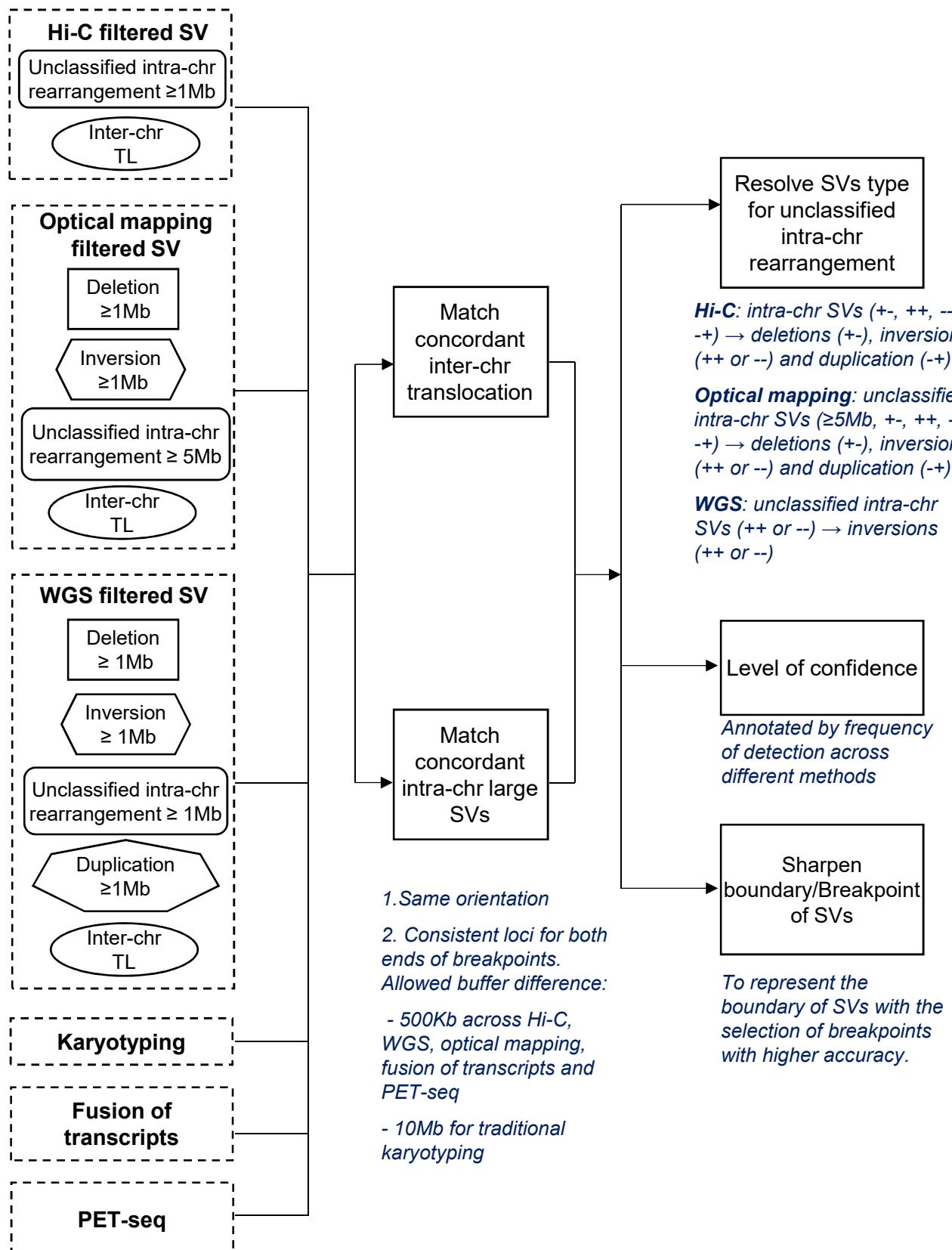


Supplementary Figure 8 | Sensitivity and internal reproducibility of rearrangements identified by Hi-C. **a.** Hi-C data from human chromosome 21 Tc1 cells. Dotted lines indicate regions shown in panels e and f. **b.** Sensitivity of Hi-C to detect gold standard SV calls at different sequencing depths. 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.



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.

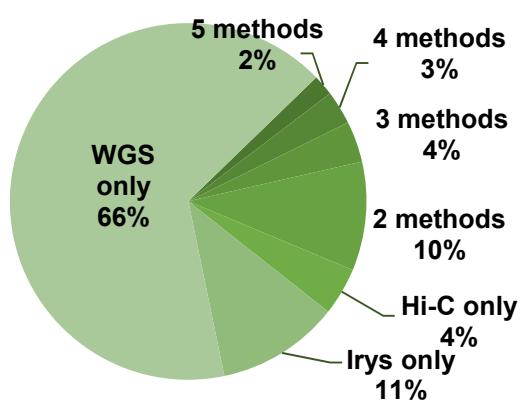
Comparative integration of inter-chr translocation and large intra-chr SVs ($\geq 1\text{Mb}$)



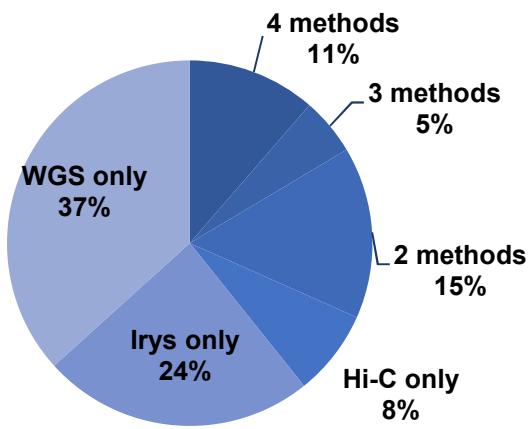
Supplementary Figure 10 | Comparison and integration of inter-chromosomal translocations and large intra-chromosomal SVs ($\geq 1\text{Mb}$). We compared the SV calls by Hi-C, optical mapping and WGS, and we also included SV calls from additional methods, including karyotyping, fusion transcripts, and 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



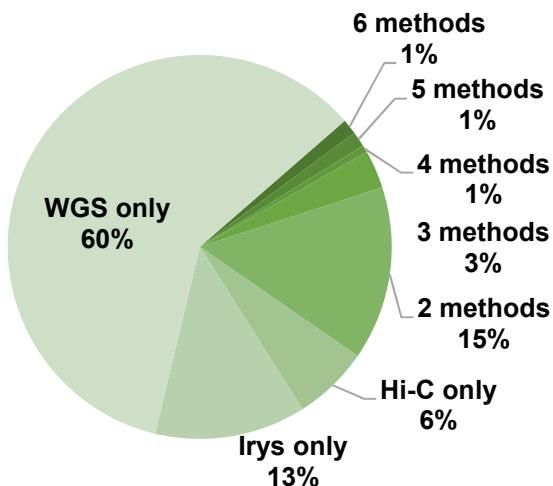
80 Intra-chr large SVs ($\geq 1\text{Mb}$)



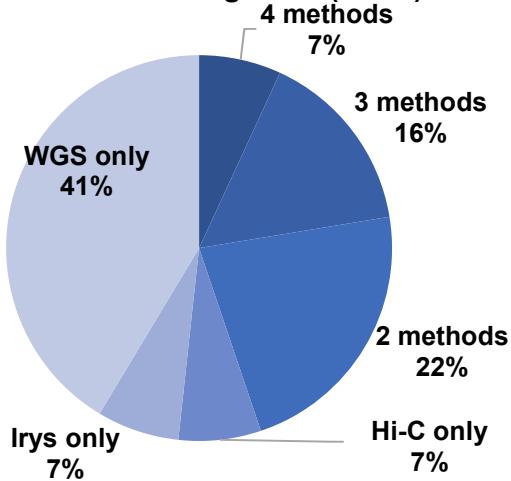
b.

K562:

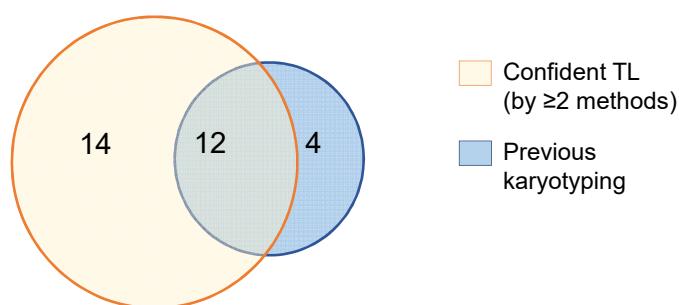
178 Inter-chr translocation



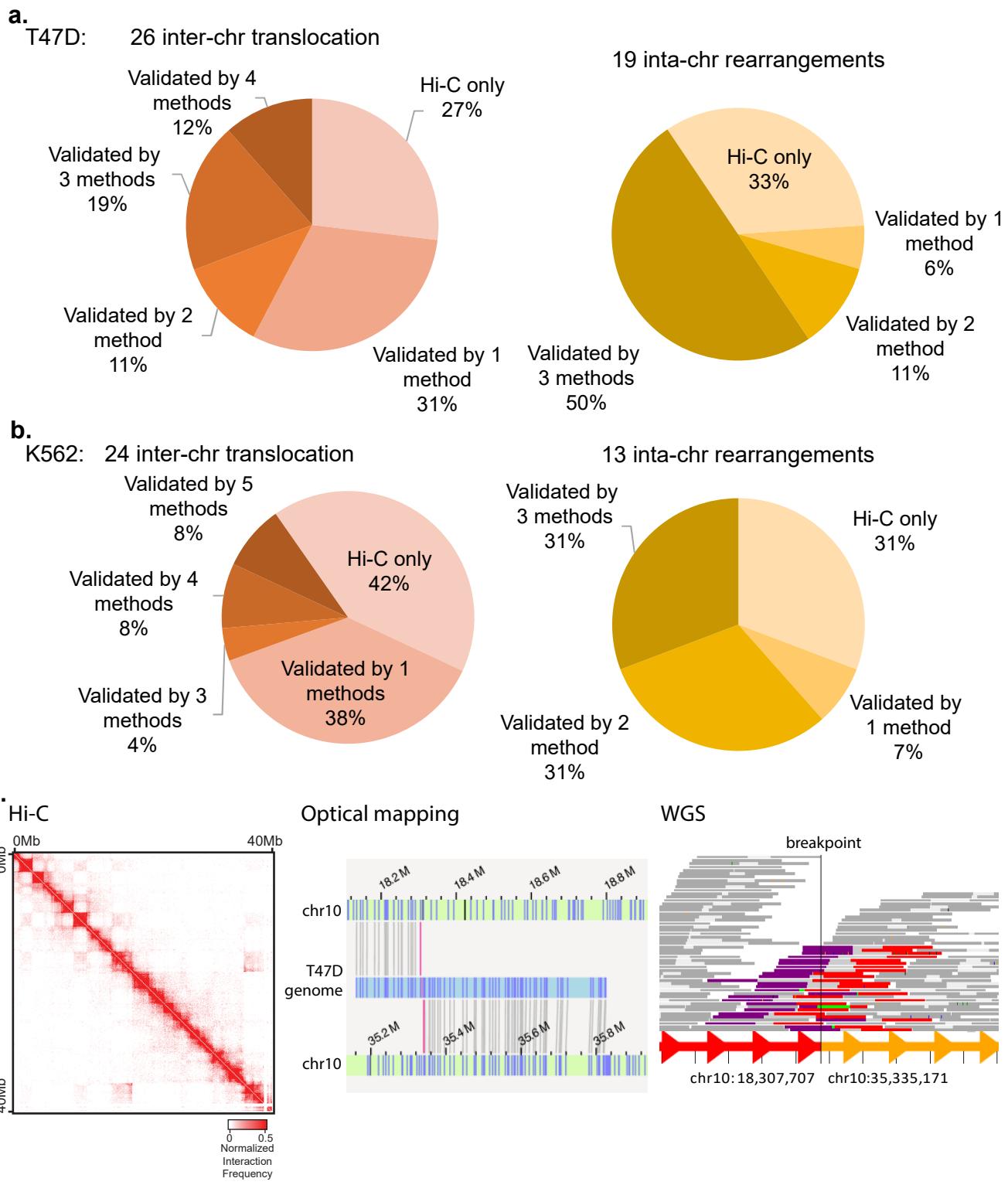
58 Intra-chr large SVs ($\geq 1\text{Mb}$)



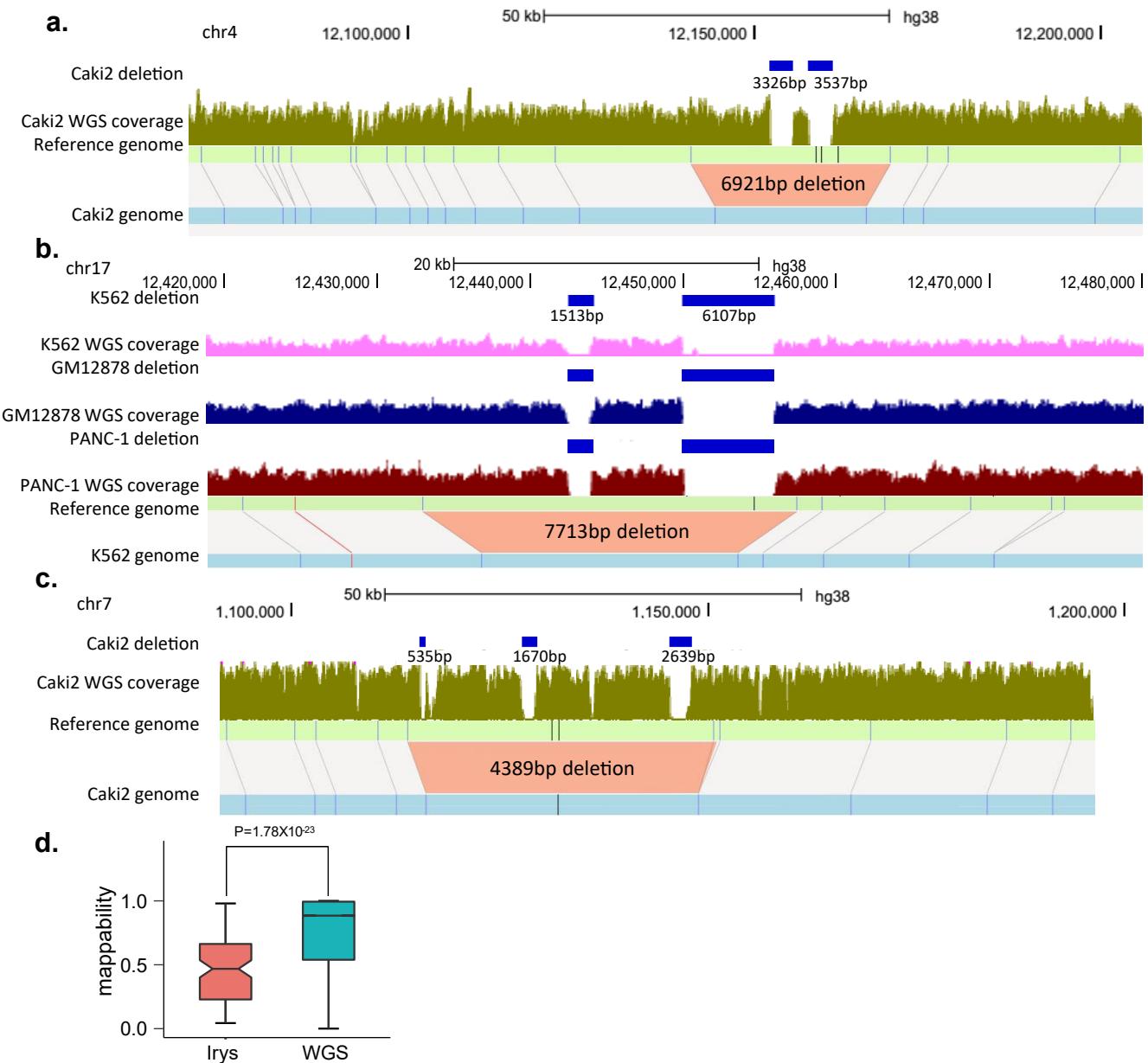
c. T47D



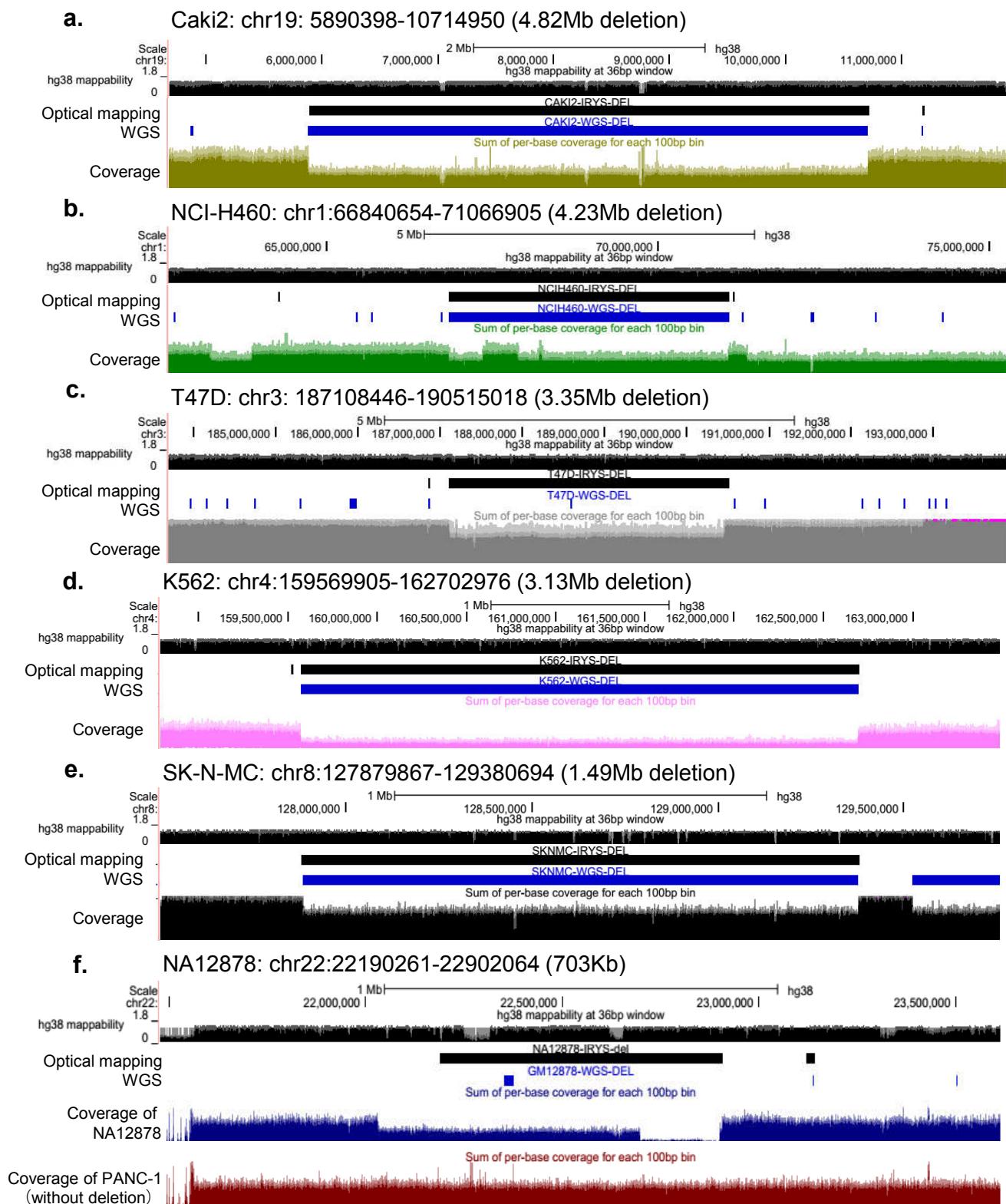
Supplementary Figure 11 | Contribution to the overall profile of large structural variants in T47D and K562 cells by different methods. a,b. We first compared inter-chromosomal translocations and large intra-chromosomal rearrangements ($\geq 1\text{Mb}$) detected by Hi-C, optical mapping and WGS and 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.



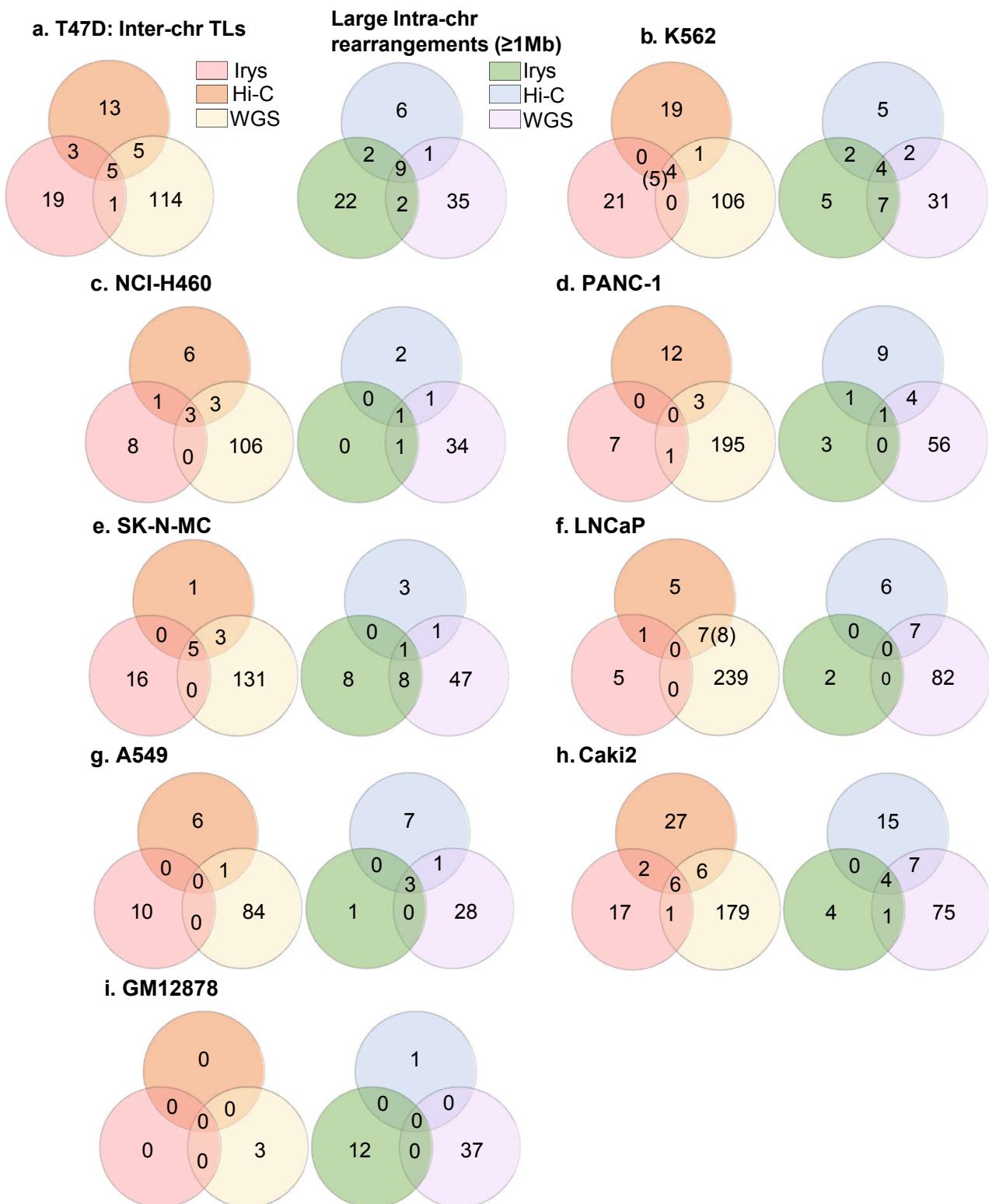
Supplementary Figure 12 | Cross-platform validation of structural variants detected by Hi-C using other five methods. Inter-chromosomal translocations and large intra-chromosomal rearrangements ($\geq 1\text{ Mb}$) detected by Hi-C are validated by WGS, optical mapping, karyotyping, fusion of transcripts 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.

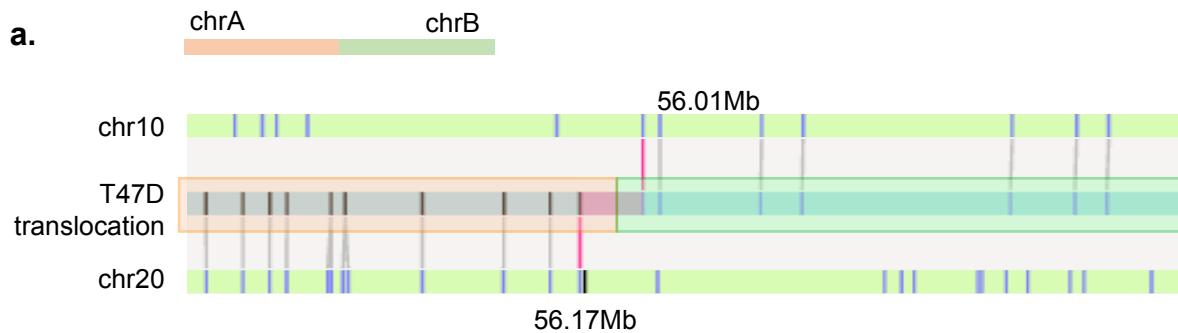


Supplementary Figure 14 | Examples of large deletions in normal and cancer genomes. **a-e.** Large deletions in cancer genomes, detected by both optical mapping and WGS coverage. **f.** A 703kb deletion in GM12878 cell line, which overlaps with IGL locus and represents potential V(D)J recombination.

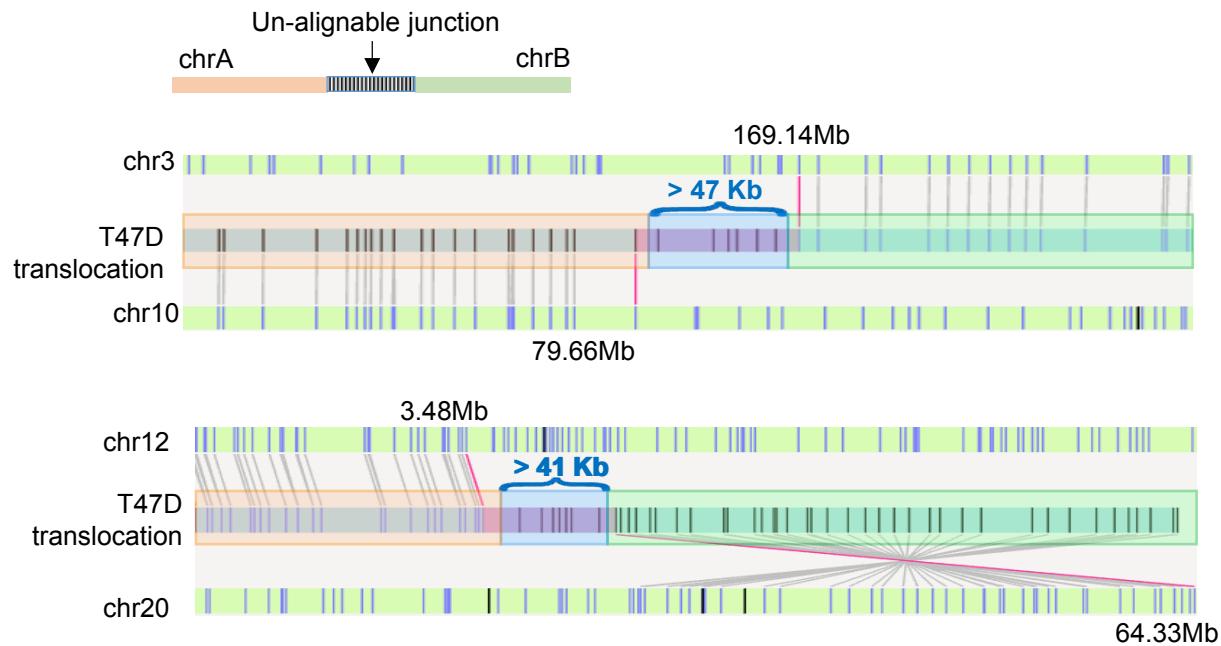


Supplementary Figure 15 | Overlap of large SVs detected by Hi-C, optical mapping, and WGS. Number of inter-chromosomal translocations (left panel) and large intra-chromosomal rearrangements ($\geq 1\text{Mb}$, right panel) detected by optical mapping, Hi-C, and WGS in T47D (a), K562 (b), NCI-H460 (c), 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

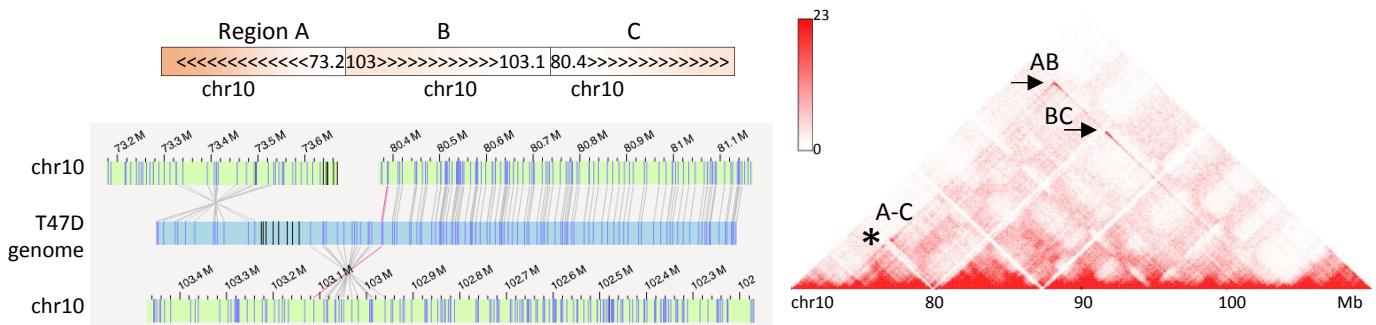


Complex translocations detected by Hi-C and optical mapping with un-alignable junctions

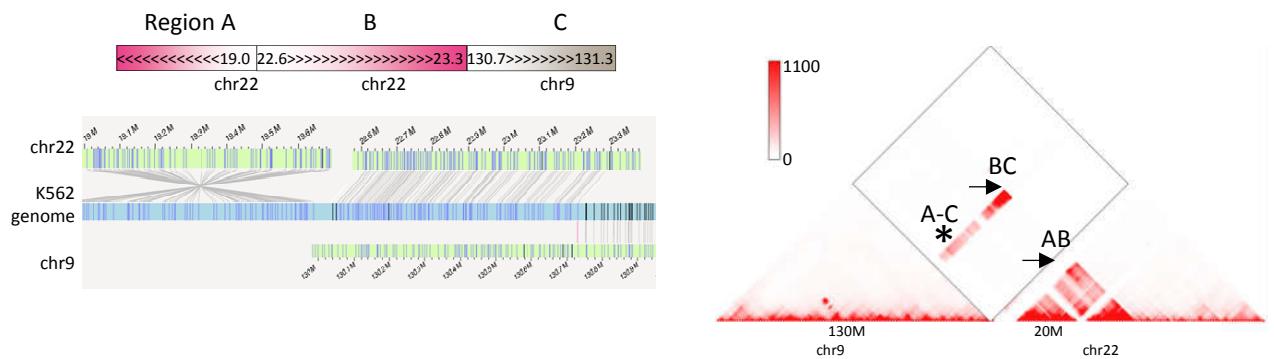


Supplementary Figure 16 | Hi-C and optical mapping detect translocations with unalignable junctions. **a.** An example of a simple translocation detected by WGS, Hi-C, and optical mapping. The predicted breakpoint is located between the two labels (nicking enzymes) and there is no unalignable region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

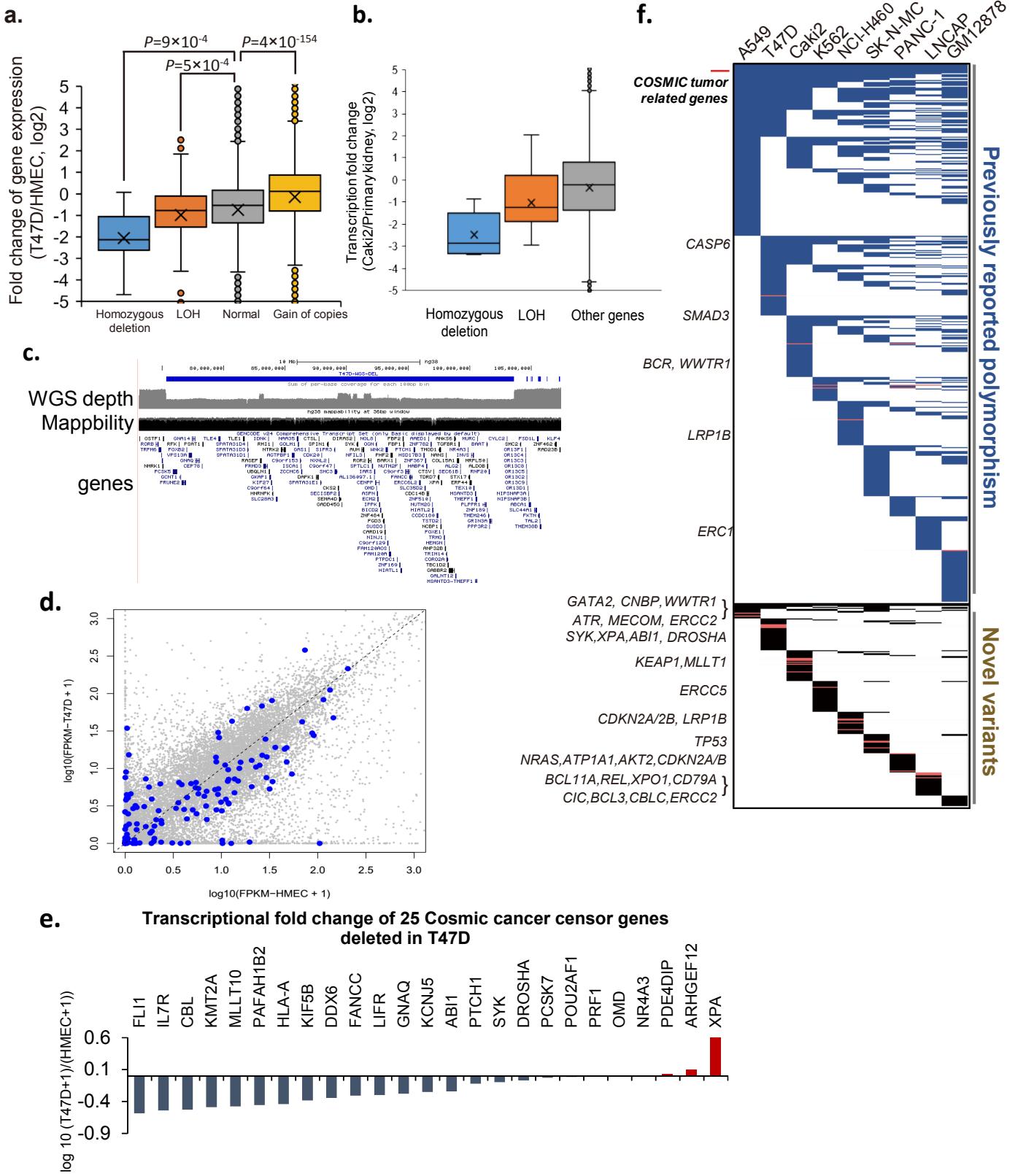
a. T47D



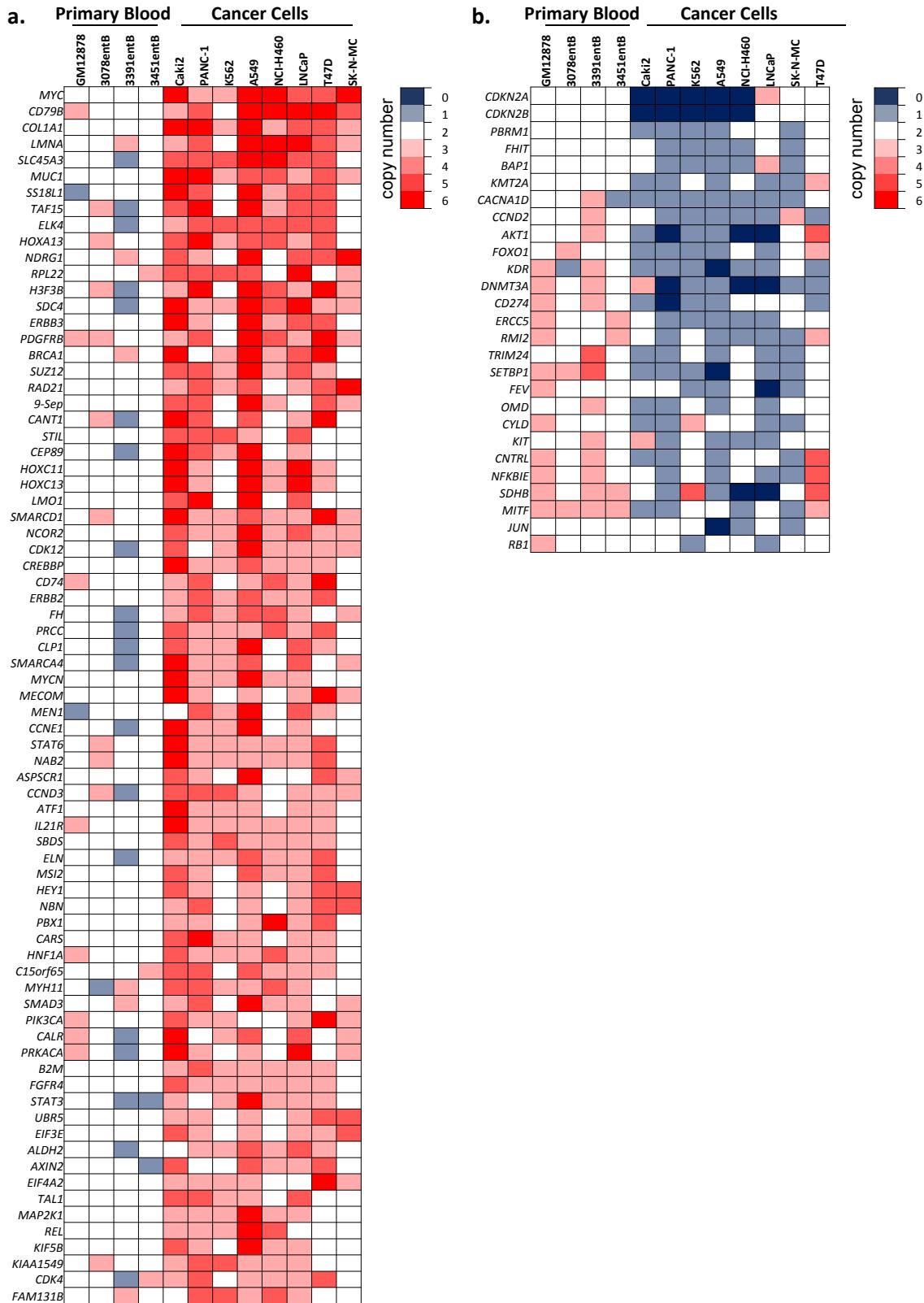
b. K562



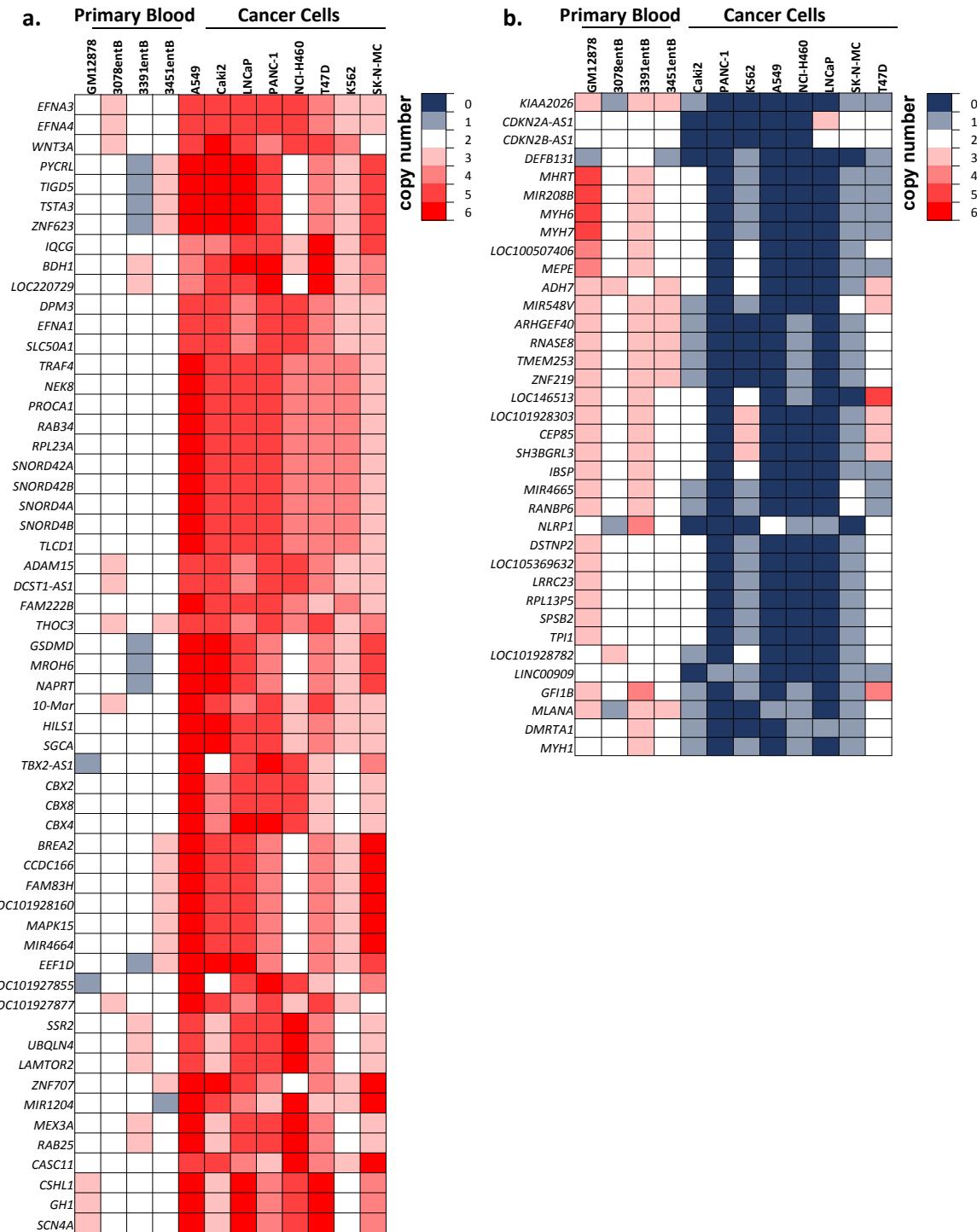
Supplementary Figure 17 | Examples of using Hi-C and optical mapping to reconstruct the overall structure of complex translocations. Similar to Fig. 3d, Arrow (->) indicates directly jointed translocation and asterisk (*) marks the linked adjacent SV. **a.** Schematic of the local chromosome structure 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.



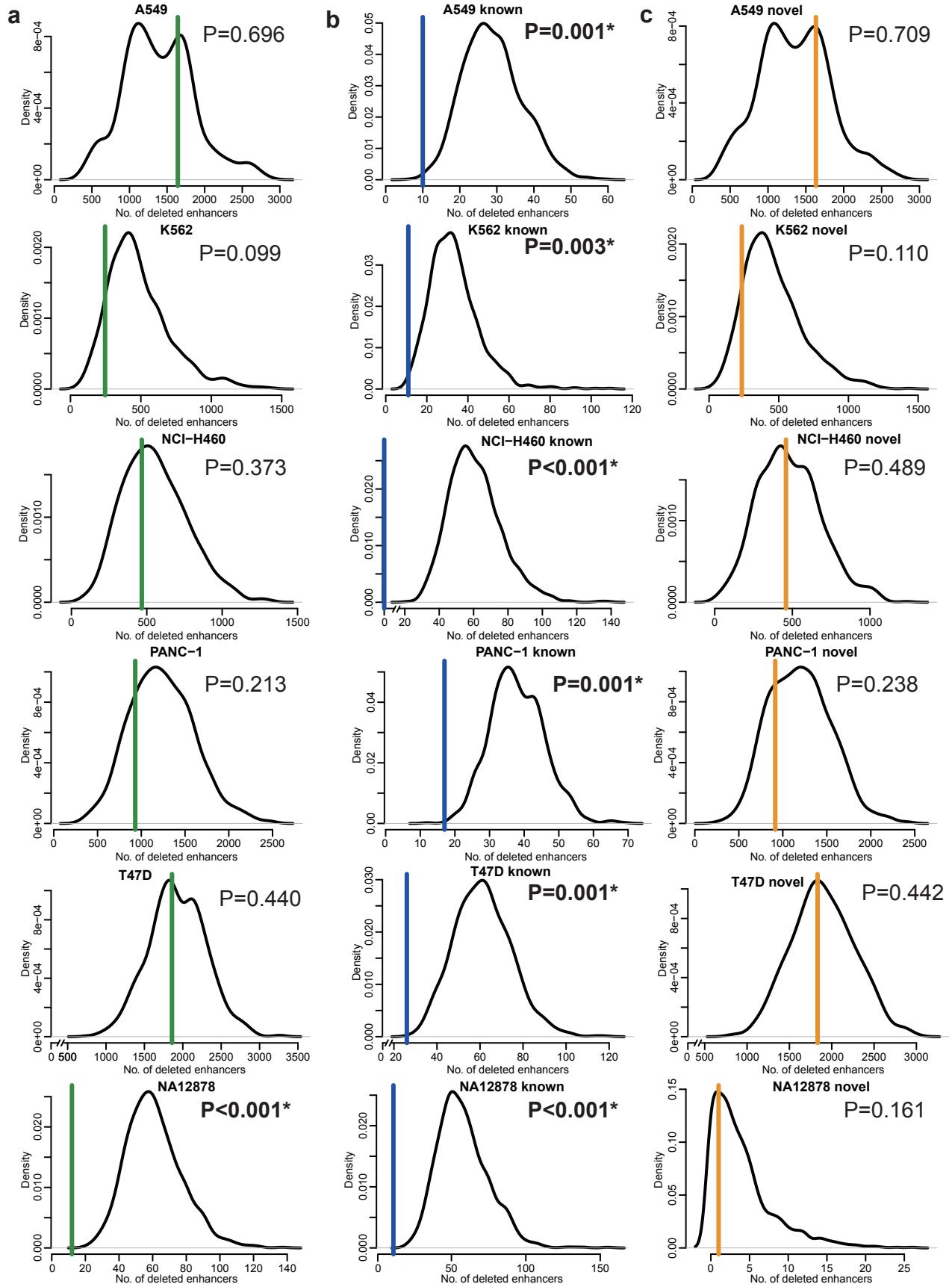
Supplementary Figure 18 | Impact of exon deletion and copy loss on gene expression. **a.** Compared with HMEC cells, expressed genes (FPKM>1 in HMEC cells) with homozygous deletions (n=10) and LOH (n=325) in T47D cells show reduced expression compared to copy-neutral genes (n=5113, P = 0.009 and 0.003 respectively, two-sided Wilcoxon rank sum test), and compared to gain of copy genes (n=6413, p=4×10-79, 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).



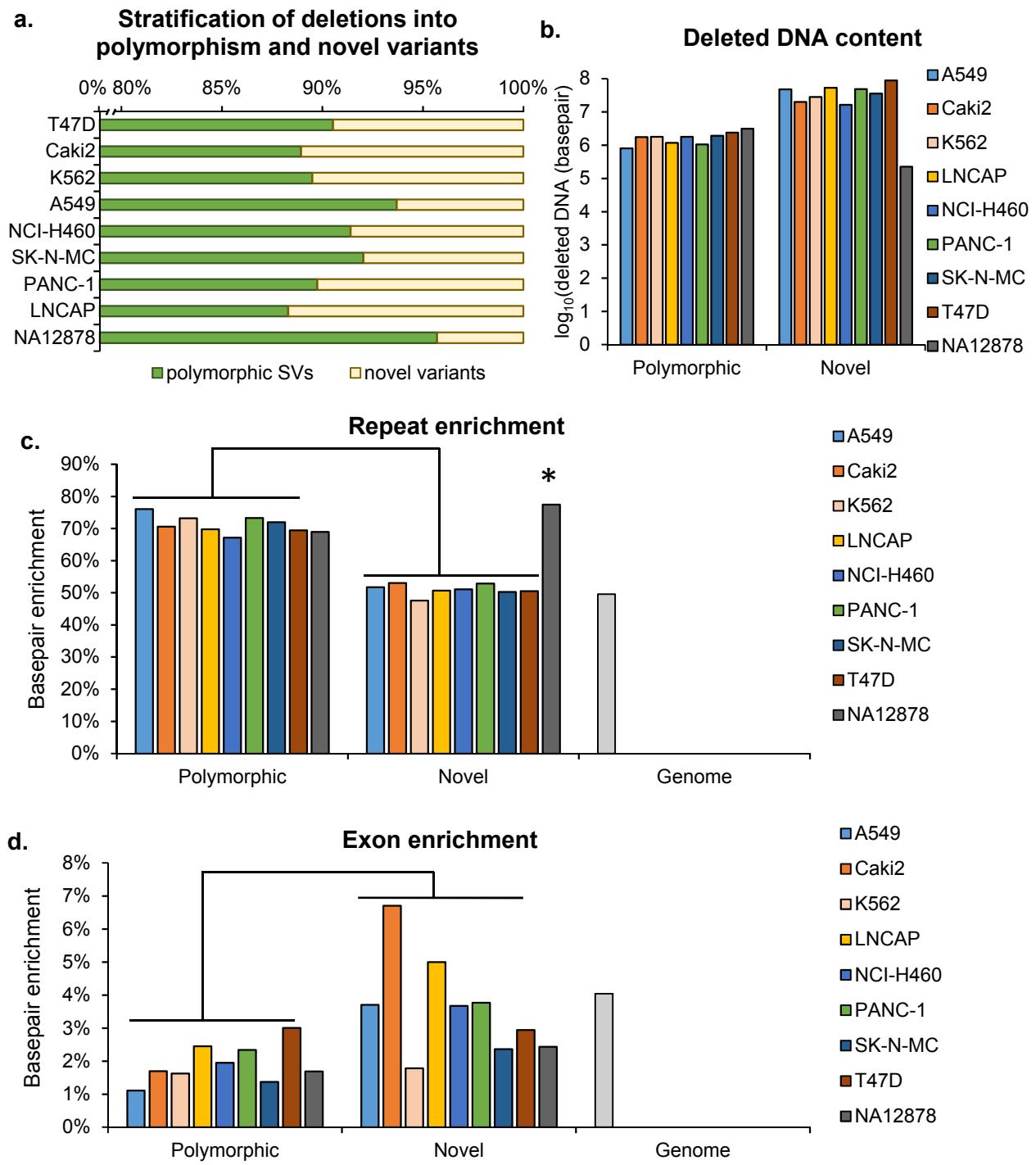
Supplementary Figure 19 | Copy number alterations of COSMIC tumor-related genes, which are computed based on its surrounding 50 kb regions by optical mapping. **a.** COSMIC tumor-related genes with extensive gain of copies in cancer cell lines. **b.** COSMIC tumor-related genes with significant loss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



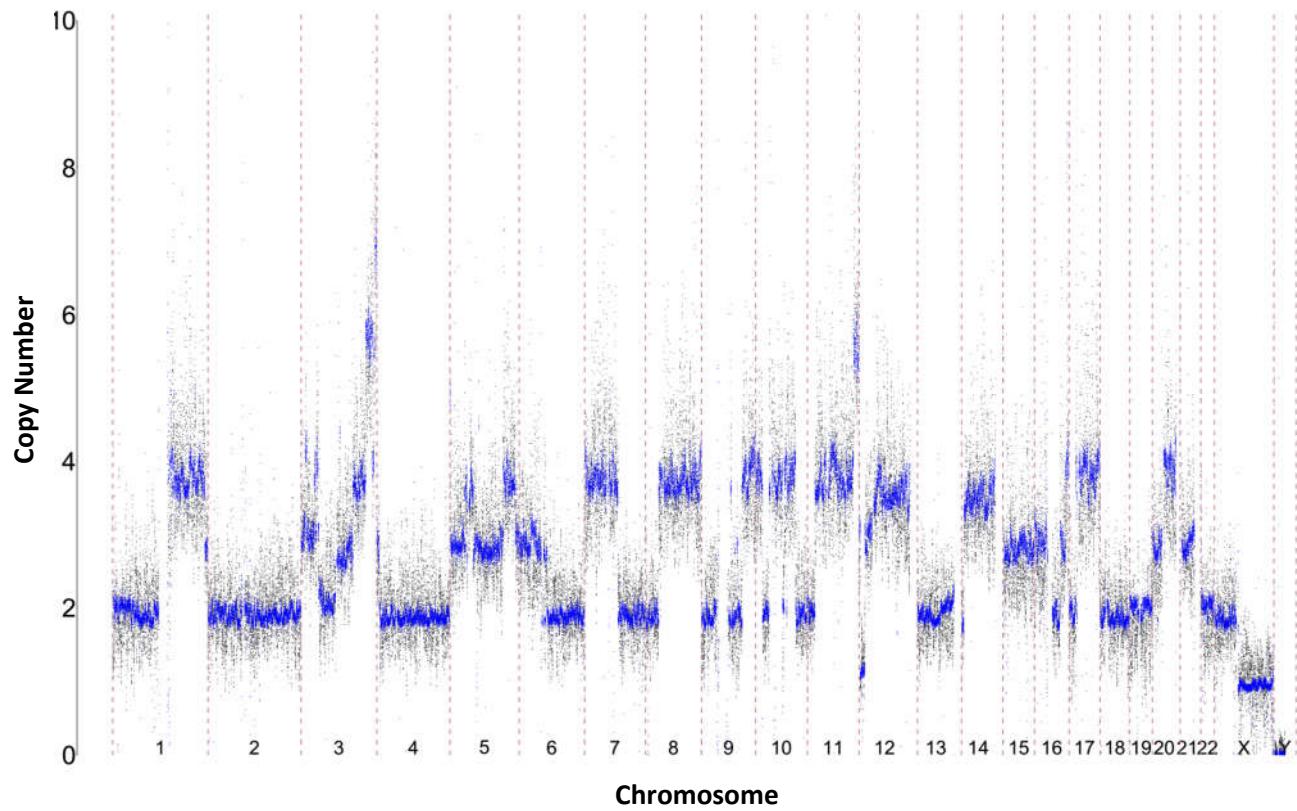
Supplementary Figure 21 | Comparison of the frequency of enhancer disruptions versus expectation. **a.** Overall, we found that deletions in normal cell types (GM12878) are less likely to delete enhancer that would be expected at random, while the enrichment level of deletions in enhancers in cancer 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.



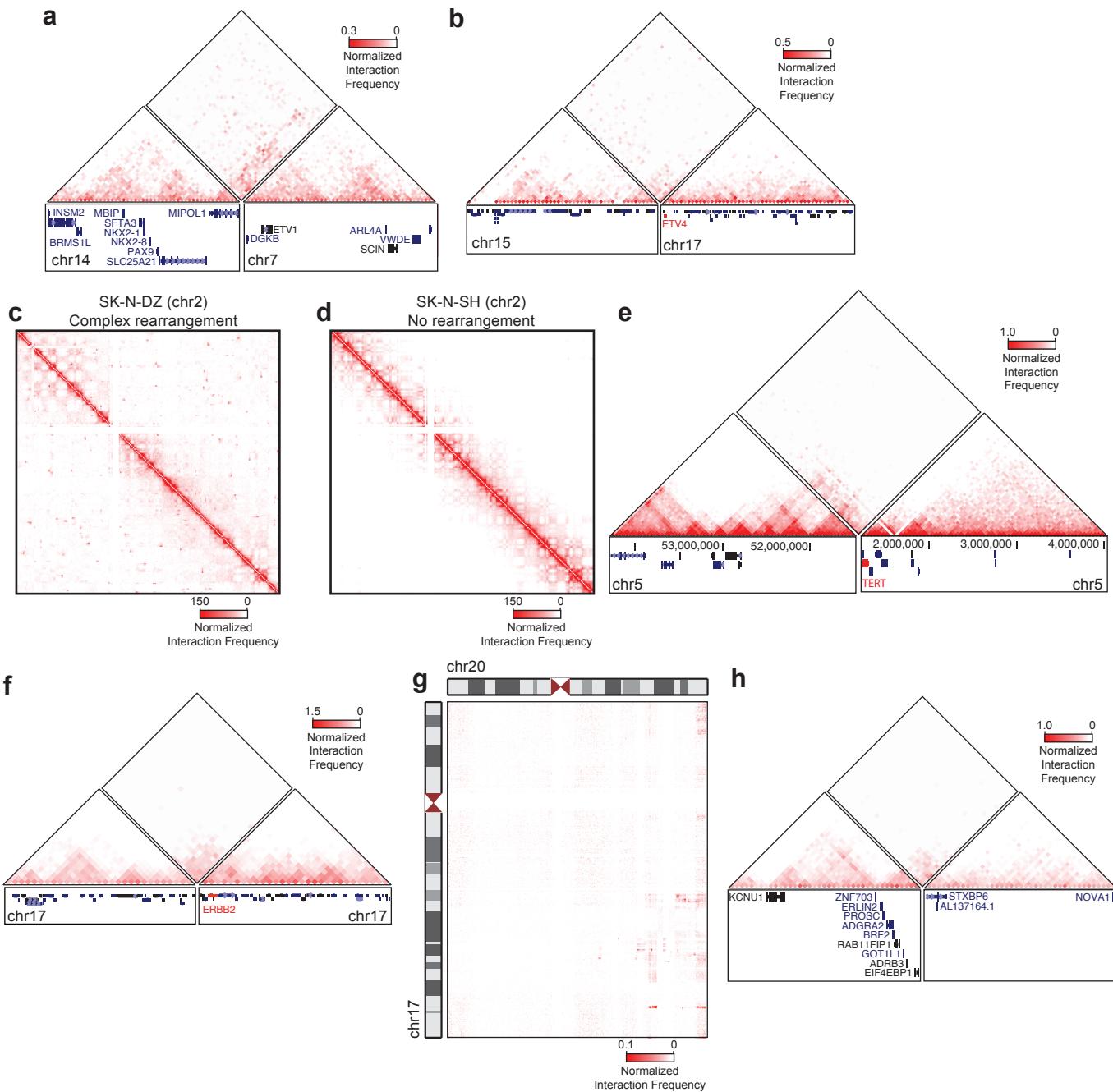
Supplementary Figure 22 | Characterization of known polymorphic deletions and novel deletions. We stratified the deletions into two categories by comparing with DGV database: polymorphic deletions and novel deletions. **a.** Only 5% of deletions in GM12878 cells are novel variants, whereas on average 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.