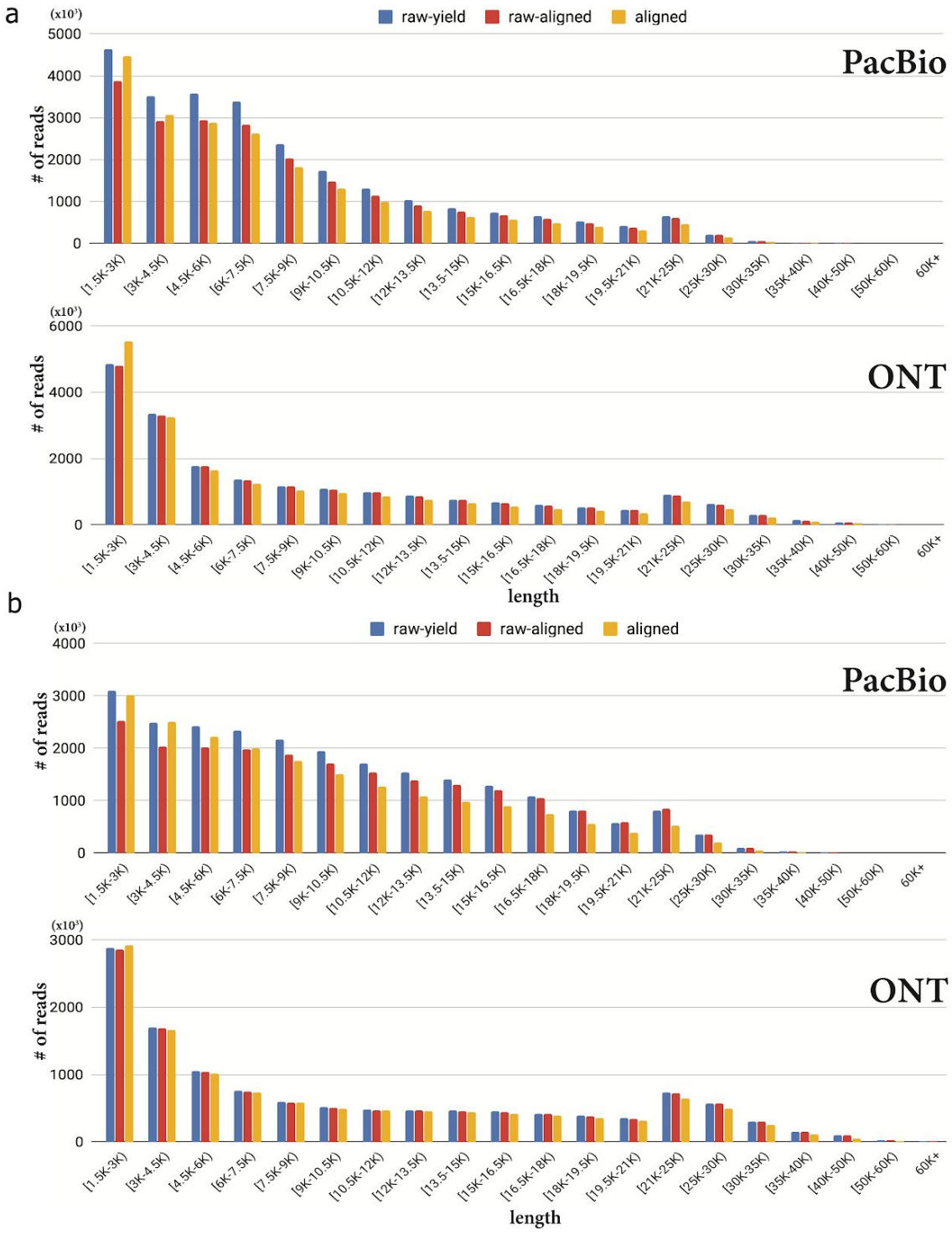
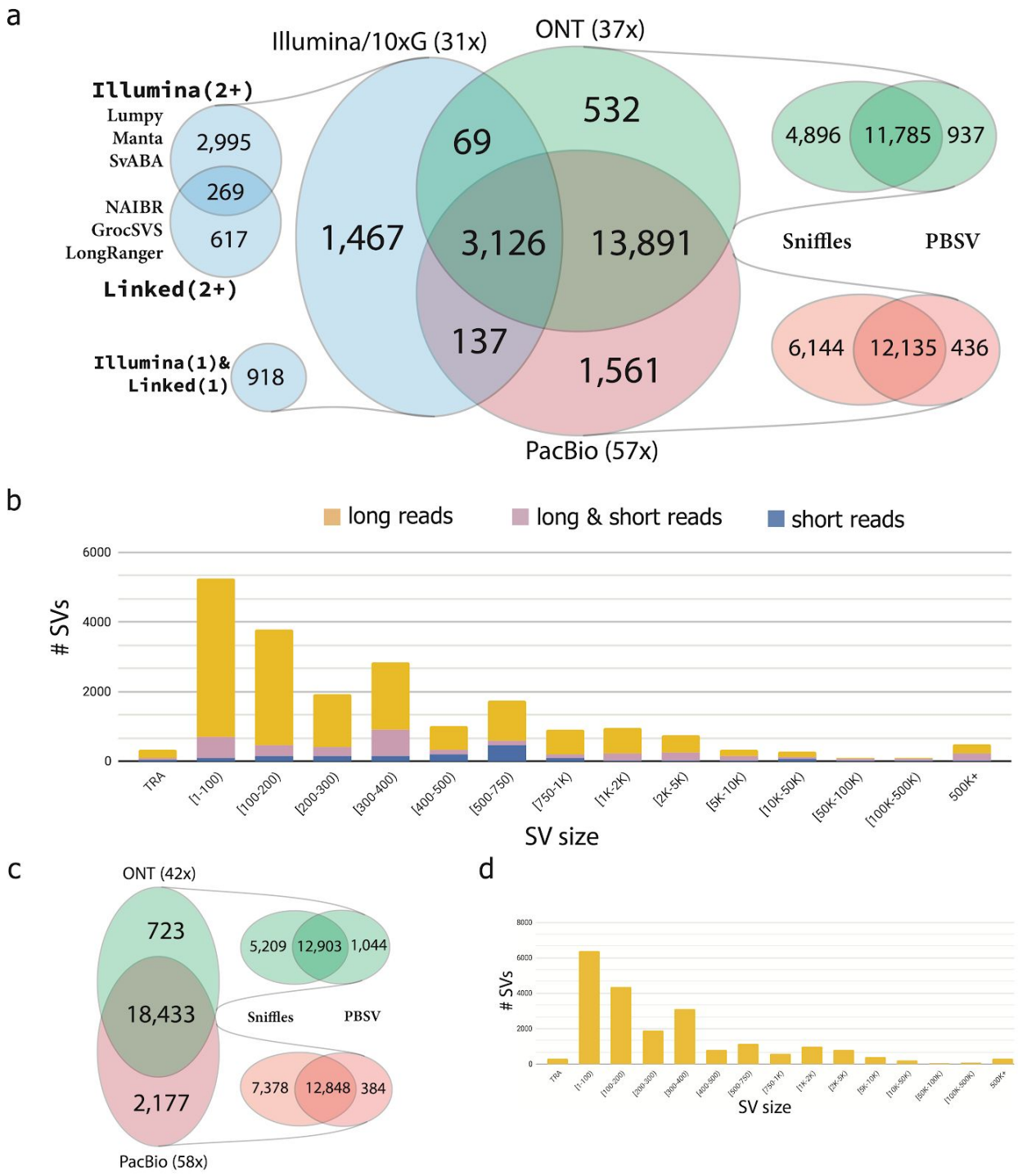
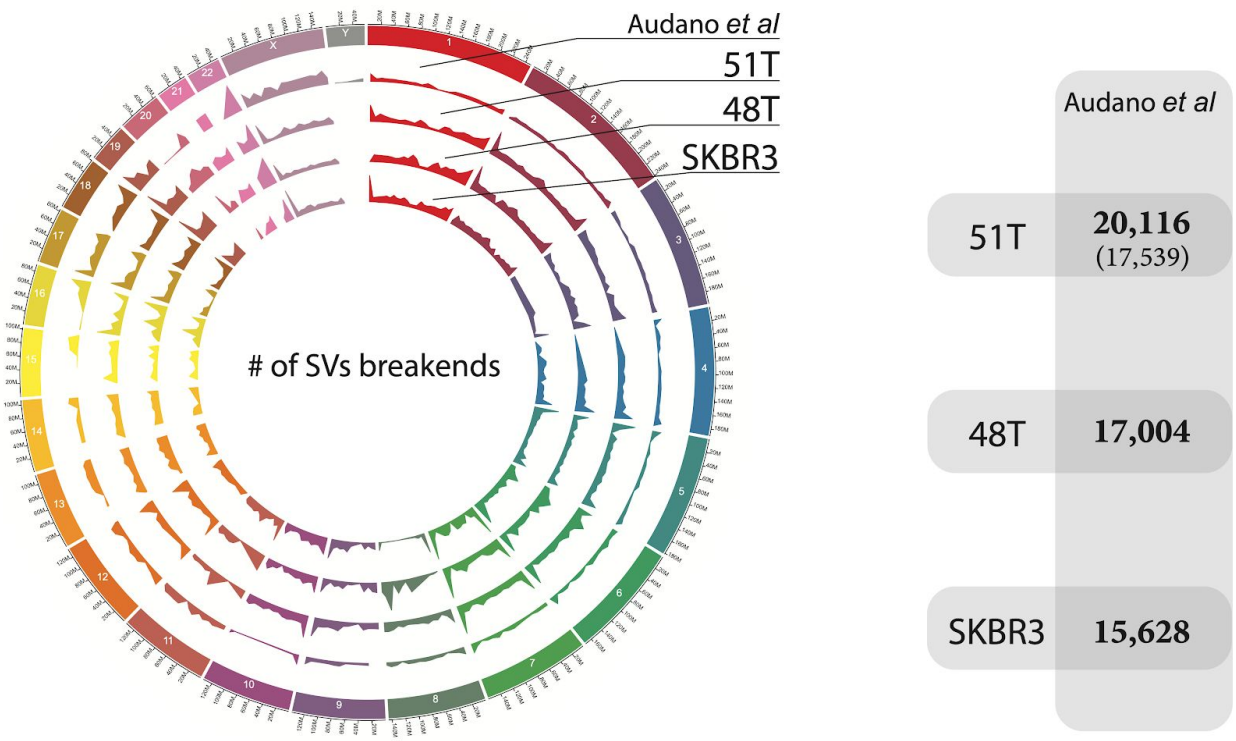
Supplemental Figures



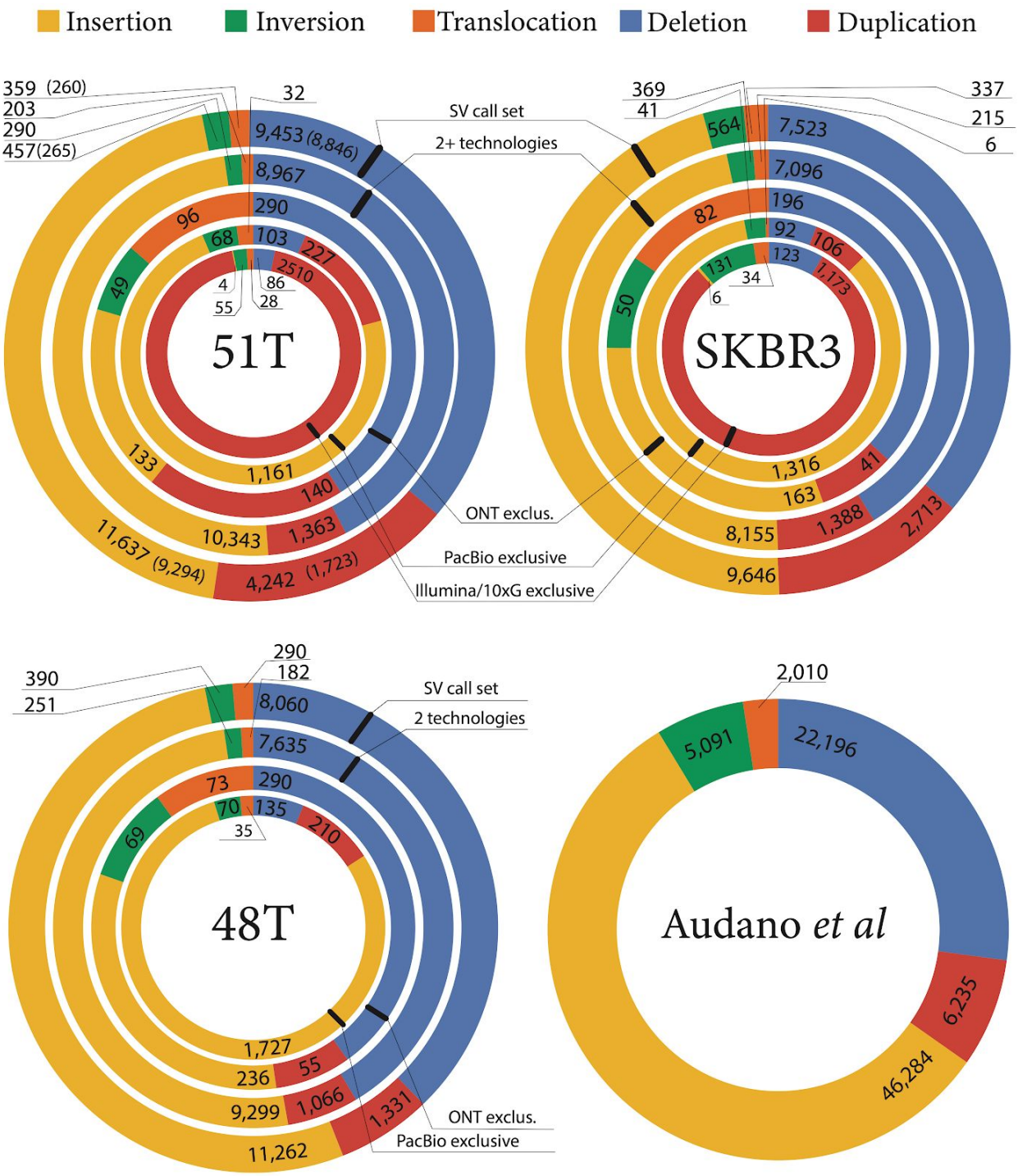
Supplemental Figure 1. | ONT and PacBio reads lengths distribution. *raw-yield* corresponds to lengths of raw sequenced reads, *raw-aligned* corresponds to lengths of raw read that had any alignment inferred for them and *aligned* corresponds to lengths of aligned parts of sequenced reads. **a)** Lengths distributions for reads of length *1.5+kbp* from PacBio and ONT sequencing runs for samples 48T. **b)** Lengths distributions for reads of length *1.5+kbp* from PacBio and ONT sequencing runs for samples SKBR3.



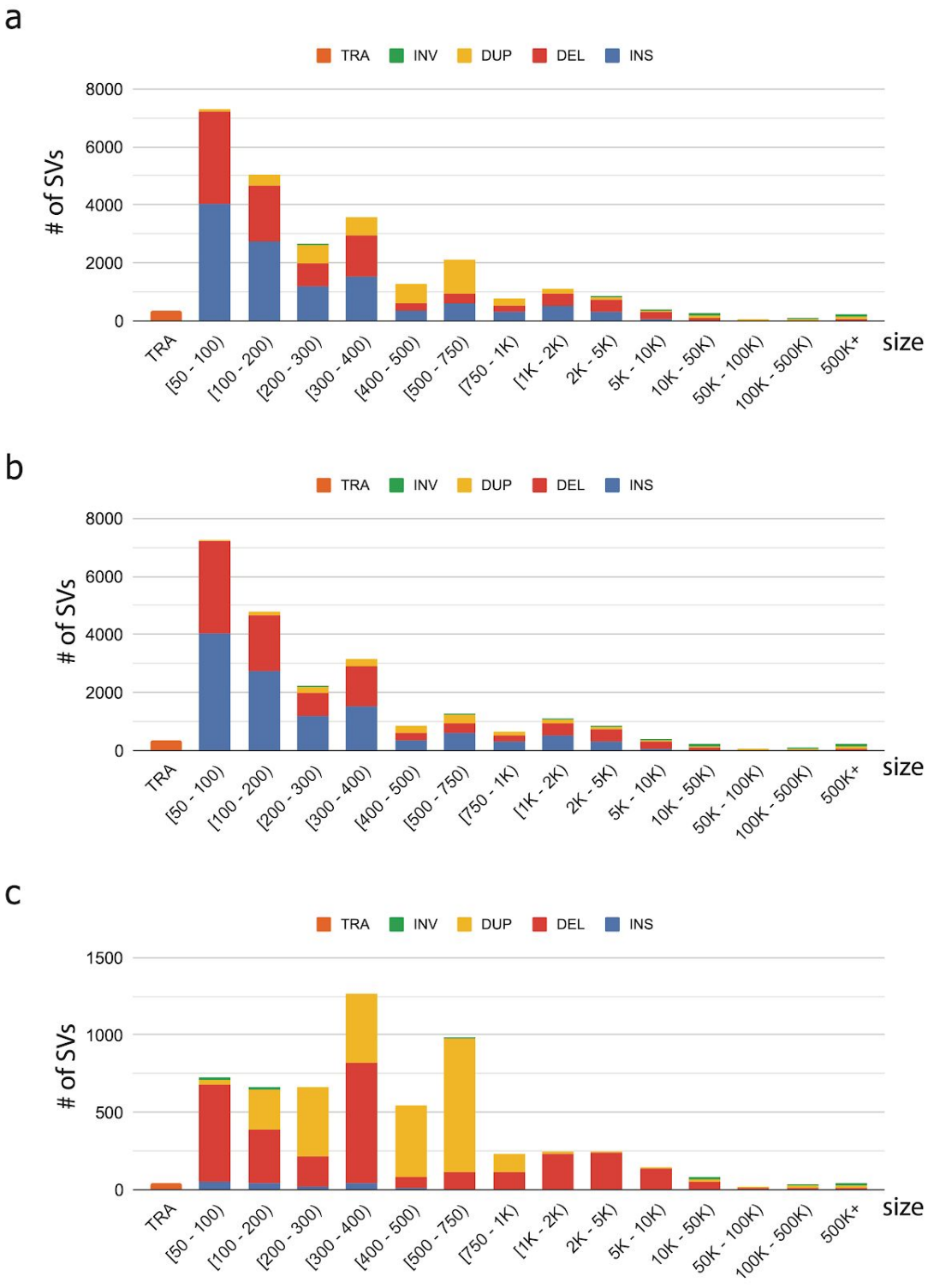
Supplemental Figure 2. | Structural variations inference across Illumina/10xG, ONT, and PacBio sequencing platforms for samples SKBR3 and 48T. a) and **c)** SV inference comparison across SVs inferred from *Platform (x)* sequencing experiments, where *Platform* corresponds to sequencing technology, and *(x)* determines the average alignment read-depth coverage. Methods-specific breakdown is provided for every sequencing technology. **b)** and **d)** Size distribution for SVs in samples SKBR3 and 48Twith SVs being either exclusively inferred from either long-reads (either ONT, or PacBio, or both), or exclusively from Illumina/10xG short-reads, or supported by both long and short-reads.



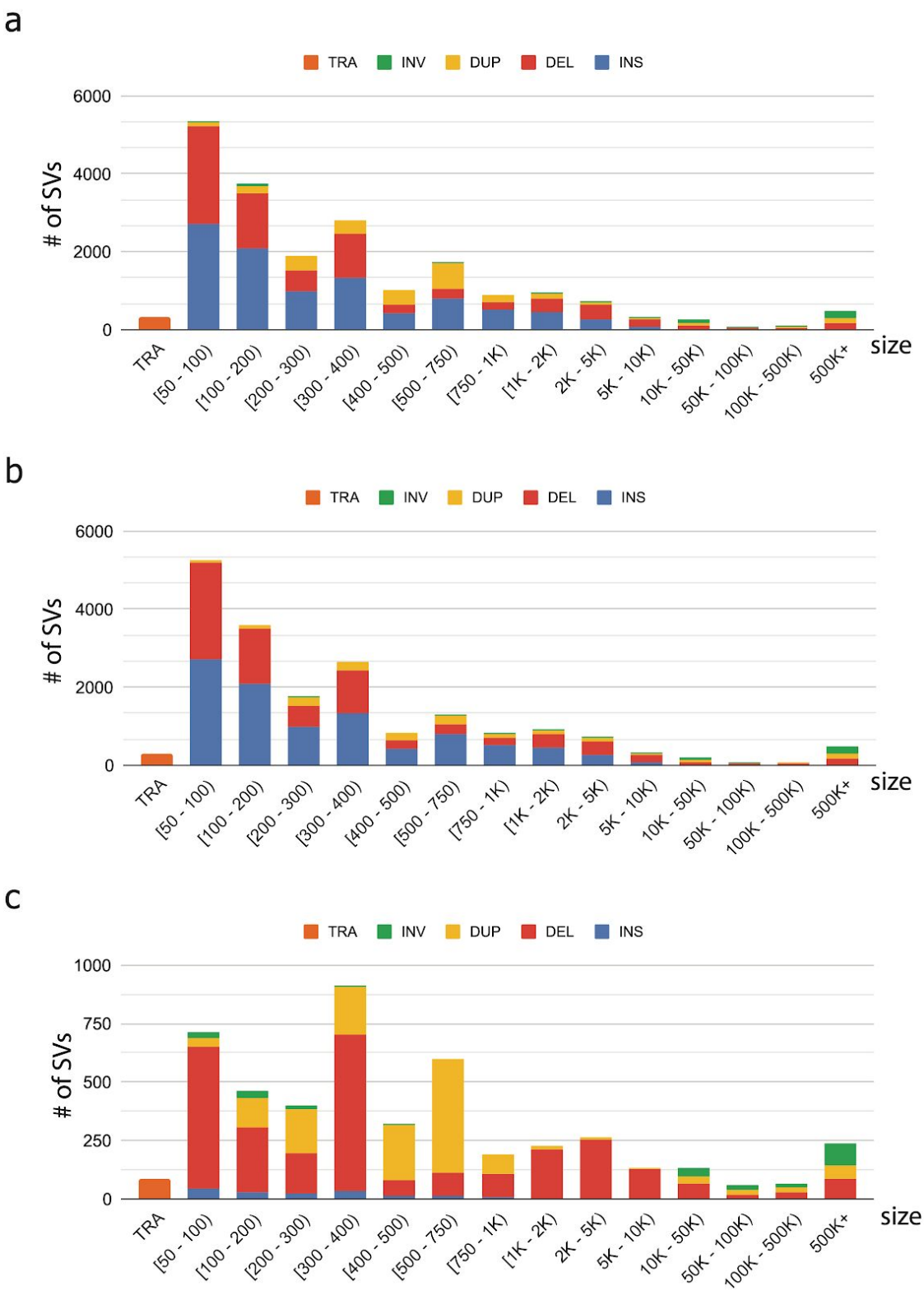
Supplemental Figure 3. | Structural Variations in samples 51T(N), 48T, SKBR3, and in Audano *et al* ***(Audano et al. 2019)*** **dataset.** Circos plot on the left shows the SVs breakends distributions across genome chromosomes. Every track is dataset-specific shows the total number of SVs' breakends over 5MB segment-length windows. Panel on the right shows intersection of SVs across observed cancer datasets (with matching normal SVs shown in parentheses) and the healthy SV set generated from 15 samples from Audano *et al*.



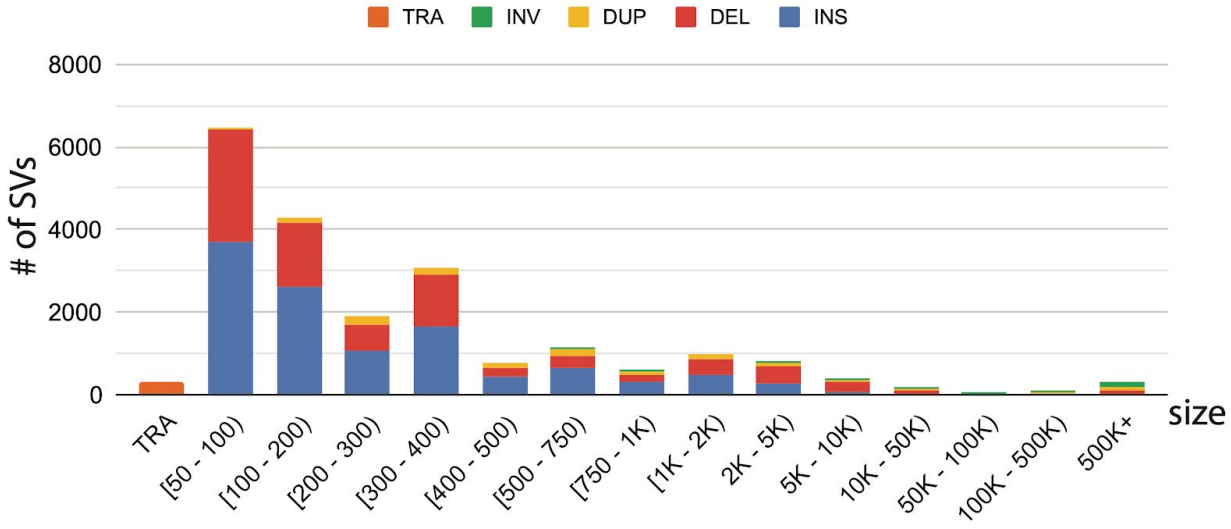
Supplemental Figure 4. |SV type breakdown over supporting sequencing technologies. Quantification of SVs types determined by involved intra-chromosomal breakends' orientations (*deletion*, *duplication*, *inversion*), inter-chromosomal breakends (*translocation*), and intra-chromosomal SVs with novel sequence (*insertions*) across different combinations of sequencing technologies in samples 51T, SKBR3, 48T, and in the healthy SV set (i.e., the union of SVs across 15 healthy PacBio sequenced samples from the Audano *et al* study dataset).



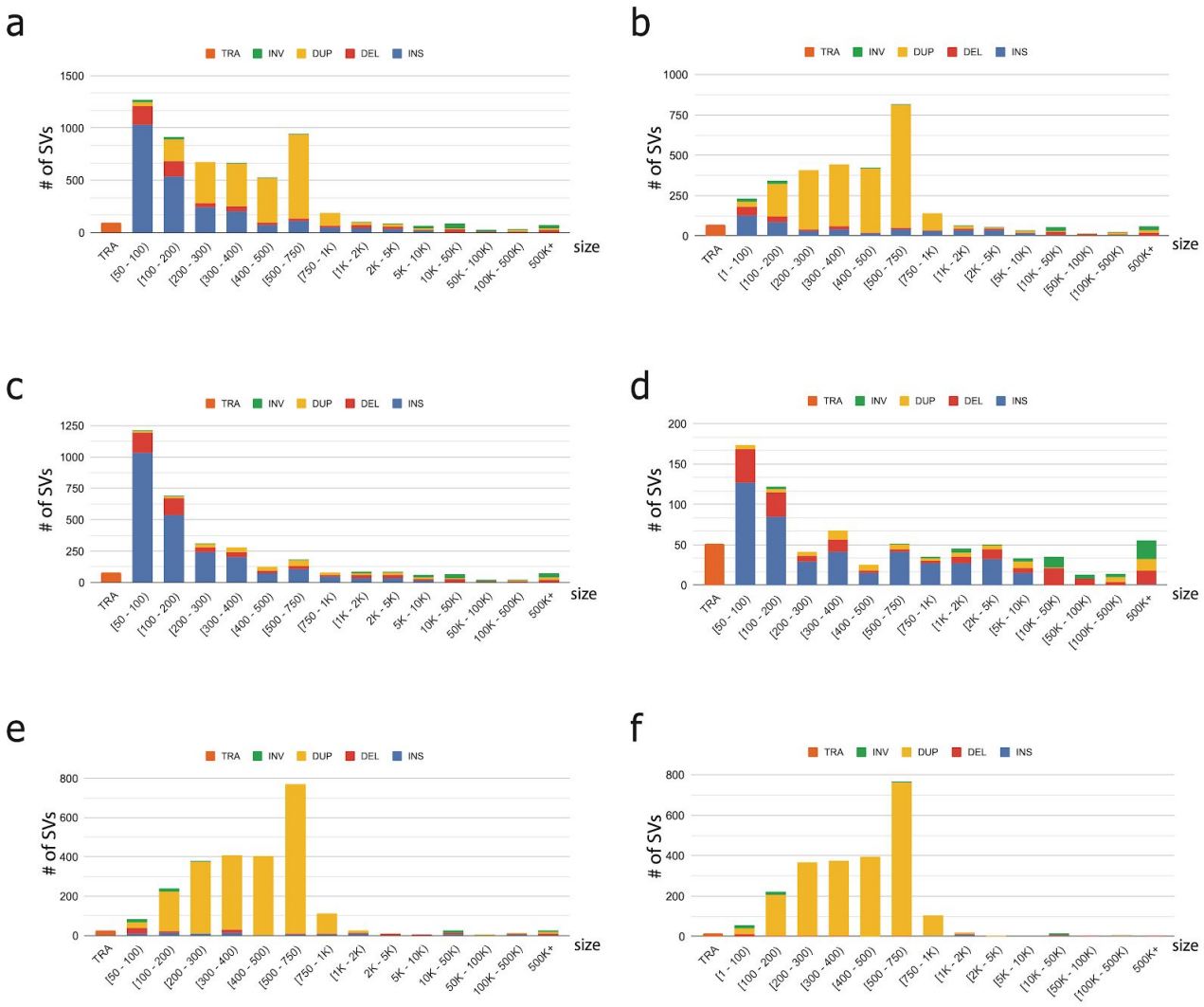
Supplemental Figure 5. |**Size distribution for SVs in 51T**. a) SVs supported by both Illumina/10xG short-reads and ONT/PacBio long-reads b) SVs supported by long-reads (either ONT, or PacBio, or both) c) SVs supported by Illumina/10xG short-reads



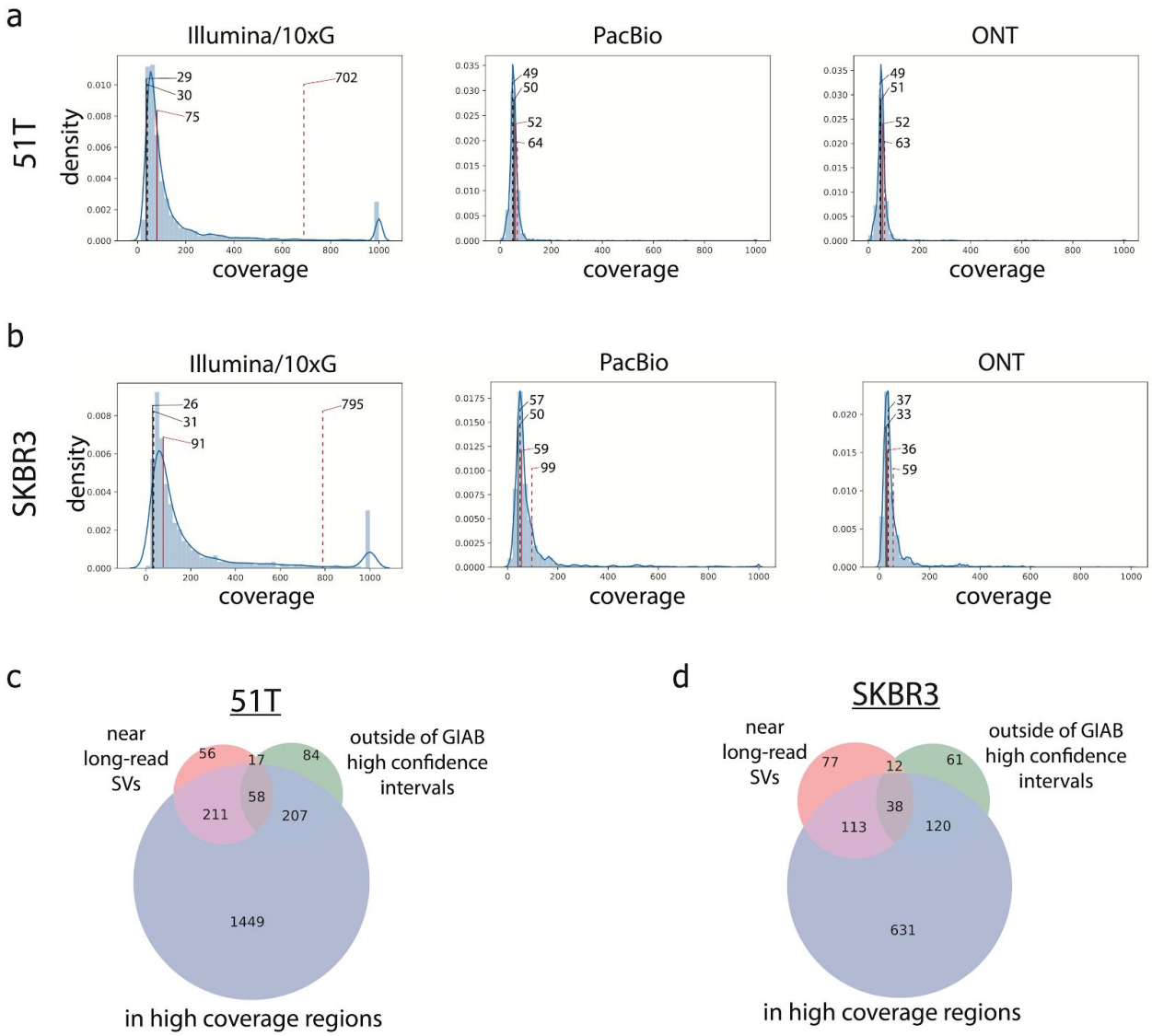
Supplemental Figure 6. |Size distribution for SVs in SKBR3. **a**) SVs supported by either Illumina/10xG short-reads and ONT/PacBio long-reads **b)** SVs supported by long-reads (either ONT, or PacBio, or both) **c)** SVs supported by Illumina/10xG short-reads



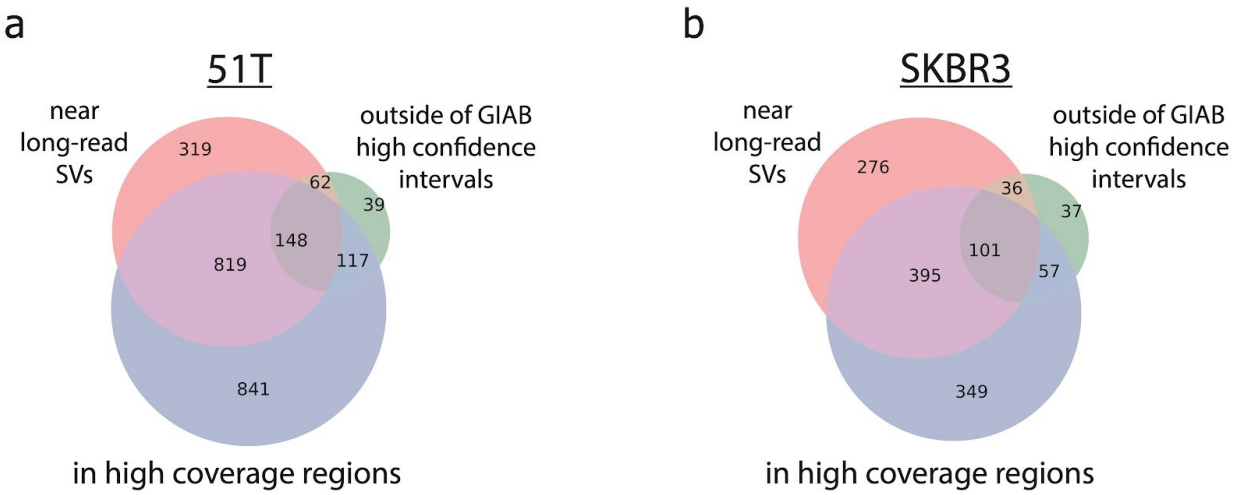
Supplemental Figure 7. |Size distribution for SVs in 48T. SVs are supported by long-reads (either ONT, or PacBio, or both).



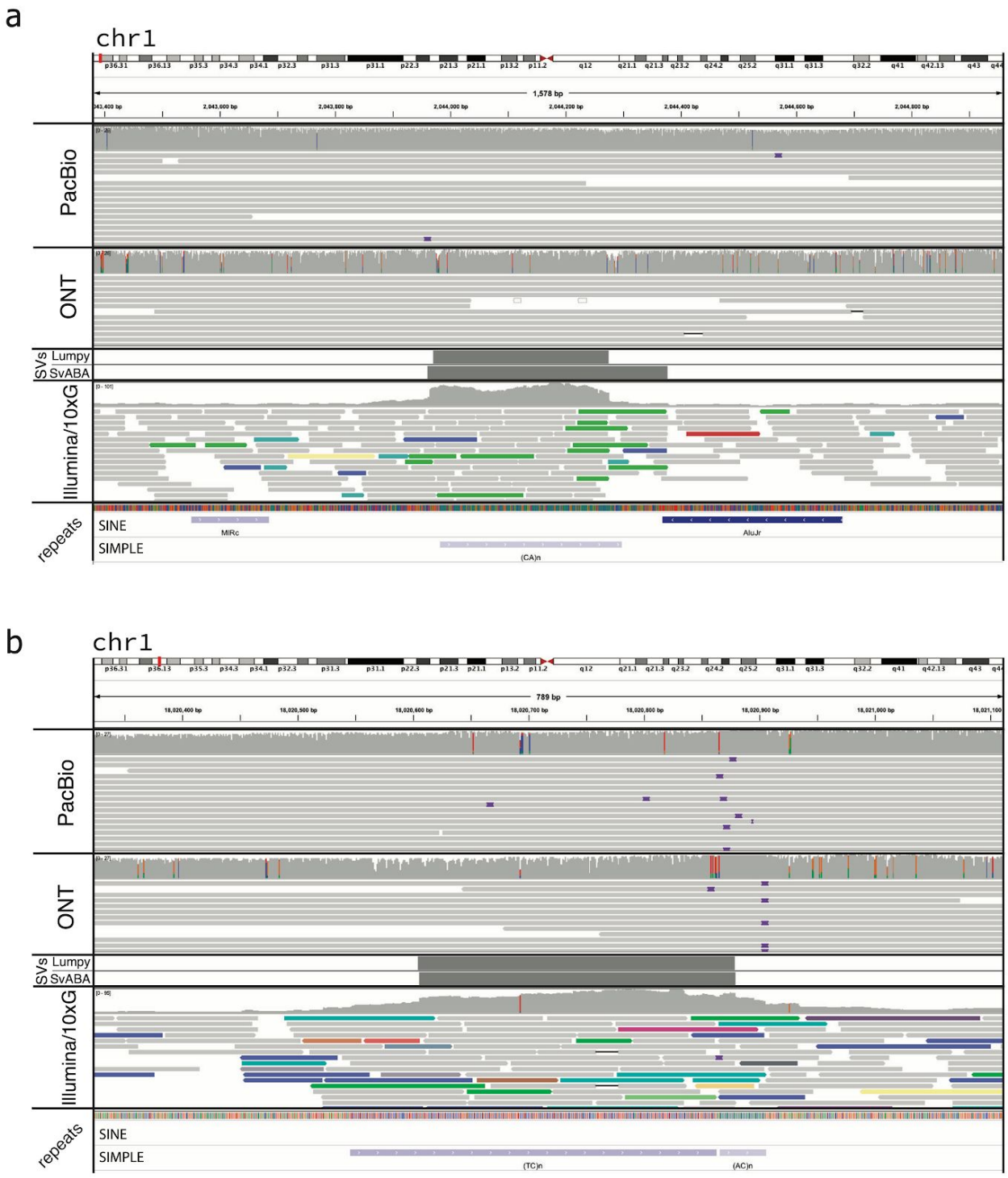
Supplemental Figure 8. |**Size distribution for somatic SVs in 51T**. Left panels (**a**, **c**, and **e)** show SVs that are present in 51T but are absent in 51N; Right panels (**b**, **d**, and **f)** show SVs that are present in 51T, absent in 51N, and are absent across 15 healthy genomes from (Audano et al. 2019). Panels **a**) and **b)** show SVs supported by either Illumina/10xG short-reads, ONT/PacBio long-reads, or both; Panels **c)** and **d)** show SVs supported by long-reads (either ONT, or PacBio, or both); Panels **e) and f)** show SVs supported by Illumina/10xG short-reads;



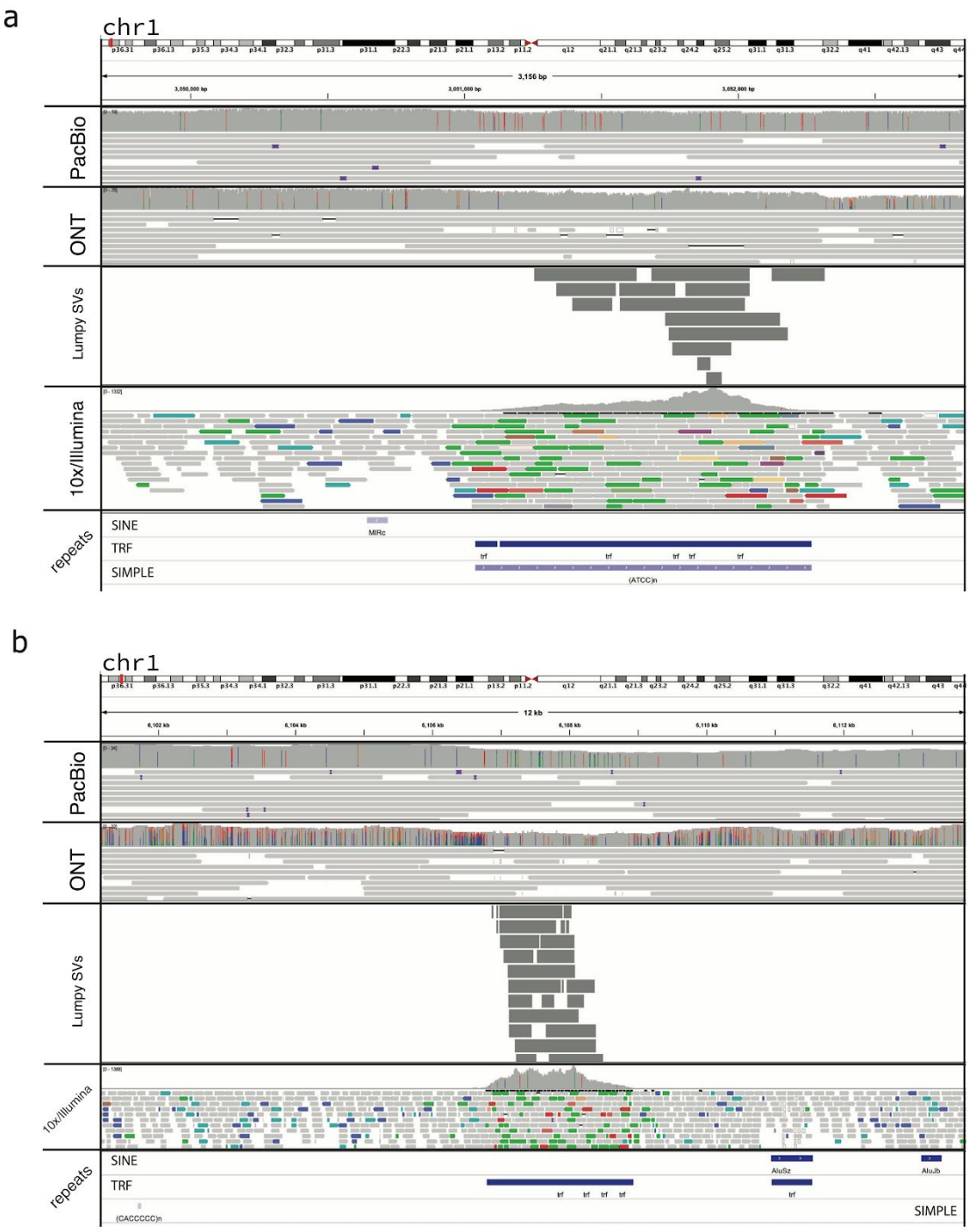
Supplemental Figure 9. |Short-reads exclusive SVs in 51T and SKBR3. Coverage distributions in 500bp windows around breakpoints regions of short-reads exclusive SVs in Illumina/10xG, PacBio, and ONT reads in patients 51T **(a)** and SKBR3 **(b)**. Solid lines indicate median level of coverage and dashed lines show mean level of coverage. Black lines correspond to genome-wide statistics, while red lines depict coverage statistics in 500bp windows around short-read exclusive breakpoints. Coverage values greater than 1000 are set to 1000. Panels (**c)** and **(d)** show breakdowns of genetics and variation context (nearby presence of *specific* long-read SVs, breakpoints falling outside of GIAB high-confidence intervals for SV detection, and location within a high-coverage Illumina/10xG 500bp window) for short-reads exclusive SVs in 51T **(c)** and SKBR3 **(d)** with a *specific* long-read SV set.



Supplemental Figure 10. |**Short-reads exclusive SVs in 51T and SKBR3.** Breakdowns of genetics and variation context (nearby presence of *sensitive* long-read SVs, breakpoints falling outside of GIAB high-confidence intervals for SV detection, and location within a high-coverage Illumina/10xG 500bp window) for short-reads exclusive SVs in 51T **(a)** and SKBR3 **(b)** with a *specific* long-read SV set.



Supplemental Figure 11. | Examples of short-reads exclusive SVs situated in high-coverage Illumina/10xG regions in 51T. Short-read exclusive duplication SVs predicted by Lumpy and SvABA, situated in high-coverage Illumina/10xG regions (~100x coverage) without **(a)** and with **(b)** nearby long-read supported insertion SV.



Supplemental Figure 12. | Examples of Lumpy predicted nested SVs situated in high-coverage regions in 51T. Duplication (**a**) and deletion (**b**) nested SVs predicted by Lumpy, situated in high-coverage Illumina/10xG region (>1000x) without nearby long-read supported SV call and with no coverabe abnormalities in either PacBio or ONT long-read alignments.

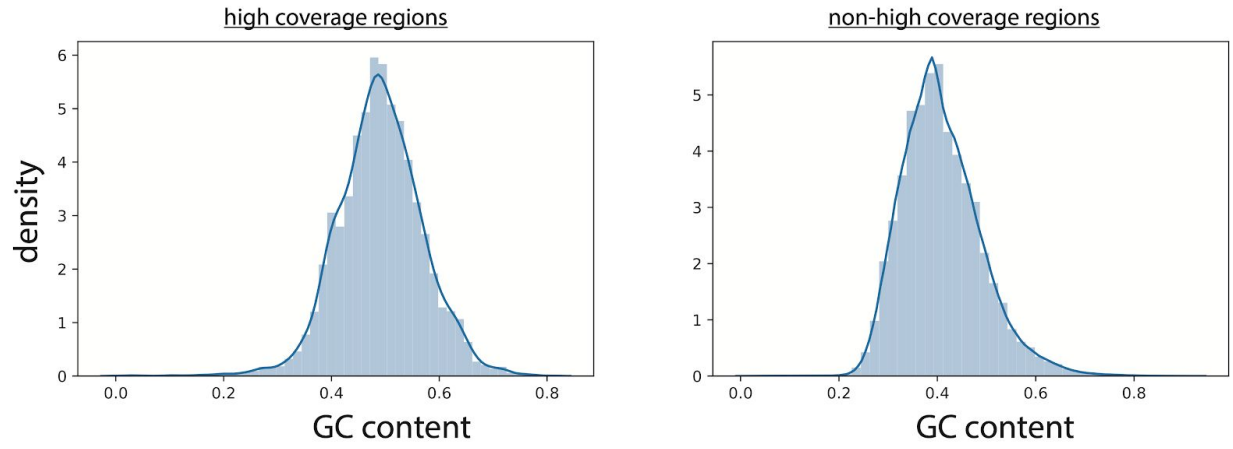
a

sample	all	in GIAB HC
51T	58630	32306
51N	51094	34118
SKBR3	42020	18081
NA12878	46705	18263
HG002	43026	14703

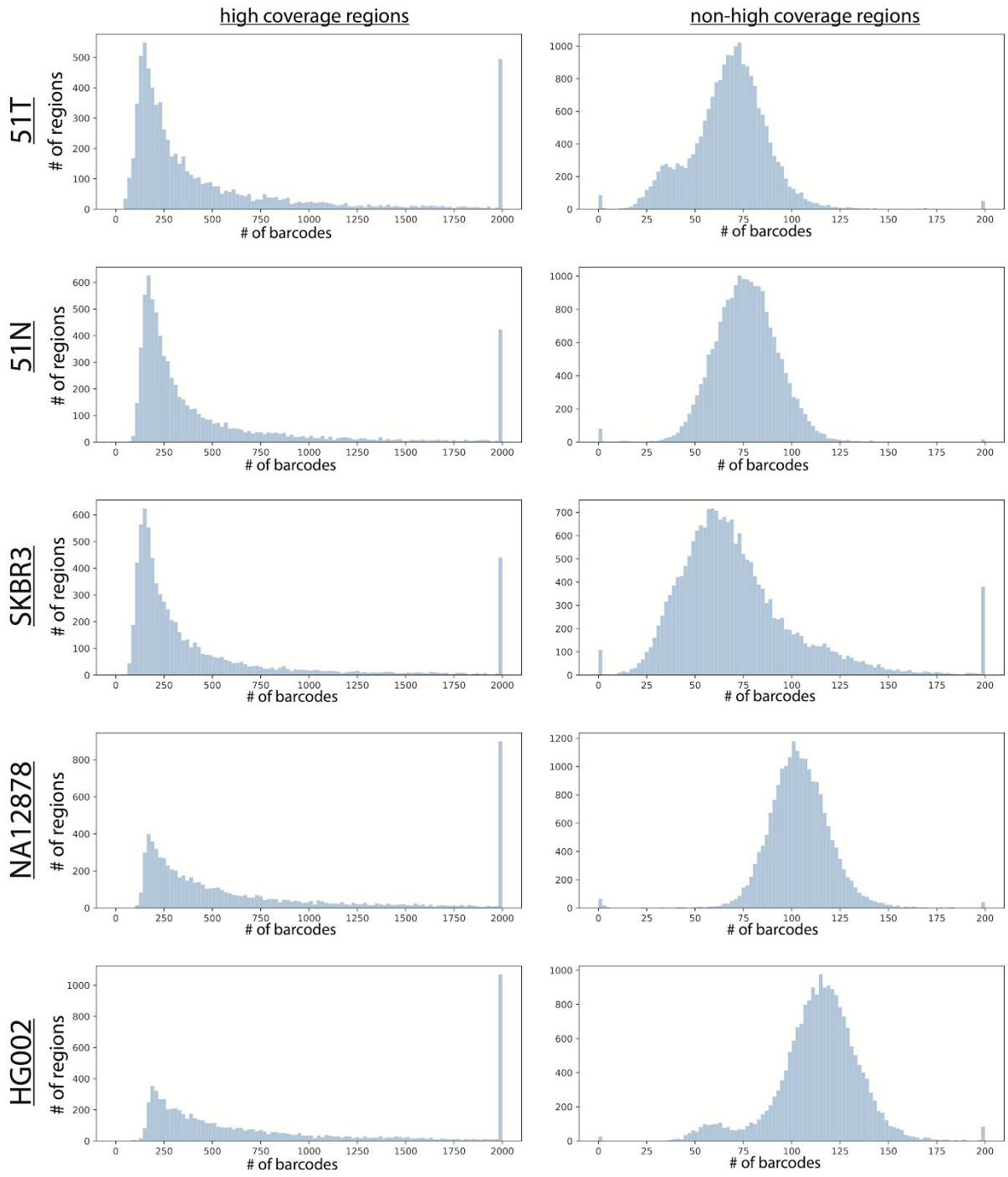
b

# of samples	shared
1	43536
2	7748
3	3773
4	2745
5	7228

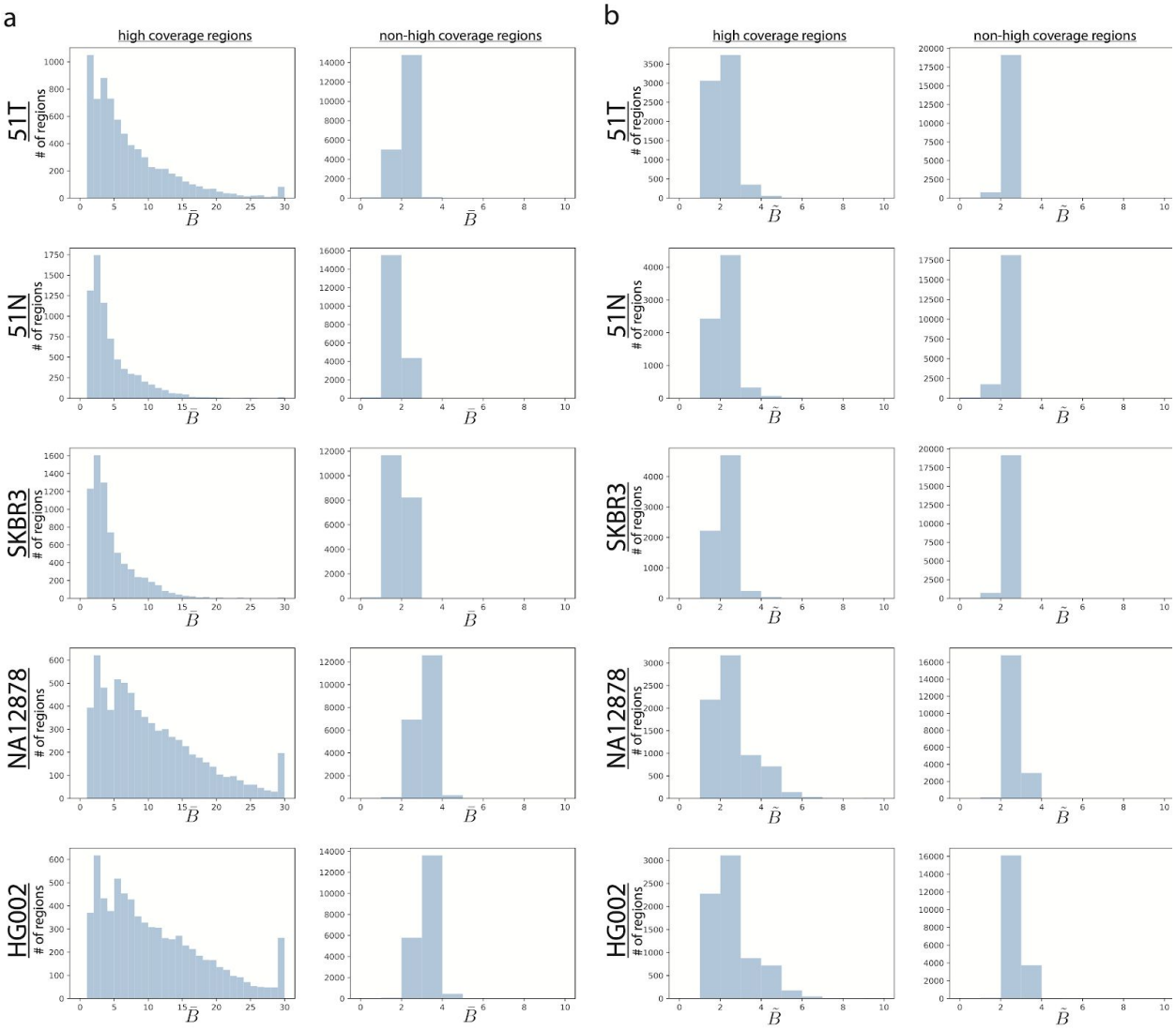
c



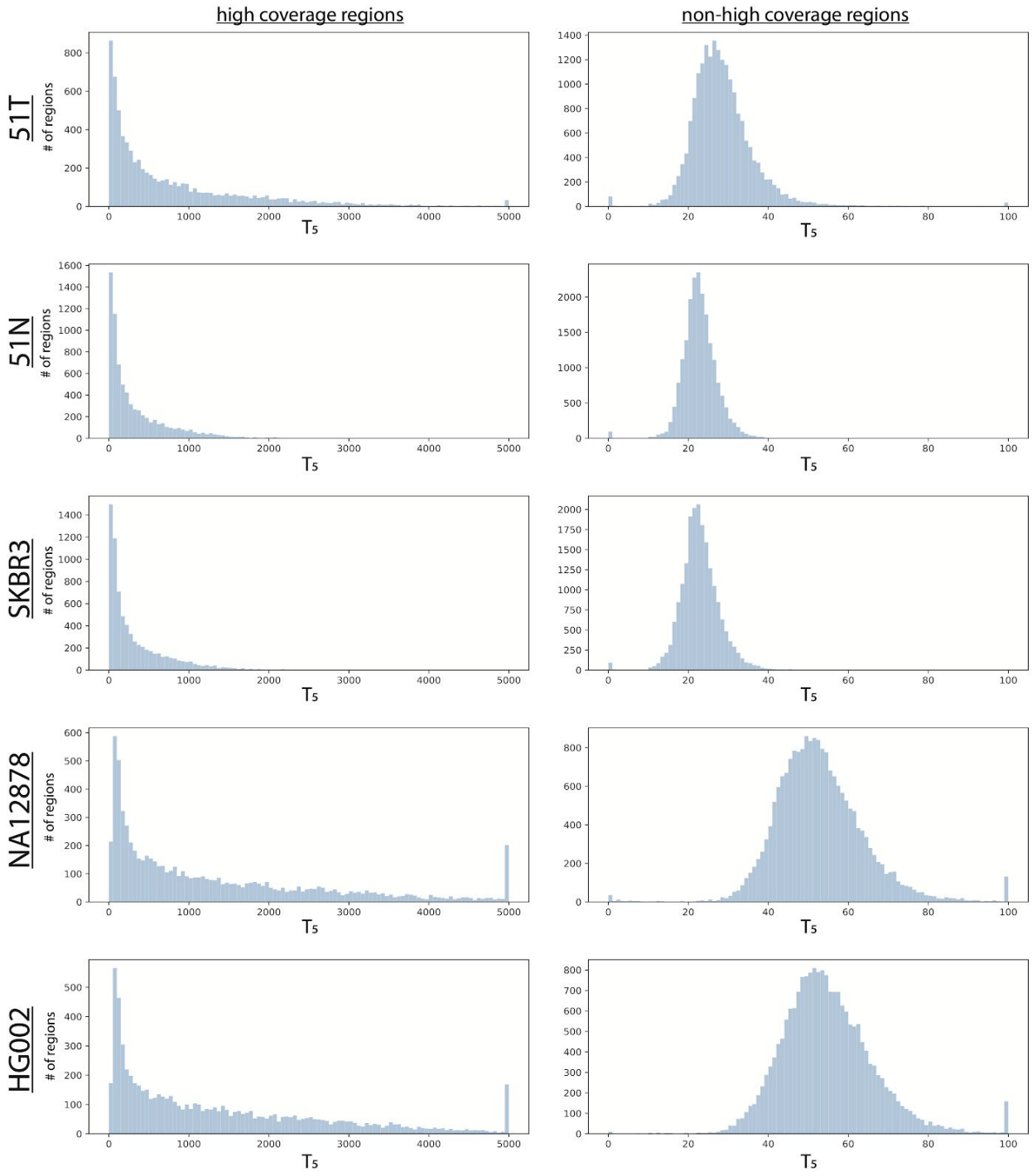
Supplemental Figure 13. | High coverage 10x Genomics regions. a) number of 500bp regions with high coverage in 10x Genomics datasets across 5 samples both across the whole genome and inside GIAB high confidence regions for SV detection **b)** number shared high coverage 10xG regions shared across samples **c)** GC content inside and outside of 500bp regions with high coverage in all 5 considered 10x Genomics datasets.



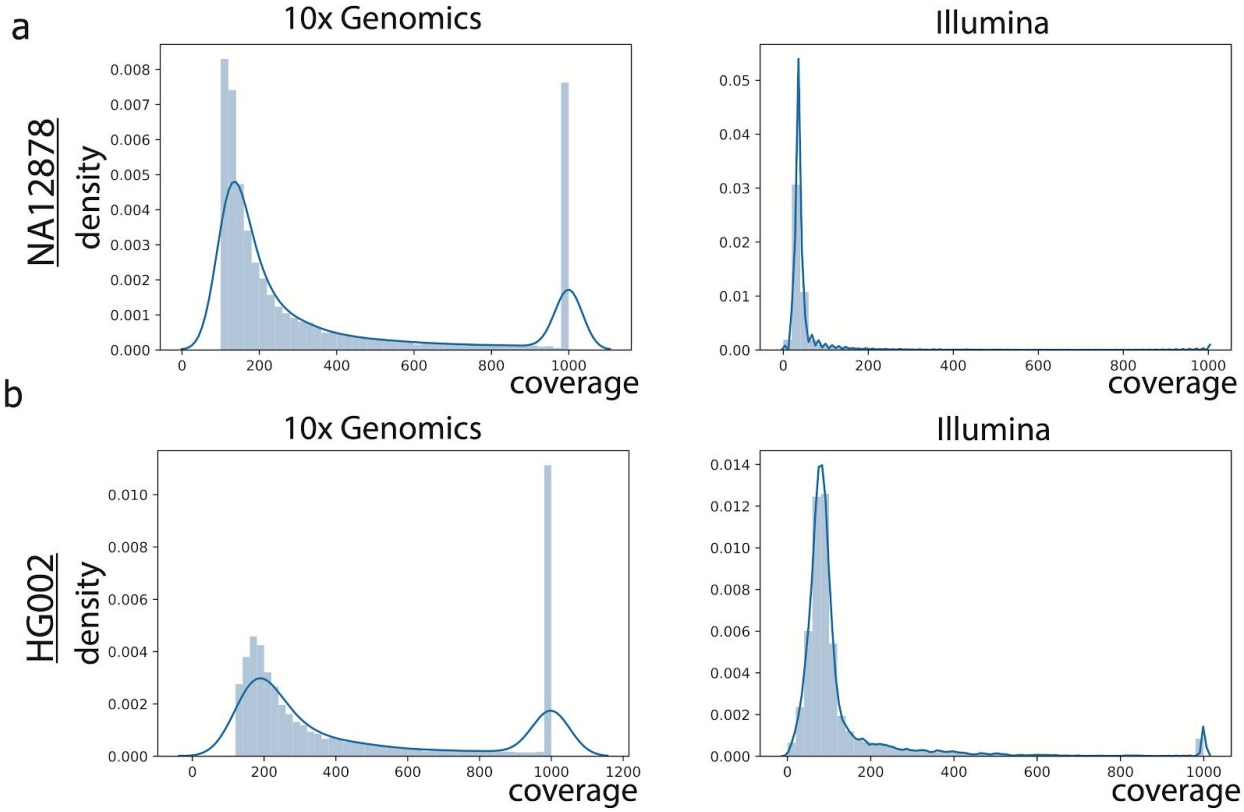
Supplemental Figure 14. | Distribution of the numbers of unique molecular barcodes with sequenced reads aligning within 500bp regions with (left) and without (right) average high coverage in 10x Genomics datasets. Values greater than 2000 (left) and 200 (right) are set to 2000 and 200 respectively.



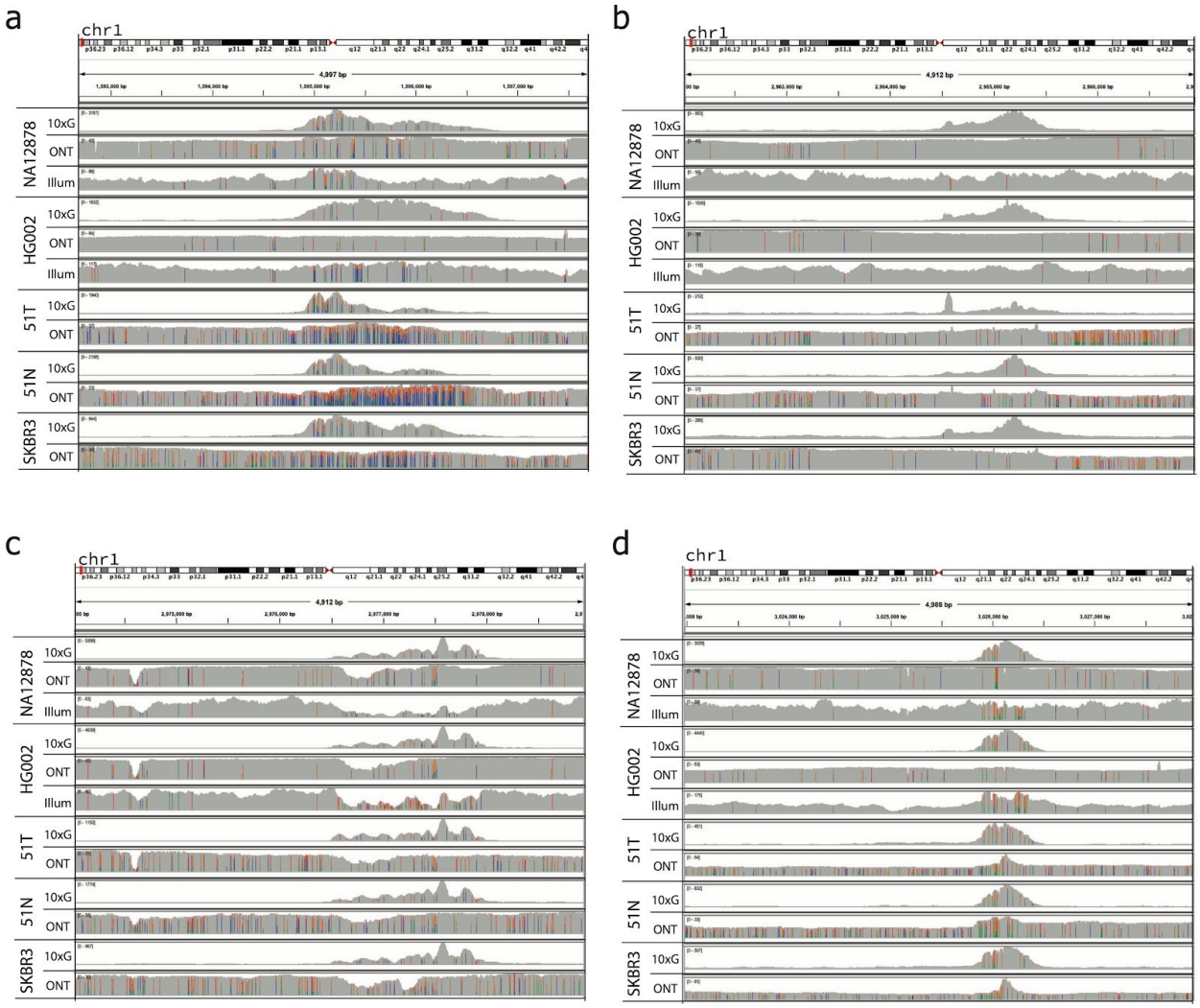
Supplemental Figure 15. | Barcode-read yield statistics in 10x Genomic datasets. a) Average number $\bar{B}$ of reads coming from a uniquely barcoded molecule(s) within distinct 500bp regions. b) Median number $\tilde{B}$ of reads coming from uniquely barcoded molecule(s) within distinct 500bp regions. $\bar{B}$ and $\tilde{B}$ values that are greater than the maximum shown x-axis value are set to the respective maximum x-axis values.



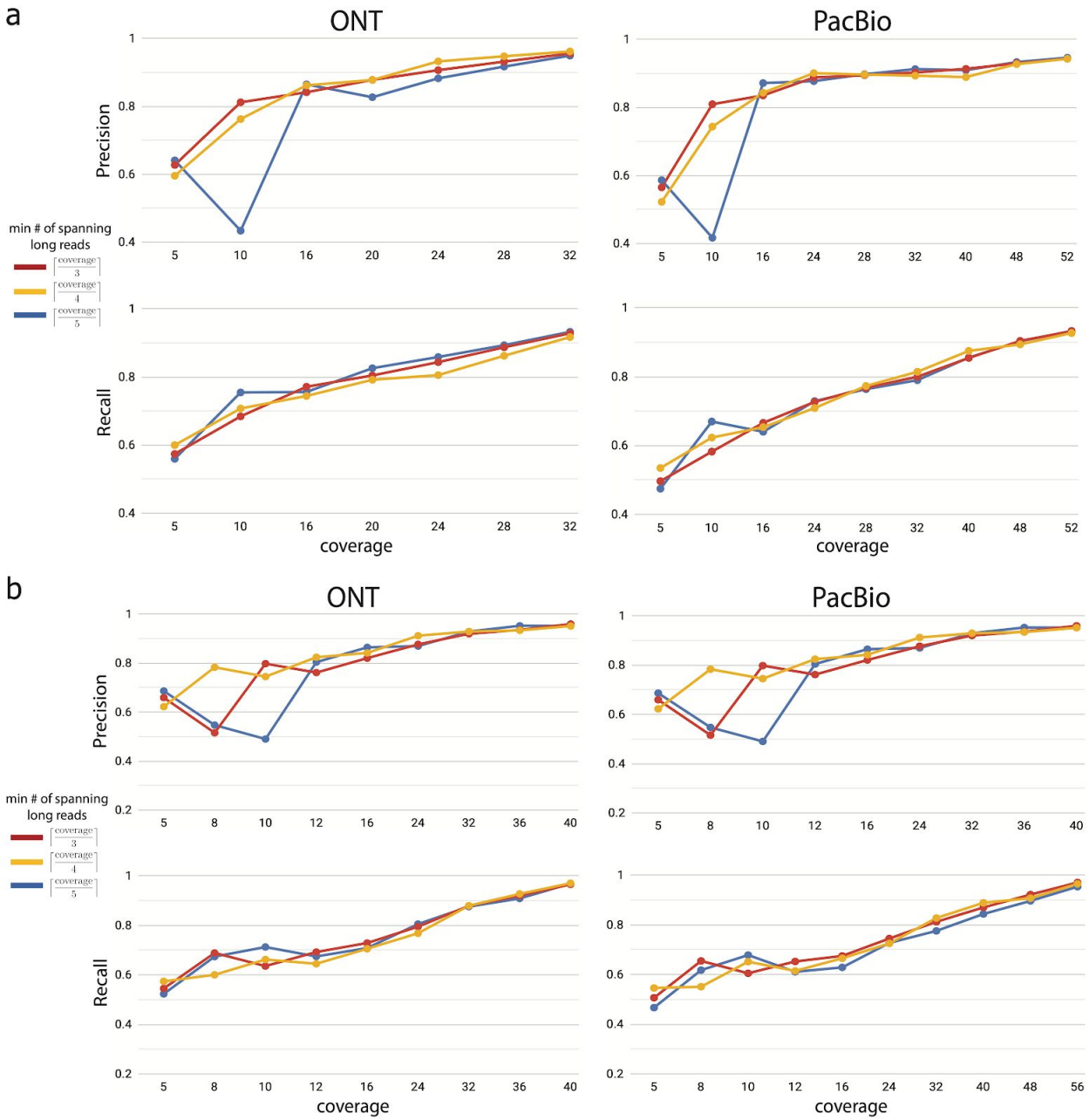
Supplemental Figure 16. | Distributions of the numbers T_5 of reads sequenced from 5 most read-represented barcoded molecules within 500bp regions with (left) and without (right) high-coverage abnormalities in 10x Genomics datasets. T_5 values greater than 5000 (left) and 100 (right) are set to 5000 and 100 respectively.



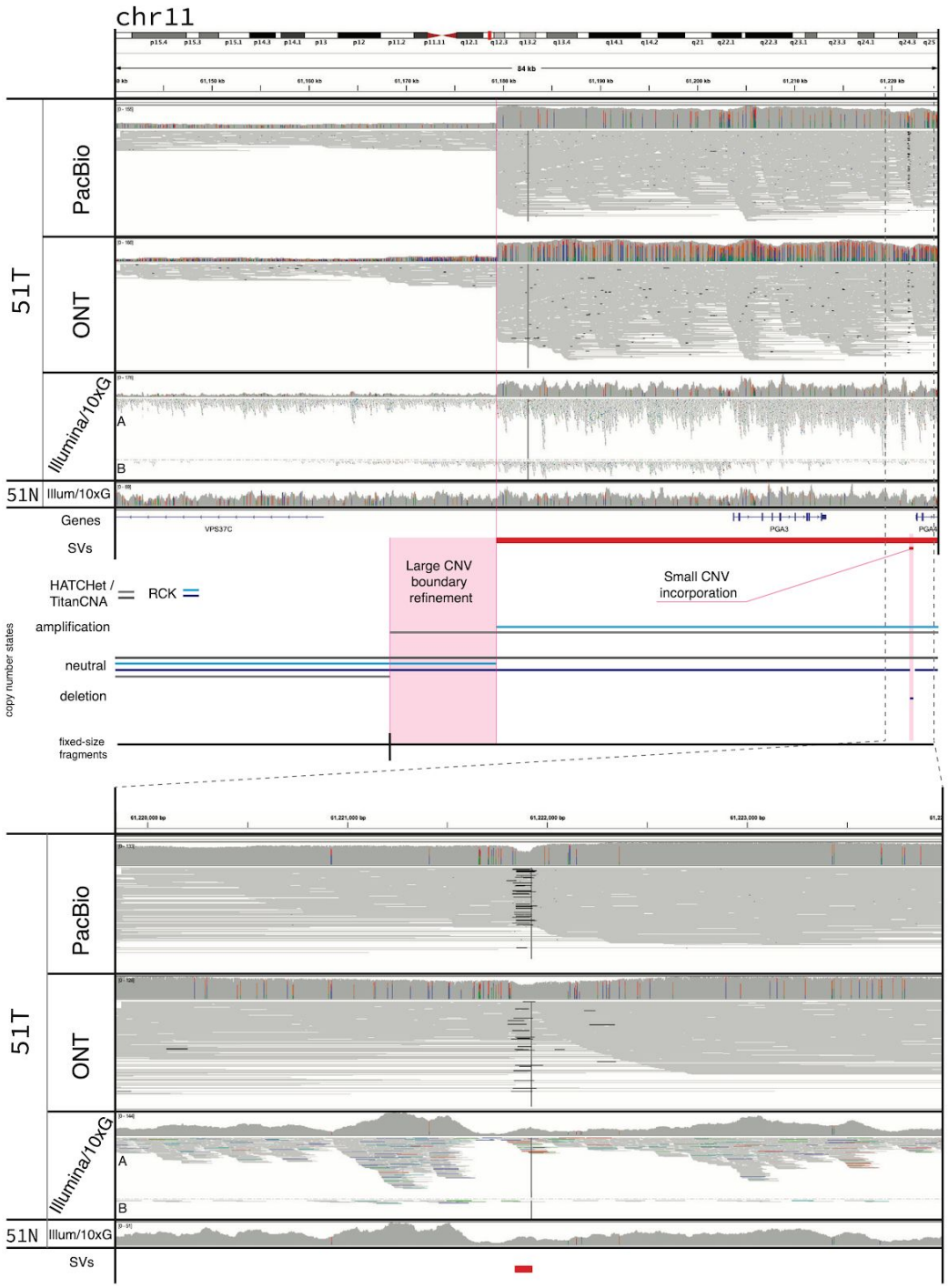
Supplemental Figure 17. | Mean read depth coverage in 10x Genomics and regular Illumina datasets in NA12878 and HG002 samples across 500bp regions that have a high coverage in 10x Genomics datasets. Coverage values that are greater than 1000 are set to 1000.



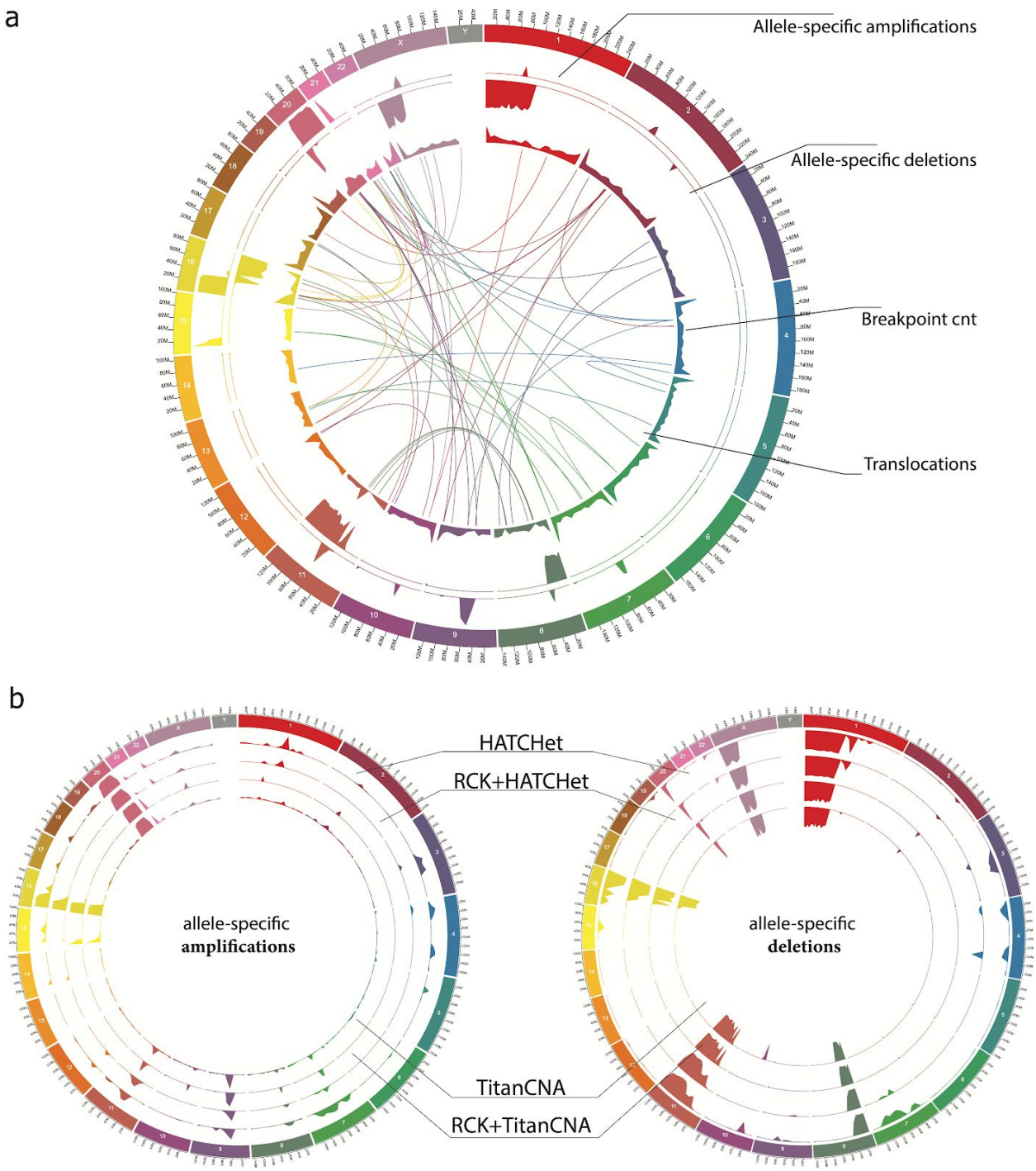
Supplemental Figure 18. | IGV screenshots of regions with extremely (>1000x) high coverage in 10x Genomics alignments alongside ONT and regular Illumina alignments with relatively uniform coverage.



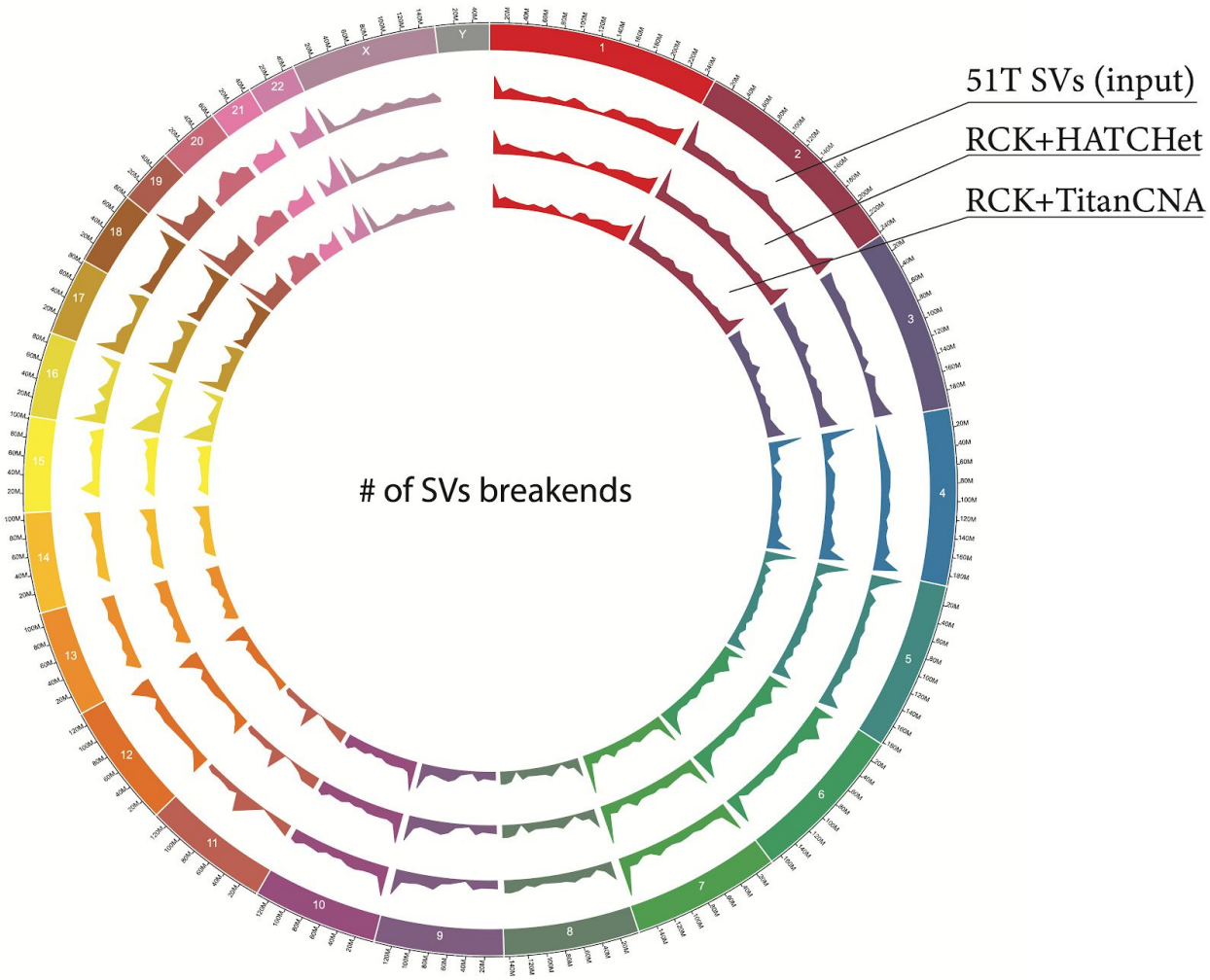
Supplemental Figure 19. | Concordance between downsampled and full coverage SKBR3 and 48T datasets with distinct minimum fractional *x/y* read support for an SV to be considered. a) Precision and Recall for SVs inferred on downsampeld ONT and PacBio dataset for sample SKBR3, with the SVs inferred on the full coverage dataset used as the ground truth. **b)** Precision and Recall for SVs inferred on downsampeld ONT and PacBio dataset for sample *48T*, with the SVs inferred on the full coverage dataset, at the matching support threshold, used as the ground truth.



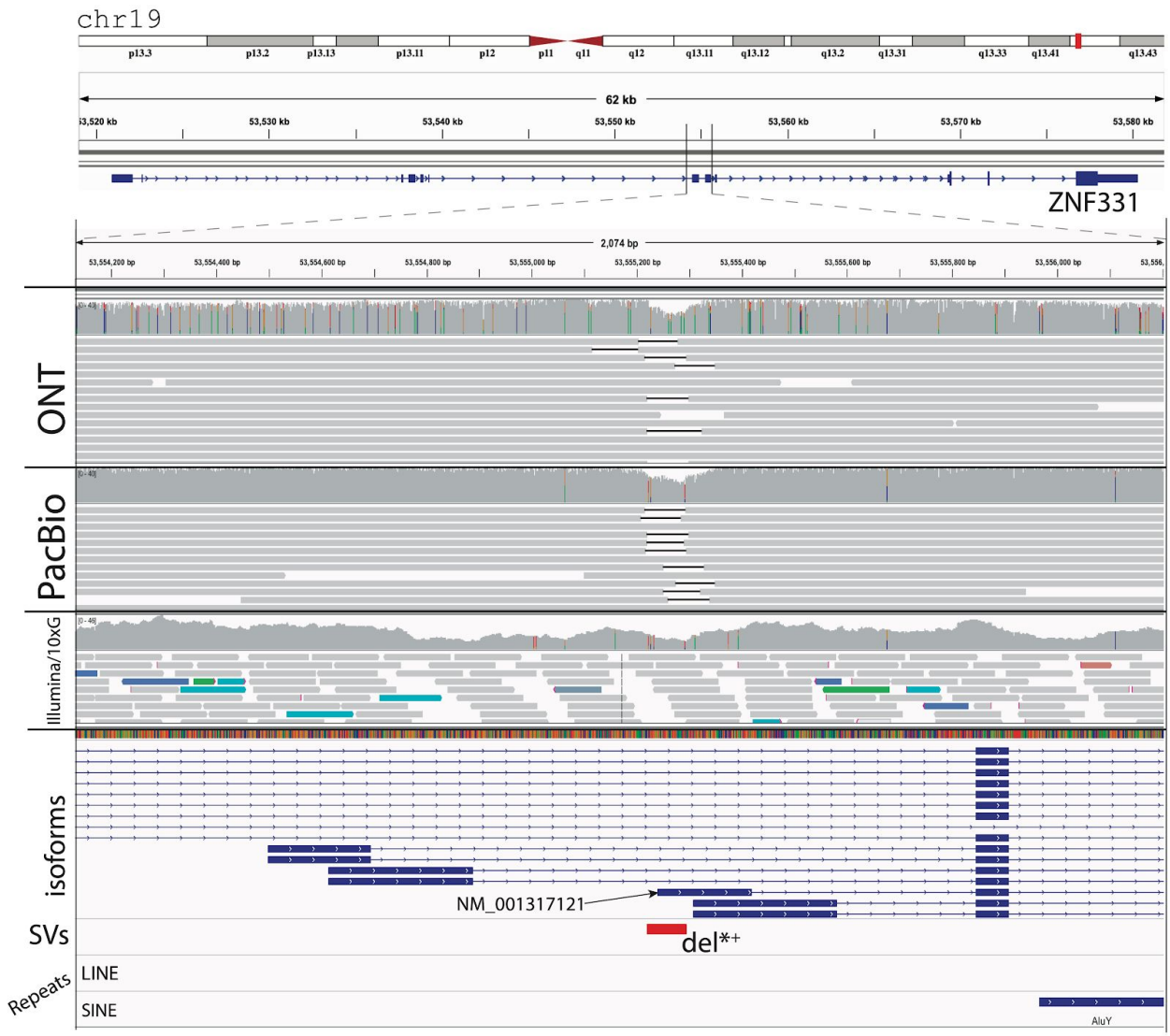
Supplemental Figure 20. | **Large-scale CNV's boundary refinement by RCK via a supporting SV and incorporation of a smaller CNV via a supporting SV on Chr11 in 51T.**



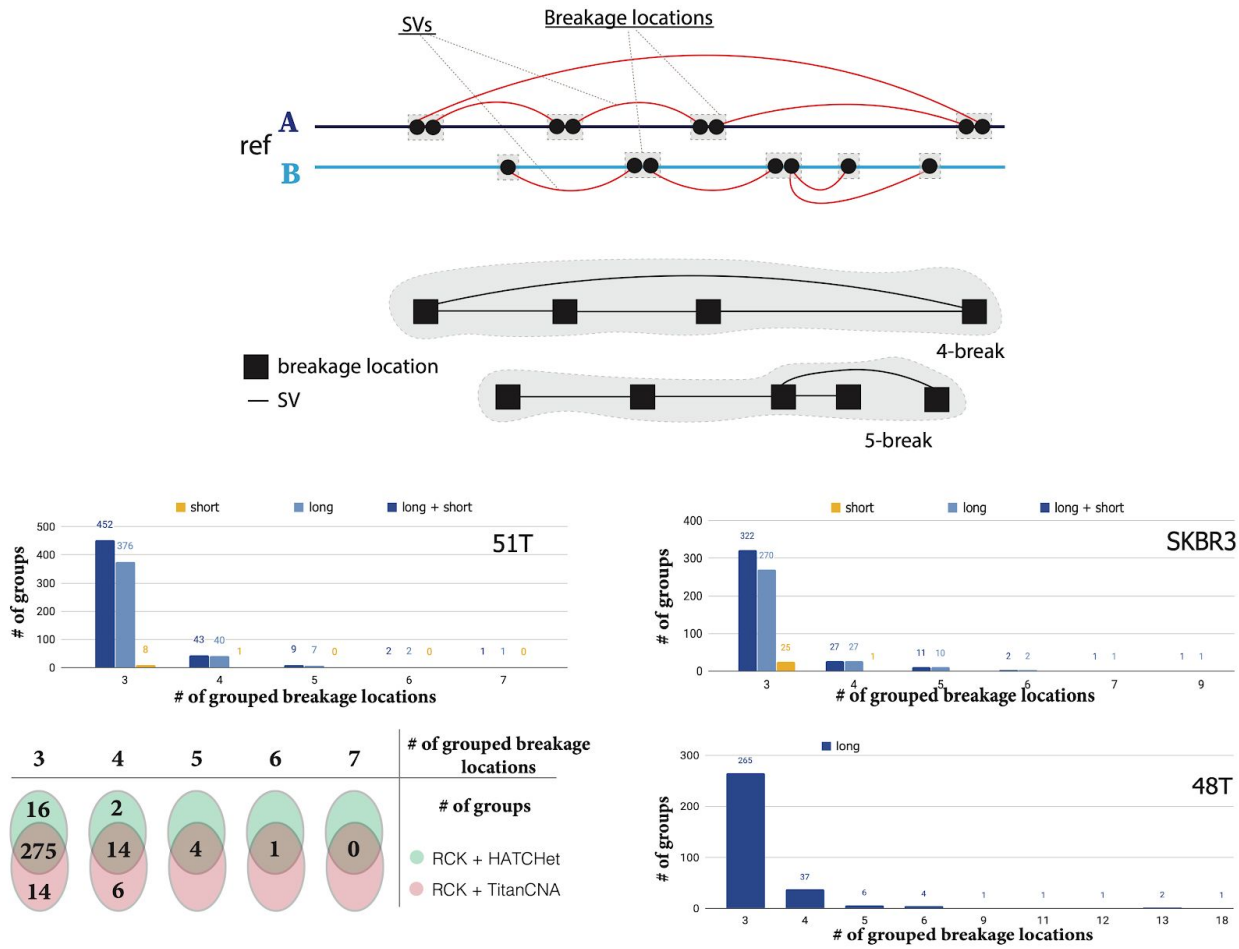
Supplemental Figure 21. | **CNVs comparison for cancer genome in sample 51T**. **a)** Circos plot of the cancer CNVs and SVs from karyotype-graph inferred by RCK for patient 51 with TitanCNA segment copy number (CN) input. Top two tracks corresponding to fractions *x/y* of the total length *x* of either amplified (CN ≥ 1) or deleted (CN = 0) fragments over the $y=5x10^6$ long windows. Breakend track shows the total number (with 590 being the maximum value shown) of breakends inferred by RCK as being present. Translocation track shows inter-chromosomal SVs inferred by RCK as being present. **b)** Circos plots of the allele-specific amplifications (left) and deletions (right) across fragments of the $y=5x10^6$ length in raw HATCHet(top track), RCK+HATCHet (second track), RCK+TitanCNA (third track), and raw TitanCNA (fourth track) CNV profiles.



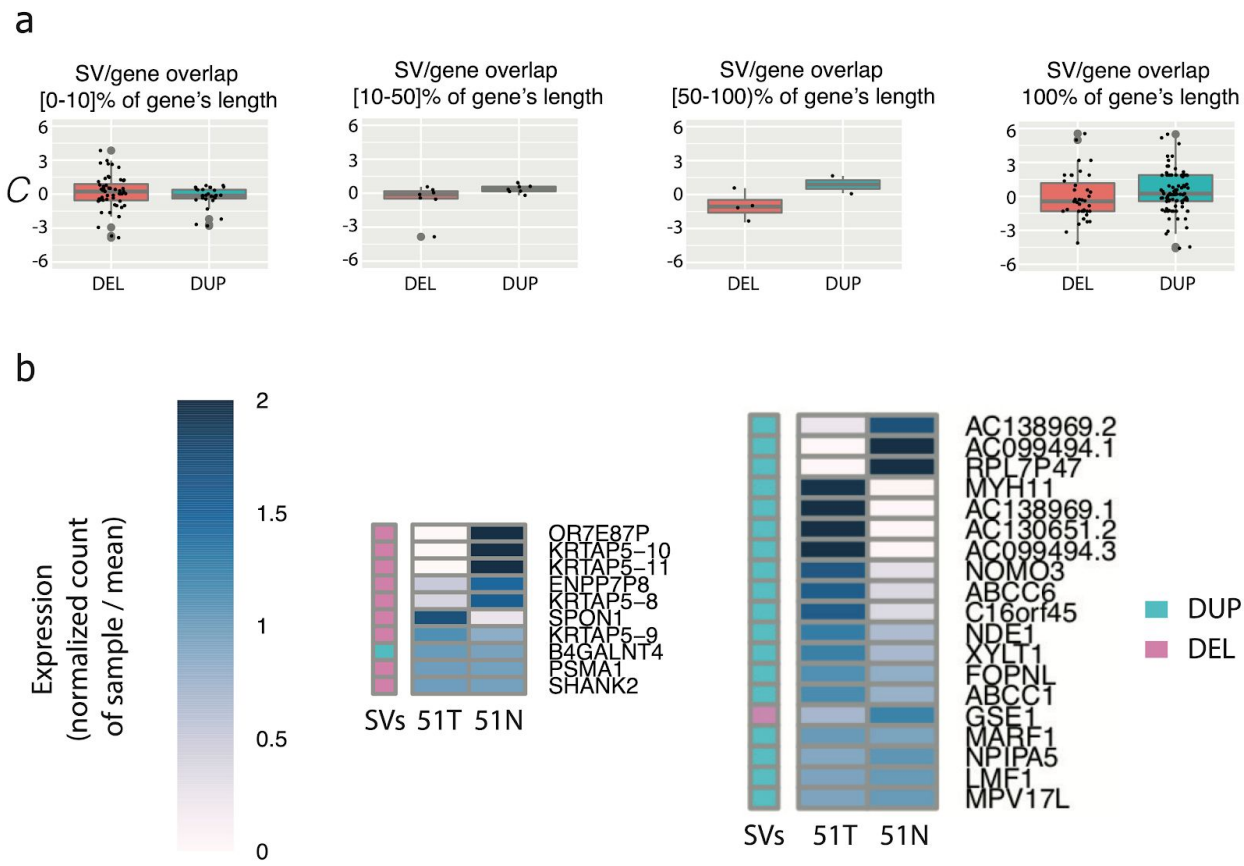
Supplemental Figure 22. | Breakpoints distribution in both raw and karyotype-graph SV callsets for sample 51T. Circos plot of the SVs breakends distributions across genome chromosomes. Every track is dataset-specific shows the total number of SVs' breakends over 5MBp segment-length windows in the 51T specific SV callset (top track), and SVs utilized in the karyotype-graph reconstructed by RCK with HATCHet (second track) and TitanCNA (third track) allele-specific CNVs input.



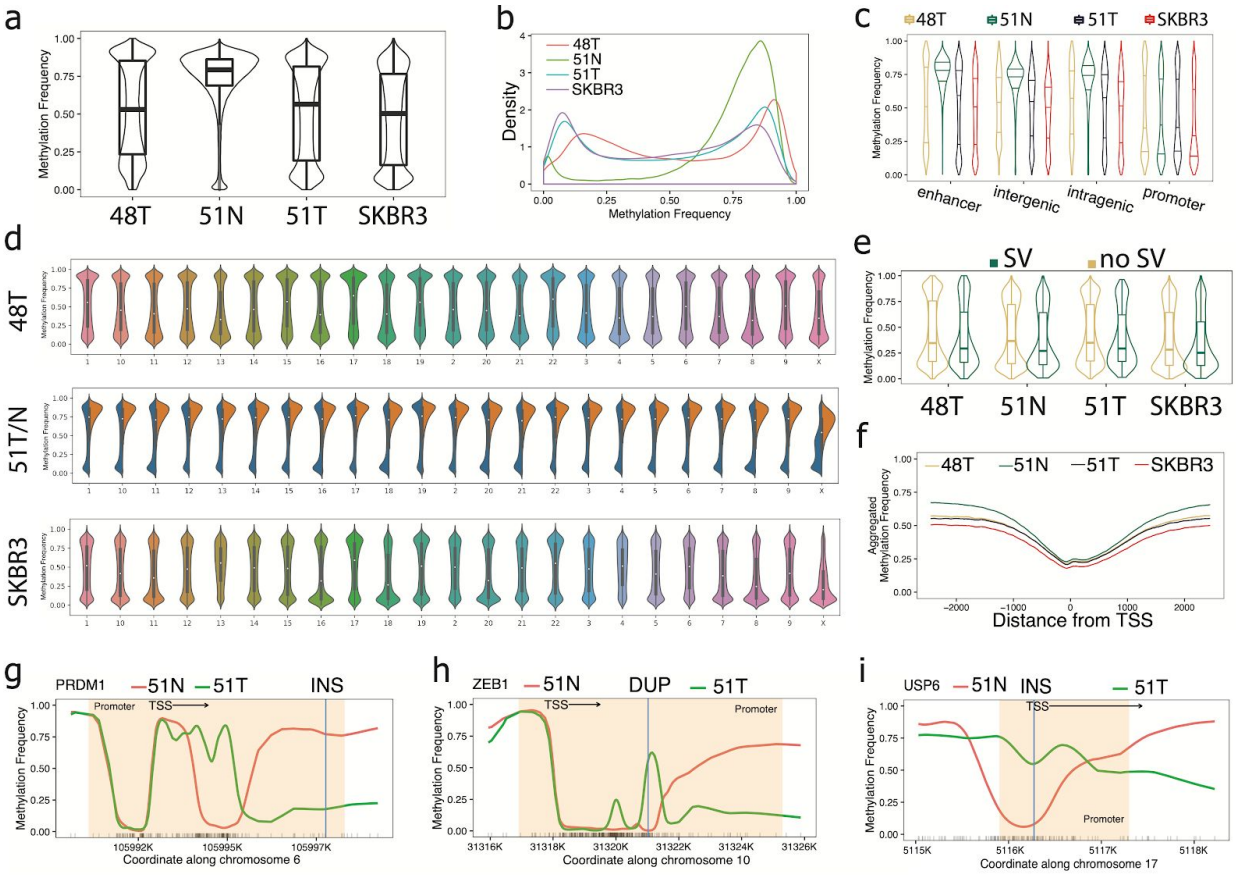
Supplemental Figure 23. | A deletion SV identified in COSMIC census gene ZNF331 in patient 51. A deletion in the *ZNF331* gene within an exon in the *NM_001317121* transcript, and genotyped in < 1% of 1KGP project samples.



Supplemental Figure 24. | SVs breakend grouping 51T, SKBR3 and 48T cancer samples. (Top) Signature of two complex k-break rearrangements (i.e., 4- and 5-breaks) as evident by breakage locations linked via reciprocal SVs. Statistics over k-breakend groups identified in the input SV callset for samples 51T, SKBR3, and 48T (with a breakdown over long, short, and long+short read support) and in RCK+HATCHet and RCK+TitanCNA karyotype-graphs for sample 51T.



Supplemental Figure 25. | RNA-seq expression analysis in patient 51 for genes overlapped by SVs present only in sample 51T and not in 51N and not in the union of SVs from 15 healthy genomes. a) Changes in expression for genes overlapped by deletion and duplication SVs quantified by the percentage of genes' lengths spanned by SVs. Every gene g is represented as a dot, with $C(g) = (T(g) + 0.5) - (N(g) + 0.5)$, where $T(g)$, and $N(g)$ are expression count for g in 51T and 51N samples respectively. b) Groups of genes affected by Deletions and Duplications. **b)** examples of long DUP/DEL SVs spanning multiple genes and changes in respective genes' expression levels in 51T vs 51N samples respectively.



Supplemental Figure 26. |Methylation analysis on cancer samples 48T, 51N, 51T, and SKBR3. a) Genome-wide distribution of methylation frequencies. **b)** Genome-wide methylation frequency density functions. **c)** Methylation frequency distributions over 1kbp windows that overlap enhancer, intergenic, intragenic, and promoter regions respectively. **d)** Per-chromosomal methylation frequency distributions. **e)** Methylation frequency distributions in the 1kbp windows which overlap promoter regions and either contain or do not contain SVs breakends respectively. **f)** Averaged aggregated methylation frequencies around Transcription Start Sites (TSS). **g)** Methylation frequencies in *PRDM1* promoter region with transition between hypermethylation in 51N and hypomethylation in 51T around identified insertion. **h)** Methylation frequencies in *ZEB1* promoter region with hypermethylated region in 51T coinciding with identified duplication SV. **i)** Methylation frequencies in *USP6* promoter region with identified insertion coinciding with blocking of the TSS from demethylation in 51T.