Cell Genomics, Volume *3*

Supplemental information

Long-read sequencing of diagnosis and post-therapy medulloblastoma reveals complex rearrangement patterns and epigenetic signatures

Tobias Rausch, Rene Snajder, Adrien Leger, Milena Simovic, Mădălina Giurgiu, Laura Villacorta, Anton G. Henssen, Stefan Fröhling, Oliver Stegle, Ewan Birney, Marc Jan Bonder, Aurelie Ernst, and Jan O. Korbel

This supplementary information is structured in figures and tables.

Supplementary Figures

S21. Comparison of the somatic variant allele frequency in the primary and relapse

S22. Relapse-specific frameshift insertion in SUFU.

S23. Templated insertion rearrangement graphs

S24. Templated insertion threads across 2,569 cancer genomes

S25. TI thread in a primary dedifferentiated liposarcoma sample P1

S26. TI thread in liposarcoma metastasis sample P2

S27. Multiple integration sites of a TI thread structure in P2

S28. Telomere sequences associated with rearranged genomic regions

S29. Comparing 450k array versus Nanopore methylation calls

S30. Length of differentially methylated regions

S31. Methylation array coverage of differentially methylated regions

S32. Differential promoter methylation compared to differential expression

S33. Differential methylation of TBX1

S34. Methylation of templated insertion threads

S35. Overview of the BASP 1 locus including expression and splicing.

S36. Gene expression effects of templated insertion threads

S37. Allele copy number ratios for chromosome 19 with telomere associated SV

S38. PHRED quality score and read length histograms

S39. Germline structural variant size distribution

S40. Comparison of targeted and *de novo* assembly

S41. Short-read simulation experiment for TI threads

S42. Long-read simulation experiment for TI threads

Supplementary Tables

Supplementary Figures

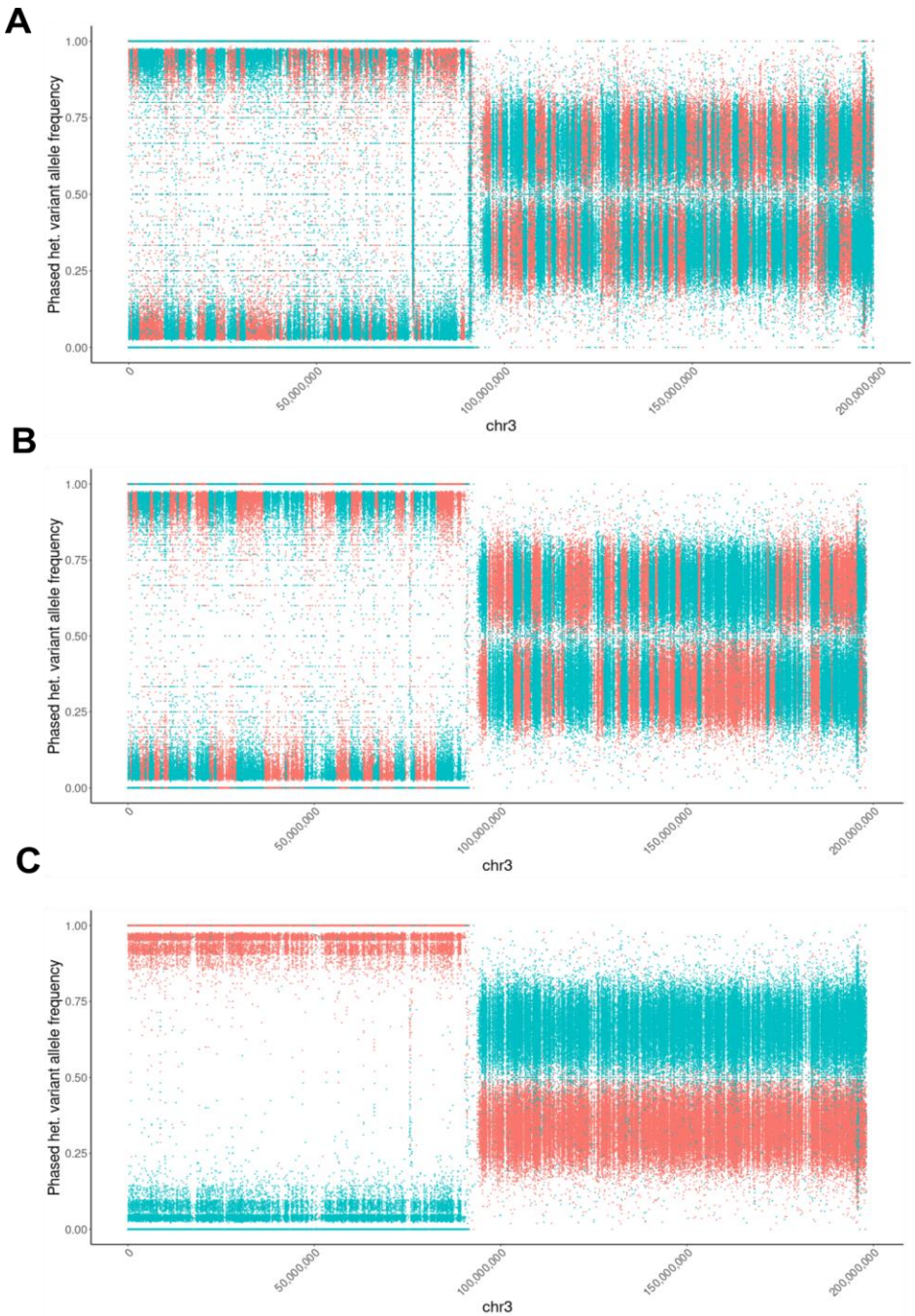


Figure S1. Haplotype phasing approach. To account for the high error-rate and relatively low germline coverage of only ~15x, a three-step phasing approach was developed. **(A)** In step one, single-nucleotide variants and short insertions and deletions were phased using Whatshap[1], yielding an N50 phased block length of 2.29Mbp for chromosome 3. Each dot is a phased heterozygous variant colored by haplotype with the tumor variant allele frequency on the y-axis. **(B)** Statistical phasing using ShapeIt[2] with the long-read phased blocks as input increased the N50 phased block length to 4.68Mbp. **(C)** The unequal haplotype dosage of chromosome 3 in the primary tumor (haploid p-arm, triploid q-arm) was used to correct remaining switch errors on chromosome 3. Related to STAR Methods.

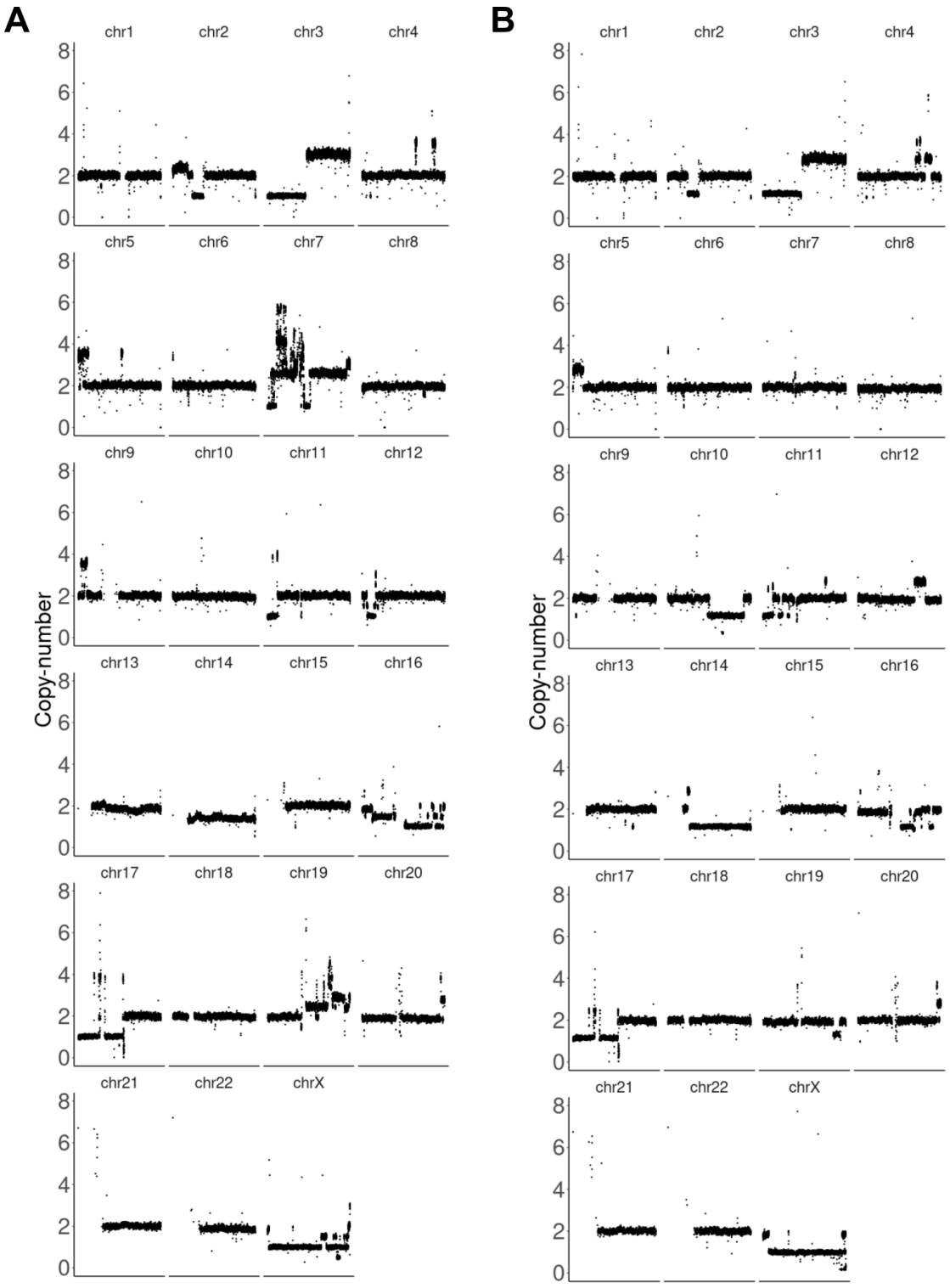


Figure S2. Copy-number profiles. The normalized copy-number (y-axis) of each chromosome (x-axis) using a 10kbp window length of uniquely mappable positions and GC-fragment corrected read counts from delly's cnv mode[3]. Panel **(A)** shows the primary tumor and panel **(B)** the relapse sample. Related to section "Haplotype-phased assembly of complex somatic rearrangements", related to Figure 1.

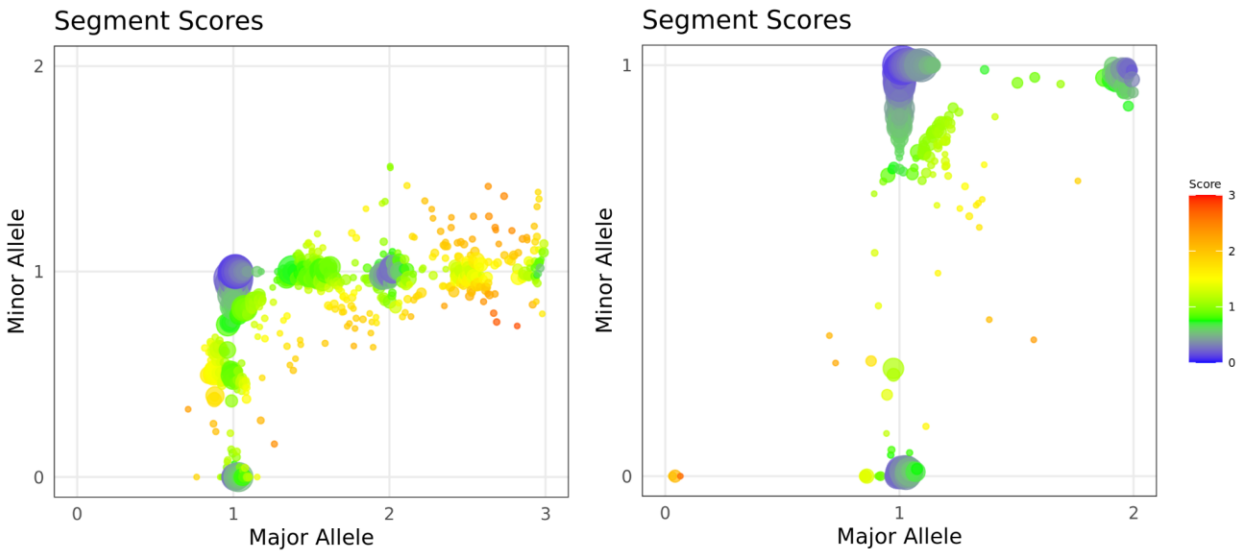


Figure S3. Major and minor copy-number for all genomic segments in the primary and relapse. Major and minor allele of each copy-number segment as computed by Purple[4] for the primary tumor (left) and relapse sample (right). Many segments in the primary tumor do show a non-integer copy-number as is the case for subclonal changes. Related to section “Haplotype-phased assembly of complex somatic rearrangements", related to Figure 1.

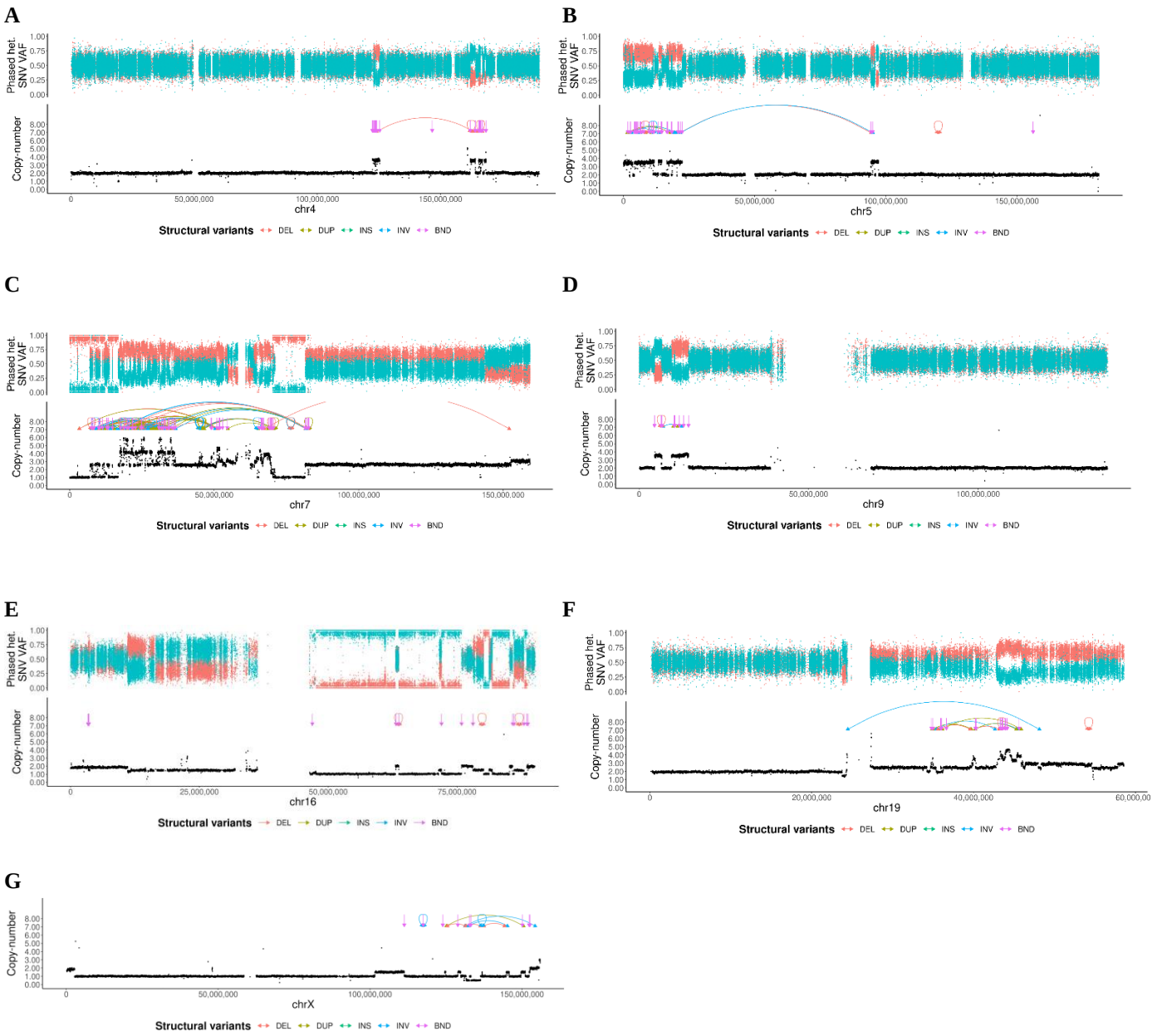
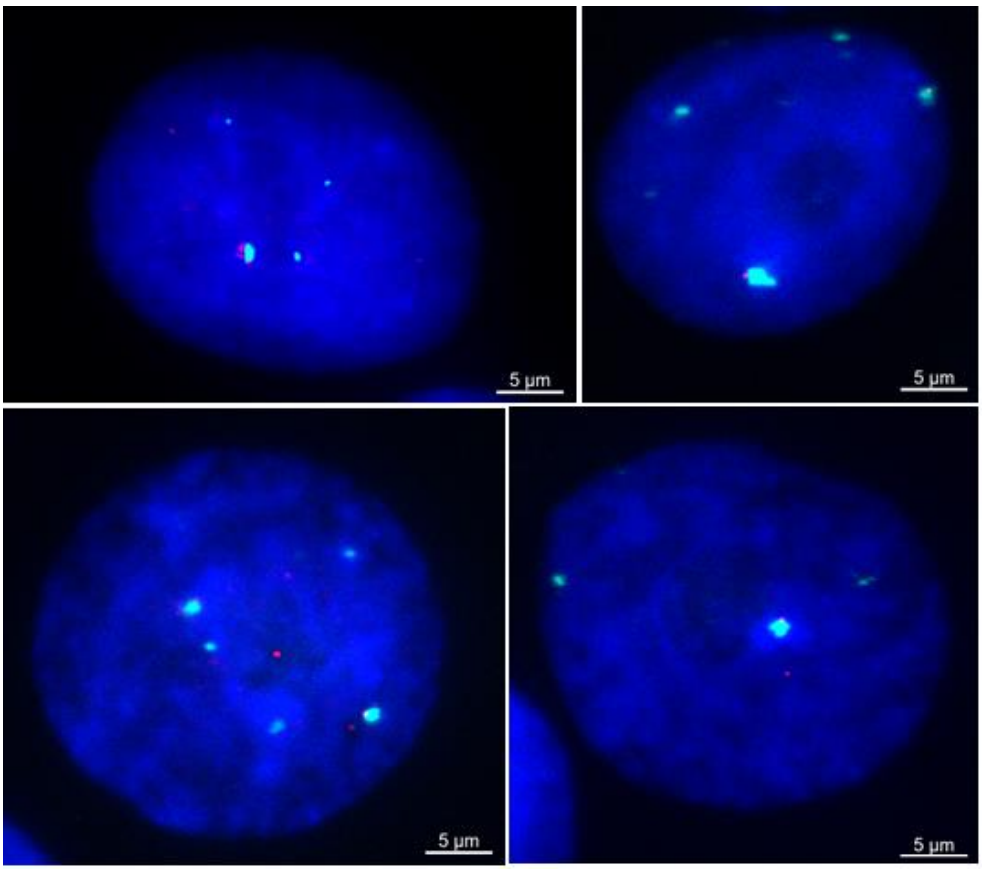


Figure S4. Copy-number plots of chromothripsis chromosomes. The upper panel shows the single-nucleotide variant (SNV) allele frequency (VAF) for phased heterozygous germline variants colored by haplotype. The lower panel shows the normalized copy-number (y-axis) of each chromosome (x-axis) using a 10kbp window length. Overlaid are somatic structural variant (SV) calls colored by SV type. Panel **(A)** shows chromosome 4, **(B)** shows chromosome 5, **(C)** shows chromosome 7, **(D)** shows chromosome 9, **(E)** shows chromosome 16, **(F)** shows chromosome 19 and panel **(G)** shows the haploid chromosome X (male sample). Related to section "Haplotype-phased assembly of complex somatic rearrangements", related to Figure 1.

A)



B)

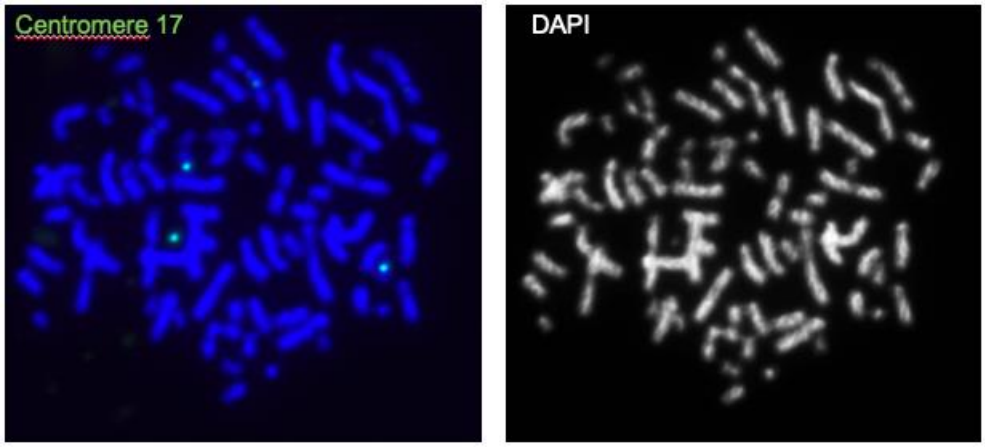


Figure S5. FISH analysis. (A) FISH pictures of the red RP11-651L9 probe (chr17:16,169,409-16,359,715) and the green centromere 17 probe showing distinctive intra-tumor heterogeneity for the CS11-17 structure. Signals for the RP11-651L9 probe vary from 2 to 10 occurrences with frequent co-localization to the centromere 17 probe. **(B)** Using metaphase spreads from patient-derived xenografted cells signals for the centromere 17 probe (left) and DAPI (right) suggest an absence of double-minute chromosomes. From the three centromere signals detected here, two are on chromosomes and one is on a chromosome fragment (lower left, putative marker chromosome or ring chromosome). Related to section “Haplotype-phased assembly of complex somatic rearrangements", related to Figure 1.

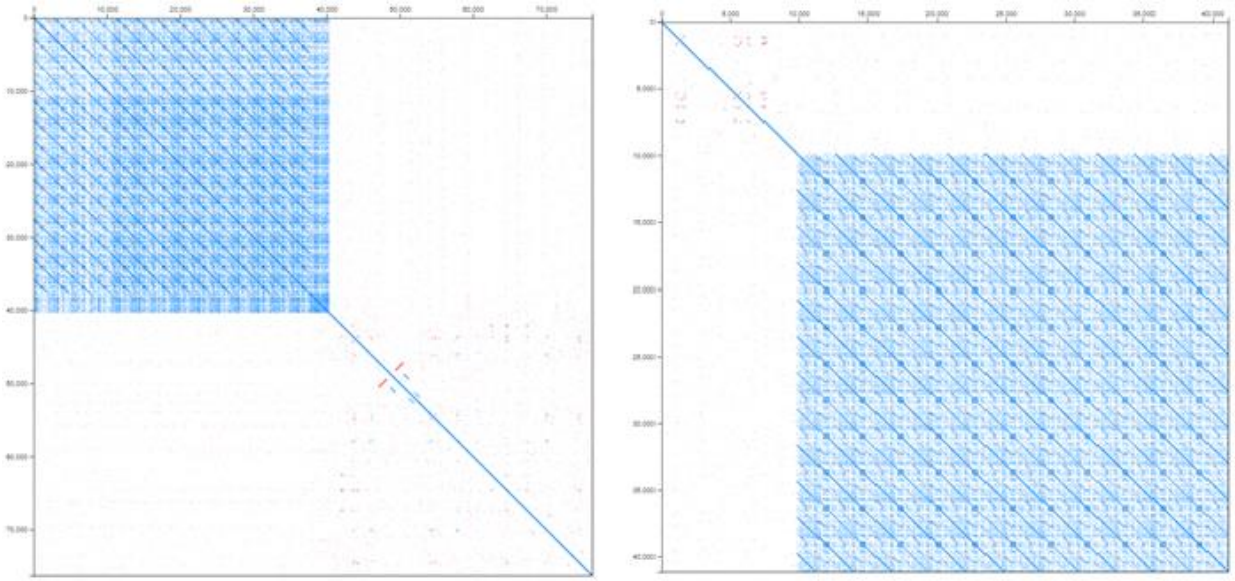


Figure S6. Repetitive embedding of contig 2 of the CS11-17 assembly. Dot plots of the longest read against itself at the left and right end of contig 2 of the CS11-17 assembly. Reads start in a unique sequence context of contig 2 but then extend into repetitive sequences on either end. Related to section “Haplotype-phased assembly of complex somatic rearrangements", related to Figure 1.

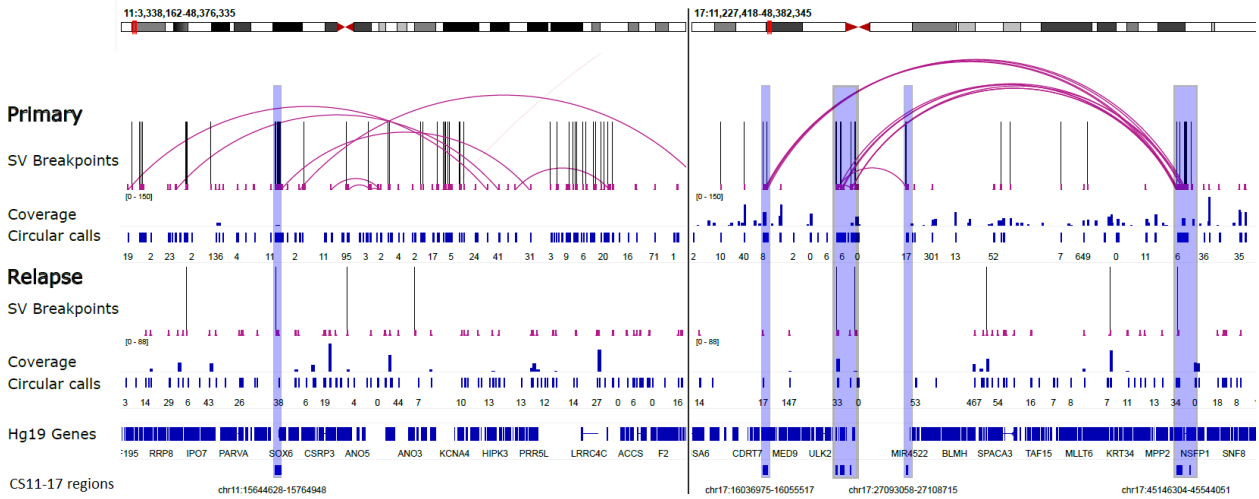


Figure S7. Circle-Seq analysis of CS11-17. Circle-Seq coverage for primary (top panel) and relapse (bottom panel) is overlaid with regions that are part of the CS11-17 assembly (bottom track). The Circle-seq data shows focal DNA enrichment for regions overlapping with CS11-17 (Circular calls track), as well as multiple SVs connecting those parts on the targeted assembly, suggesting a possible circularized structure for CS11-17. Related to section “Haplotype-phased assembly of complex somatic rearrangements", related to Figure 1.

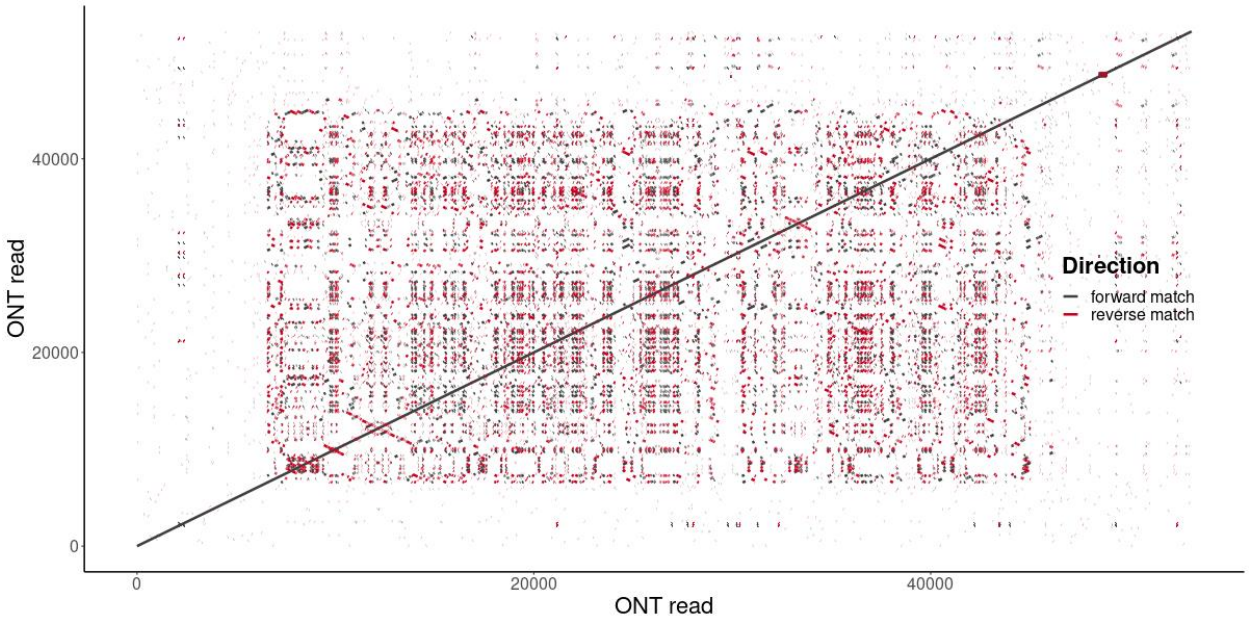


Figure S8. Self-alignment of the templated insertion thread. Forward and reverse matches of a single ONT read aligned against itself for the 2nd instance of a templated insertion thread in the primary tumor, as in Figure 2A for the 1st instance. Related to section “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread", related to Figure 2.

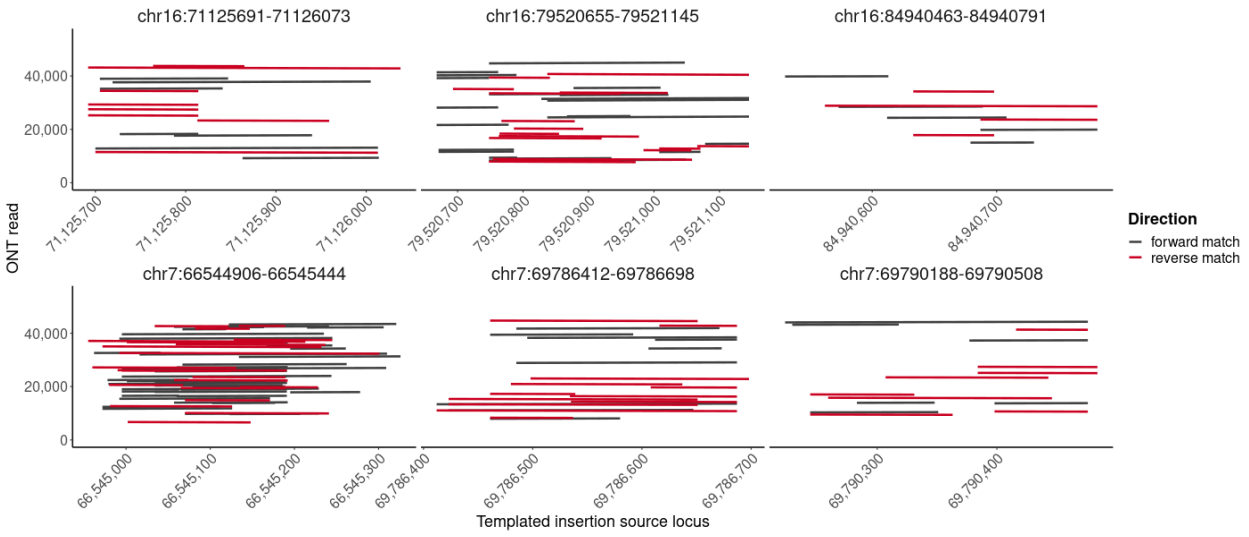


Figure S9. Templated insertion thread alignments. Similar to Figure 2B, forward and reverse matches for a subset of the templated insertion source sequences of the 2nd instance of a templated insertion thread. Related to section “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread", related to Figure 2.

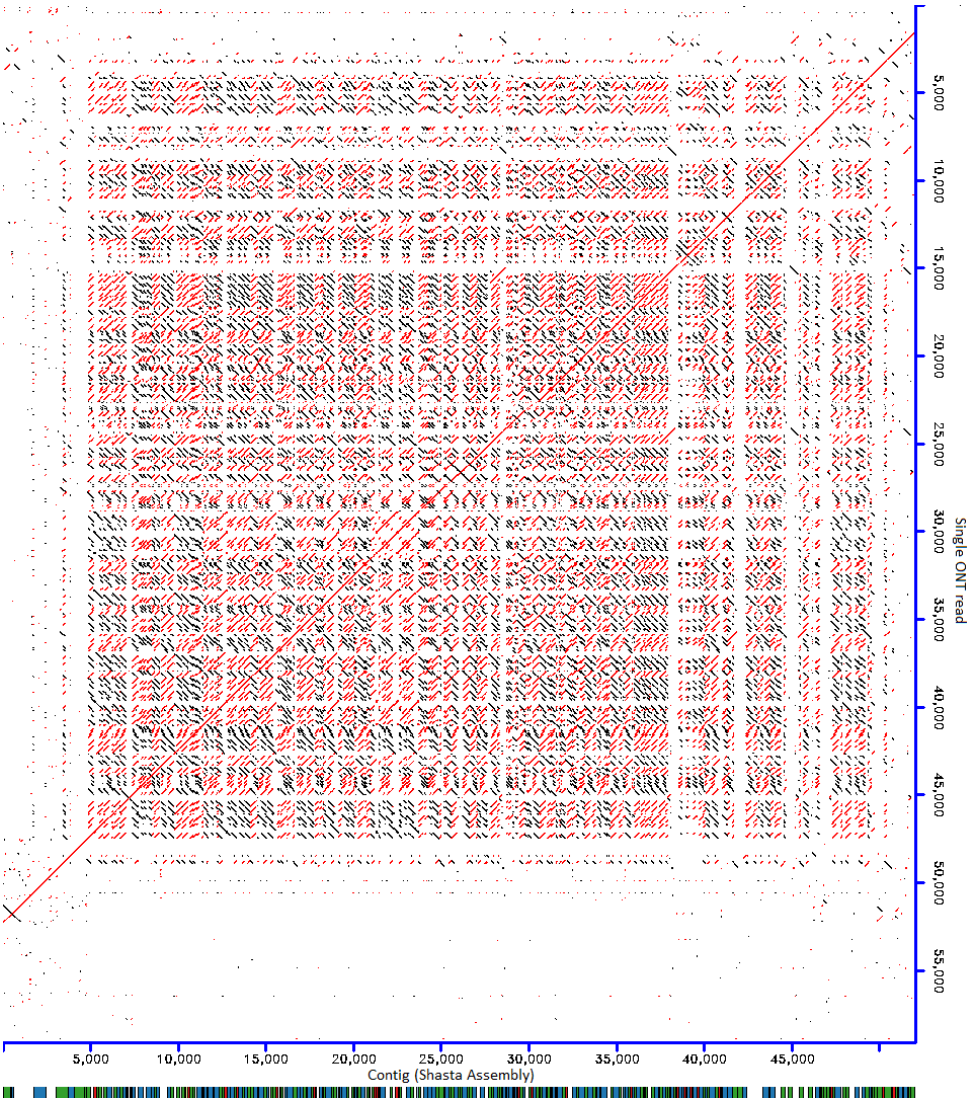


Figure S10. TI threads are supported by *de novo* assembly. TI thread structure is supported by a raw long read (y-axis) and a *de novo* assembled Shasta[5] contig (x-axis). Colored bars at the bottom indicate unique segment alignments to GRCh38 with different colors representing different chromosomes from which the TIs were derived. Related to section “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread", related to Figure 2.

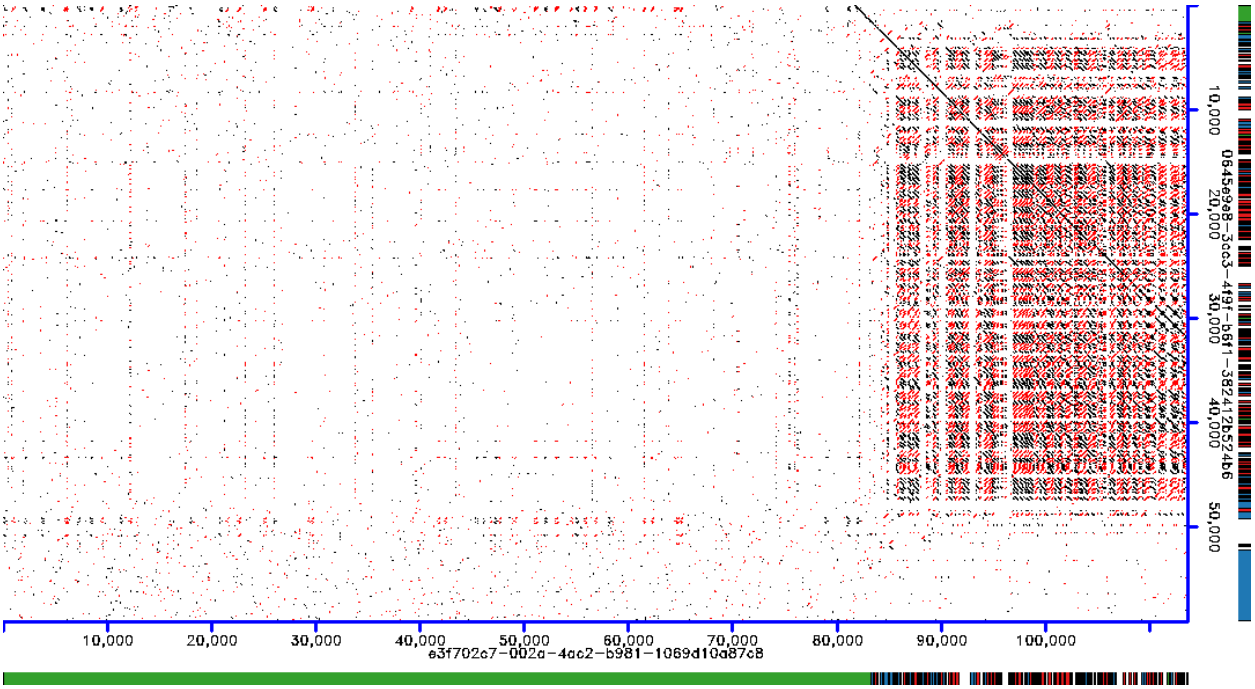


Figure S11. Patient-derived xenograft (PDX) data confirms TI thread. Dot plots of a PDX derived raw ONT read (x-axis) against a raw ONT read of the primary tumor (y-axis) presented in Figure 2 of the main manuscript. Both reads support the TI structure with matches to GRCh38 highlighted in the plot margins and colored by source chromosome. Related to section “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread", related to Figure 2.

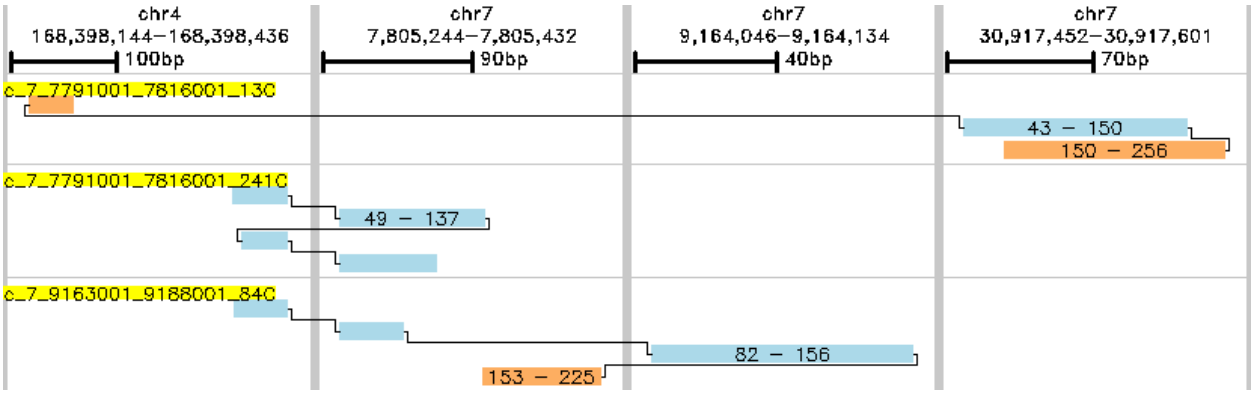


Figure S12. Short-read SvABA contigs confirm TI thread breakpoints. Short-read contigs computed by SvABA[6] confirm individual SV breakpoint junctions of TI threads. Each of the 3 contigs highlighted in yellow (top, middle and bottom panel) support at most 3 SV breakpoints but the characteristic self- and cross-linking of TI threads is evident. Related to section “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread", related to Figure 2.

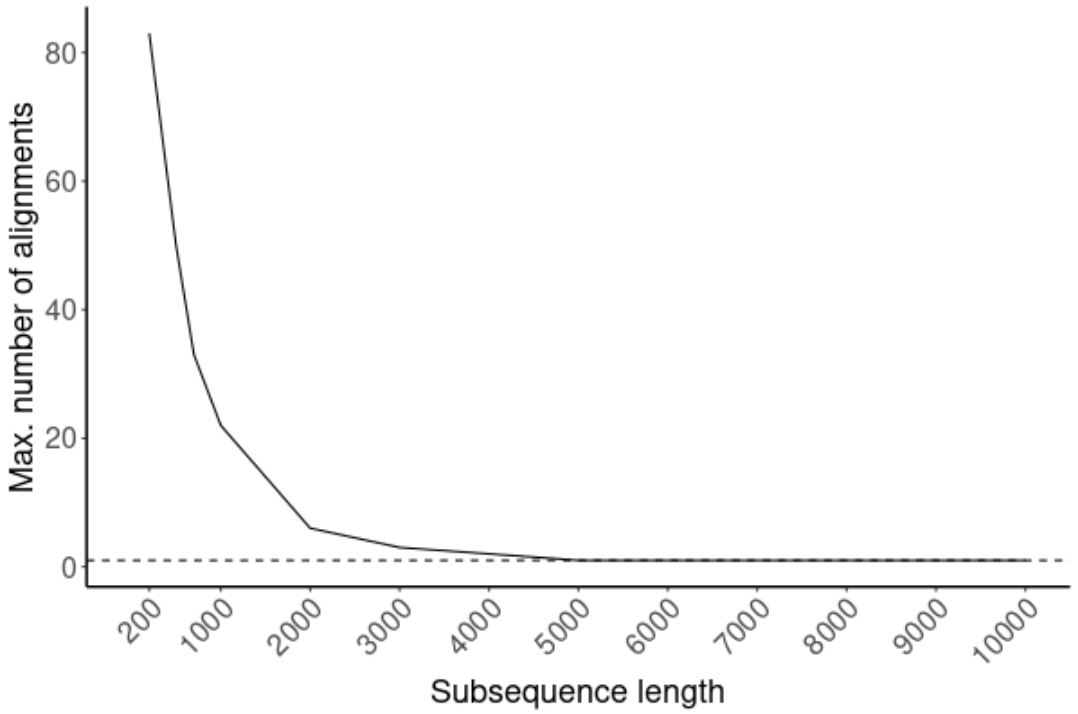


Figure S13. Subsampling analysis of ONT reads spanning templated insertion threads. To identify possible higher-order repeating structures, we randomly sampled sub-sequences of different length (x-axis). Each sub-sequence was iteratively aligned to the original ONT read, masking all previous alignments until no further confident alignments could be identified. The dashed line corresponds to the original sampling location without any further alignments (y=1). Related to section “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread", related to Figure 2.

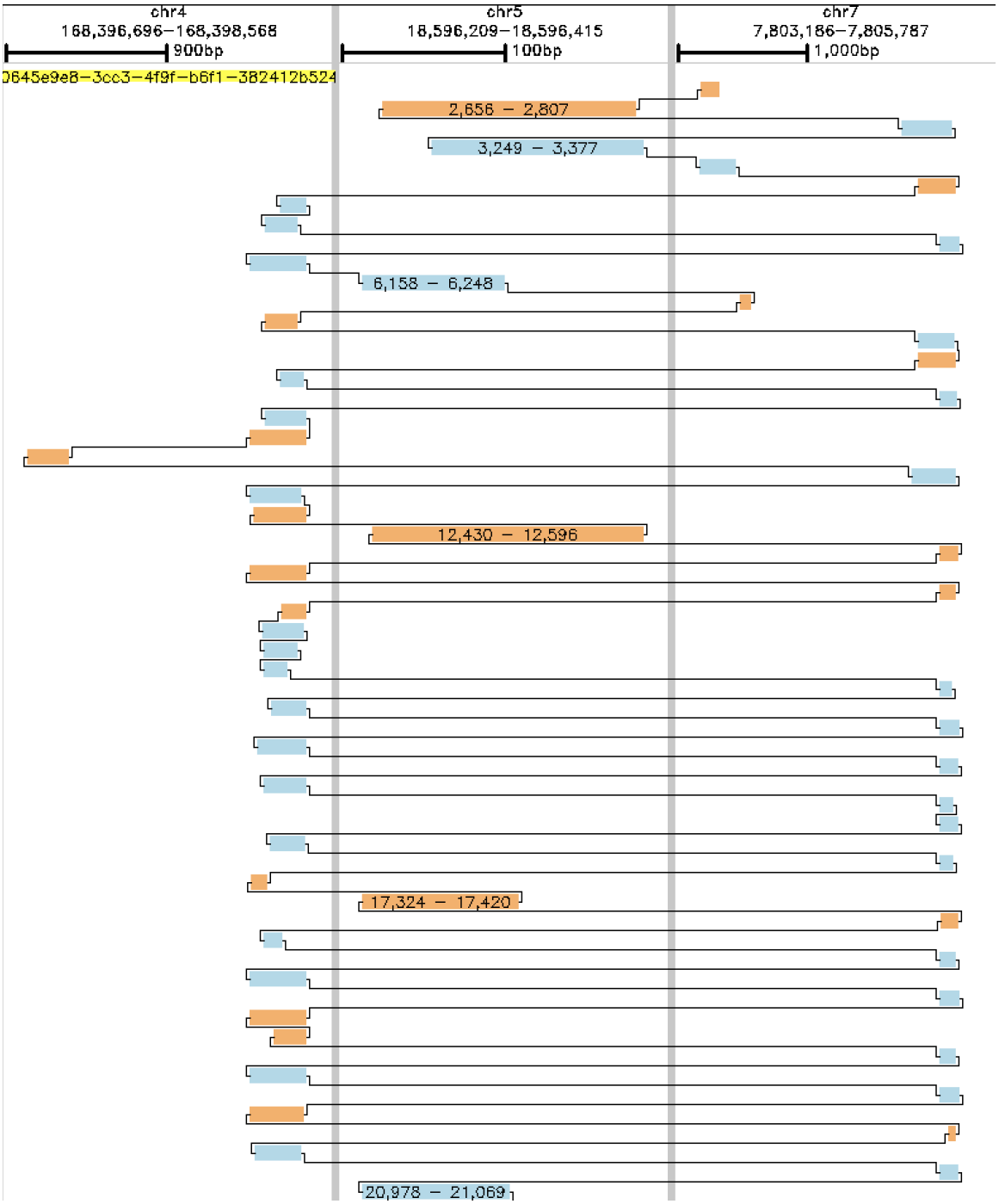


Figure S14. Genomic matches of a single ONT read with a TI thread. Forward matches are in blue, reverse matches in ochre and the black line traces the matches along the read, concatenating in a zig-zag fashion the templated insertion source sequences. Each vertical panel is a separate genomic alignment region specified in the header. Only a subset of the matches in the first 22kbp of the read are shown for illustration reasons, the full alignment of this read is available as a separate file in our Zenodo package (https://doi.org/10.5281/zenodo.7658888). Related to section “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread", related to Figure 2.

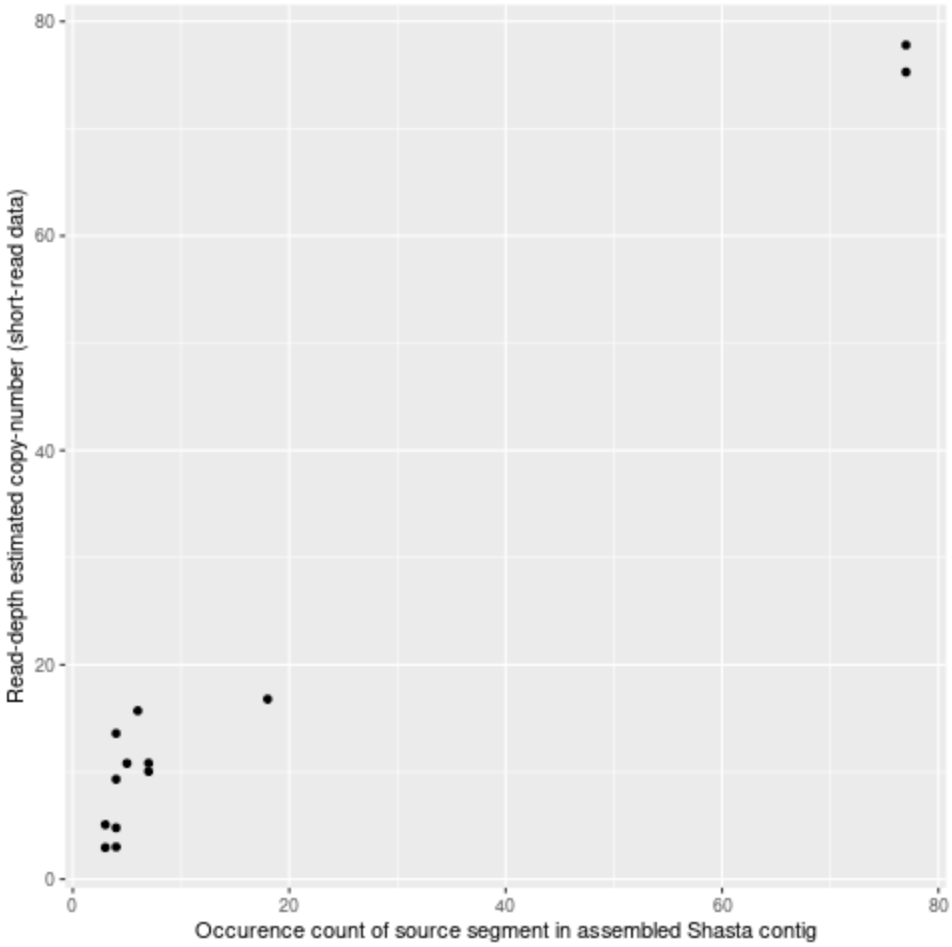
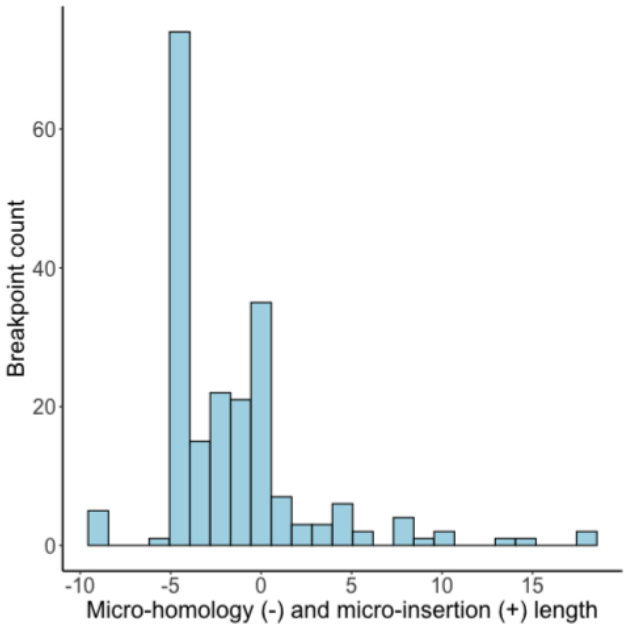


Figure S15. Correlation analysis of TI segment occurrence count and estimated copy-number. Each segment of a TI thread with a unique alignment in GRCh38 is annotated with its occurrence count in the TI thread and the read-depth based estimated copy-number in the matched short-read illumina data. Overall correlation is 0.99 (Pearson) and 0.85 (Spearman). Related to section “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread", related to Figure 2.



chr4:168,398,369-168,398,412 (+) GCTGTTTCCAGACATTGAGTGCCAGTGAGCATGGGAGGGAggccaaagtgggggctaaga
TI-thread breakpoint sequence GCTGTTTCCAGACATTGAGTGCCAGTGAGCATGGGAGGGAGGCCCCACCACCCTGGAGCCTCTCTGTGCCCAAGTTTTAA
chr7:7,805,260-7,805,299 (+) tagaagaactattgatacctGGCCCCACCACCCTGGAGCCTCTCTGTGCCCAAGTTTTAA

Figure S16. Micro-homology and micro-insertion breakpoint junction analysis. Breakpoint microhomology length is plotted on the negative scale, microinsertions as positive lengths. The bottom panel shows a characteristic micro-homology length of 4 (GGCC) for a junction that occurs 58 times in the TI thread. Related to section “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread", related to Figure 2.

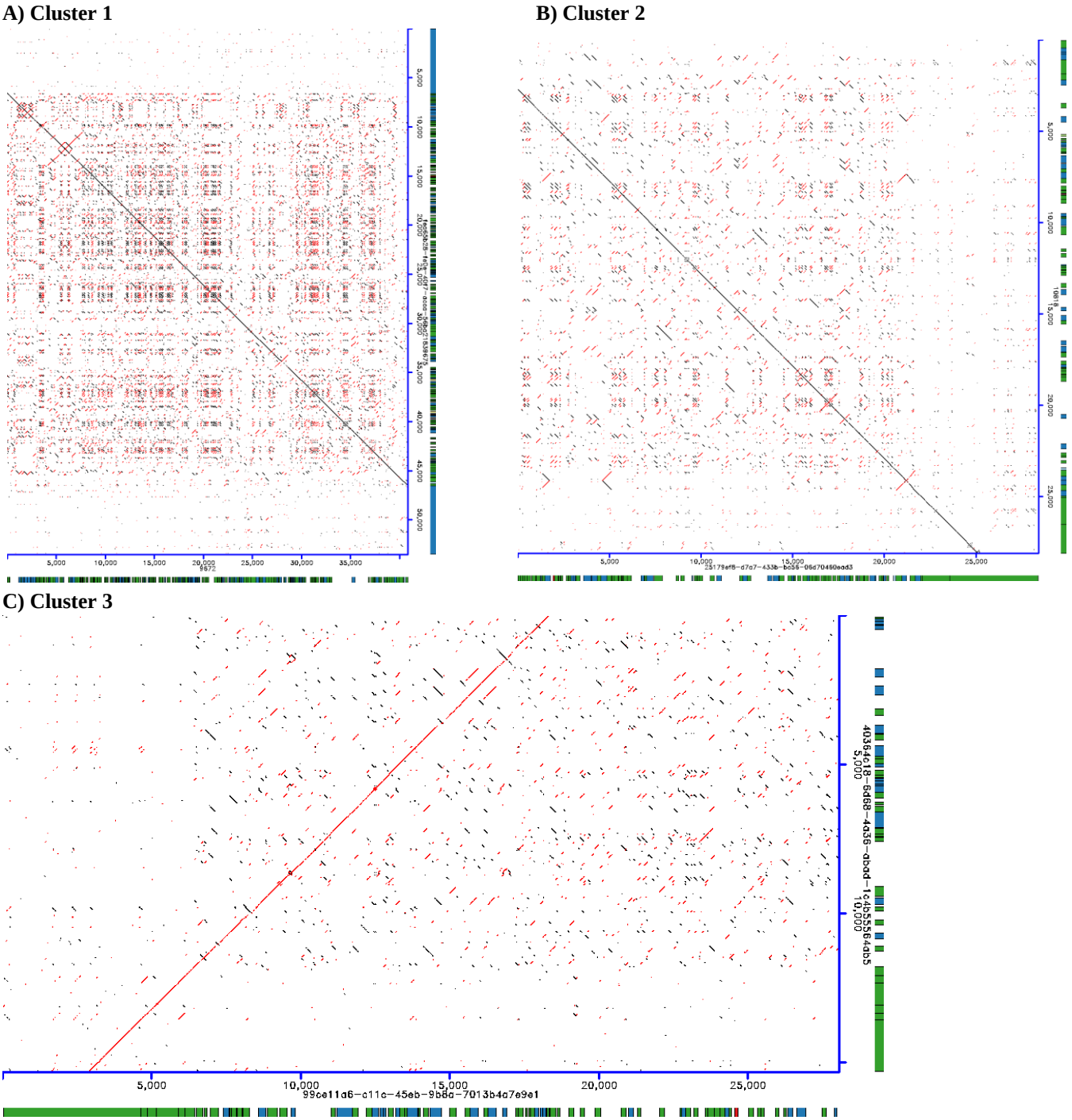


Figure S17. Tumor heterogeneity of TI threads. Templated insertion threads show signals of tumor heterogeneity. **(A)** Pairwise alignment of a contig (9672) from the Shasta[5] assembly to a raw ONT read supporting the contig. Forward matches are in black, reverse matches in red. The plot margins show alignment matches to GRCh38 with different colors indicating different chromosomes. **(B)** Pairwise alignment of a contig (10818) from the Shasta assembly to a raw ONT read supporting the contig. **(C)** Pairwise alignment of two ONT reads supporting a third configuration of templated insertions that was not assembled by Shasta. Alignment matches in the plot margins are interspersed with unmapped sequences, suggesting non-templated DNA synthesis or very short templated sequences that cannot be mapped unambiguously in GRCh38. Related to section “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread", related to Figure 2.

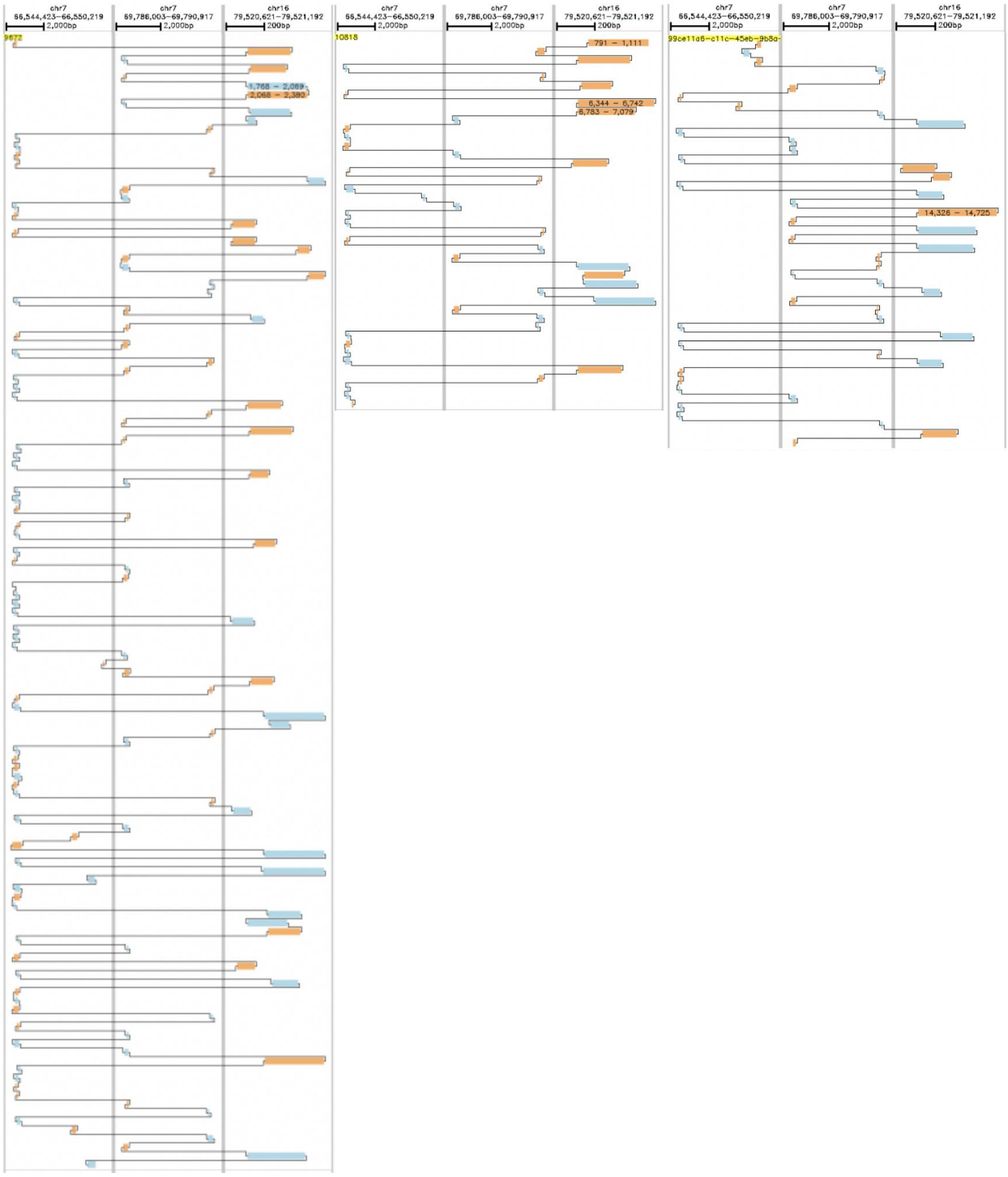


Figure S18. Different architectures of TI threads share common source segments. The 3 clusters from Figure S17 are aligned side-by-side from left to right: Shasta contig 9672 (left panel), Shasta contig 10818 (middle panel) and the additional raw ONT read (right panel). Each panel is subdivided into 3 alignment regions representing the 3 most common TI source segments occurring in all 3 clusters. Each contig or read supports a different TI thread architecture with apparent large differences of how often the same TI source segments are concatenated and in what order and orientation. Related to section “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread", related to Figure 2.

Figure S19. Alignment of raw ONT reads back to the *de novo* assembled TI thread. An alignment of ONT raw reads back to the assembled Shasta contig of the TI thread shown in Figure 2A reveals a subclonal ~9kbp deletion from 33kbp to 42kbp. Related to section “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread", related to Figure 2.

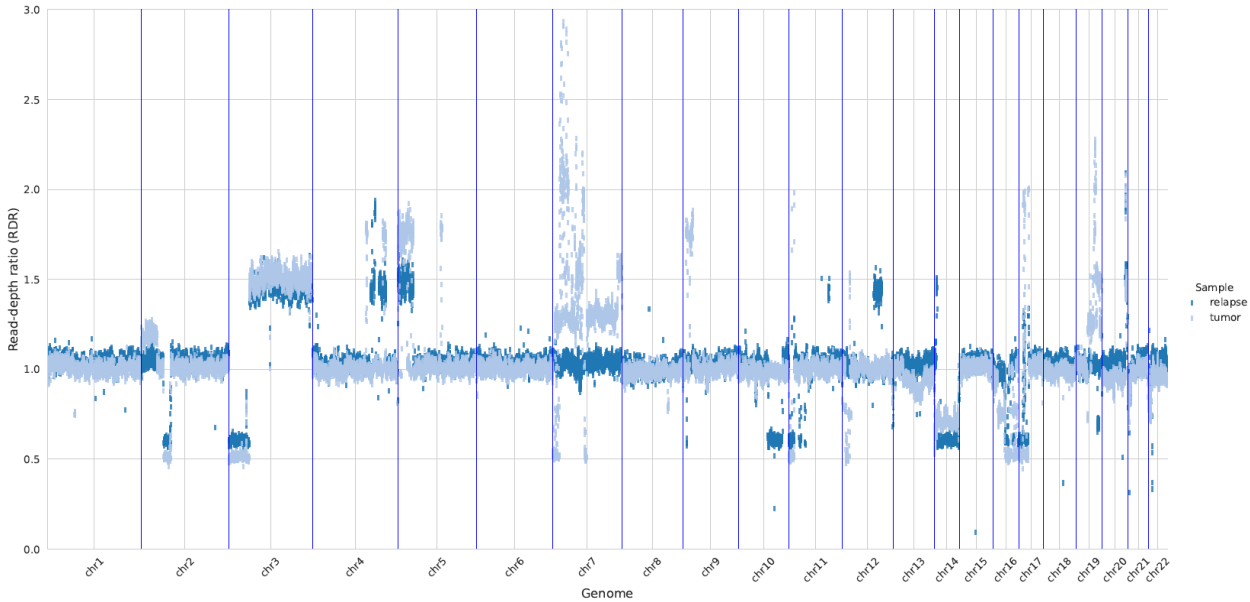


Figure S20. Copy-number plot of the primary tumor and relapse. Read-depth ratio plots of the primary tumor (light blue) and relapse (dark blue) computed by HaTCHet[7] show that many complex rearrangements are lost in the relapse, in particular chromosome 7 and 19 show an overall diploid copy-number. Related to section “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread", related to Figure 2.

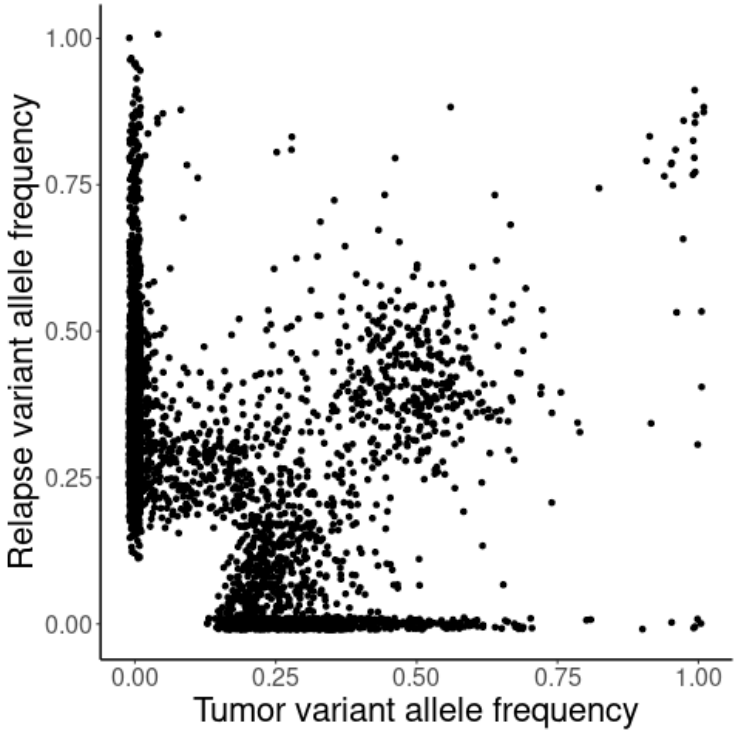


Figure S21. Comparison of the somatic variant allele frequency in the primary and relapse. Each dot is a somatic SNV detected in one of the tumor samples. Placement of dots has been randomized by +/-0.01 in x- and y-direction to better highlight the variants that have been lost (18%) or acquired (48%) in relapse. 34% of all somatic SNVs are shared. Related to section “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread", related to Figure 2.

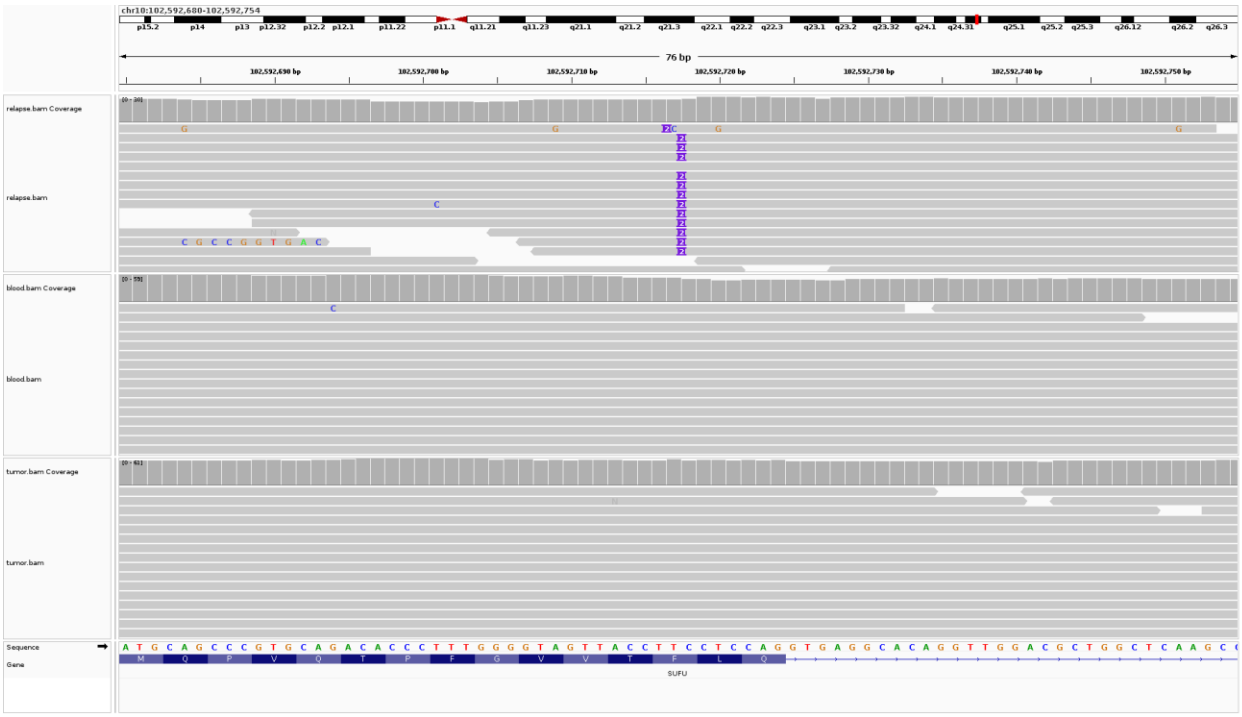


Figure S22. Relapse-specific frameshift insertion in *SUFU*. A homozygous 2bp frameshift insertion in *SUFU* that is present in the relapse sample (top panel) but absent in blood (middle panel) and the primary tumor (bottom panel), visualized using IGV[8]. Related to section “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread", related to Figure 2.

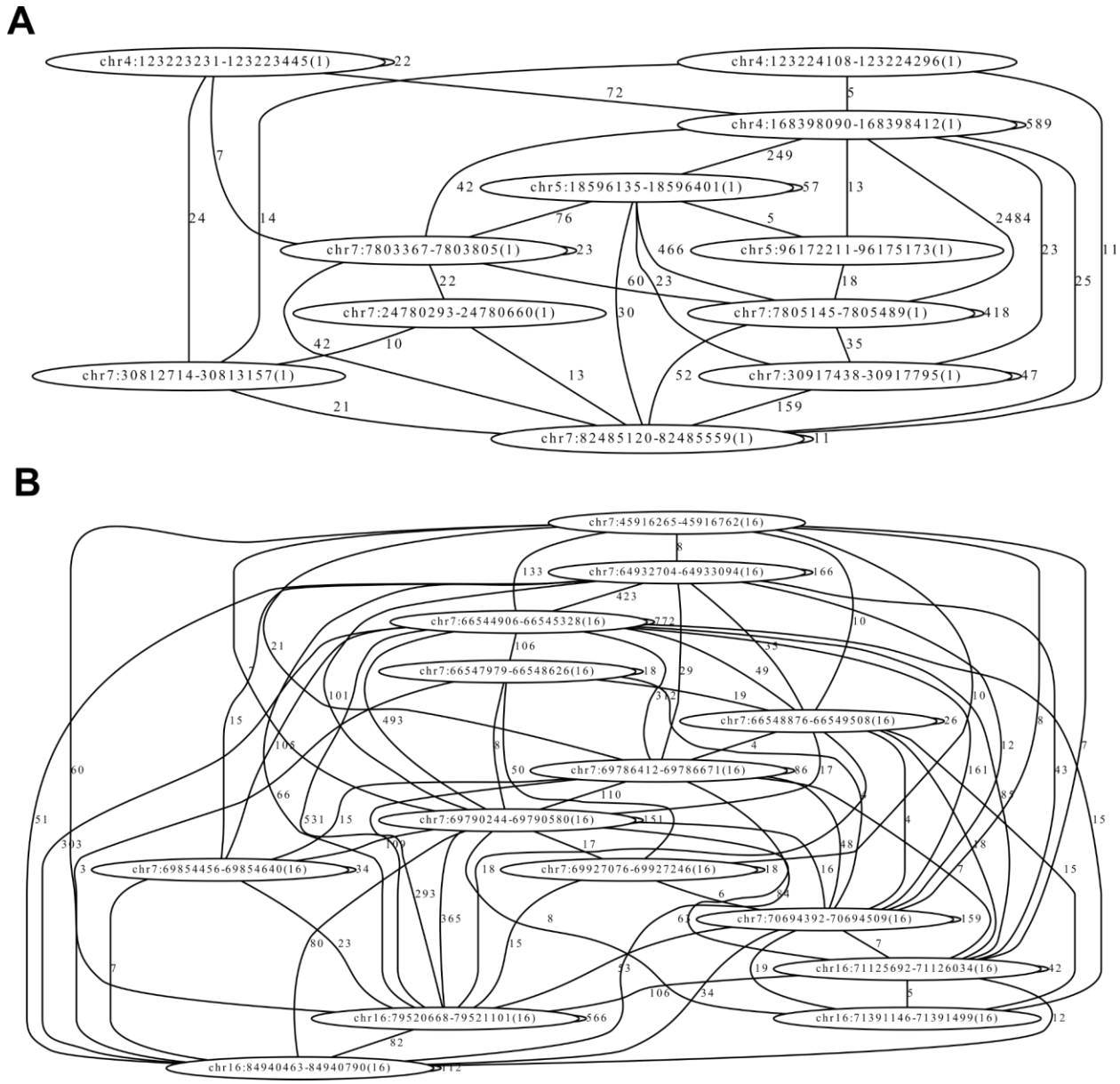


Figure S23. Templated insertion rearrangement graphs. Nodes of the graph represent templated insertion source sequences. Edges of the graph represent split-read and paired-end connections between different templated insertions, including self edges. The edge weight specifies the support for the rearrangement junction. The 1st instance in panel **(A)** corresponds to Figure 2AB and the 2nd instance in panel **(B)** corresponds to Figure S8 and S9. Related to section “Graph-based discovery of TI threads in Illumina WGS data", related to STAR Methods.

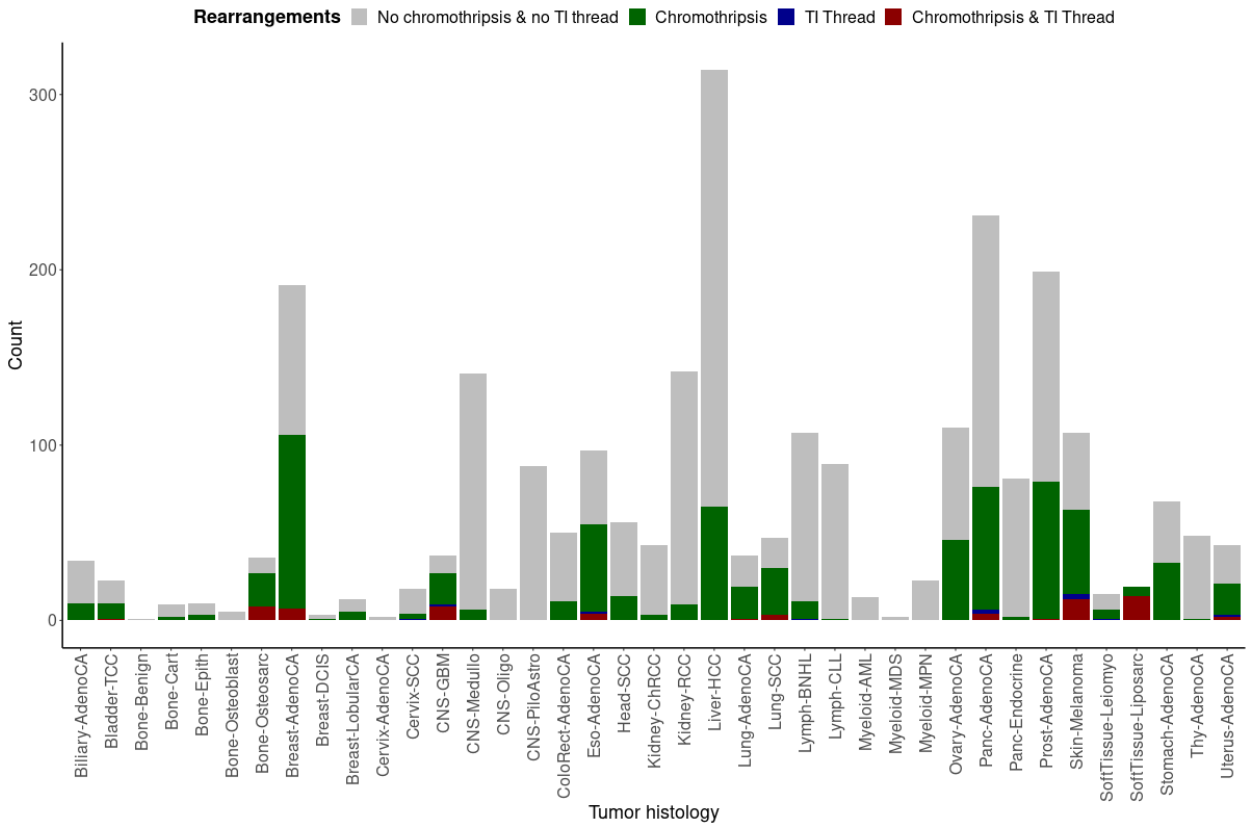


Figure S24. Templated insertion threads across 2,569 cancer genomes. A stacked histogram of the distribution of templated insertion threads (TI Threads) and chromothripsis across all PCAWG samples stratified by tumor histology. Related to section: “Pan-cancer landscape of TI threads”, Related to Figure 2.

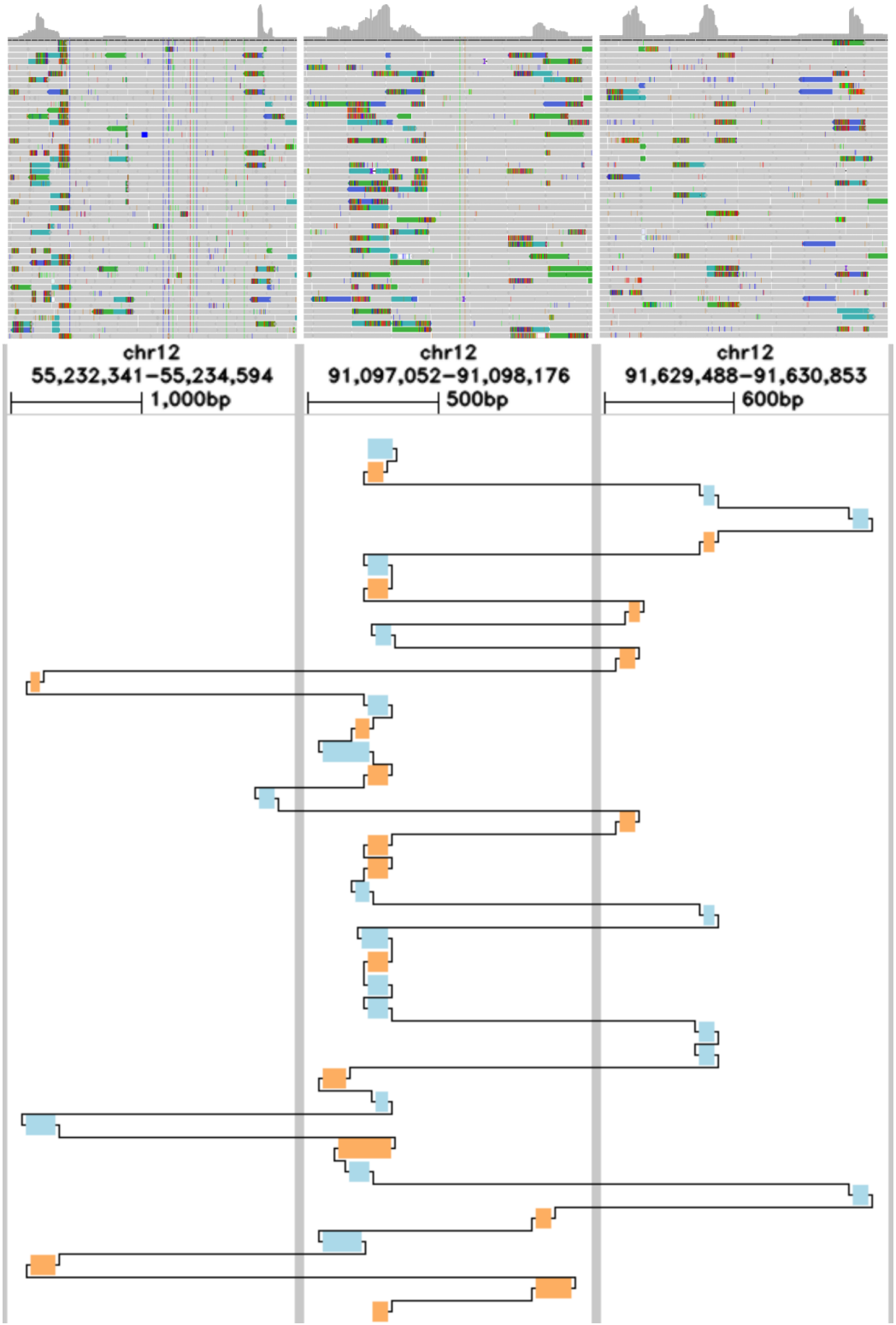


Figure S25. TI thread in a primary dedifferentiated liposarcoma sample P1. A chained alignment view for a raw ONT read supporting a TI thread (bottom panel) in P1. Alignment matches to GRCh38 are colored in blue (forward) or ochre (reverse). Aligned segments show strong coverage increases in the matched illumina short-read data (top panel) as well as SV-supporting split-reads with soft-clips. Related to section: “Pan-cancer landscape of TI threads”, Related to Figure 2.

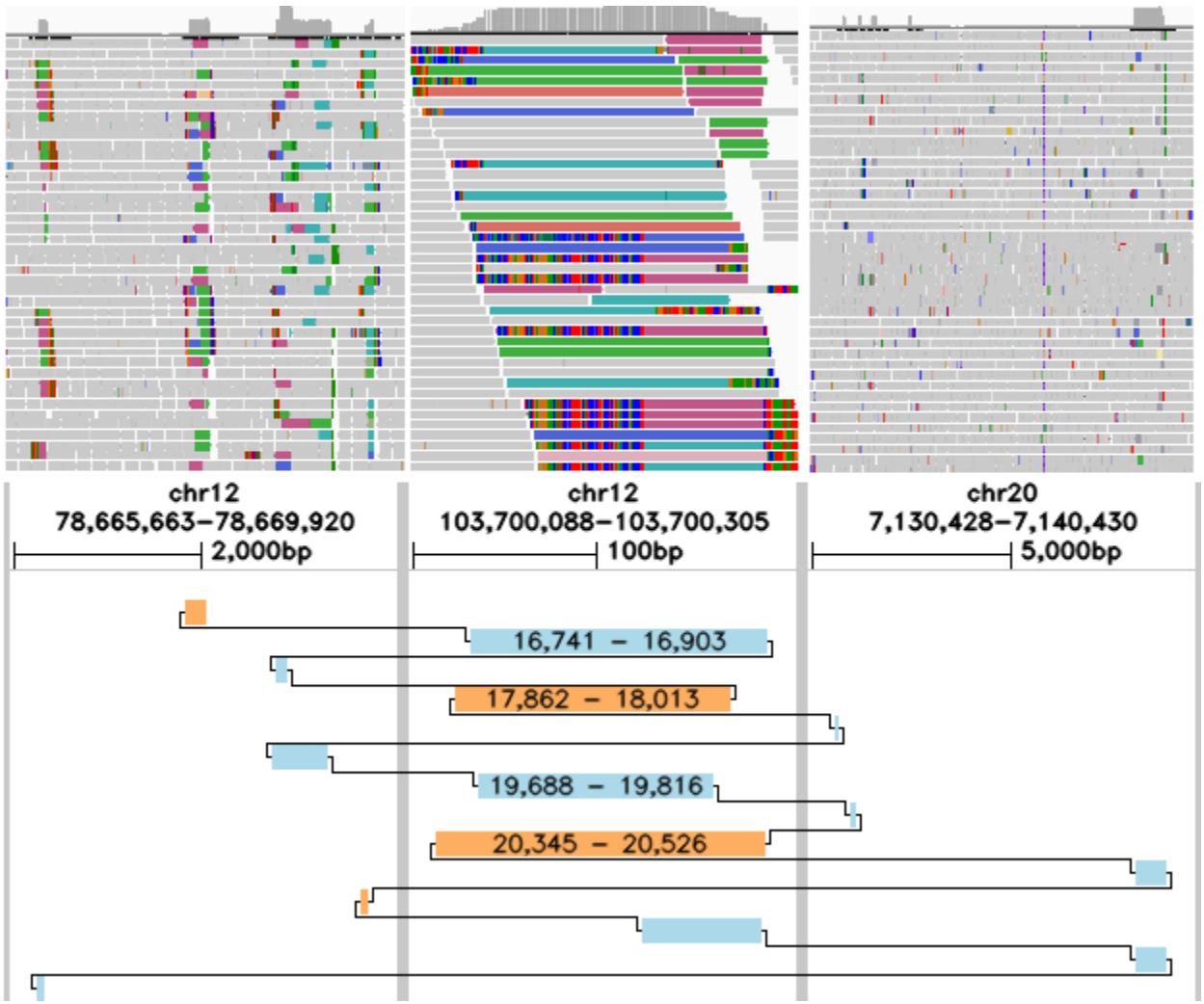
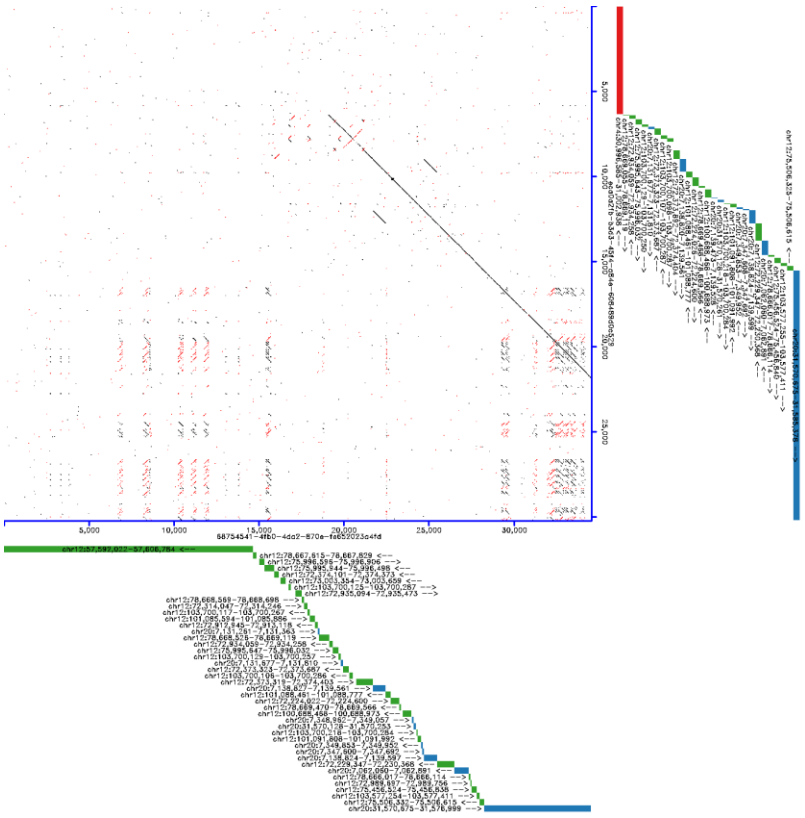


Figure S26. TI thread in liposarcoma metastasis sample P2. A chained alignment view for a raw ONT read supporting a TI thread (bottom panel) in P2. Alignment matches to GRCh38 are colored in blue (forward) or ochre (reverse). Only regions with at least 4 matches are shown and aligned segments show strong coverage increases in the matched illumina short-read data (top panel) as well as SV-supporting split-reads with soft-clips. Related to section: “Pan-cancer landscape of TI threads”, Related to Figure 2.

A)



B)

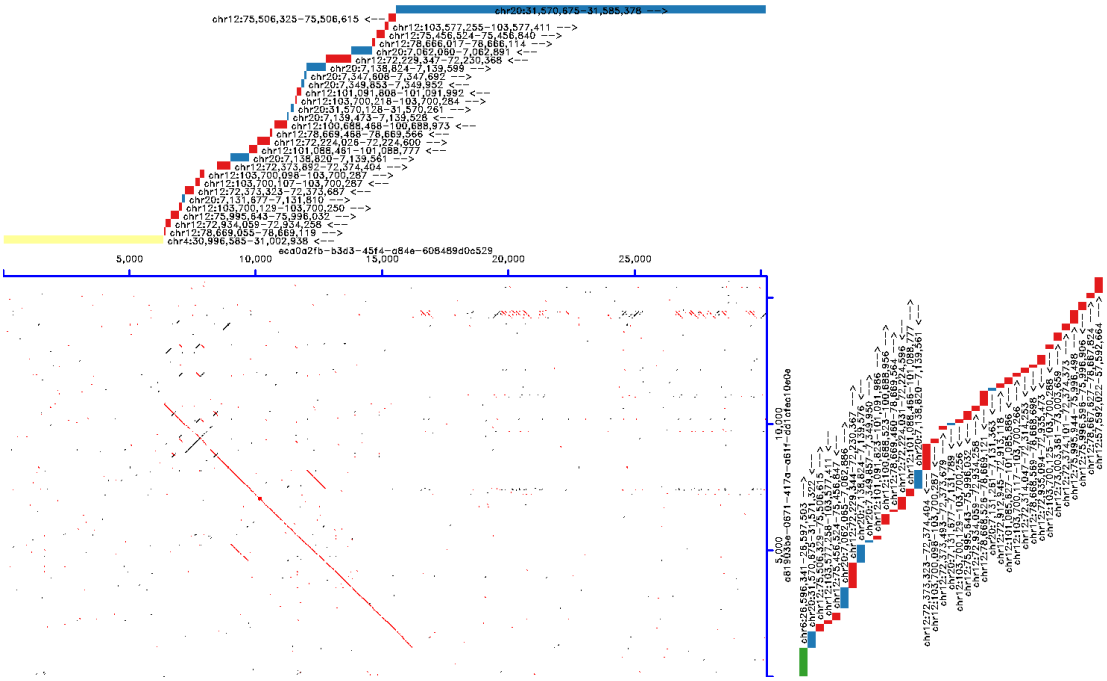


Figure S27. Multiple integration sites of a TI thread structure in P2. A pair-wise dot plot of ONT reads supporting a TI thread reveals different adjacent genomic segments in panel A and B. The dot plots show forward matches in black and reverse matches in red. Matches to GRCh38 are highlighted in the plot margins and colored by the source chromosome. Arrows indicate forward or reverse matches to GRCh38. Related to section: “Pan-cancer landscape of TI threads”, Related to Figure 2.

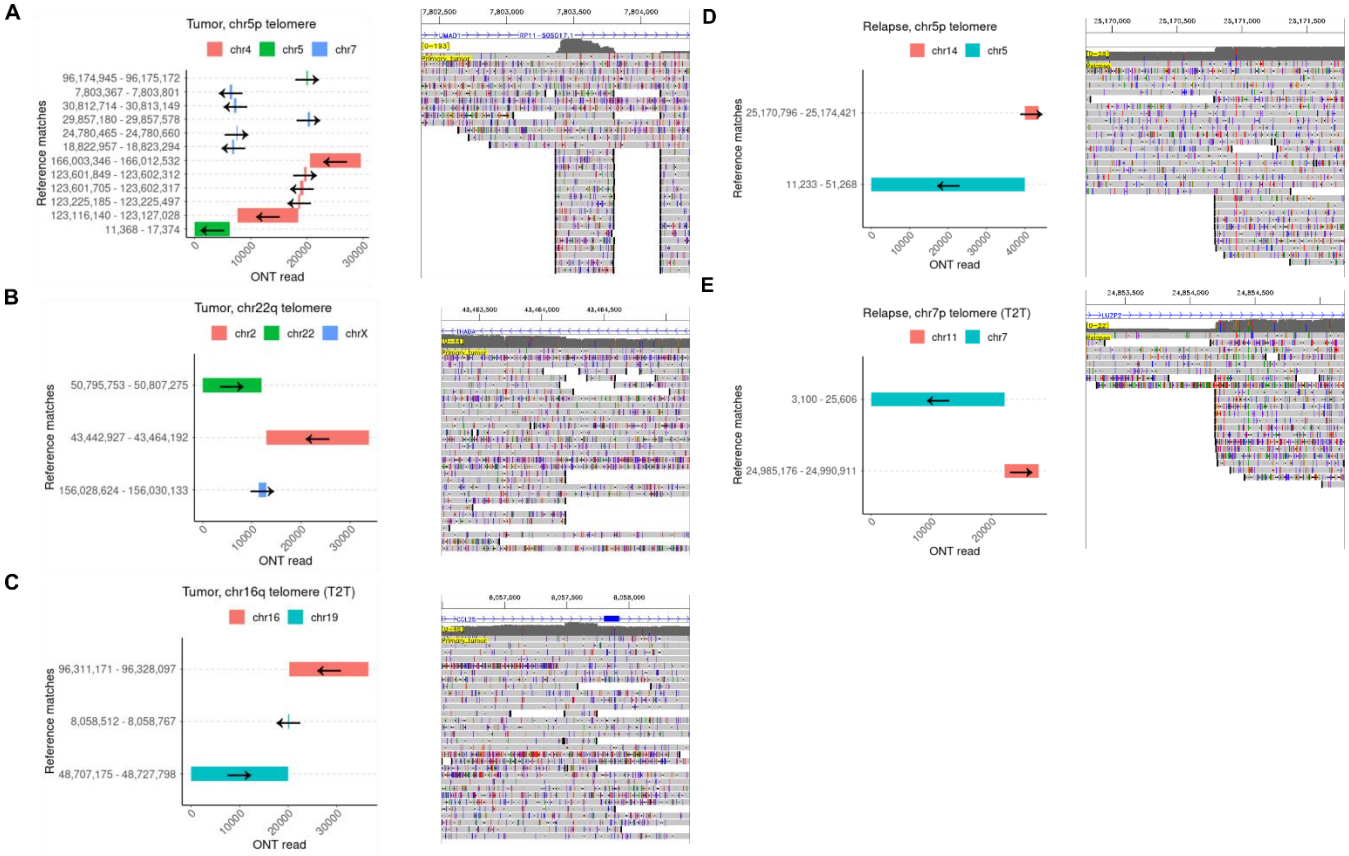
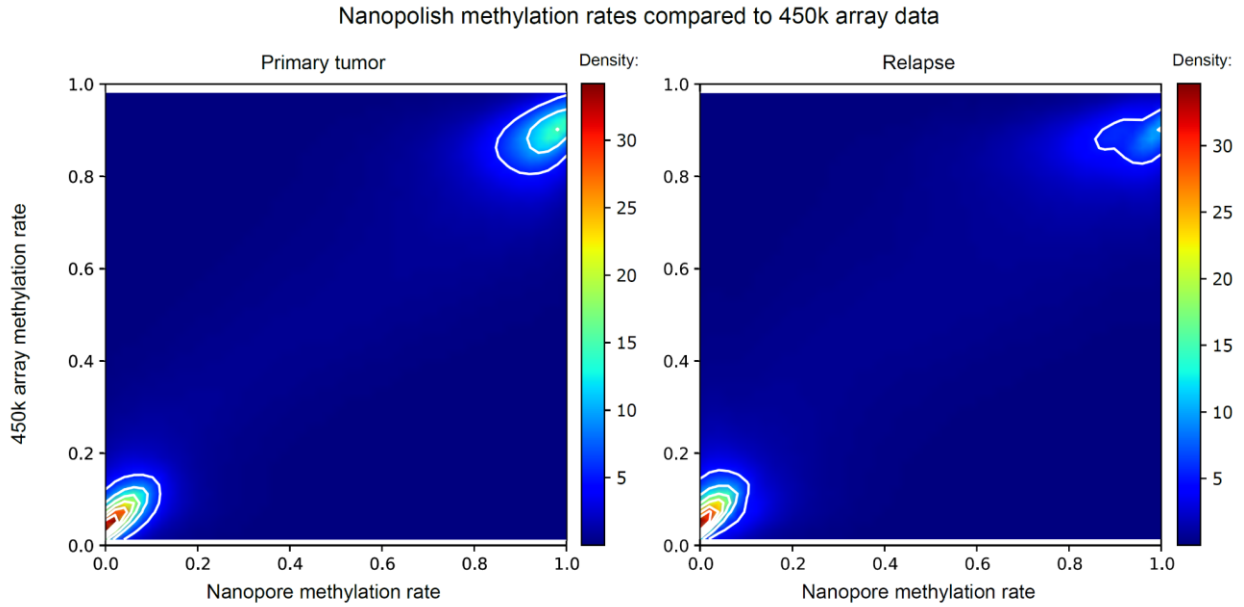
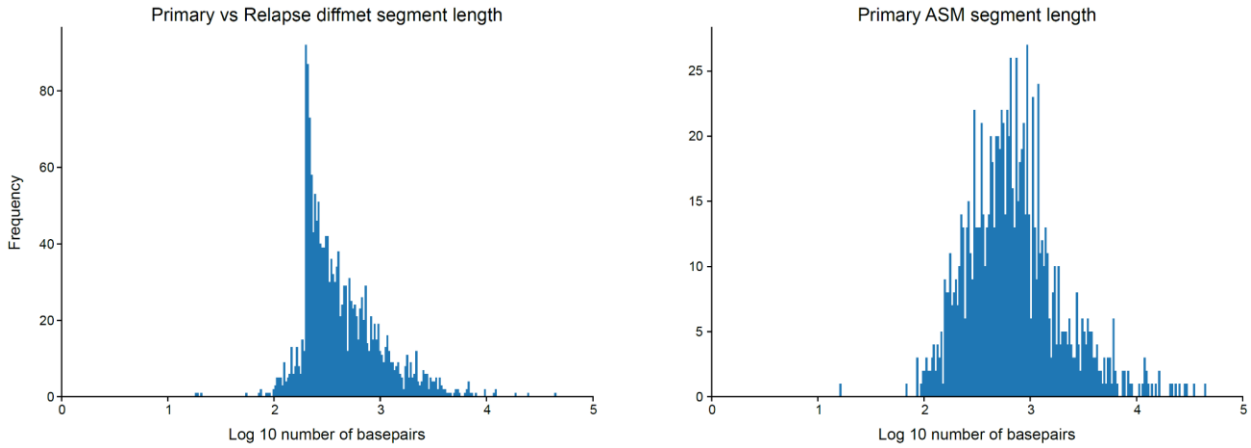


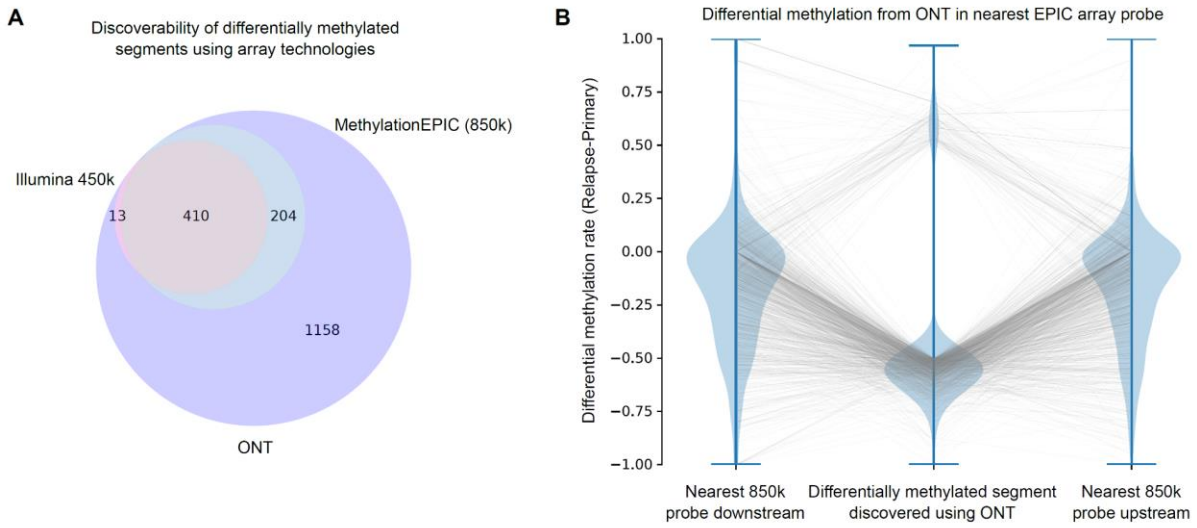
Figure S28. Telomere sequences associated with rearranged genomic regions. Panels (**A-E**) show five telomere sequences associated with SVs, three in the primary tumor (**A-C**) and two in relapse (**D-E**), using a representative ONT read (x-axis) and the genomic mapping locations (y-axis) colored by chromosome (left panels). Coordinates are in GRCh38, except for rearrangements mapped to the telomere to telomere assembly (T2T) to resolve alignment ambiguities (panel C and E). The panels on the right show a selected non-telomeric SV breakpoint spanned by the respective read from the left panel. Some of these breakpoints are within an intron of a gene (panel A, B, C, and E) or show alignment signals characteristic for templated insertions (panel A and C). Related to section: “Telomere analysis of derivative chromosomal segments”, Related to Figure 2 and STAR Methods.



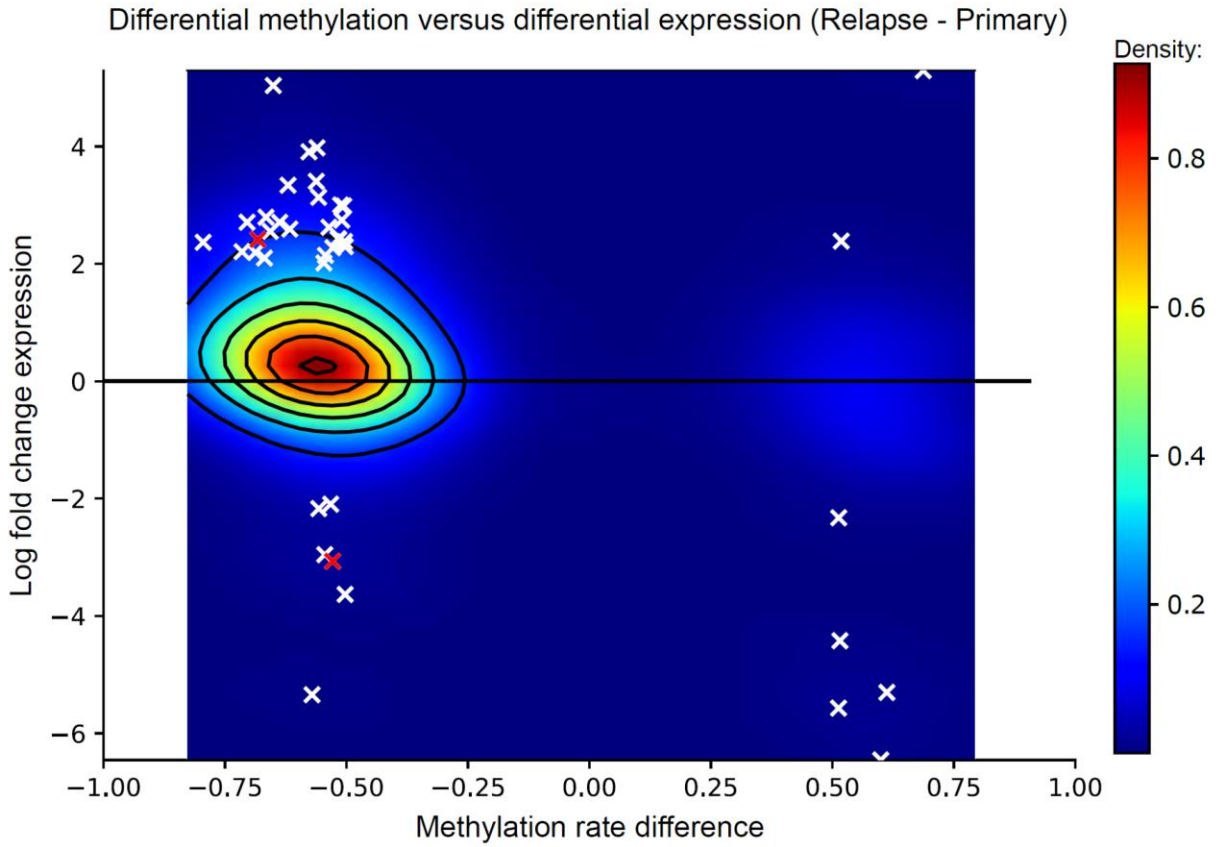
Supplementary Figure S29. Validation of nanopolish methylation calls using Illumina 450k array data. Plotting each array probe's methylation rate against the mean methylation rate in the matching region from nanopolish methylation calls (log-likelihood ratio threshold 2.0). Pearson correlation r^2=0.9111 for primary tumor and r^2=0.8500 for the relapse sample. Related to section: "Differential methylation from long-read data", Related to Figure 3 and STAR Methods.



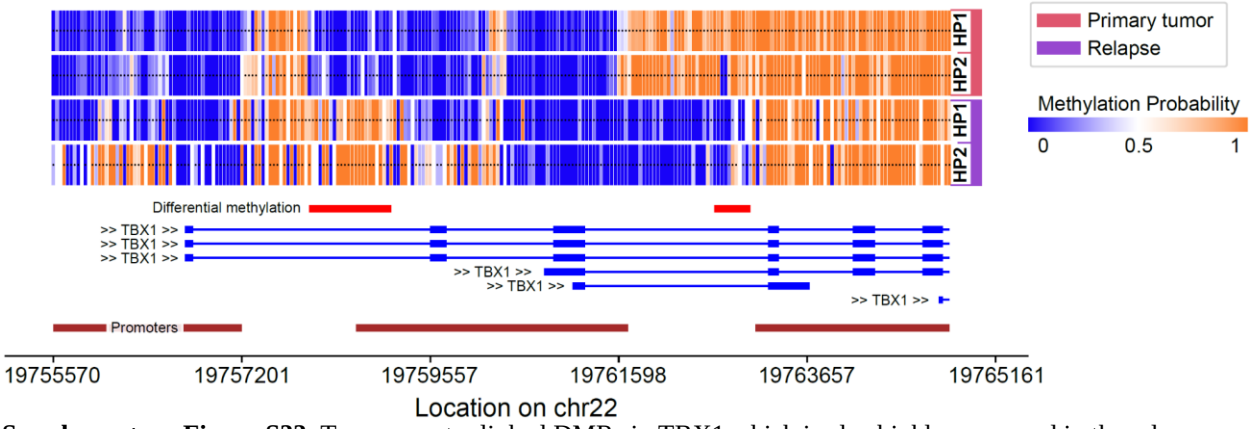
Supplementary Figure S30. Length of differentially methylated segments (A) in the primary tumor vs relapse sample comparison and (B) allele specific methylation analysis in the primary tumor sample. Related to section: "Differential methylation from long-read data", Related to Figure 3.



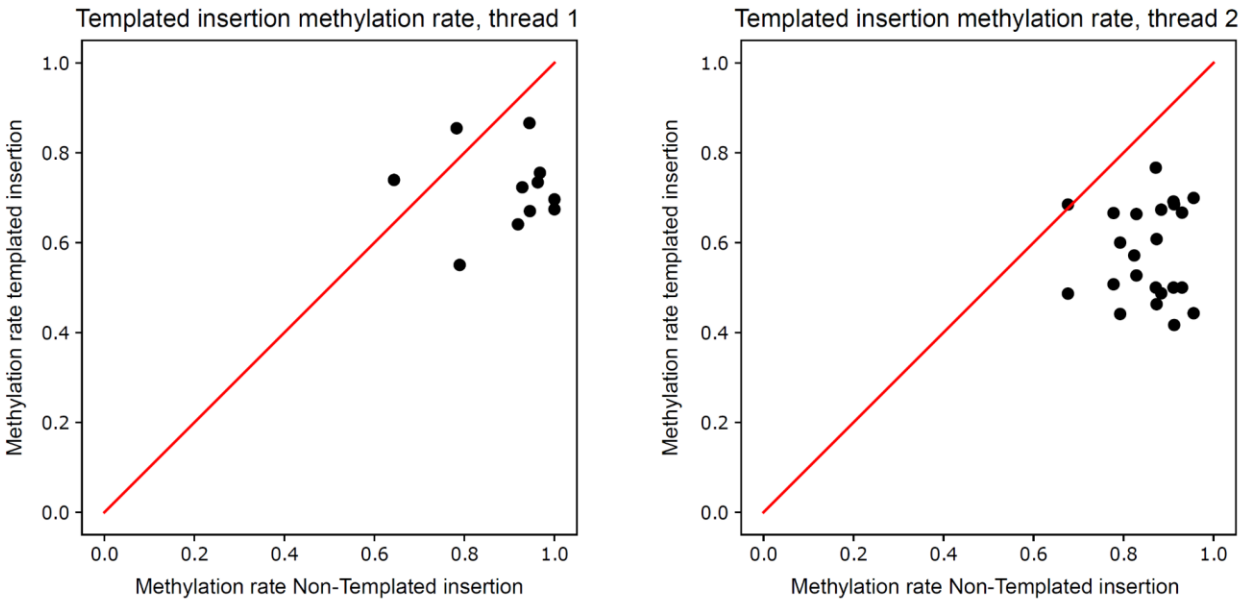
Supplementary Figure S31. Differentially methylated regions between primary tumor and relapse sample discovered from ONT (abs methylation rate difference > 0.5) have been tested for discoverability using Illumina 450k array and MethylationEPIC (850k) array technologies. A) Counting the number of segments for which an array probe at least partially overlapped a segment found as differentially methylated in ONT. B) For the 1158 segments not covered by a probe in the MethylationEPIC array, we investigate the methylation as measured from ONT at the location of the nearest array probes upstream and downstream from the differentially methylated segment. Related to section: “Differential methylation from long-read data”, Related to Figure 3 and STAR Methods.



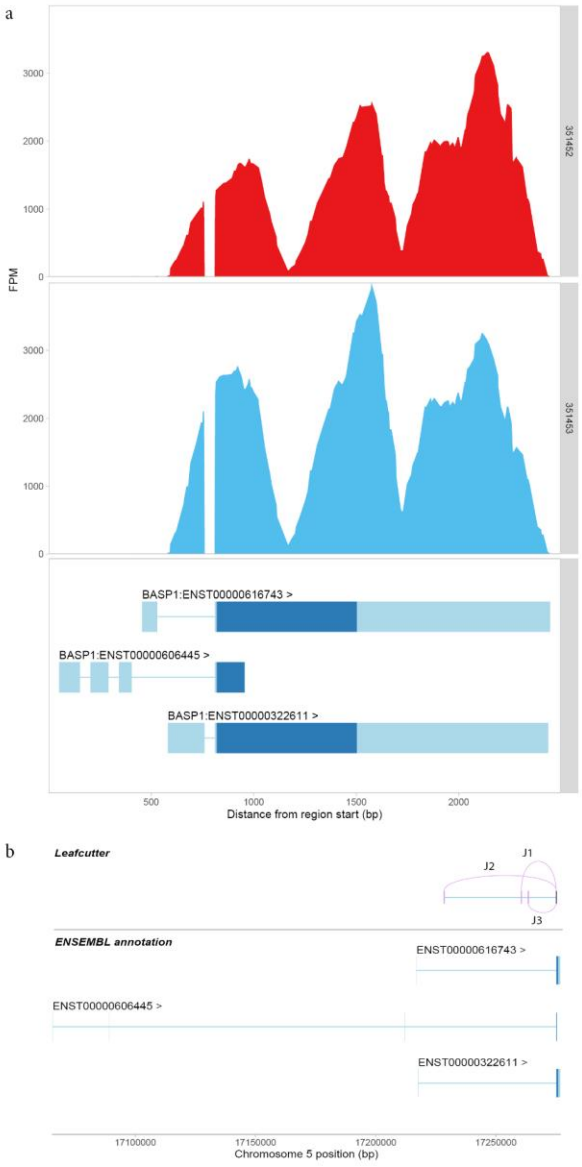
Supplementary Figure S32. Differential promoter methylation between primary tumor and relapse. Crosses represent genes which also have high expression change (log2-fc > 2). The two red crosses are genes which also have strong copy number changes (higher copy number in primary). Related to section: "Differential methylation from long-read data", Related to Figure 3.



Supplementary Figure S33. Two promoter linked DMRs in TBX1 which is also highly expressed in the relapse sample. The DMR demethylated in primary tumor precedes the short form transcript TBX1-206, whereas the DMR highly methylated in primary tumor directly follows the transcription start site of said transcript. Related to section: "Differential methylation from long-read data", Related to Figure 3.



Supplementary Figure S34. A number of high coverage templated insertions from two templated insertion threads. Methylation rates from templated insertions are plotted against methylation rates from the unmodified haplotype at the insertions location of origin. A reduction in methylation rate (by about 0.18 in thread 1 and about 0.28 in thread 2) can be observed in templated insertions. Related to sections: “Functional annotation of the TI threads and telomere SVs” & “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread”, Related to Figure 3 & Figure 4.



Supplementary Figure S35. Overview figure of the expression changes in the BASP1 gene. a) Expression levels, features per million, of BASP1 in primary (top) and relapse (middle) and the protein coding transcript derived from Ensembl (bottom), scaled to exon lenghts. b) Main differentially splicing events in BASP1. In the bottom the three protein coding isoforms are shown, based on Ensembl annotation, on top the three differentially spliced junctions are illustrated, read count information can be found in supplementary table S10. All leaf-cutter junctions are indicative of non-annotated exons, as none are overlapping with the three protein coding transcripts. In the relapse sample junction 1 and junction 3 (J1 & J3, totaling 3% of the reads) are used, while J2 is not used (0% junction reads), while in the primary tumor it is the other way around (J2 corresponds to 1% of the junction reads, J1 and J3 0%). Related to sections: “Functional annotation of the TI threads and telomere SVs” & “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread”, Related to Figure 4.

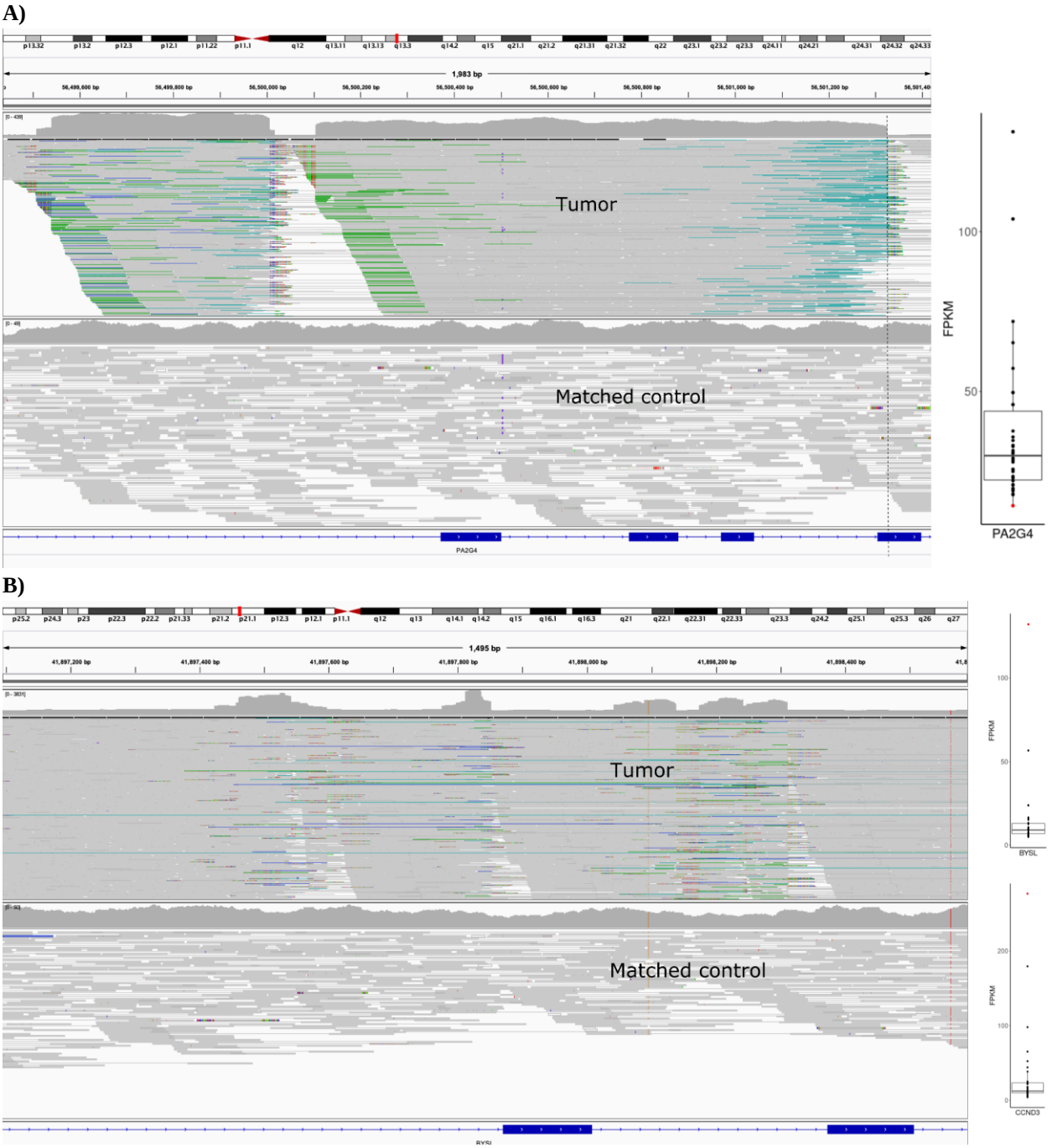
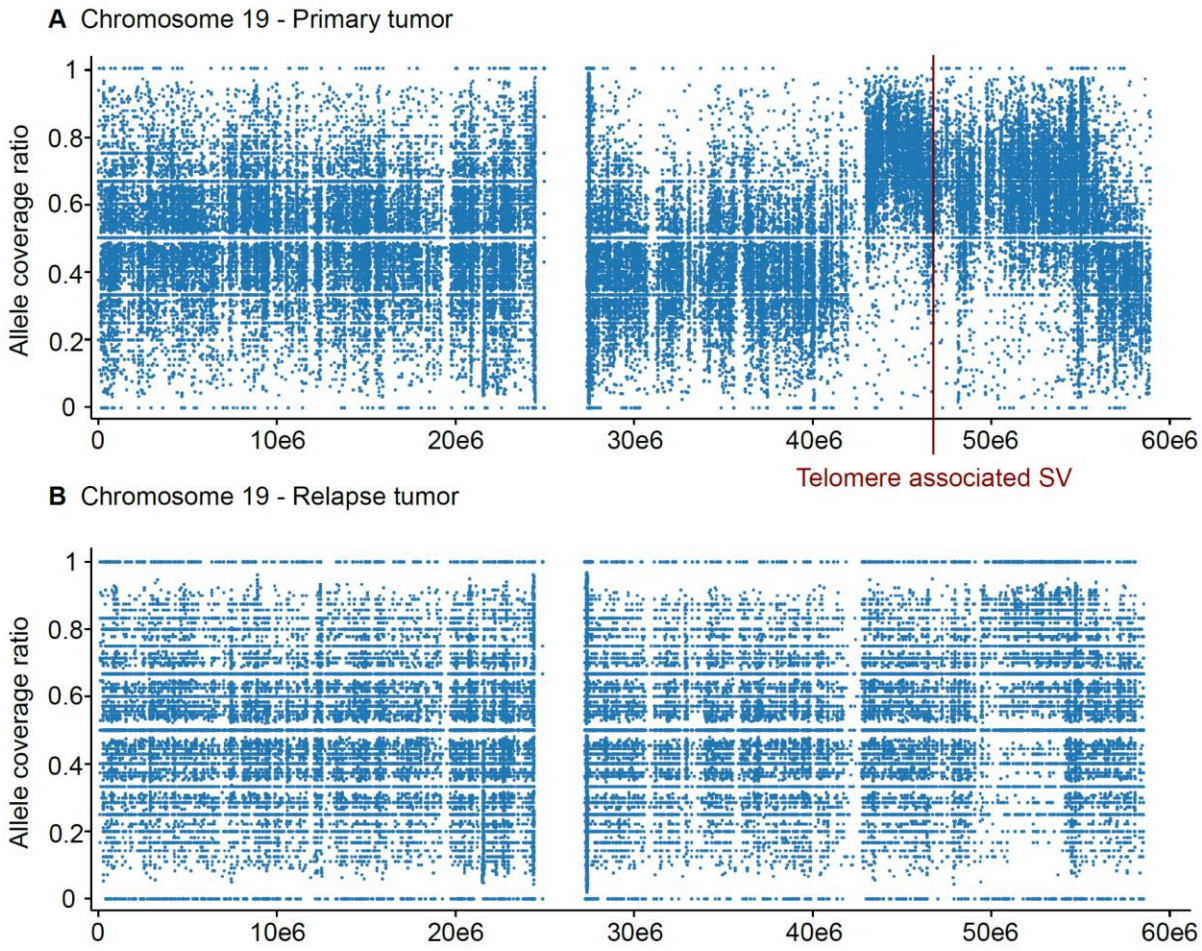


Figure S36. Gene expression effects of templated insertion threads. (A) Example of a templated insertion thread identified by *rayas* that intersects a coding exon of PA2G4. The boxplot on the right shows the FPKM distribution of all sarcoma samples in PCAWG with the tumor shown on the left having the lowest expression among all samples, likely due to the exon intersecting templated insertion thread. **(B)** Overexpression of BYSL and CCND3 relative to other sarcoma samples (right boxplots) and the alignment view of a templated insertion thread overlapping multiple introns of BYSL (left alignment view). Related to sections: “Functional annotation of the TI threads and telomere SVs” & “ONT sequencing reveals a novel complex rearrangement pattern denoted TI thread”, Related to Figure 4.



Supplementary Figure S37. Allele coverage ratio from WhatsHap phased ONT reads of chromosome 19 in primary (**A**) and relapse (**B**) tumor. The telomere associated SV observed in primary tumor coincides with a haplotype specific amplification of chromosome 19q. Related to STAR Methods.

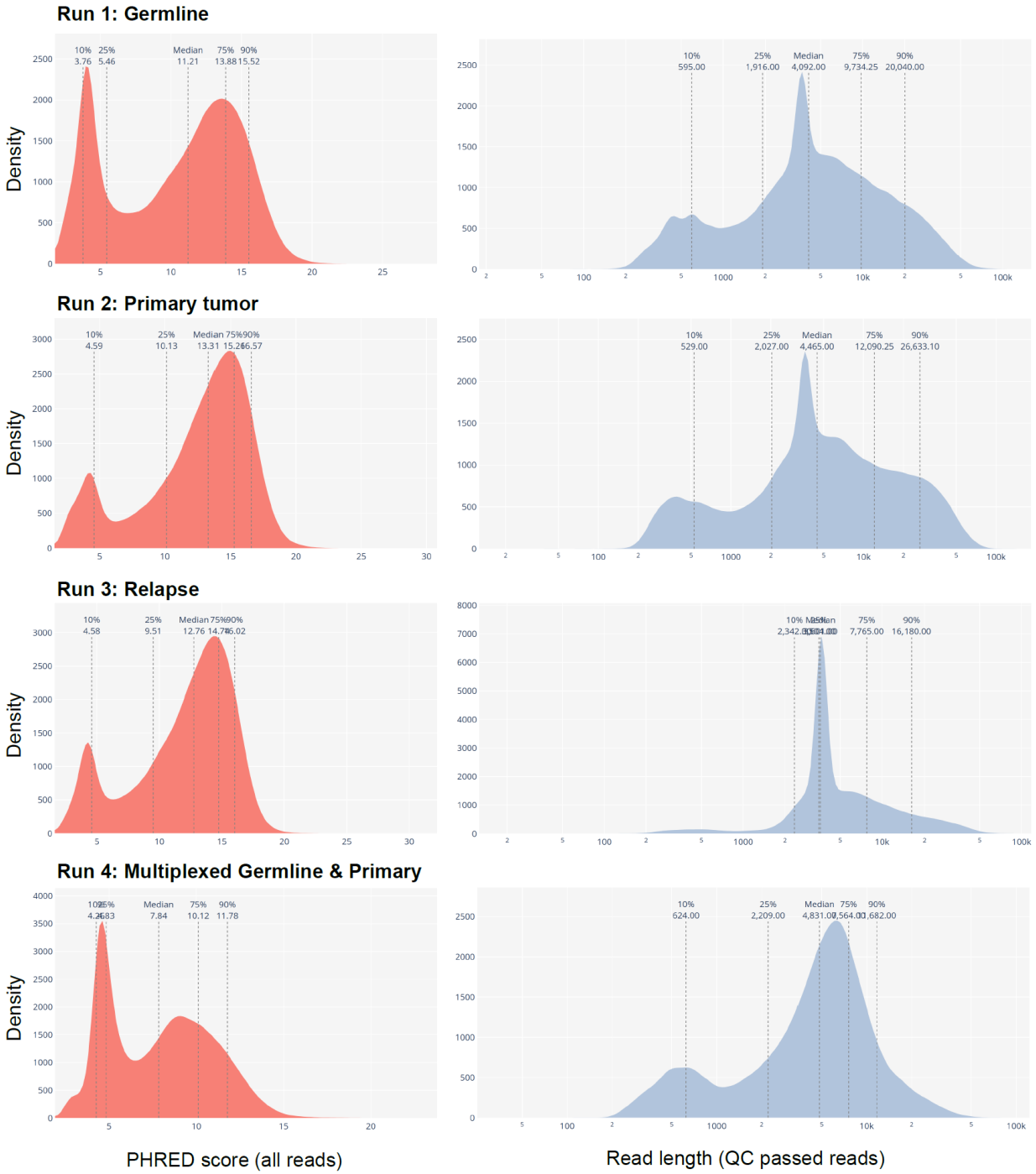


Figure S38. PHRED quality score and read length histograms. PHRED quality scores as reported by guppy 6.1.7 (left panel) and read length histograms after QC filtering of reads with average PHRED score of 7 or higher (right panel). Related to STAR Methods.

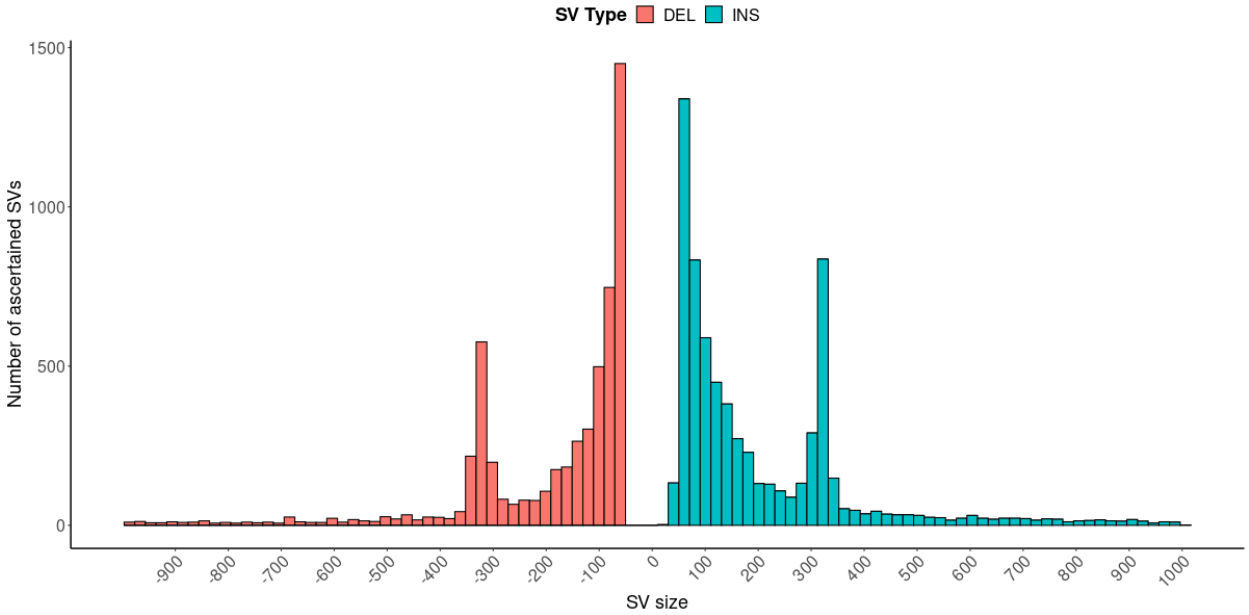


Figure S39. Germline structural variant size distribution. Germline structural variants >50bp binned by size and type. Deletions are shown to the left (negative size) and insertions to the right (positive size). Insertions and deletions show the characteristic ALU peak at approximately 300bp, indicative of transposable element insertions and deletions relative to the reference. Related to STAR Methods.

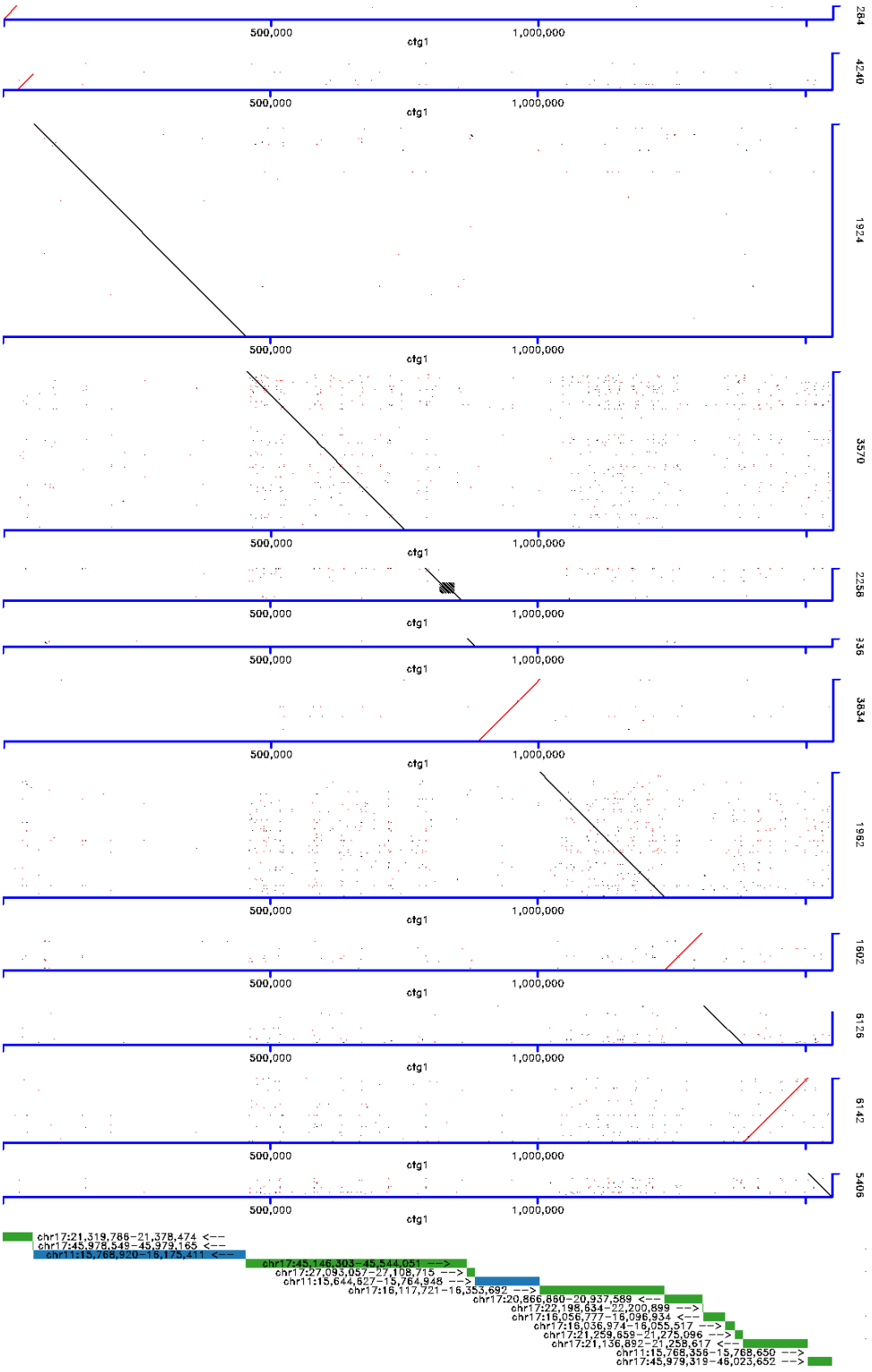


Figure S40. Comparison of targeted and *de novo* assembly. Contigs from a *de novo* Shasta assembly[5] (y-axis) aligned to contig 1 (x-axis) of the targeted CS11-17 assembly. Multiple contigs fully overlap the CS11-17 assembly but the frequent breaks illustrate the need for targeted haplotype-resolved assembly approaches in cancer genomics. Colored bars at the bottom indicate unique segment alignments to GRCh38 with different colors representing different chromosomes from which the genomic segments were derived. Related to STAR Methods.

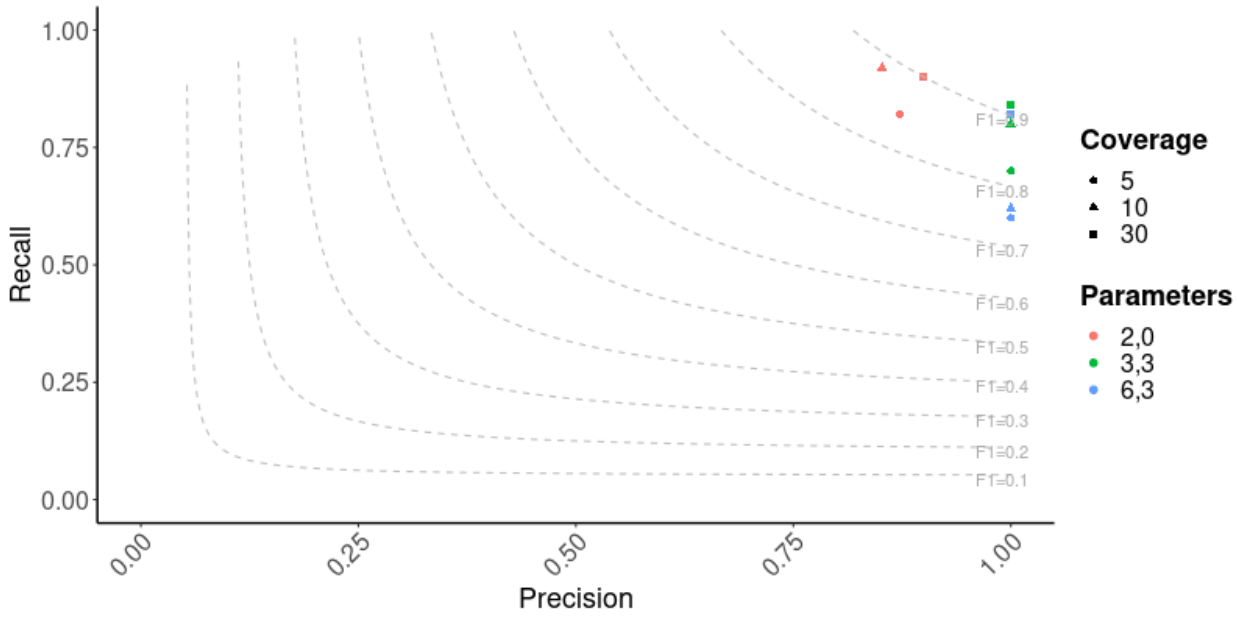


Figure S41. Short-read simulation experiment for TI threads. We simulated 10 TI threads on chromosome 18 at coverages 5x, 10x and 30x with 5 randomly sampled source segments of a size smaller than 1kbp that were concatenated and copied in random order to a TI thread with 50 segments. We then applied rayas using a range of parameters denoted as (A,B) in the legend that control the required increase in coverage (A) and the minimum number of required split-reads at the boundary of a segment (B). The default of rayas uses A=3 and B=3 (green plotting symbols). Related to STAR Methods.

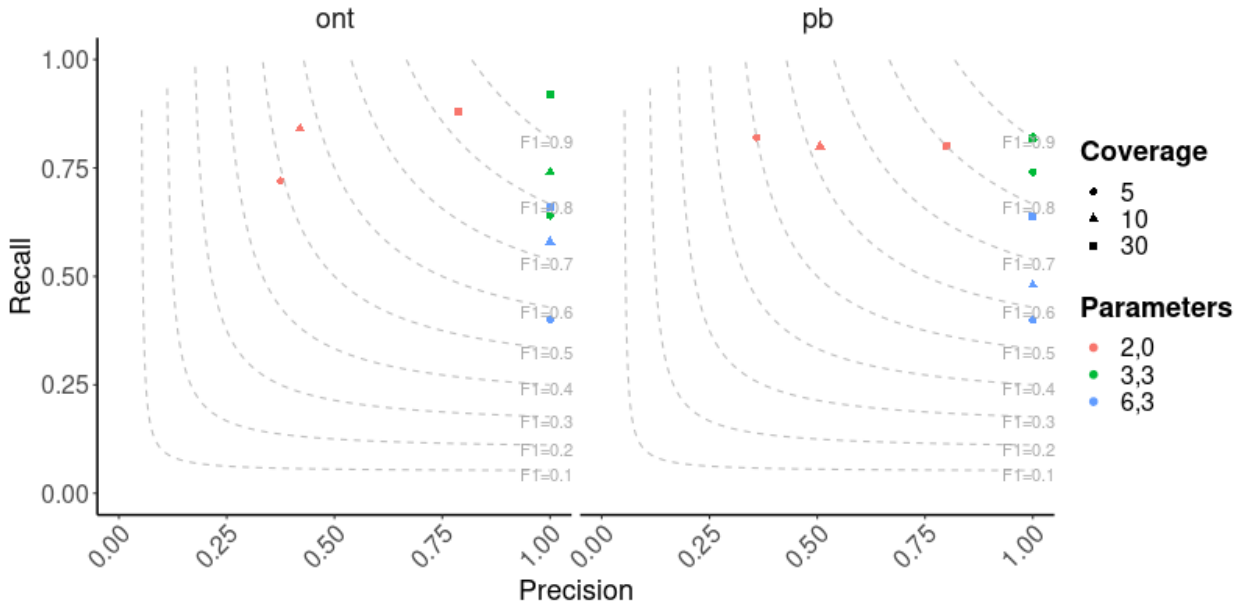


Figure S42. Long-read simulation experiment for TI threads. We simulated 10 TI threads on chr18 at 5x, 10x and 30x coverage, each with 5 randomly sampled source segments of a size smaller than 1kbp that were then concatenated and copied in random order to a final TI thread with 50 segments. We then applied lorax using a range of parameters, denoted as (A,B) in the legend. These parameters control the required increase in coverage (A) and the minimum number of required split-reads at the boundary of a segment (B). The default of lorax uses A=3 and B=3 (green plotting symbols). Simulated data with a typical error profile for ONT is on the left and simulated PacBio data on the right. Related to STAR Methods.

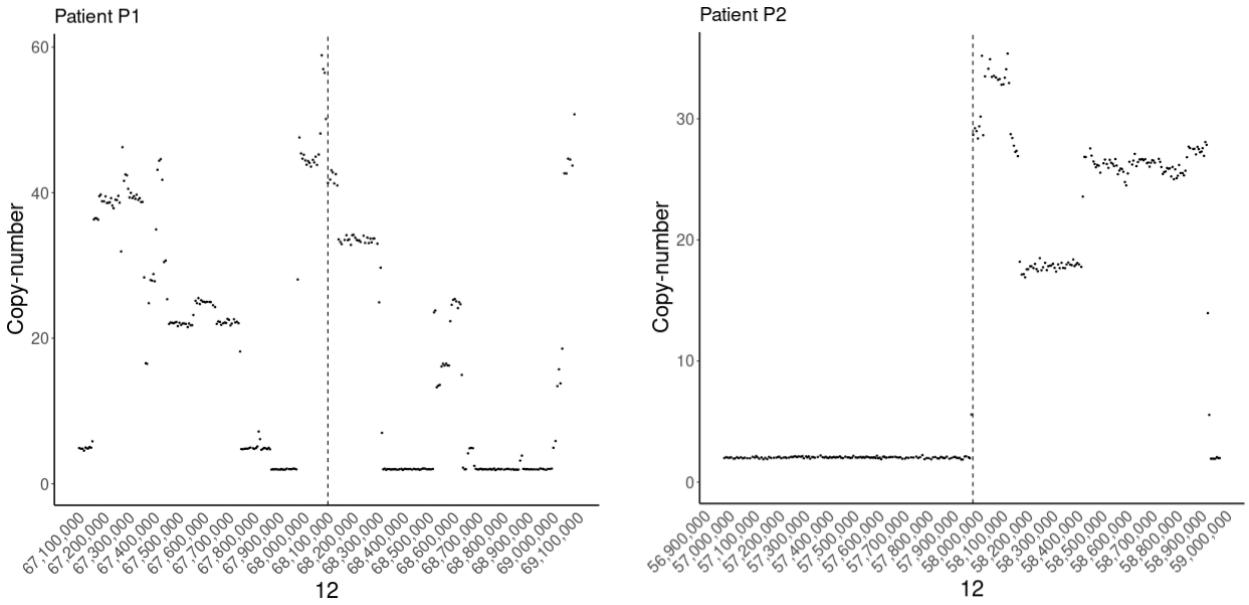


Figure S43. TI thread integration sites for patient P1 and P2. TI threads (dashed vertical line) colocalize with rearrangement breakpoints on chr12 where copy number changes occur. Coordinates are in GRCh37 and patient P1 is shown in the left panel and patient P2 in the right panel. Related to STAR Methods.

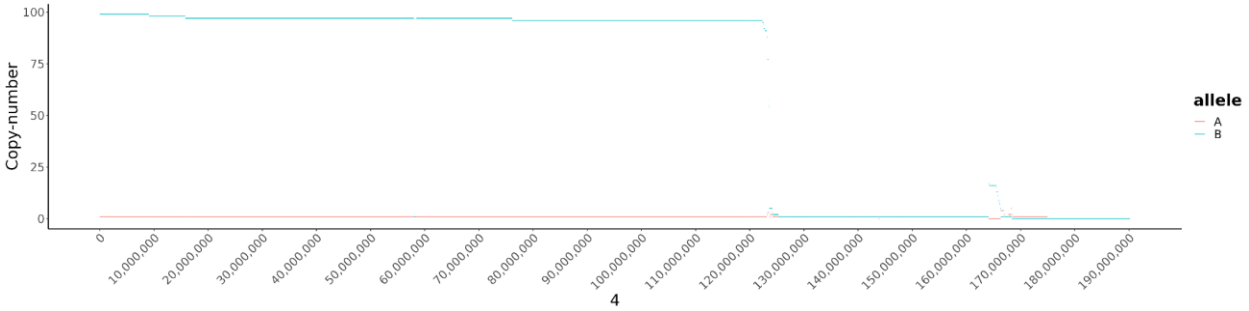


Figure S44. Copy-number reconstruction of chromosome 4 using RCK. A and B allele copy-number reconstruction output from RCK[9]. Presumably the large number of somatic SVs incurred by templated insertions near chr4:168Mbp and chr4:123Mbp in conjunction with their expected absence in the output of allele-specific copy-number callers like HaTCHet[7] caused an incorrect reconstruction of the A and B allele copy-number of chr4 compared to Figure S4. Related to STAR Methods.

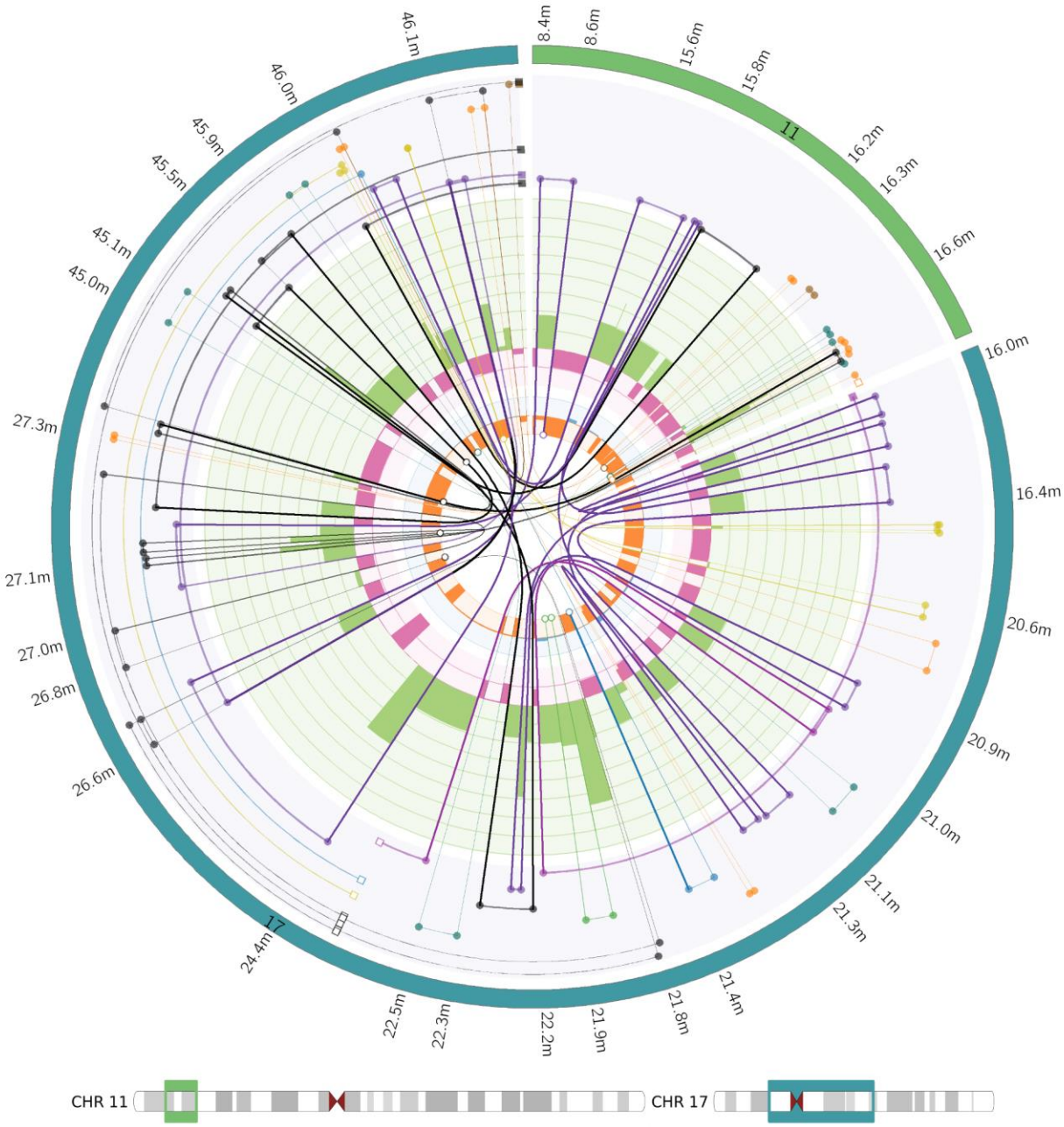


Figure S45. Short-read based reconstruction of CS11-17. Linx[4] visualization of the complex event overlapping the CS11-17 structure with chromosomal segments on the outer ring, chaining of variants as colored lines and copy-number gains (green) and losses (red) in the inner circle. For a detailed guide see Shale et al. 2022[4]. Related to STAR Methods.

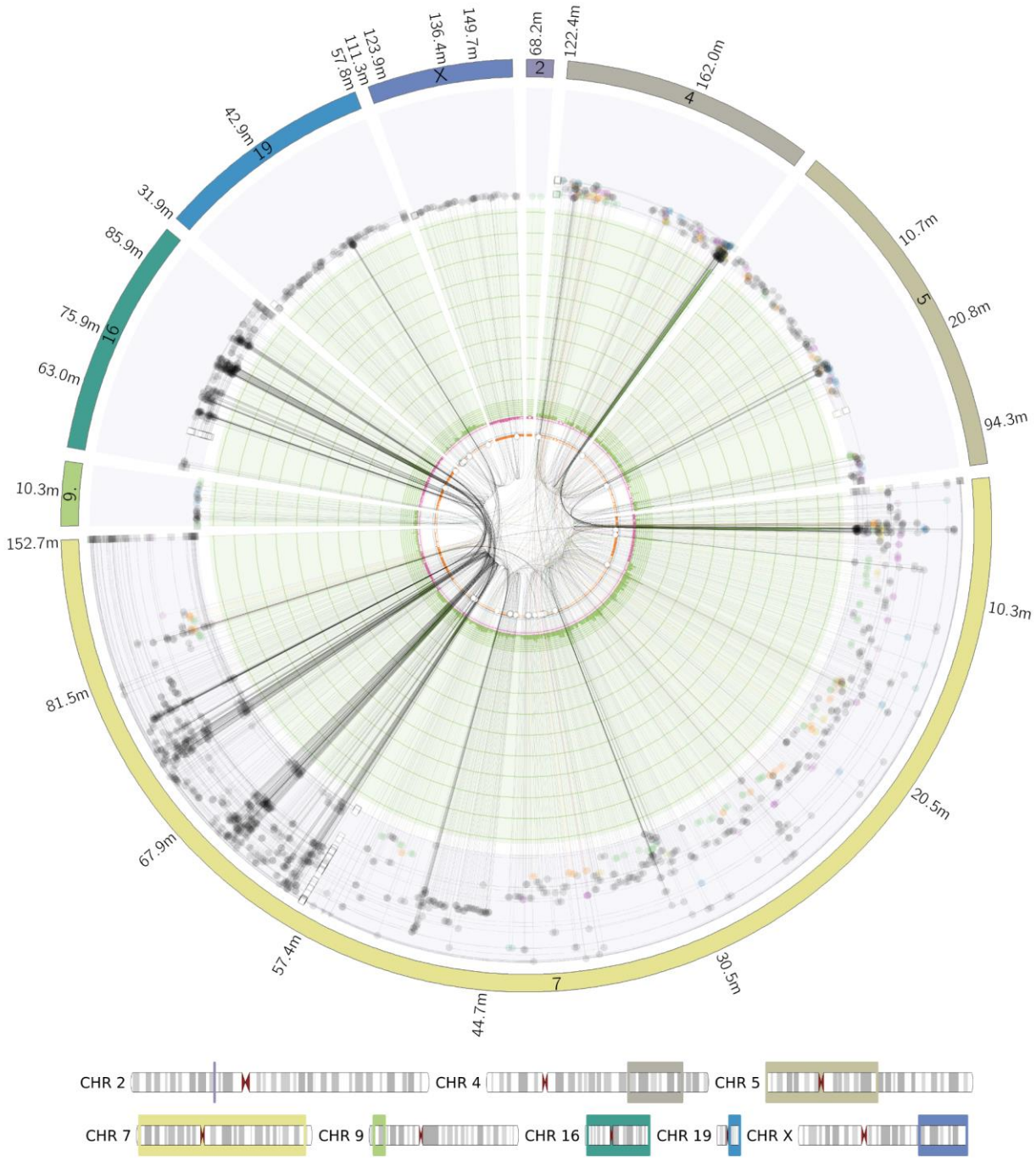


Figure S46. Short-read based reconstruction of a TI thread. Linx visualization of the complex event overlapping the TI thread in Figure 2. For a detailed guide of the plot see Shale et al. 2022[4]. Related to STAR Methods.

Supplementary Tables

Table S1. Long-read sequencing statistics on the four ONT sequencing runs. Read length histograms are in Figure S38. Related to STAR Methods.

Sequencing run	Reads	Bases (mb)	Median read length	N50	Median PHRED	Coverage
Run1: **Germline**	6,030,382	46,778	4,090	15,600	12.890	~15x
Run 2: **Primary**	9,222,301	87,564	4,490	21,800	13.941	~29x
Run 3: **Relapse**	8,078,296	57,294	3,600	10,800	13.476	~19x
Run 4: Multiplexed **Germline** & **Primary**	2,916,725	17,255	4,830	8,230	9.792	~6x

Table S2. Short-read sequencing statistics. Related to STAR Methods.

	Germline	**Primary tumor**	**Relapse**
Coverage	48x	45x	47x
Seq. mode	2 * 150bp	2 * 150bp	2 * 150bp
Insert size	373bp	387bp	406bp

Table S3. FISH analysis using probes for RP11-651L9 and centromere 17 (primary tumor). Related to section “Haplotype-phased assembly of complex somatic rearrangements", related to Figure 1.

Number of signals	**651L9**	**Centr.17**
0	2	0
1*	23	4
2**	36	20
3	20	30
4	16	28
5	3	15
6		2
7		1

*Due to the tissue cutting, nuclei with only one signal can be explained by signal truncation of tumor nuclei during the sectioning process

**Normal cell nuclei are included

Table S4. Combined FISH analysis using probes RP11-651L9 and centromere 17. Related to section “Haplotype-phased assembly of complex somatic rearrangements", related to Figure 1.

Signals	**651L9 / Centr.17**
0/1	1
0/4	1
1/1	2
1/2	3
1/3	9
1/4	7
1/5	2
2/2	13
2/3	9
2/4	9
2/5	3
2/6	1
2/7	1
3/1	1
3/2	2
3/3	7
3/4	5
3/5	4
3/6	1
4/2	2
4/3	5
4/4	2
4/5	5
4/6	1
4/7	1
5/4	2
5/5	1

Table S5. FISH on metaphase spreads from matched patient derived xenografts in 2 replicates. Related to section “Haplotype-phased assembly of complex somatic rearrangements", related to Figure 1.

Number of signals	**Probe 651L9 (Replicate 1)**	**Probe 651L9 (Replicate 2)**
1	11	4
2	28	22
3	11	18
4	12	15
5	0	6
6	2	4
7	0	1
Clusters*	36	30

*Clusters of signals close to each other

Table S6. Interphase nuclei of the PDX primary (two independent experiments, 100 nuclei each). Related to section "Haplotype-phased assembly of complex somatic rearrangements", related to Figure 1.

651L9 (non-colocalized)	Centr. 17 (non-colocalized)	Colocalization	Frequency
0	3	0	1
0	4	0	6
0	5	0	2
1	1	0	1
1	4	0	2
2	4	0	3
2	5	0	1
4	4	0	1
3	4	0	1
4	4	1	1
0	2	1	3
0	3	1	14
0	4	1	5
0	1	1	2
1	1	1	2
1	2	1	11
1	3	1	19
1	4	1	6
1	6	1	2
2	1	1	2
2	2	1	7
3	3	1	2
2	3	1	15
4	3	1	1
3	4	1	3
2	4	1	5
0	0	2	2
0	1	2	3
0	2	2	9
0	4	2	1
0	3	2	14
1	1	2	2
1	2	2	5
1	3	2	3
1	4	2	6
2	3	2	6
2	2	2	8
3	4	2	1
0	0	3	1
0	1	3	3
0	2	3	10
0	4	3	1
1	2	3	2
1	3	3	3
1	4	3	1
2	3	3	1

References

1. Patterson, M., Marschall, T., Pisanti, N., van Iersel, L., Stougie, L., Klau, G.W., and Schönhuth, A. (2015). WhatsHap: Weighted Haplotype Assembly for Future-Generation Sequencing Reads. J. Comput. Biol. *22*, 498–509.

2. Delaneau, O., Zagury, J.-F., Robinson, M.R., Marchini, J.L., and Dermitzakis, E.T. (2019). Accurate, scalable and integrative haplotype estimation. Nat. Commun. *10*, 5436.

3. Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics *28*, i333–i339.

4. Shale, C., Cameron, D.L., Baber, J., Wong, M., Cowley, M.J., Papenfuss, A.T., Cuppen, E., and Priestley, P. (2022). Unscrambling cancer genomes via integrated analysis of structural variation and copy number. Cell Genomics *2*, 100112.

5. Shafin, K., Pesout, T., Lorig-Roach, R., Haukness, M., Olsen, H.E., Bosworth, C., Armstrong, J., Tigyi, K., Maurer, N., Koren, S., et al. (2020). Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. Nat. Biotechnol. *38*, 1044–1053.

6. Wala, J.A., Bandopadhayay, P., Greenwald, N.F., O'Rourke, R., Sharpe, T., Stewart, C., Schumacher, S., Li, Y., Weischenfeldt, J., Yao, X., et al. (2018). SvABA: genome-wide detection of structural variants and indels by local assembly. Genome Res. *28*, 581–591.

7. Zaccaria, S., and Raphael, B.J. (2020). Accurate quantification of copy-number aberrations and whole-genome duplications in multi-sample tumor sequencing data. Nat. Commun. *11*, 4301.

8. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

9. Aganezov, S., and Raphael, B.J. (2020). Reconstruction of clone- and haplotype-specific cancer genome karyotypes from bulk tumor samples. Genome Res. *30*, 1274–1290.