

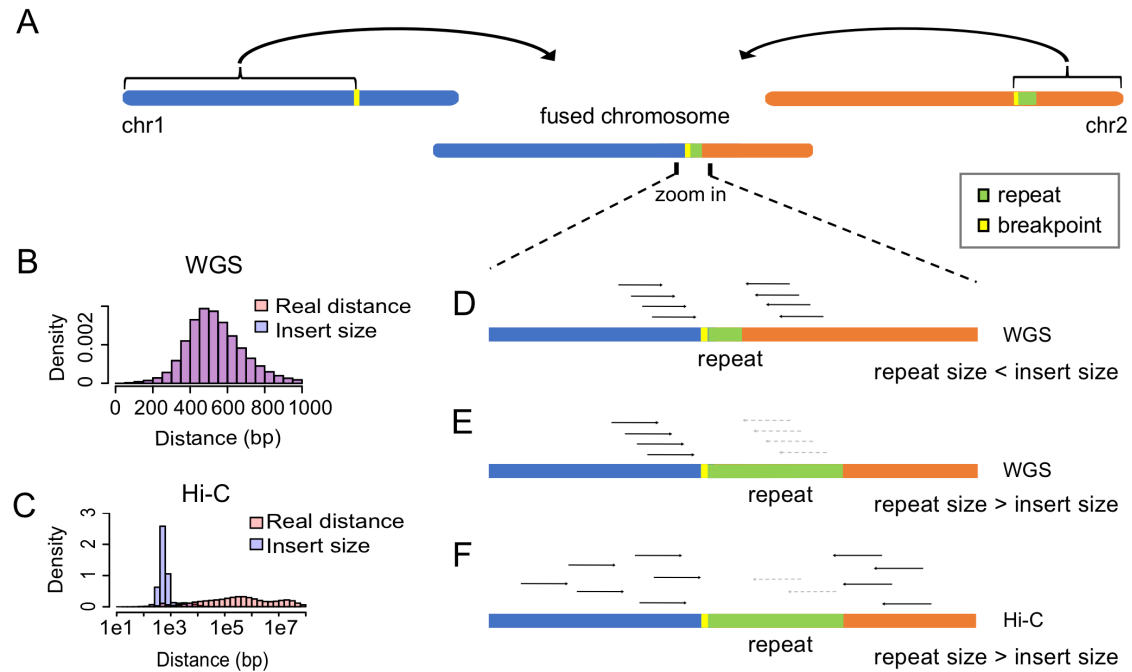


Fig. S1. Hi-C data is superior to WGS in variation detection in repetitive regions. **A**, Illustration of a fused chromosome with a breakpoint located in repetitive region. **B-C**, The distribution of the real distances (pink) between two mates, and the insert sizes (light blue) in WGS (**B**) and Hi-C (**C**). **D**, Reads can be correctly mapped to the reference genome if repeat size is less than the insert size in WGS. **E**, Reads cannot be correctly mapped to the reference genome if repeat size is larger than the insert size in WGS. **F**, Reads surrounding the repetitive regions can be used to detect the breakpoint in Hi-C.

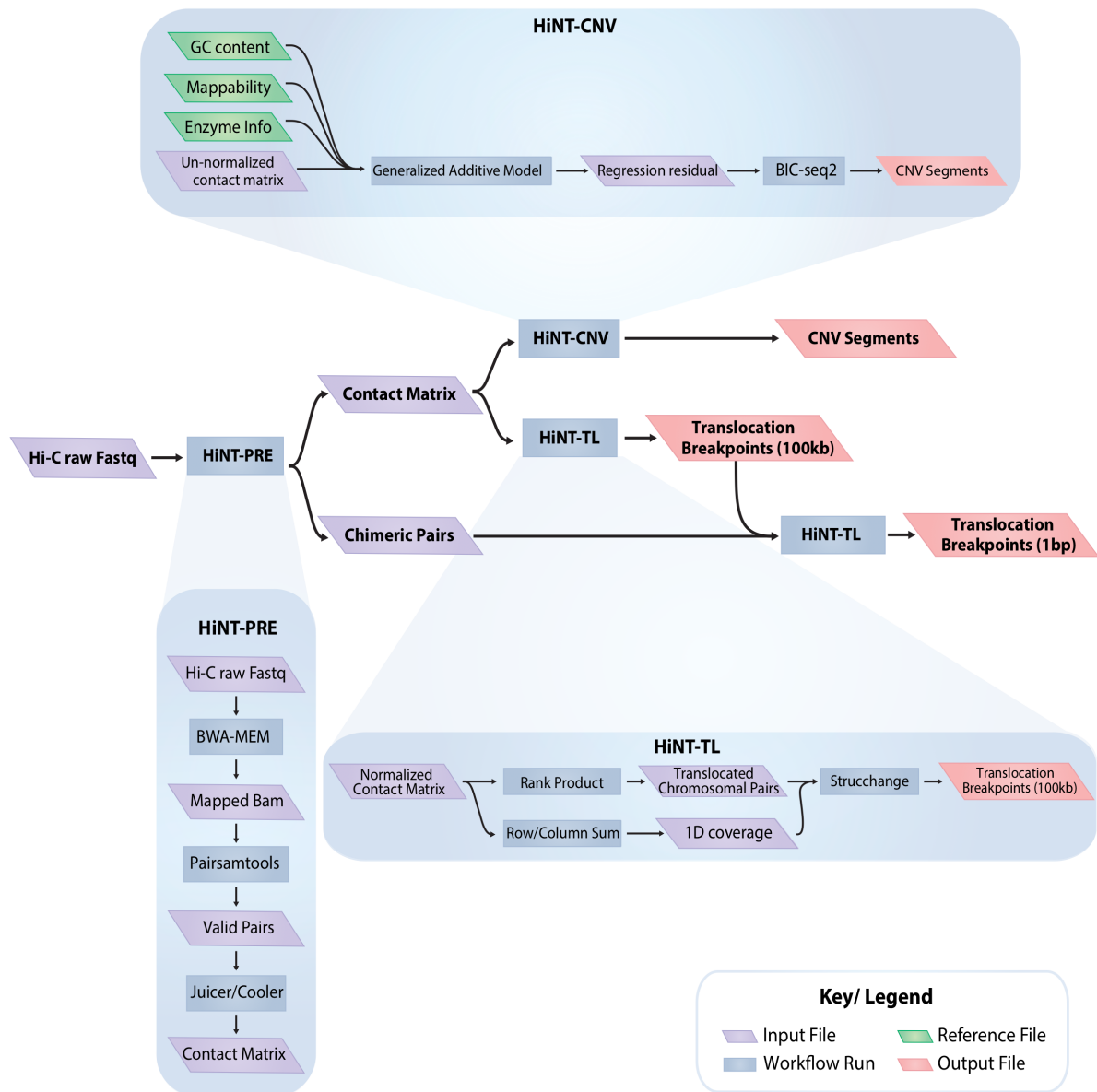


Fig. S2. Overview of the HiNT workflow. HiNT has three components: HiNT-PRE, HiNT-CNV, and HiNT-TL. HiNT-PRE preprocesses Hi-C data to generate the contact matrix; HiNT-CNV performs CNV detection; and HiNT-TL detects translocation breakpoints at 100kb as well as base-pair resolution.

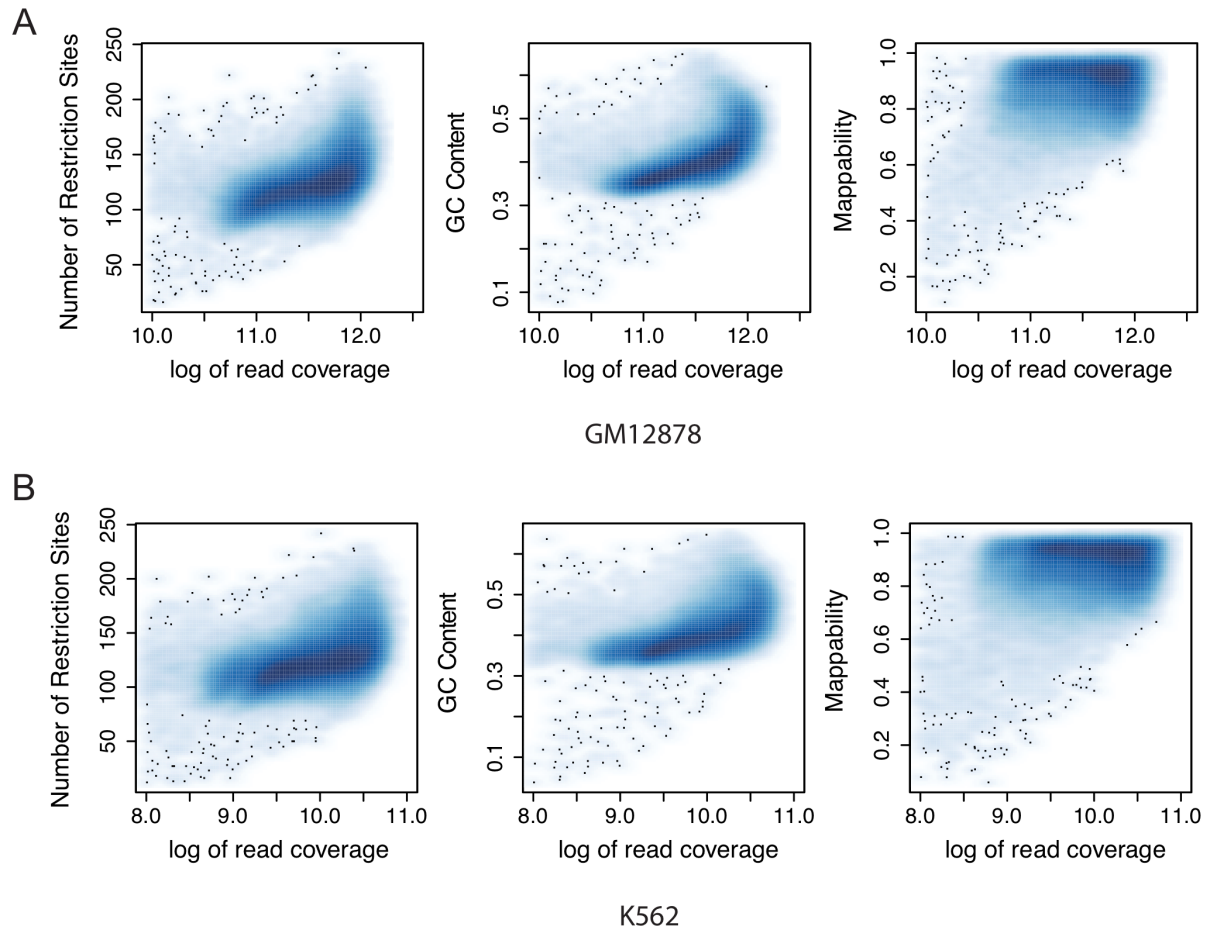


Fig. S3. Correlation between the natural log of 1D coverage and the number of restriction sites (left), GC content (middle), and mappability (right) in each 50kb bin in GM12878 (A) and K562 (B) cell.

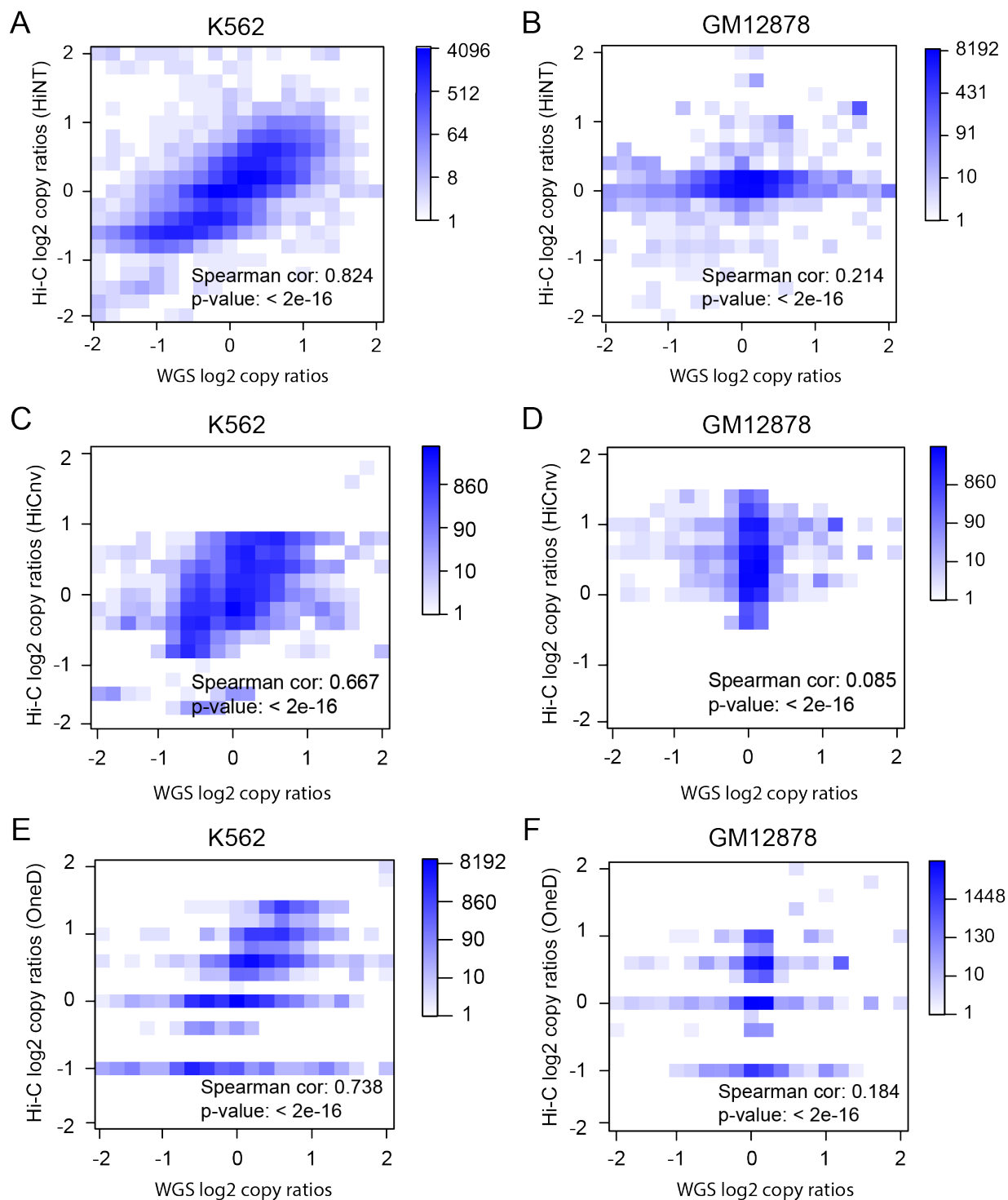


Fig. S4. CNVs detected by HiNT from Hi-C are consistent with those detected from WGS. A-B, Correlation of log2 copy ratios in each bin (50kb) detected from WGS and Hi-C (HiNT) in K562 (A) and GM12878 (B). **C-D,** Correlation of log2 copy ratios in each bin (50kb) detected from WGS and Hi-C (HiCnv) in K562 (C) and GM12878 (D). **E-F,** Correlation of log2 copy ratios in each bin (50kb) detected from WGS and Hi-C (OneD) in K562 (E) and GM12878 (F).

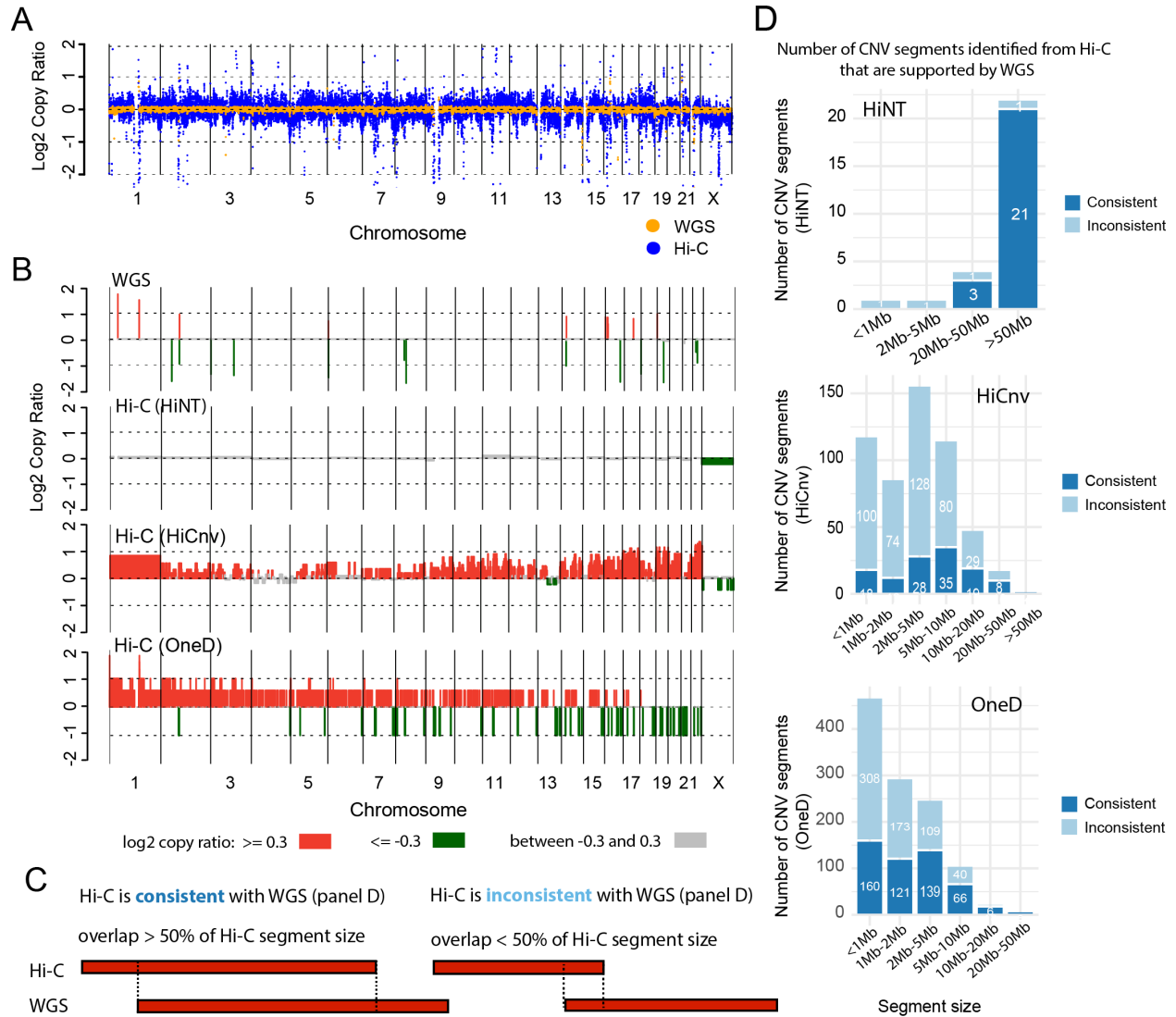


Fig. S5. Copy number inference in GM12878 cells. **A**, Comparison of log2 copy ratios calculated using regression residuals from Hi-C (blue) and using read coverage from WGS (orange). **B**, Comparison of CNV profiles from Hi-C and WGS after segmentation. Red, green and grey bars represent copy gain(log2 copy ratio > 0.3), copy loss(log2 copy ratio < -0.3), and copy neutral regions (log2 copy ratio between -0.3 and 0.3), respectively. **C**, Schematic of the consistency analysis. CNV segment detected from Hi-C is consistent with that detected from WGS if the overlapped region is larger than 50% of the original segment size, and vice versa. **D**, The number of CNV segments (categorized by size) detected from Hi-C that are also supported by WGS. Specifically, 92%, 28%, and 60% of the large CNV segments identified by HiNT, HiCnv, and OneD are supported by those from WGS, respectively. The overlap criteria for consistency are shown in panel C.

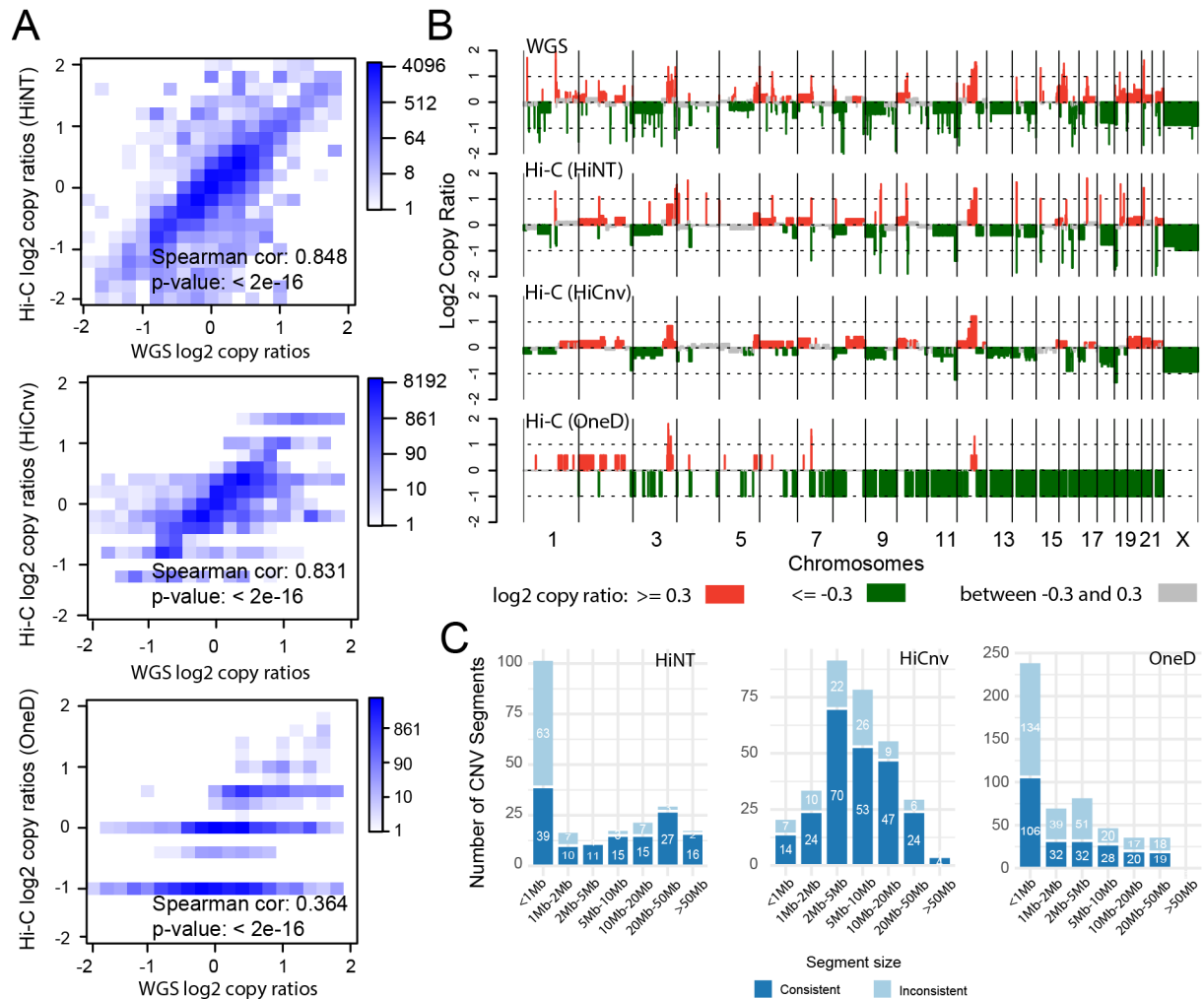


Fig. S6. Evaluation of copy number inference from Hi-C data in Caki2 cells. **A**, Correlation of log2 copy ratios in each bin (50kb) detected from WGS and HiNT, HiCnv, and OneD in Caki2. **B**, Comparison of CNV profiles from Hi-C and WGS after segmentation. Red, green and grey bars represent copy gain(log2 copy ratio > 0.3), copy loss(log2 copy ratio < -0.3), and copy neutral regions (log2 copy ratio between -0.3 and 0.3), respectively. **C**, The number of CNV segments (categorized by size) detected from Hi-C that are also supported by WGS. Specifically, 84%, 76%, and 49% of the large CNV segments identified by HiNT, HiCnv, and OneD are supported by those from WGS, respectively. The overlap criteria for consistency are shown in Supp. Fig. 5C.

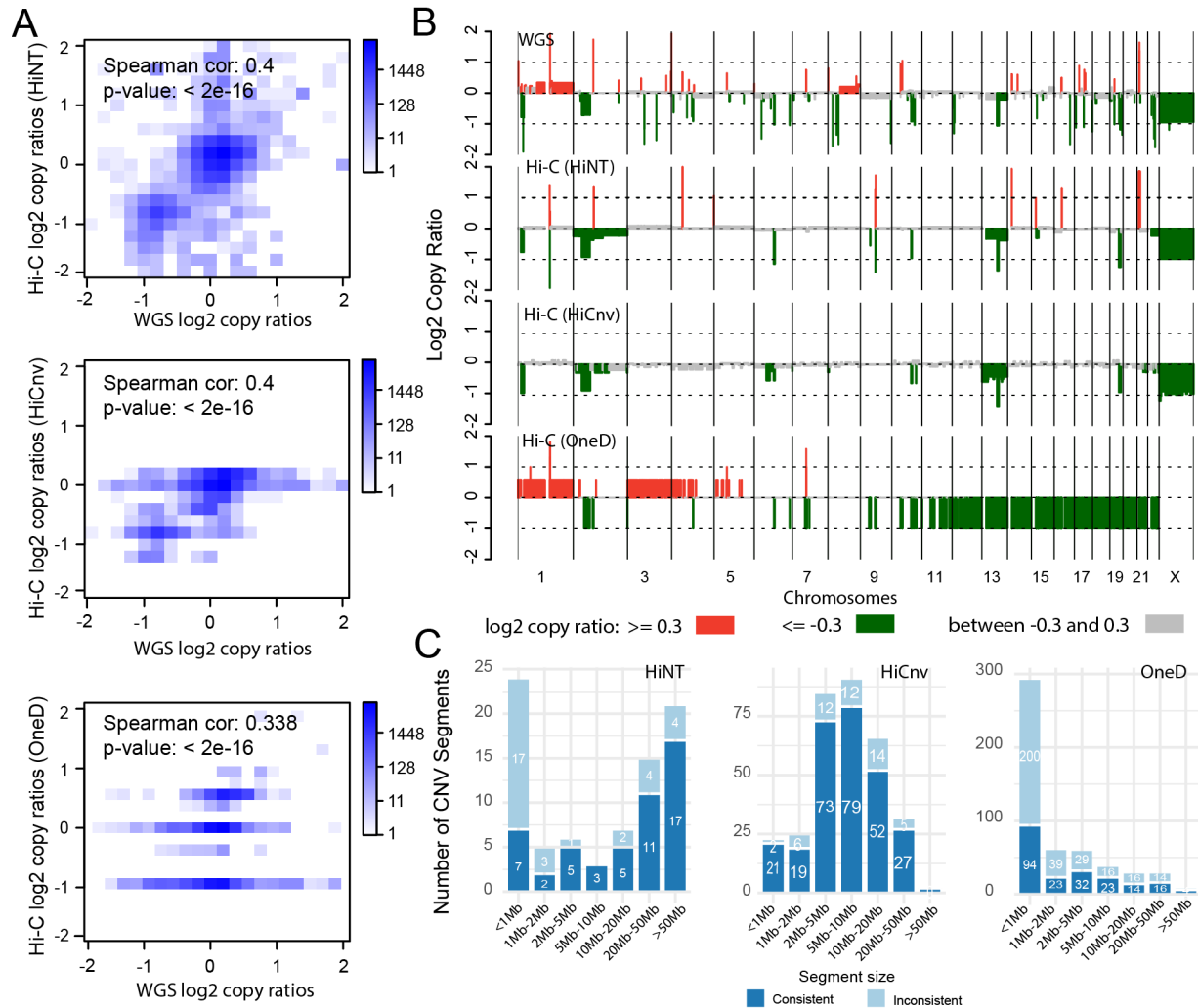


Fig. S7. Evaluation of copy number inference from Hi-C data in LNCaP cells. **A**, Correlation of log2 copy ratios in each bin (50kb) detected from WGS and HiNT, HiCnv, and OneD in LNCaP. **B**, Comparison of CNV profiles from Hi-C and WGS after segmentation. Red, green and grey bars represent copy gain(log2 copy ratio > 0.3), copy loss(log2 copy ratio < -0.3), and copy neutral regions (log2 copy ratio between -0.3 and 0.3), respectively. **C**, The number of CNV segments (categorized by size) detected from Hi-C that are also supported by WGS. Specifically, 79%, 84%, and 54% of the large CNV segments identified by HiNT, HiCnv, and OneD are supported by those from WGS, respectively. The overlap criteria for consistency are shown in Supp. Fig. 5C.

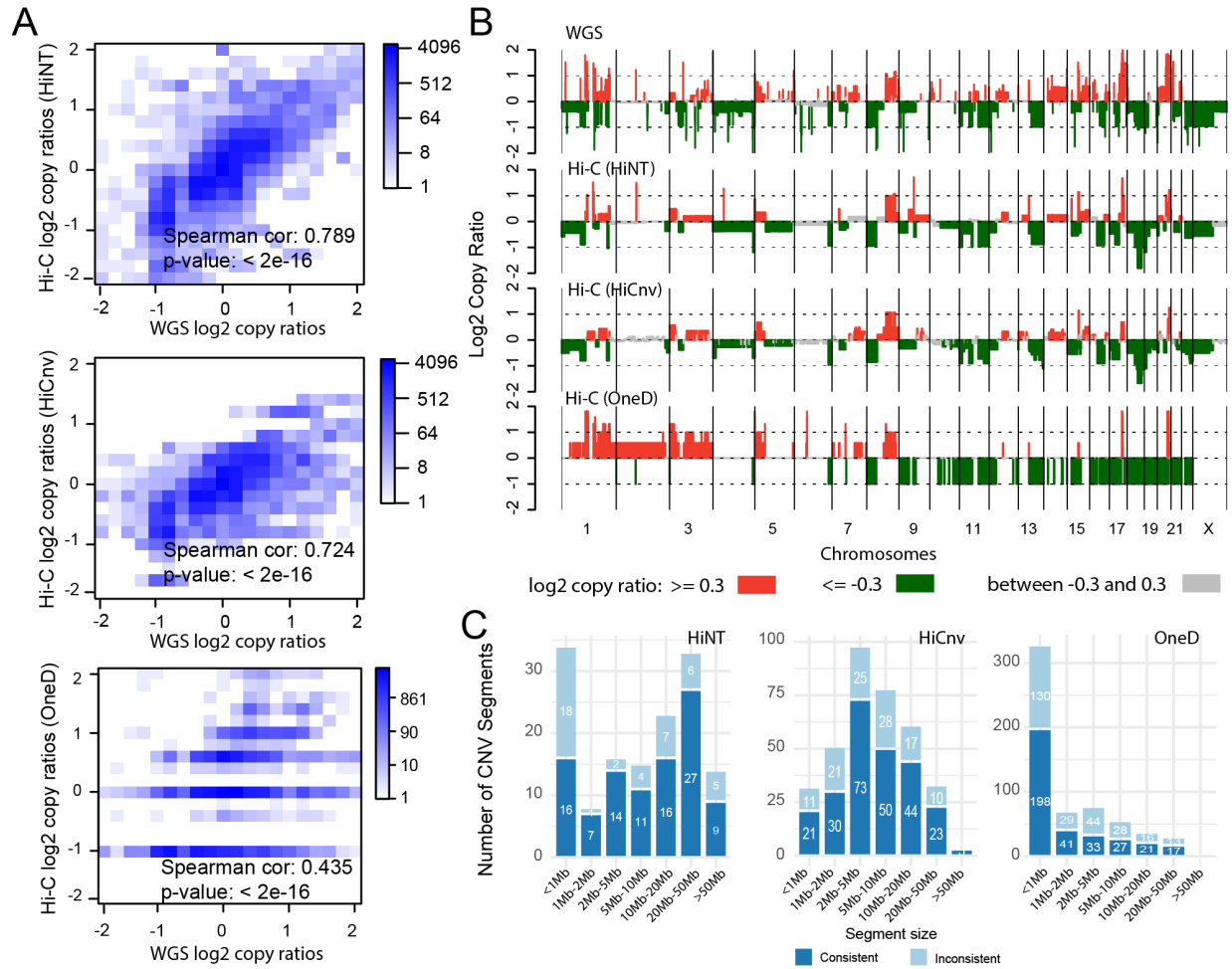


Fig. S8. Evaluation of copy number inference from Hi-C data in MCF7 cells. **A**, Correlation of log2 copy ratios in each bin (50kb) detected from WGS and HiNT, HiCnv, and OneD in MCF7. **B**, Comparison of CNV profiles from Hi-C and WGS after segmentation. Red, green and grey bars represent copy gain(log2 copy ratio > 0.3), copy loss(log2 copy ratio < -0.3), and copy neutral regions (log2 copy ratio between -0.3 and 0.3), respectively. **C**, The number of CNV segments (categorized by size) detected from Hi-C that are also supported by WGS. Specifically, 76%, 71%, and 49% of the large CNV segments identified by HiNT, HiCnv, and OneD are supported by those from WGS, respectively. The overlap criteria for consistency are shown in Supp. Fig. 5C.

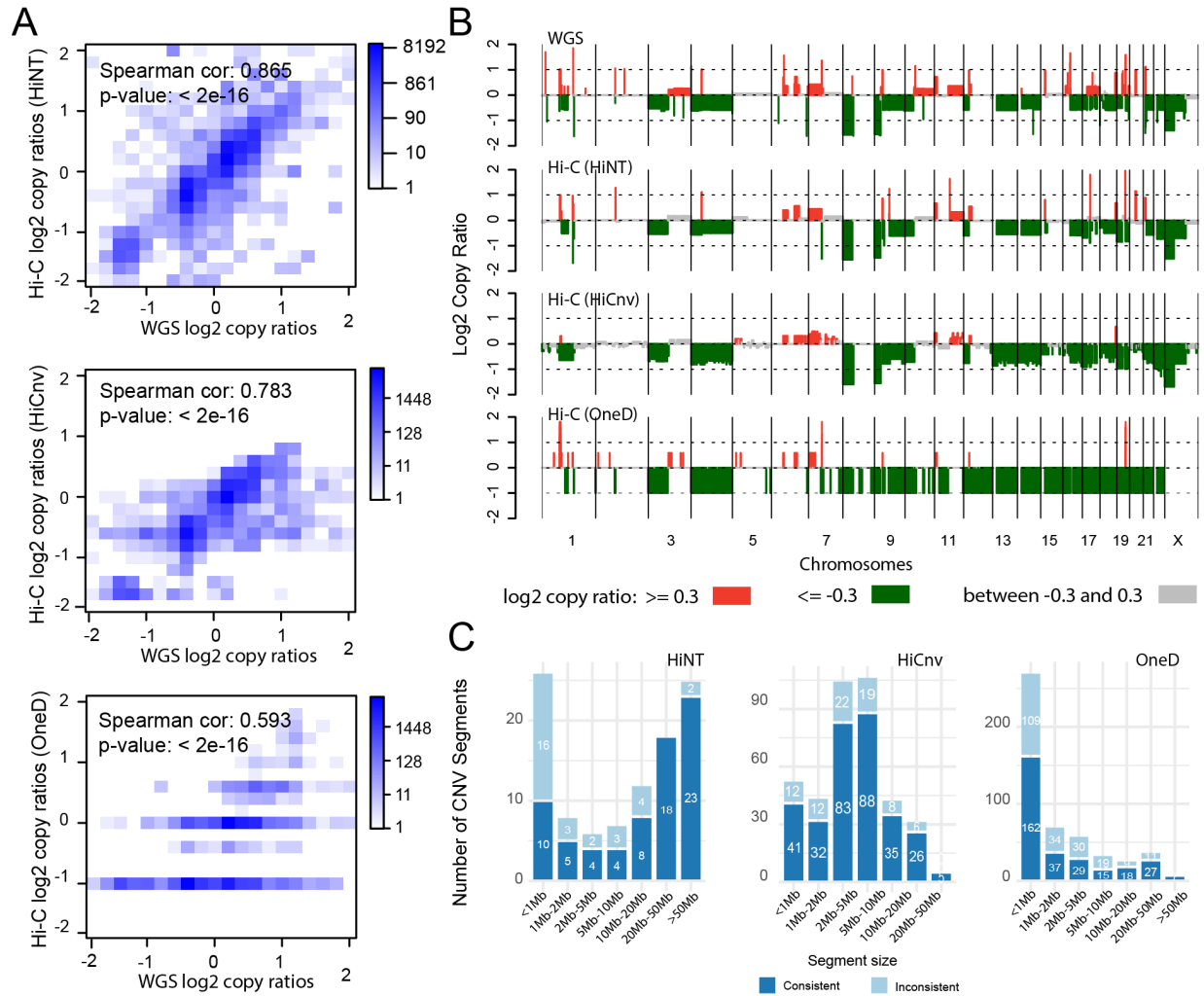


Fig. S9. Evaluation of copy number inference from Hi-C data in PANC-1 cells. **A**, Correlation of log2 copy ratios in each bin (50kb) detected from WGS and HiNT, HiCnv, and OneD in PANC-1. **B**, Comparison of CNV profiles from Hi-C and WGS after segmentation. Red, green and grey bars represent copy gain(log2 copy ratio > 0.3), copy loss(log2 copy ratio < -0.3), and copy neutral regions (log2 copy ratio between -0.3 and 0.3), respectively. **C**, The number of CNV segments (categorized by size) detected from Hi-C that are also supported by WGS. Specifically, 84%, 81%, and 58% of the large CNV segments identified by HiNT, HiCnv, and OneD are supported by those from WGS, respectively. The overlap criteria for consistency are shown in Supp. Fig. 5C.

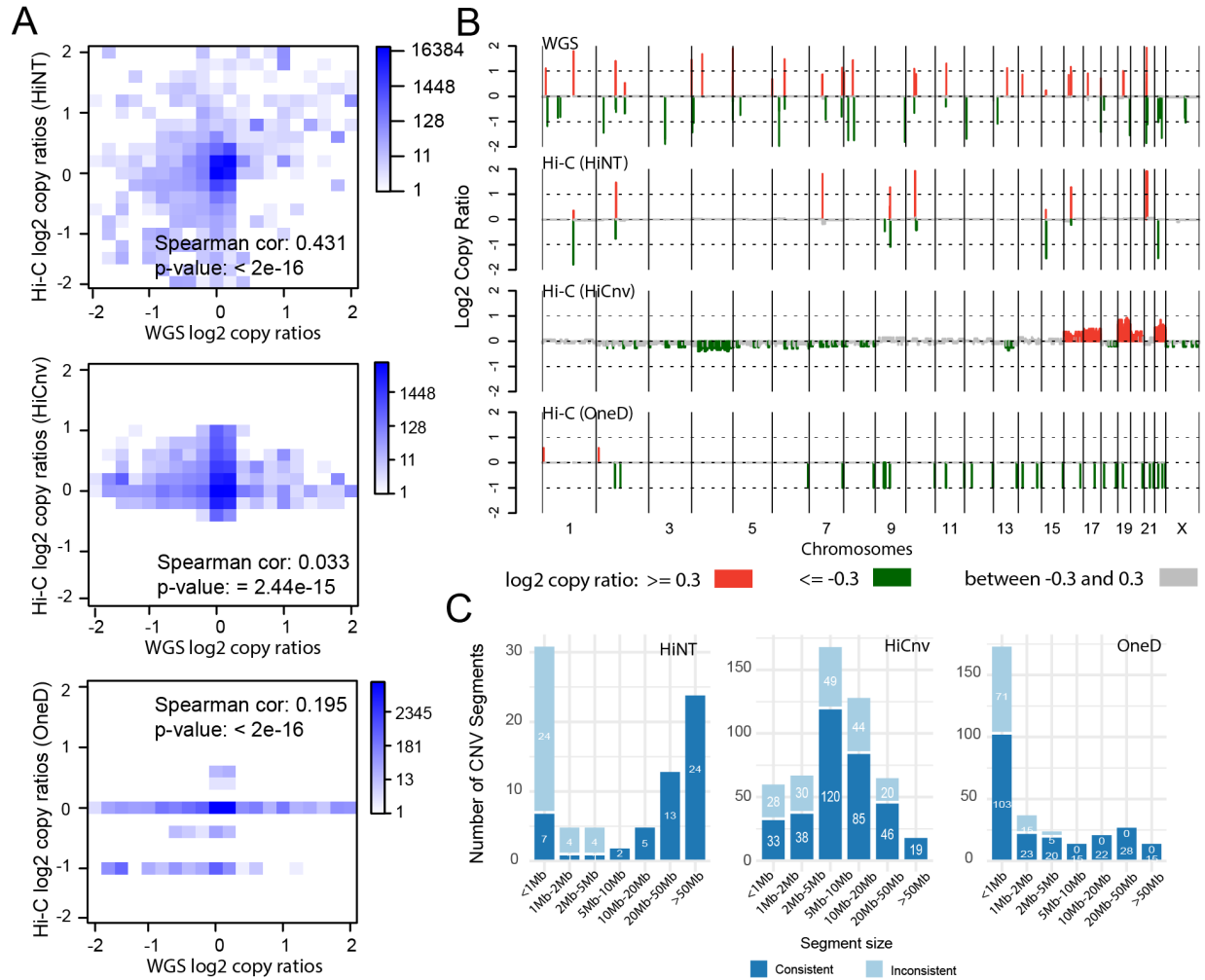


Fig. S10. Evaluation of copy number inference from Hi-C data in CHM13 cells. **A**, Correlation of log₂ copy ratios in each bin (50kb) detected from WGS and HiNT, HiCnv, and OneD in CHM13. **B**, Comparison of CNV profiles from Hi-C and WGS after segmentation. Red, green and grey bars represent copy gain(log₂ copy ratio > 0.3), copy loss(log₂ copy ratio < -0.3), and copy neutral regions (log₂ copy ratio between -0.3 and 0.3), respectively. **C**, The number of CNV segments (categorized by size) detected from Hi-C that are also supported by WGS. Specifically, 92%, 71%, and 95% of the large CNV segments identified by HiNT, HiCnv, and OneD are supported by those from WGS, respectively. The overlap criteria for consistency are shown in Supp. Fig. 5C.

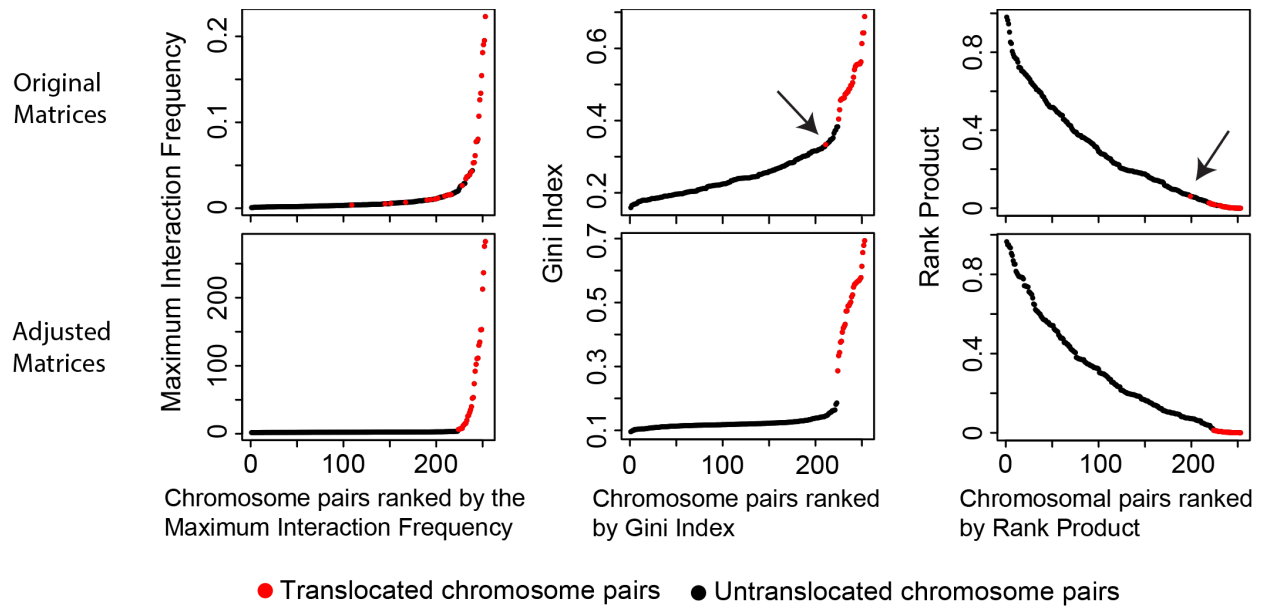


Fig. S11. Rank Product approach accurately identifies simulated translocated chromosome pairs. Distribution of the maximum interaction frequency (left), the Gini Index in an inter-chromosome contact matrix (middle), and the rank product of these two (right) in Hi-C data with simulated translocations.

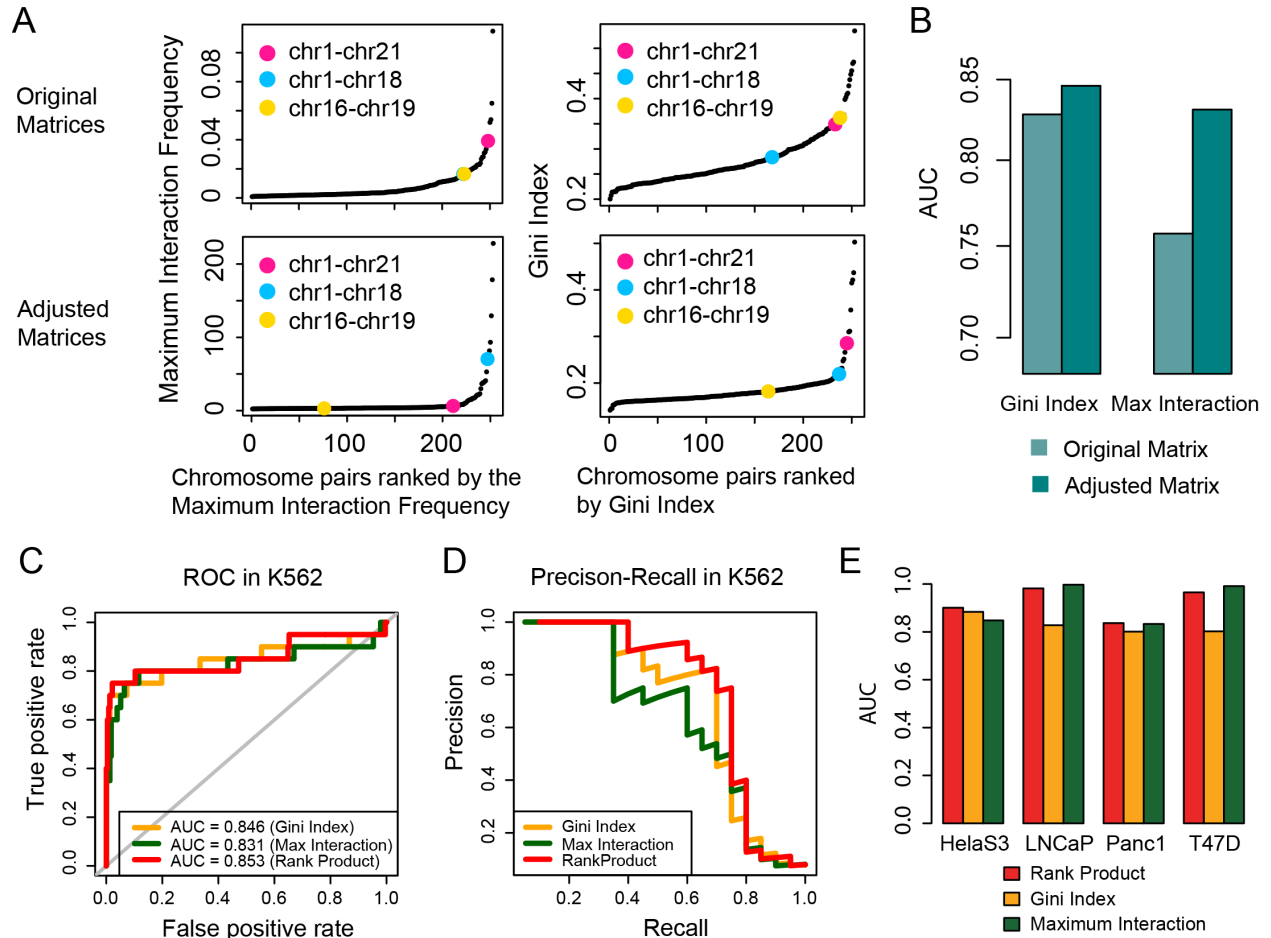


Fig. S12. Rank Product approach accurately identifies translocated chromosome pairs. **A**, The distribution of the maximum interaction frequency (MIF, left), the Gini Index (right), and the rank product of these two (figure 4A) in inter-chromosome contact matrices before (upper) and after (lower) adjustment in K562 cells. Chromosomal pairs in pink and blue correspond to two FISH-validated translocation pairs (chr1, chr21) and (chr1, chr18); the one in yellow corresponds to a chromosome pair (chr16, chr19) without translocation. **B**, AUROC values show either Gini Index or MIF perform better after the background subtraction in K562 cells. **C-D**, ROC curves (C) and precision-recall curves (D) of translocated chromosomal pairs predicted by using Gini Index only (orange), the maximum interaction only (dark green), and the rank product of these two (red) in K562 cells. **E**, Performance of rank product, Gini index, and the maximum interactions in HeLaS3, LNCaP, Panc1, and T47D cells.

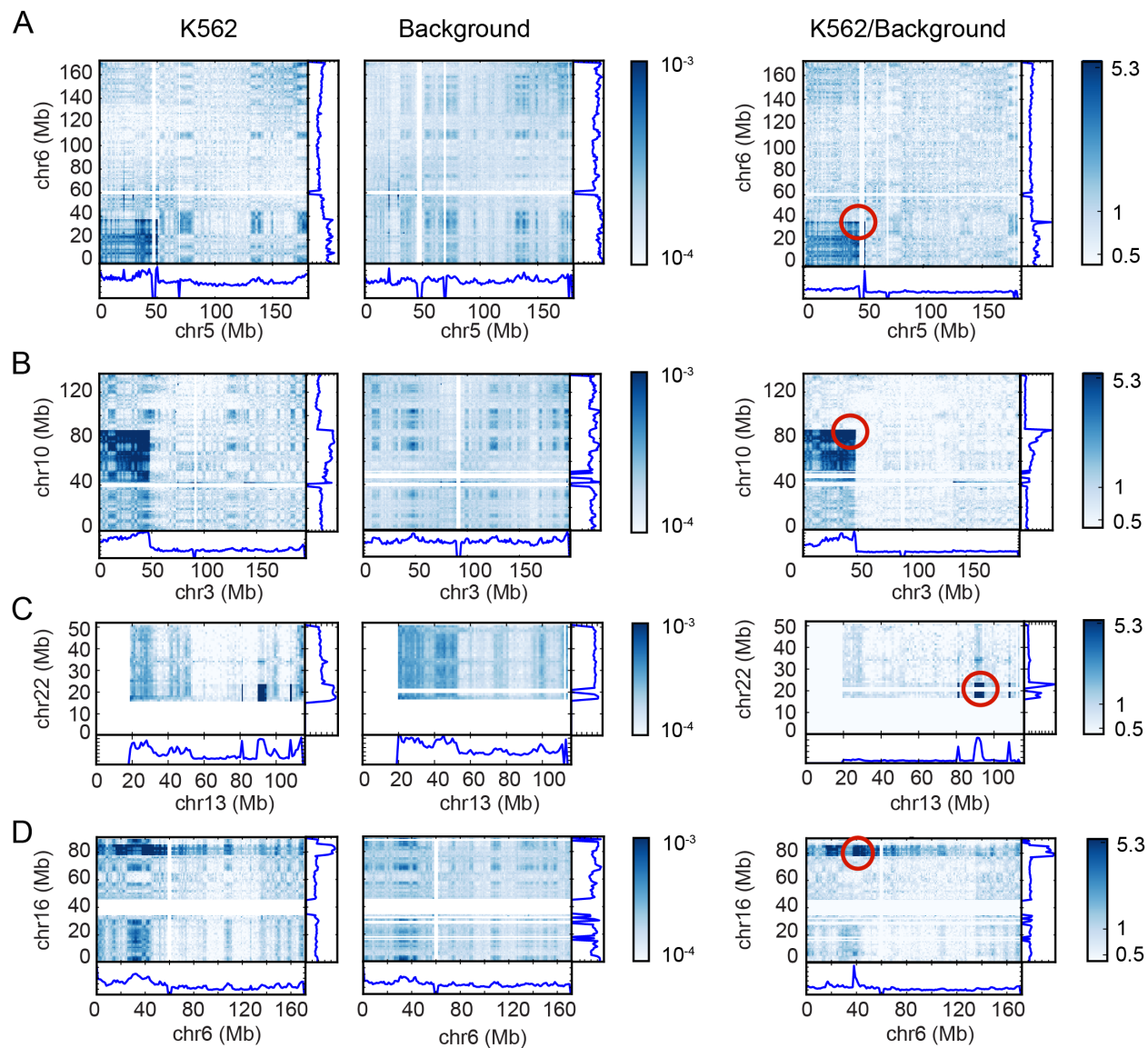


Fig. S13. Examples of chromosomal pairs with most significant rank product. A-D, Hi-C inter-chromosomal heatmaps and 1D coverage in original K562(left), background (middle), and adjusted K562 (K562/Background, right) data. Hi-C 1-D profiles (the sum of rows and columns of each inter-chromosomal interaction matrix) are shown along with the interaction maps. Translocation breakpoints are marked by red circles.

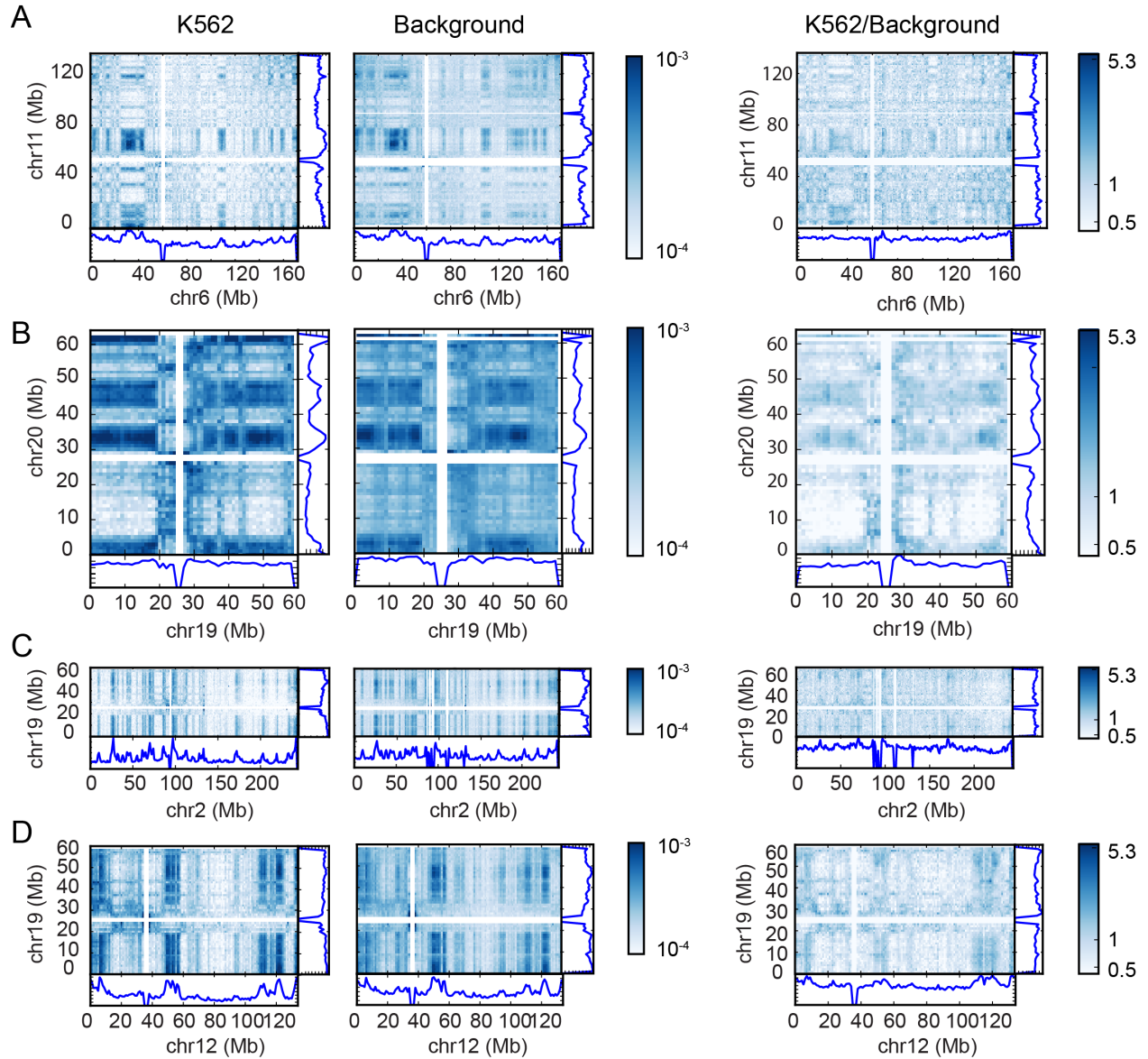


Fig. S14. Examples of missed translocated chromosomal pairs by HiNT. A-D, Hi-C inter-chromosomal heatmaps and 1D coverage in original K562(left), background (middle), and adjusted K562 (K562/Background, right) data. Hi-C 1-D profiles (the sum of rows and columns of each inter-chromosomal interaction matrix) are shown along with the interaction maps.

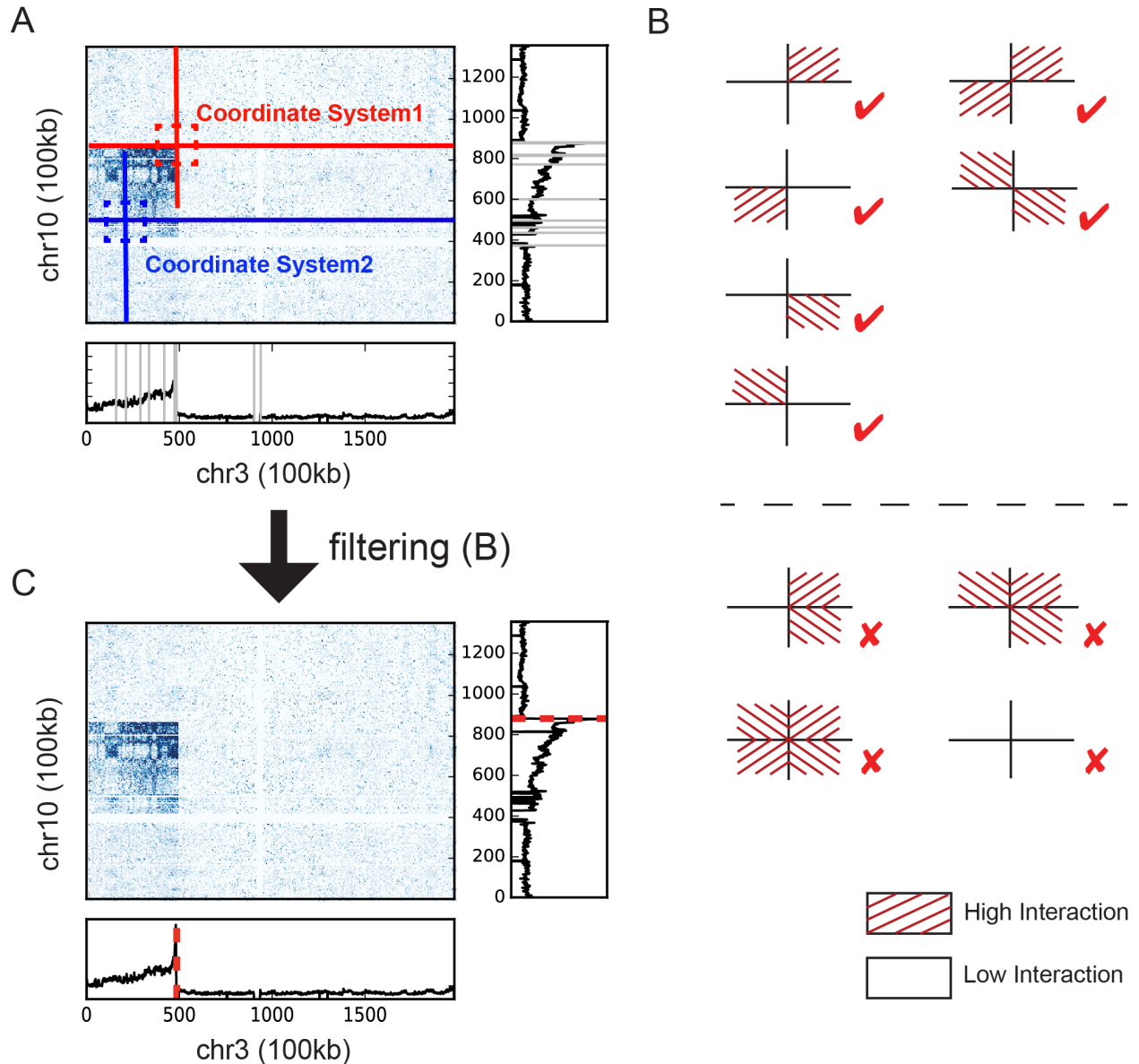


Fig. S15. Breakpoint detection and filtering. **A**, Candidate breakpoints (grey lines) detected by *strucchange* based on the 1D coverage profile (sum of rows and columns). Two-dimensional Cartesian coordinate systems originating from the intersection of each pair of candidate breakpoints are constructed; two examples are shown in the figure. **B**, Patterns of Hi-C interaction frequencies in four 5-bin-by-5-bin quadrants, that generated by the pair of breakpoints from both chromosomes. Valid translocation breakpoints are shown above the dash line, and invalid breakpoints are shown below. **C**, Translocation breakpoints (red dotted lines) after the filtering step.

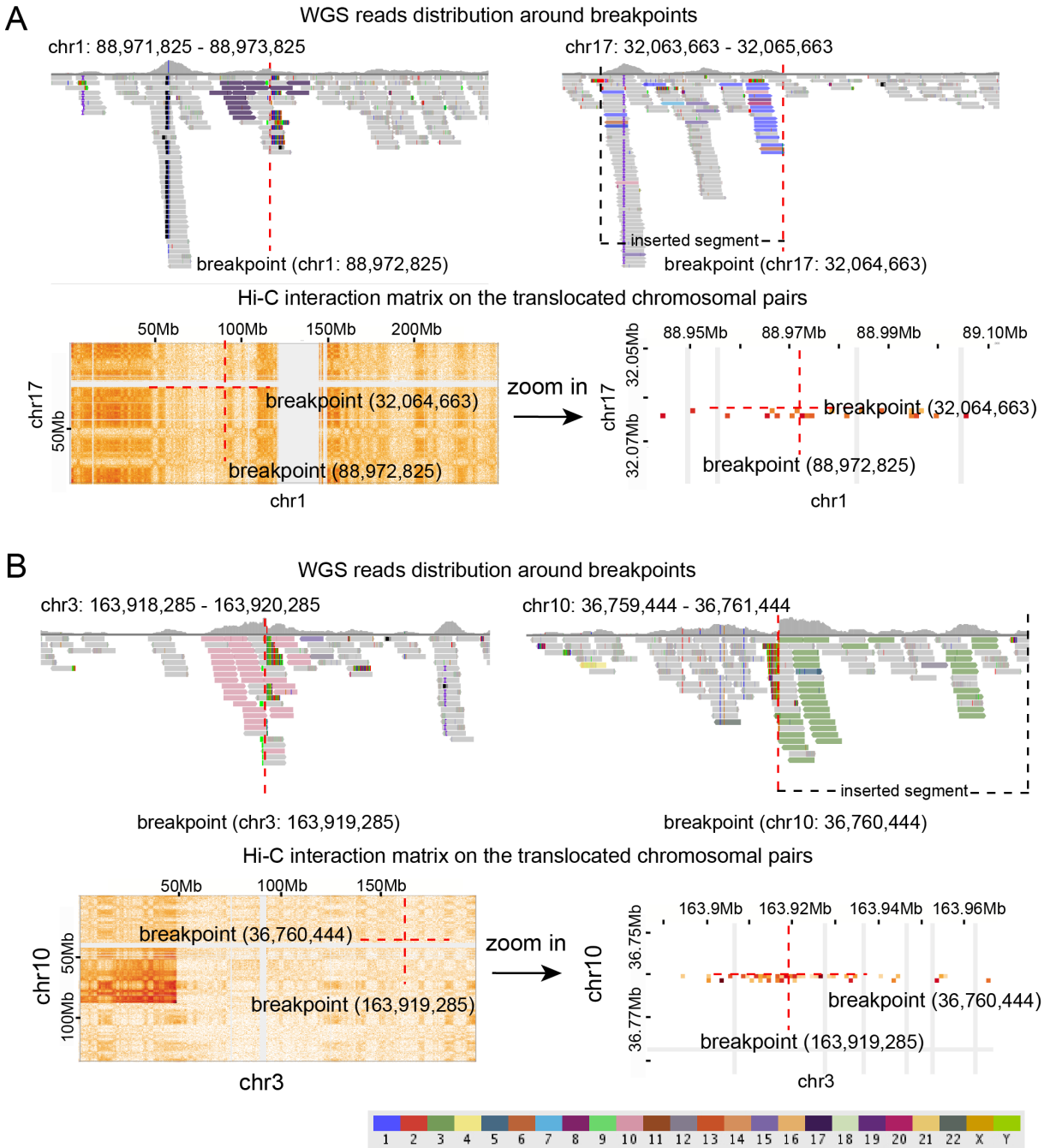


Fig. S16. Examples of the small segment inter-chromosomal insertional translocation that is detected from only WGS. A, The distribution of discordant reads and clipped reads around the translocation breakpoints detected from WGS on chr1 and chr17 (upper); Hi-C interaction heatmap across the whole chromosomes (bottom left) and regions around breakpoints (bottom right). **B,** Similar to A, but the translocation between chr3 and chr10. In the IGV screenshot (WGS reads distribution), each color bar represents a SNV (single nucleotide variant), and the colored reads are paired end reads coded by the chromosome on which their mates can be found. The color code for discordant reads is shown at the bottom.

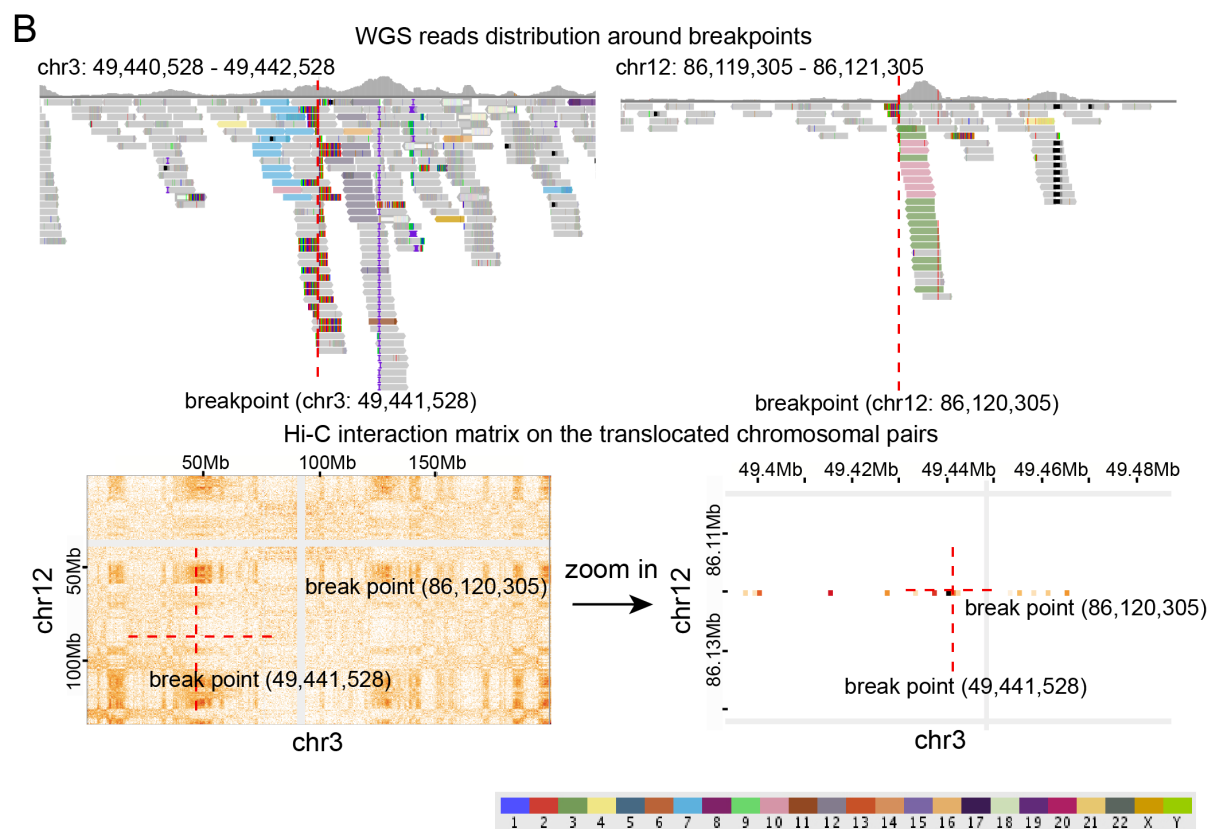
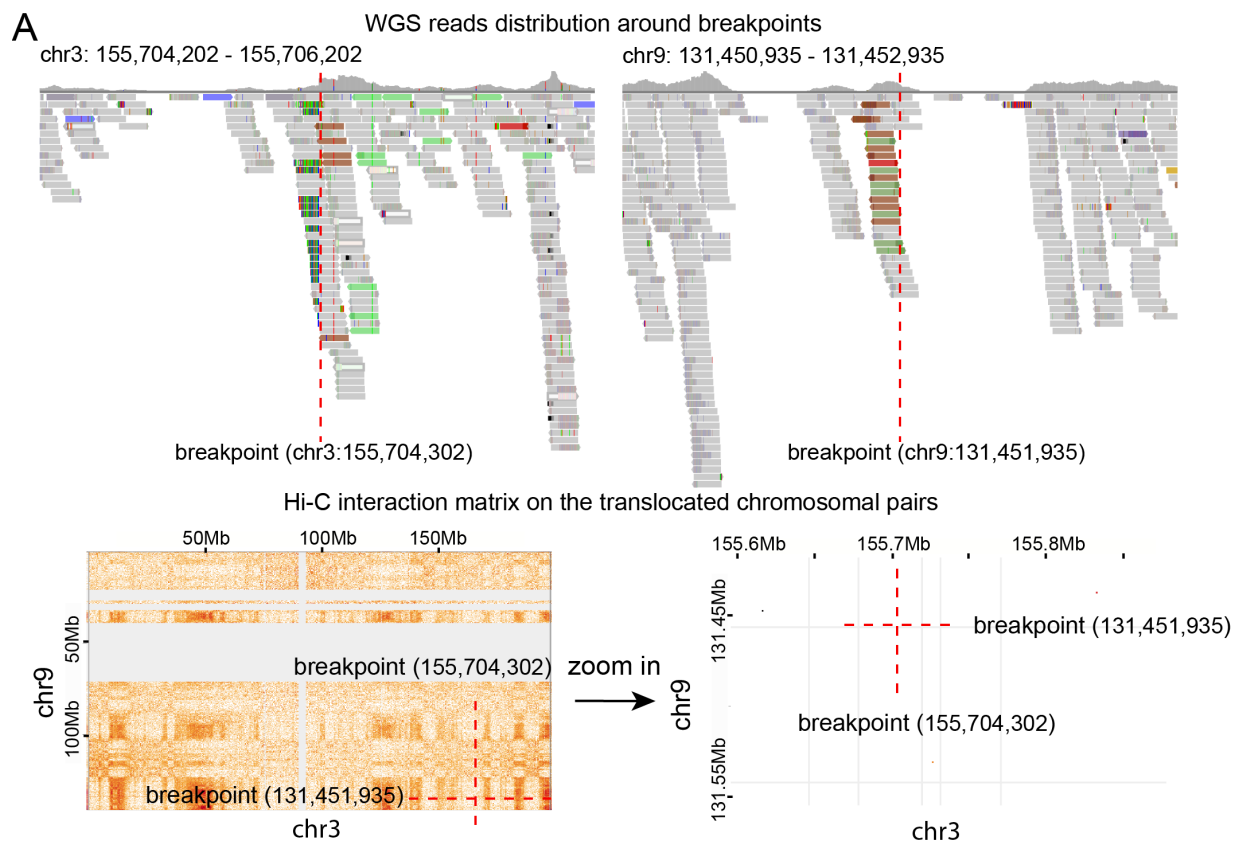
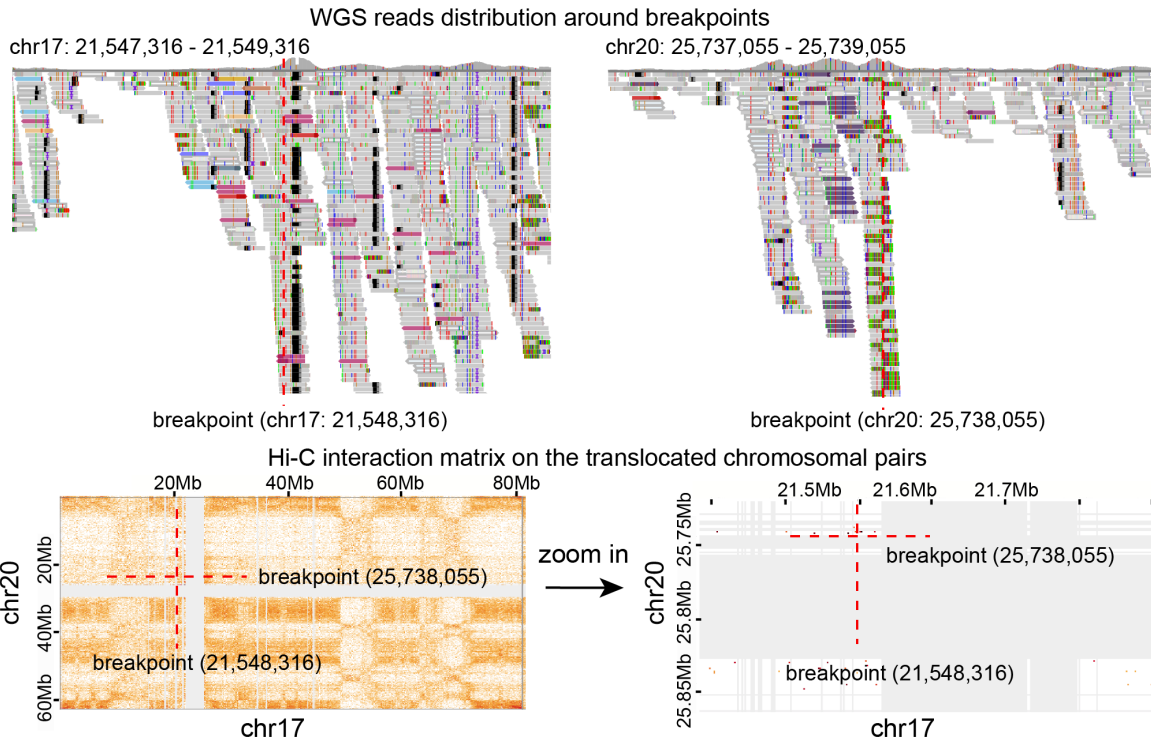


Fig. S17. Examples of the complex SVs. **A,** The distribution of discordant reads and clipped reads around the translocation breakpoints detected from WGS on chr3 and chr9 (upper); Hi-C interaction heatmap across the whole chromosomes (bottom left) and regions around breakpoints (bottom right). **B,** Similar to A, but the translocation between chr3 and chr12. In the IGV screenshot (WGS reads distribution), each color bar represents a SNV (single nucleotide variant), and the colored reads are paired end reads coded by the chromosome on which their mates can be found. The color code for discordant reads is shown at the bottom.

A



B

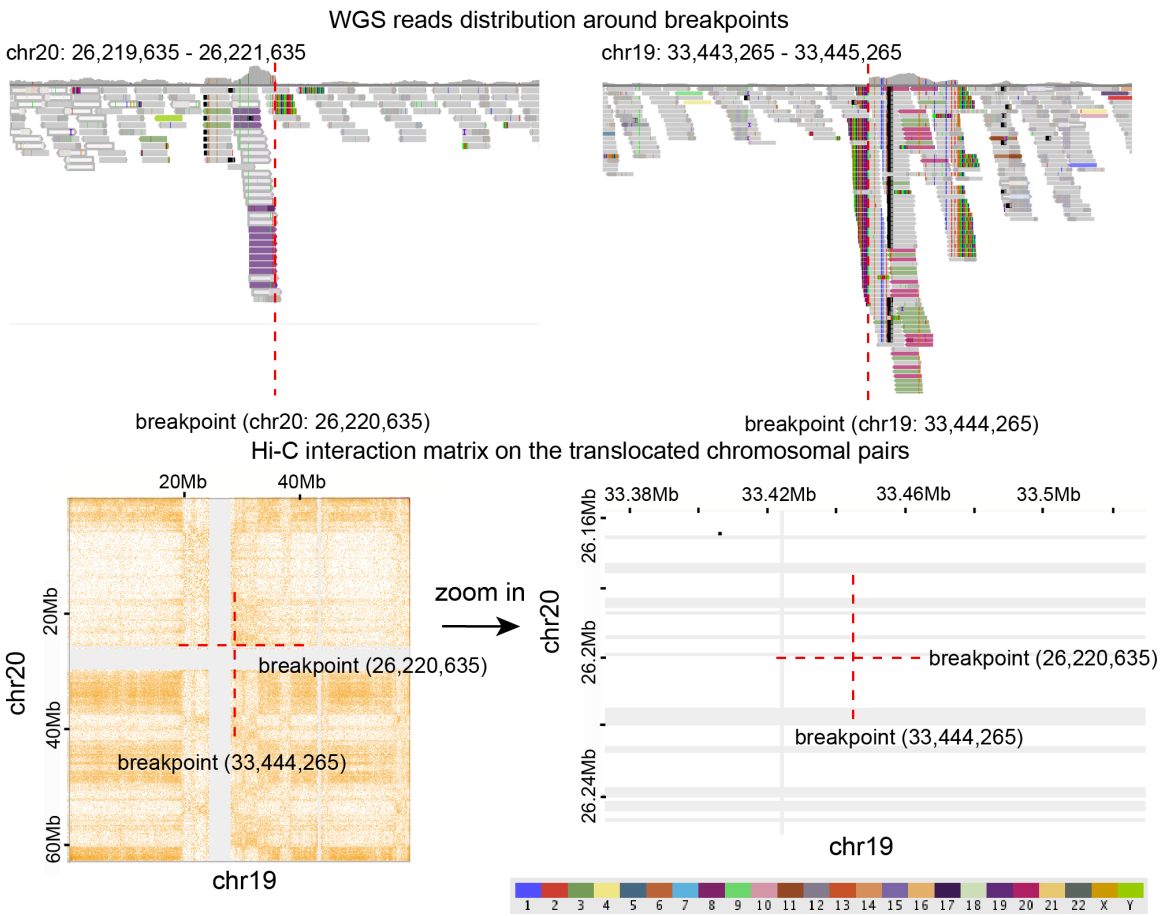


Fig. S18. Examples of the false positives that identified from WGS data. **A**, The distribution of discordant reads and clipped reads around the translocation breakpoints detected from WGS on chr17 and chr20 (upper); Hi-C interaction heatmap across the whole chromosomes (bottom left) and regions around breakpoints (bottom right). **B**, Similar to A, but the translocation between chr19 and chr20. In the IGV screenshot (WGS reads distribution), each color bar represents a SNV (single nucleotide variant), and the colored reads are paired end reads coded by the chromosome on which their mates can be found. The color code for discordant reads is shown at the bottom.

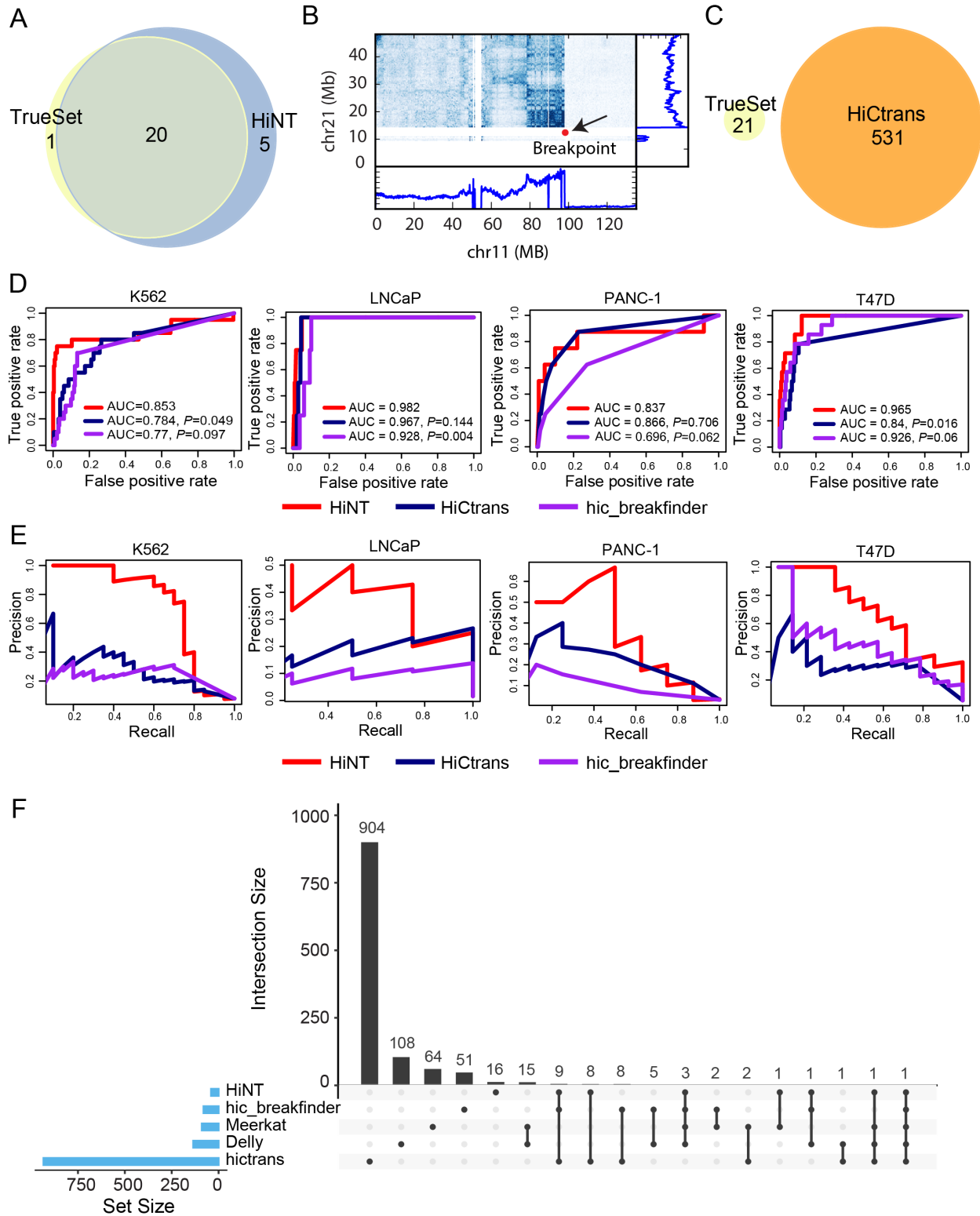


Fig. S19. HiNT outperforms existing methods on translocation breakpoints detection in both simulated and real Hi-C data. **A**, The overlap of translocation breakpoints detected by HiNT and simulated true set. **B**, Hi-C interaction heatmap for the breakpoint that was missed by HiNT, the sum of rows and columns are shown along the matrix. **C**, The overlap of translocation breakpoints detected by HiCtrans and simulated true set. **D-E**, Evaluation of the performance of HiNT (red curve), HiCtrans (navy curve), and hic_breakfinder (purple curve) on translocated chromosome pairs prediction in K562, LNCaP, PANC-1, and T47D cells by ROC curves (D) and precision-recall curves (E). *P-values* (see Methods) for the AUC comparison between HiNT and HiCtrans/OneD are labeled in the figures. **F**, Intersections of the translocation breakpoints detected by Meerkat and Delly from WGS, and HiNT, HiCtrans and hic_breakfinder from Hi-C.