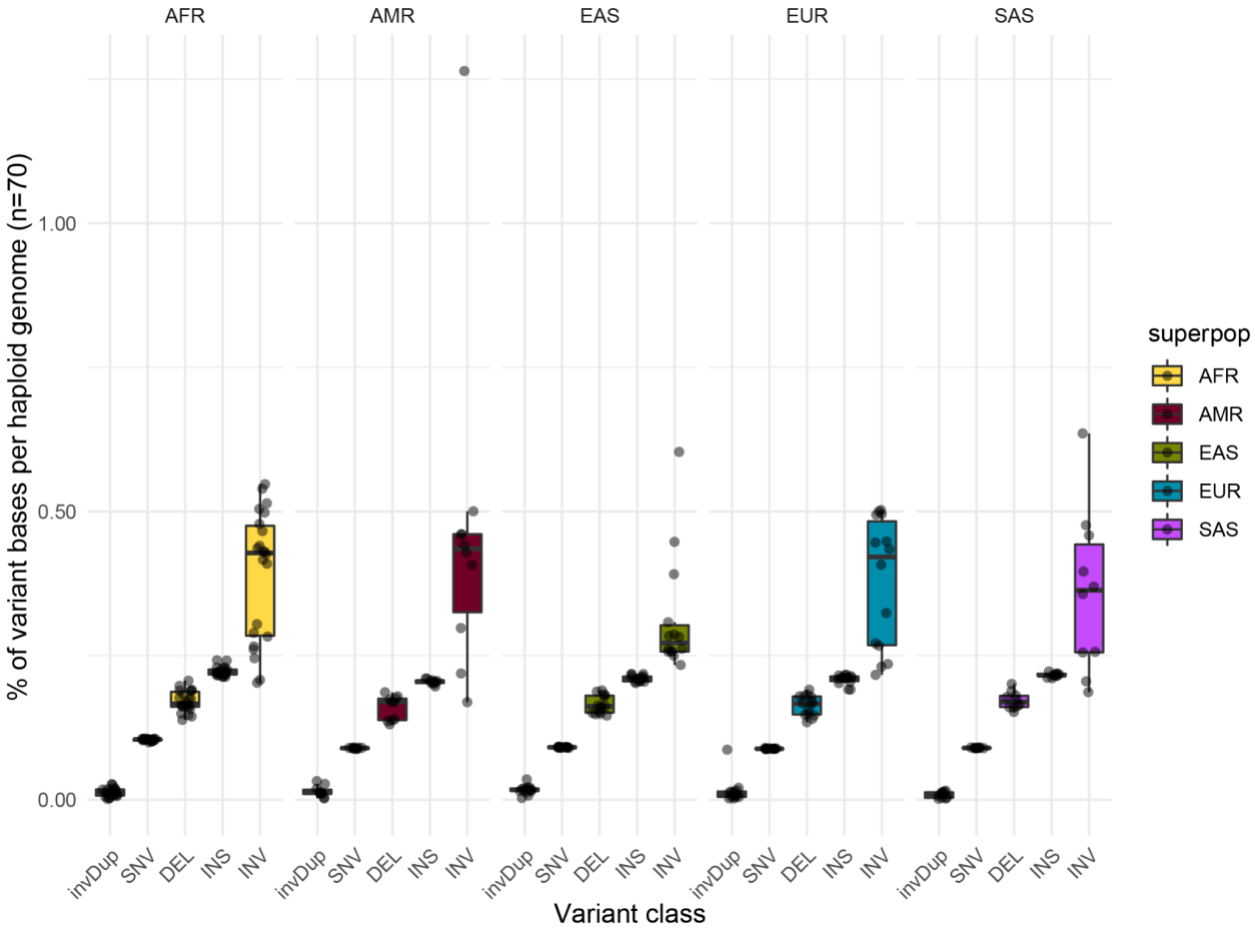
Data S1: Related to Main figures, STAR Methods, and Discussion

Table of Contents

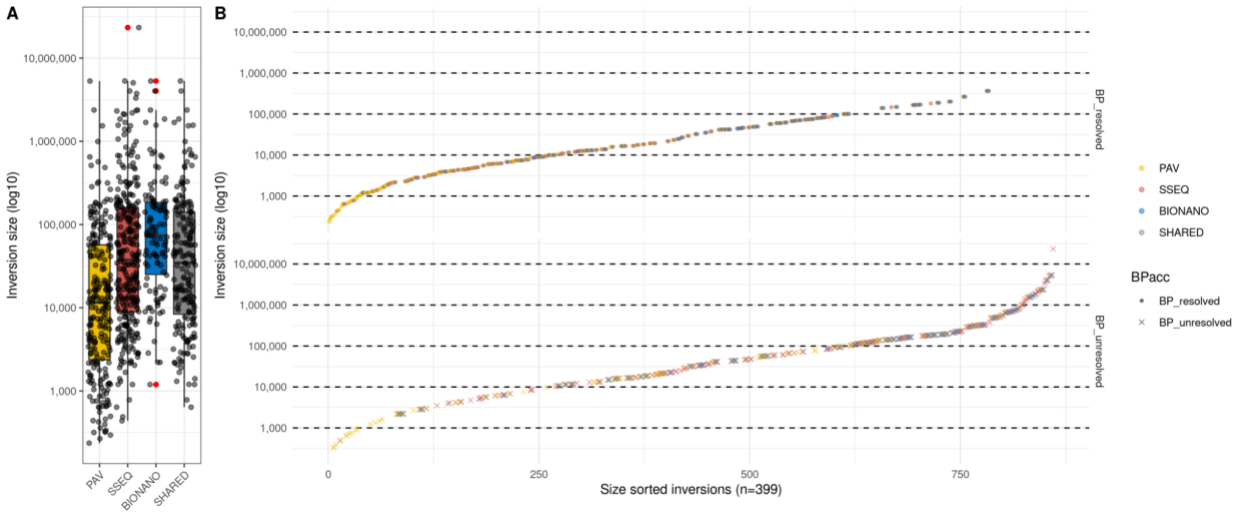
Supporting data for Main figures



Supporting data figure for Figure 1B.

Figure i: Genome-wide prevalence of different classes of human genetic variation.

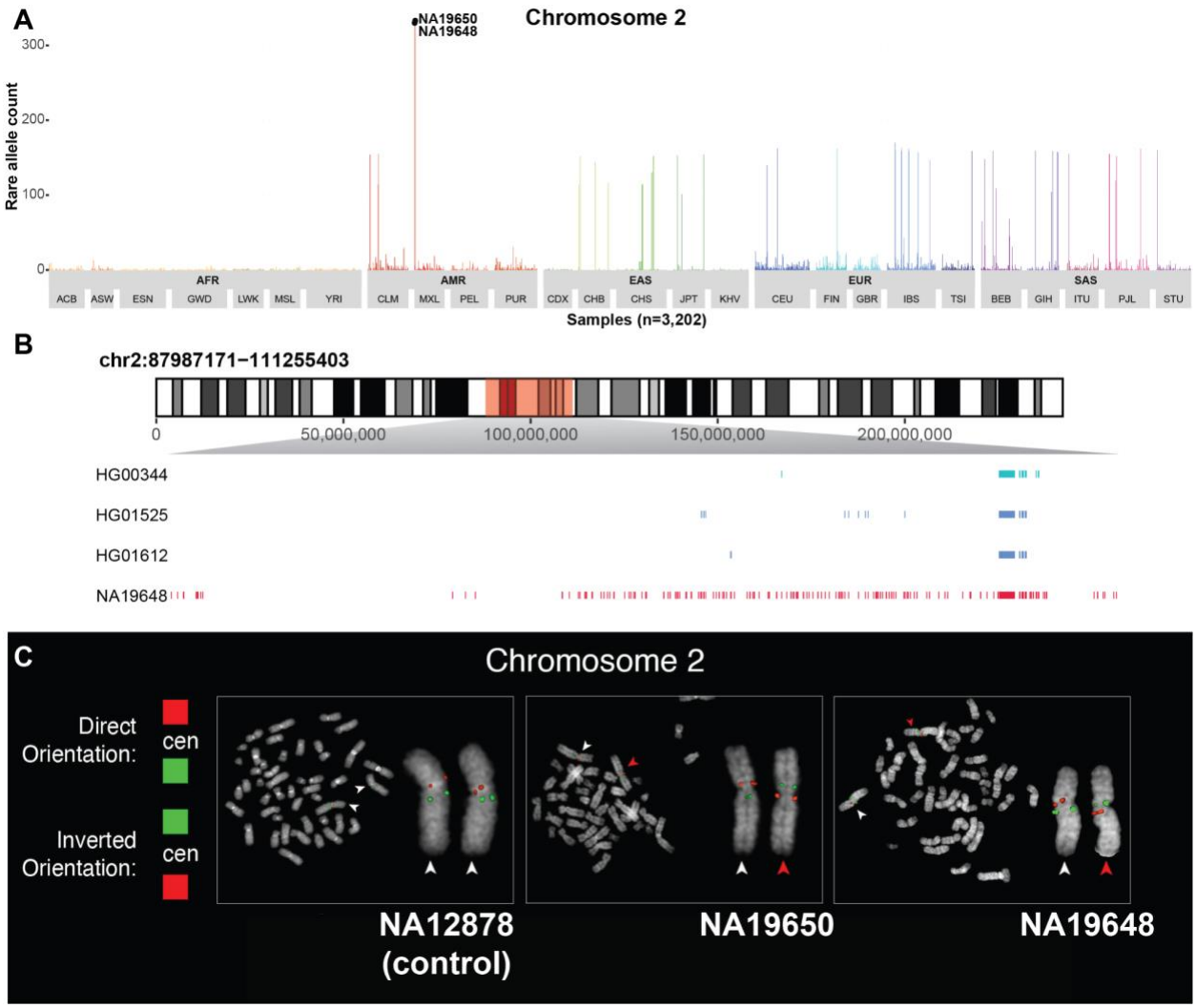
A boxplot showing the percentage of inverted bases stratified by structural variant type and by the population of origin (AFR - African, SAS - South Asian, EAS - East Asian, EUR - European, AMR - American) across 70 haploid genomes (35 human samples for which data from all genomic platforms were available).



Supporting data figure for Figure 1D.

Figure ii: Summary statistics of inversion callset for inversions outside of L1-internal sequence.

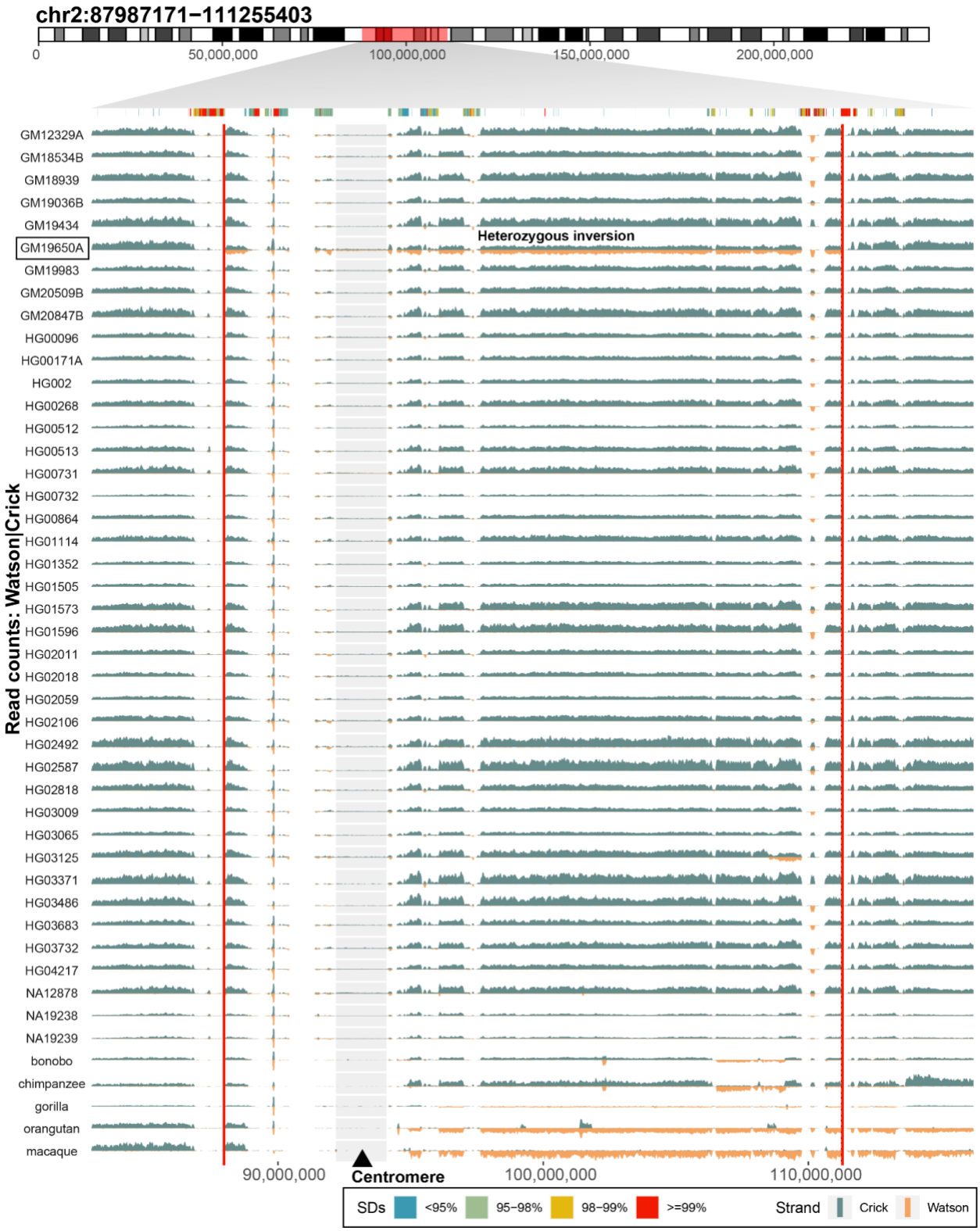
A) A boxplot showing the size distribution of inversions calls (n=399) per caller ('PAV' - phased assembly inversion caller, 'BIONANO' - Bionano optical maps, and 'SSEQ' - automated Strand-seq inversion calls) and inversions called by at least two independent technologies ('SHARED'). PAV- min: 236 bp, max: 5.3 Mbp; SSEQ - min: 438 bp, max: 23.3 Mbp; and BIONANO - min: 1,195 bp, max: 5.3 Mbp. **B)** Distribution of inversion sizes colored by respective caller. Inversions with sequence-resolved breakpoints (BP) are marked as full circles while the inversions without base-pair level breakpoint accuracy are labeled by crosses.



Supporting data figure for Figure 1F.

Figure iii: Inference in short-read data and validation of the large chromosome 2 inversion.

A) A barplot showing the number of rare alleles shared between the index individual (NA19650) and all 1KG samples (n=3,202) stratified by superpopulation (AFR, AMR, EAS, EUR and SAS) and population (same abbreviations used as in 1KG (1000 Genomes Project Consortium et al., 2015). Individuals with the largest number of shared rare alleles are highlighted by a black dot and a sample-specific identifier. **B)** A distribution of detected shared rare alleles along the inverted region on chromosome 2. The inverted region is highlighted by a transparent red rectangle. Rare alleles for the chromosome 2-specific inversion are evenly distributed along the whole inverted region only in a single sample (NA19648). **C)** FISH results of a ~23.2 Mbp inversion on chromosome 2 (chr2:88064758-111283969) are shown. Both chromosome 2 homologs have a direct orientation in the control individual (NA12878) while NA19650 and NA19648 individuals are inverted in heterozygous state. White arrowheads indicate chromosomes in direct orientation while red arrowheads indicate chromosomes with the inversion (ABC8-2121940H19 in red mapping at chr2:88223569-88269173; WI2-1849B17 in green mapping at chr2:110712025-110745244).

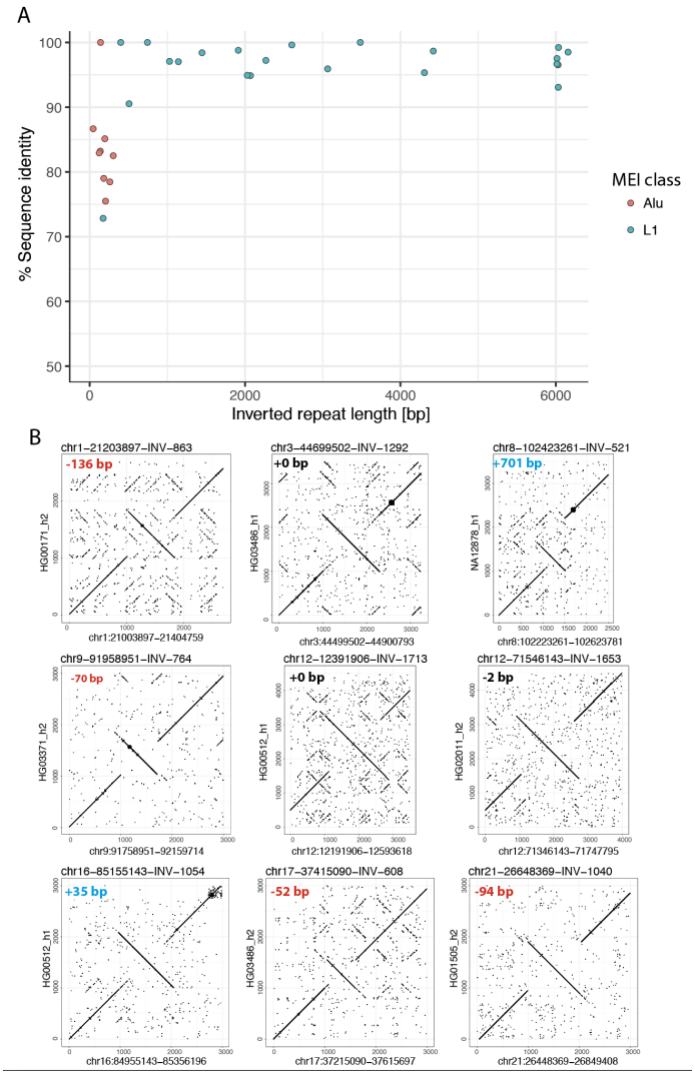


Supporting data figure for Figure 1E.

Figure iv: Large-scale inversion on chromosome 2 spanning the chromosome centromere.

Top: Chromosome 2 ideogram with inverted region highlighted in transparent red color. Below is the SD annotation for a given region represented as a set of rectangles colored by sequence identity of each SD. Lastly, underneath are the read coverage profiles of Strand-seq data for the region summarized as binned (bin size: 50,000, step size: 10,000) read counts represented as bars above

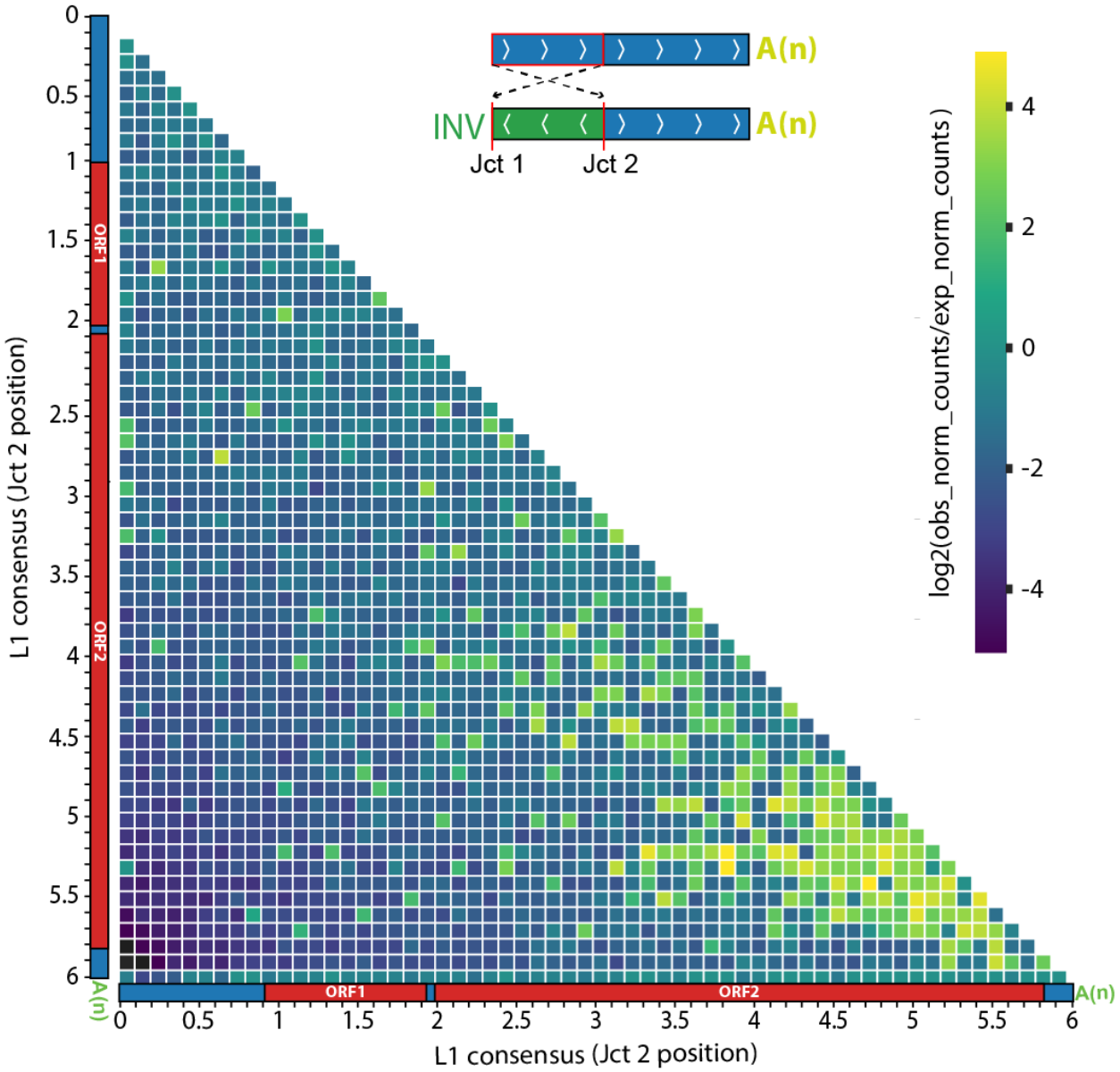
(teal; Crick read counts) and below (orange; Watson read counts) midline. A region with roughly equal coverage of Watson and Crick reads represents a heterozygous inversion as only one homologue is inverted in respect to the reference. A region with reads aligned only in Watson orientation represents a homozygous inversion as both homologs are inverted in respect to the reference while a region with purely Crick reads is represented by both homologs being in direct orientation in respect to the reference. Vertical red lines highlight the inverted region present in sample NA19650 (black rectangle) as a heterozygous inversion.



Supporting data figure for Figure 2A.

Figure v: Length and sequence identity of balanced inversion-flanking repeats mapping to mobile elements.

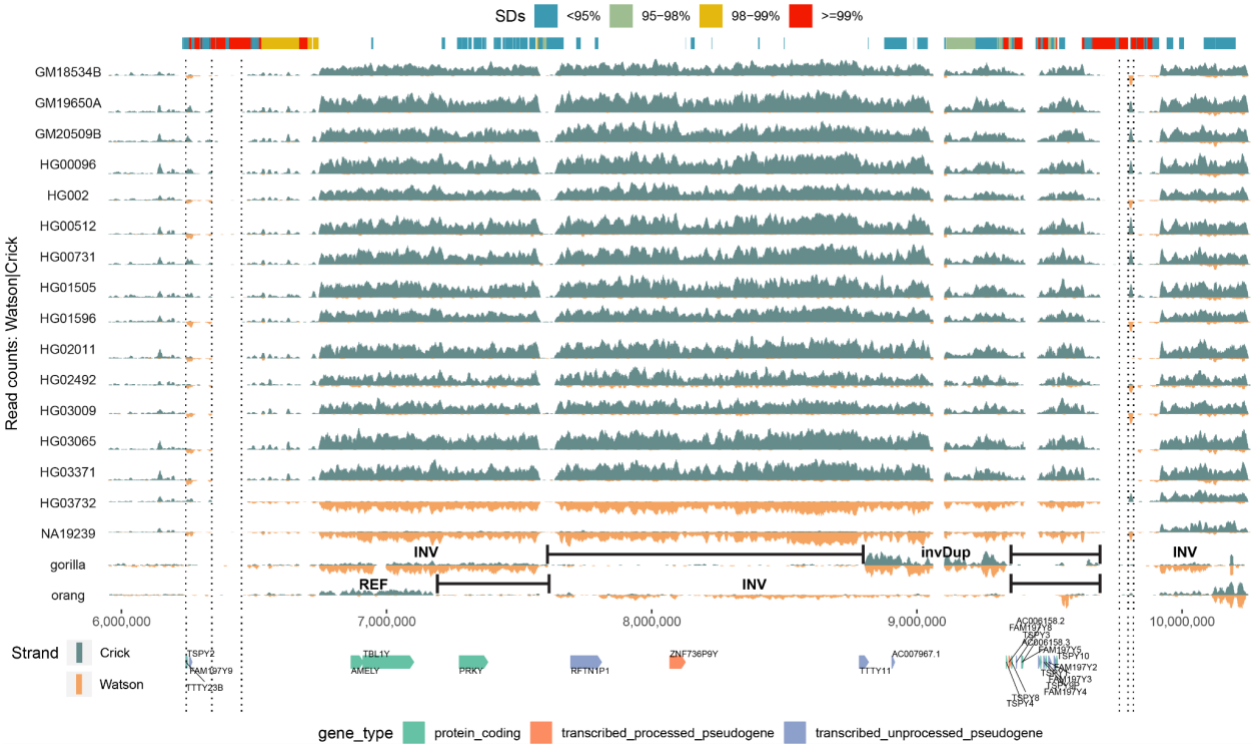
A) There were 31 inverted repeat pairs flanking balanced inversions found to map to mobile elements (MEI). In comparison with pairs of flanking Alu repeats, pairs of flanking L1 repeats are significantly longer (median: 2,435 bp and 181 bp, respectively, p = 3.1e-06, one-sided t-test) and display higher sequence identity (median: 97.2% vs. 82.92%, p = 0.00049, one-sided t-test). **B)** Dotplot visualization of the nine inversions flanked by Alu elements. The number of base pairs gained (blue) or lost (red) across the whole inversion locus in the inverted haplotype is indicated in the top left corner of each dotplot.



Supporting data figure for Figure 2F.

Figure vi: Clustering of inversion coordinates along the L1 consensus sequence.

For each inversion, the positions for the breakpoint junction 1 and 2 (Jct 1-2) were assigned to 100 bp bins along the L1 consensus. The number of inversions connecting each possible combination of bin pairs was computed both for the observed twin-priming events and an in silico generated dataset of 3,000 inverted L1s (**Supporting data for STAR Methods**). Observed (obs_norm_counts) and expected (exp_norm_counts) inversion counts per bin were scaled by the total number of counts per group, respectively. The log2 ratio between the observed and expected scaled counts was computed per bin pair and displayed as a dark to light yellow-colored gradient in the heatmap. An explanatory schematic is displayed over the heatmap.

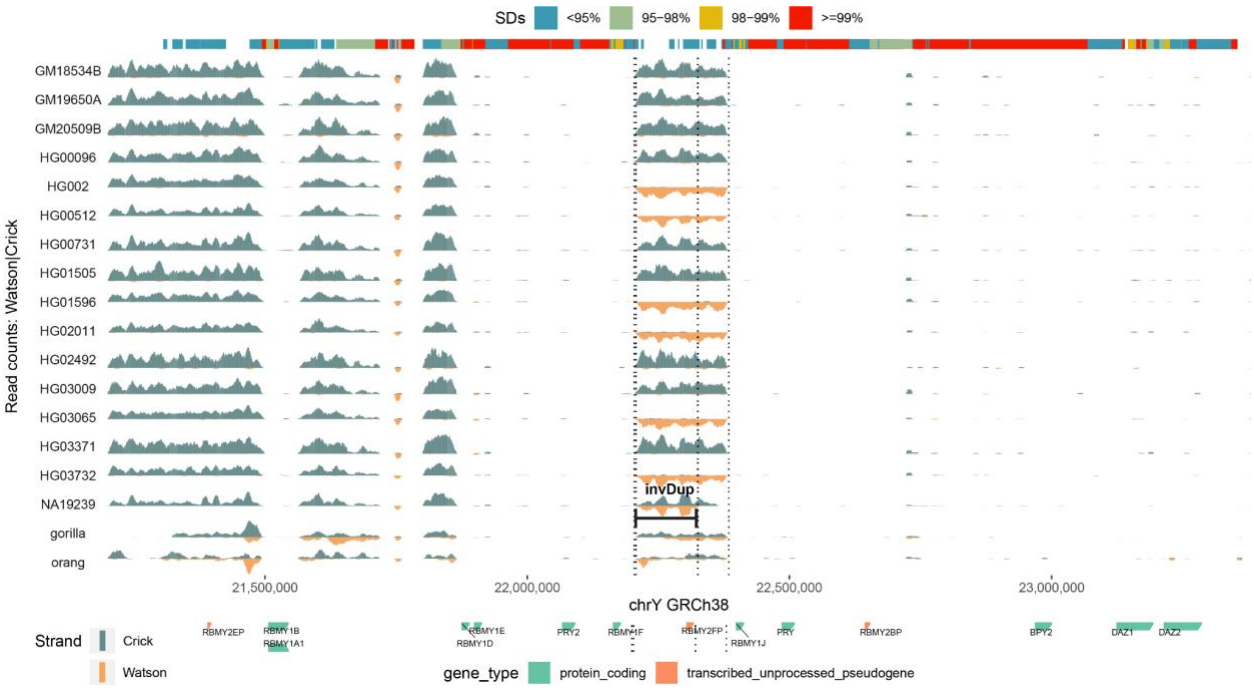


Supporting data figure for Figure 4A and STAR Methods section: Construction and dating of Y phylogeny and Y inversion rate estimation.

Figure vii: Large gene-rich ~3.3 Mbp IR3/IR3 inversion on chromosome Y.

Top: Reference genome-specific SD annotation for a given region represented as a set of rectangles colored by sequence identity. Underneath, read coverage profiles of Strand-seq data over a given region (inversion number 2; **Table S3**) are summarized as binned (bin size: 10 kbp, step size: 1 kbp) read counts and represented as bars above (teal; Crick read counts) and below (orange; Watson read counts) midline. Regions with reads aligned only in Watson orientation represent inversions. Vertical dotted lines highlight the breakpoints of inverted regions. Coverage profile for each sample is reported in separate rows. In the gorilla and orangutan samples, we highlight regions that are inverted (INV), in reference (direct) orientation (REF), or duplicated and inverted (invDup). Regions of lower mappability of nonhuman primate sequencing reads in respect to GRCh38 are highlighted as horizontal black lines. A gene annotation track (GENCODE v36) is shown at the bottom.

The ~3.3 Mbp IR3/IR3 inversion was previously reported to toggle at least 12 times in recent human history, with an estimated rate of $\geq 2.3 \times 10^{-4}$ per father-to-son Y transmission (Repping et al., 2006). We identified this inversion in an African (NA19239) and a Southeast Asian (HG03732) individual carrying Y lineages E1a2a1a1a-CTS1792 and R2-L266, respectively, closely related to those previously reported to be inverted

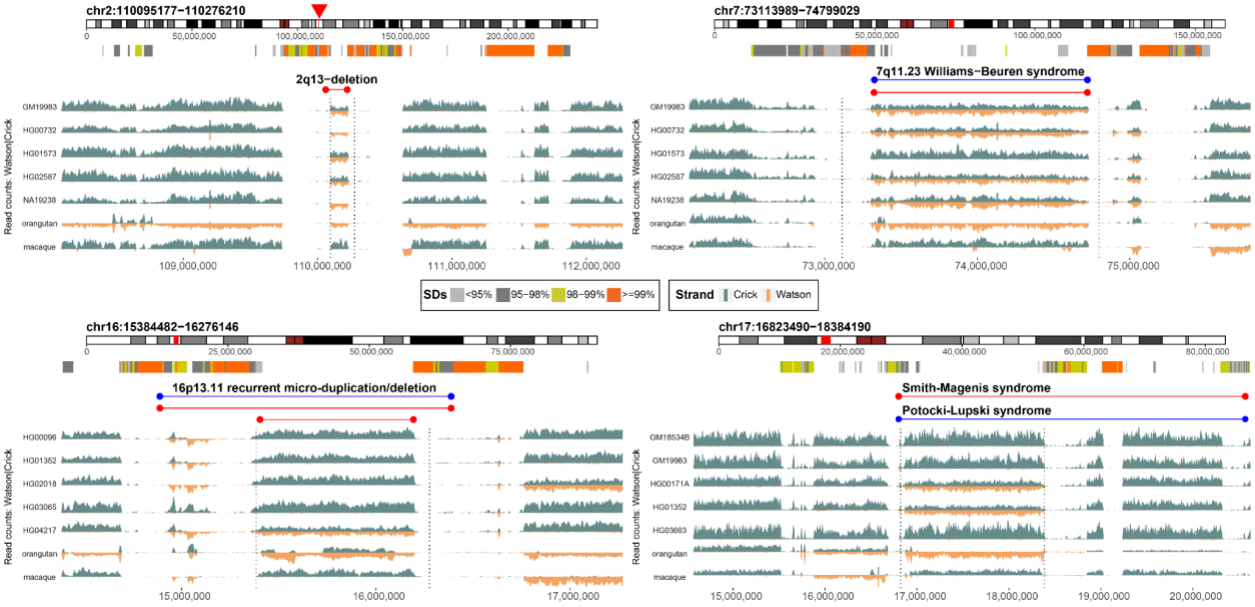


Supporting data figure for Figure 4A and STAR Methods section: Construction and dating of Y phylogeny and Y inversion rate estimation..

Figure viii: Two different Y-chromosomal inversions identified in palindrome P3.

The larger (inversion number 14; **Table S9**) ~180 kbp balanced inversion on chromosome Y was identified in six samples, whereas one individual (NA19239) showed evidence of a smaller, ~118 kbp inverted duplication (inversion number 15). Top: Reference genome-specific SD annotation represented as a set of rectangles colored by sequence identity. Underneath: Read coverage profiles of Strand-seq data over a given region summarized as binned (bin size: 10 kbp, step size: 1 kbp) read counts represented as bars above (teal; Crick read counts) and below (orange; Watson read counts) midline. Regions with reads aligned only in Watson orientation represent inversions. Vertical dotted lines highlight the breakpoints of inverted regions. A gene annotation track (GENCODE v36) is shown at the bottom.

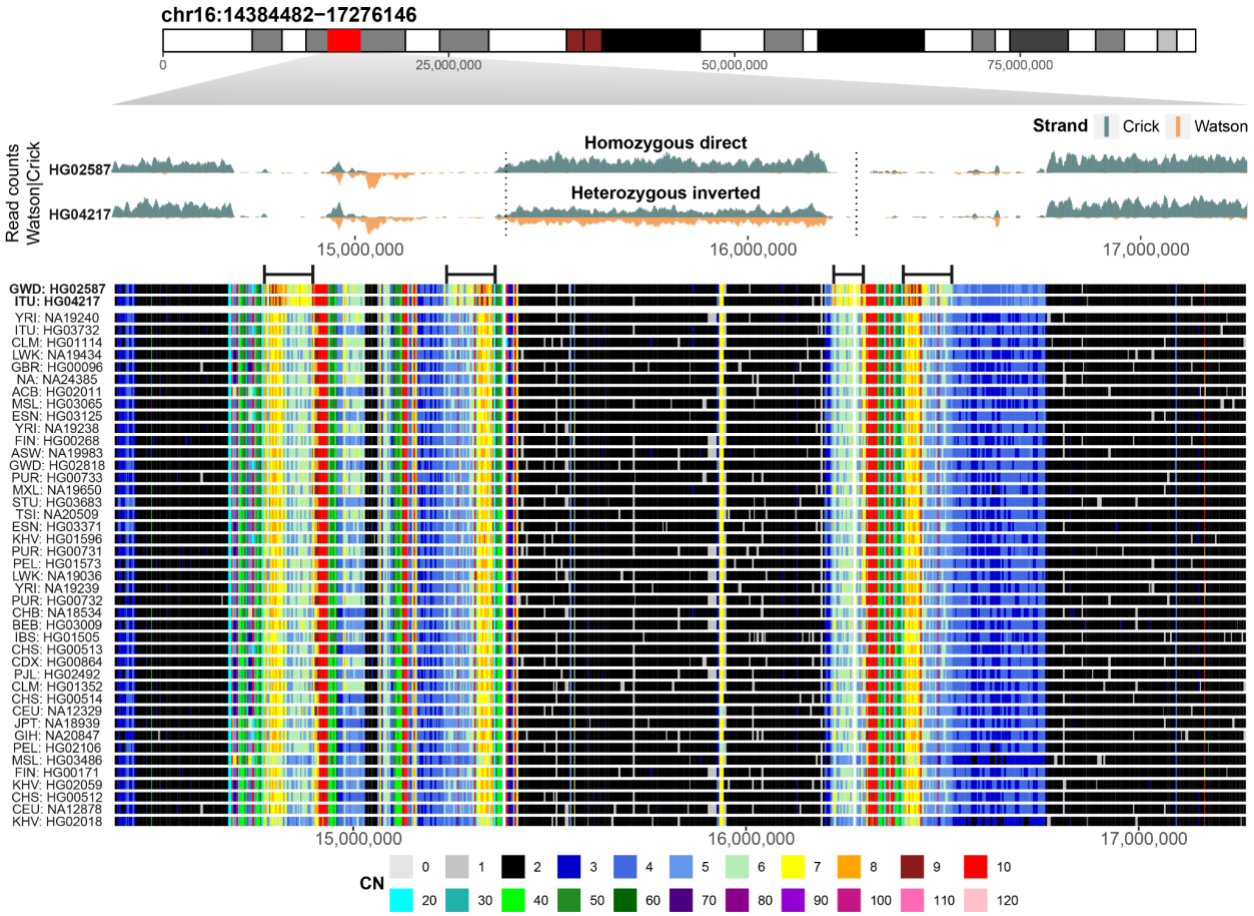
An inversion of palindrome P3, containing some of the copy number variable *RBMY1* genes, was previously reported as a single inversion event based on fiber FISH experiments in a panel of 14 male samples (Shi et al., 2019). In contrast, we identify a ~180 kbp long recurrent inversion at P3 (~284 kbp long flanking arms) with five separate inversion events. We estimate an inversion rate at P3 of 2.68×10^{-4} (95% C.I.: 2.37×10^{-4} to 3.04×10^{-4}) per father-to-son Y transmission. We further observe an inverted duplication in an African male (NA19239), affecting a ~118 kbp segment overlapping with this region, consistent with extensive structural variability of this genomic region (Shi et al., 2019). We further observe an inverted duplication in an African male (NA19239), affecting a ~118 kbp segment overlapping with this region, consistent with extensive structural variability of this genomic region (Shi et al., 2019). Inverted duplication (invDup) in NA19239 is highlighted by black line range.



Supporting data figure for Figure 5D; 5E; 6C.

Figure ix: Strand-seq read profiles for selected inversions.

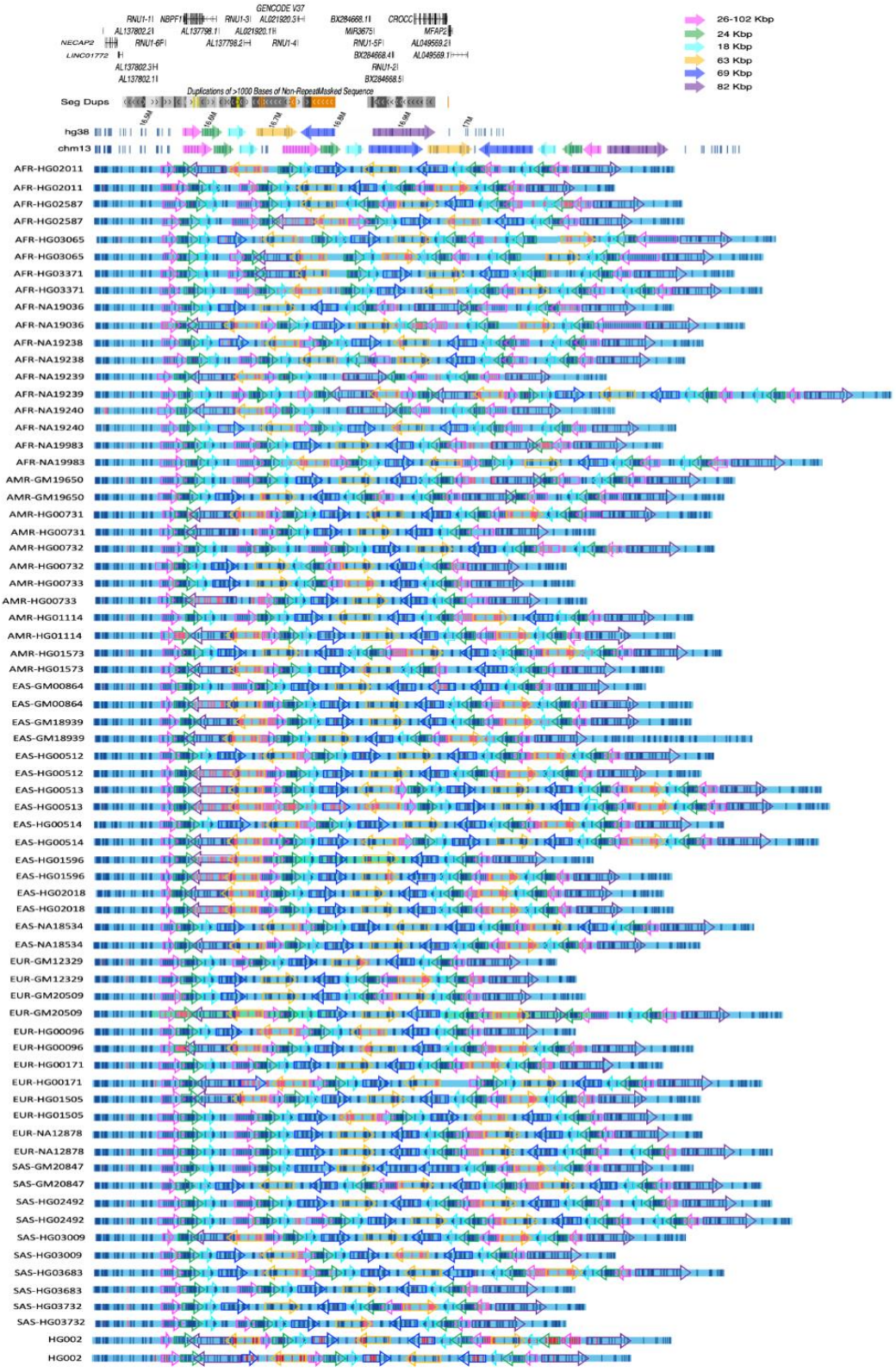
A selected set of regions (2q13, 7q11-23, 16p13-11, and 17p11-2) where inversions overlap previously defined morbid CNVs. Each plot shows a position of a given region on a chromosome-specific ideogram. Below tracks show SDs colored by increasing sequence identity and highlight the regions of previously defined morbid CNVs (red - deletion, blue - duplication). Underneath are the read coverage profiles of Strand-seq data over a given region summarized as binned (bin size: 10,000, step size: 1000) read counts represented as bars above (teal; Crick read counts) and below (orange; Watson read counts) midline. Regions with roughly equal coverage of Watson and Crick reads represent a heterozygous inversion as only one homologue is inverted with respect to the reference. Regions with reads aligned only in Watson orientation represent a homozygous inversion as both homologs are inverted with respect to the reference, while regions with purely Crick reads are in direct (reference) orientation. Vertical dotted lines highlight the inverted region of interest.



Supporting data figure for Figure 6C.

Figure x: Inversion and copy number profile analysis of the 16p13-11 region.

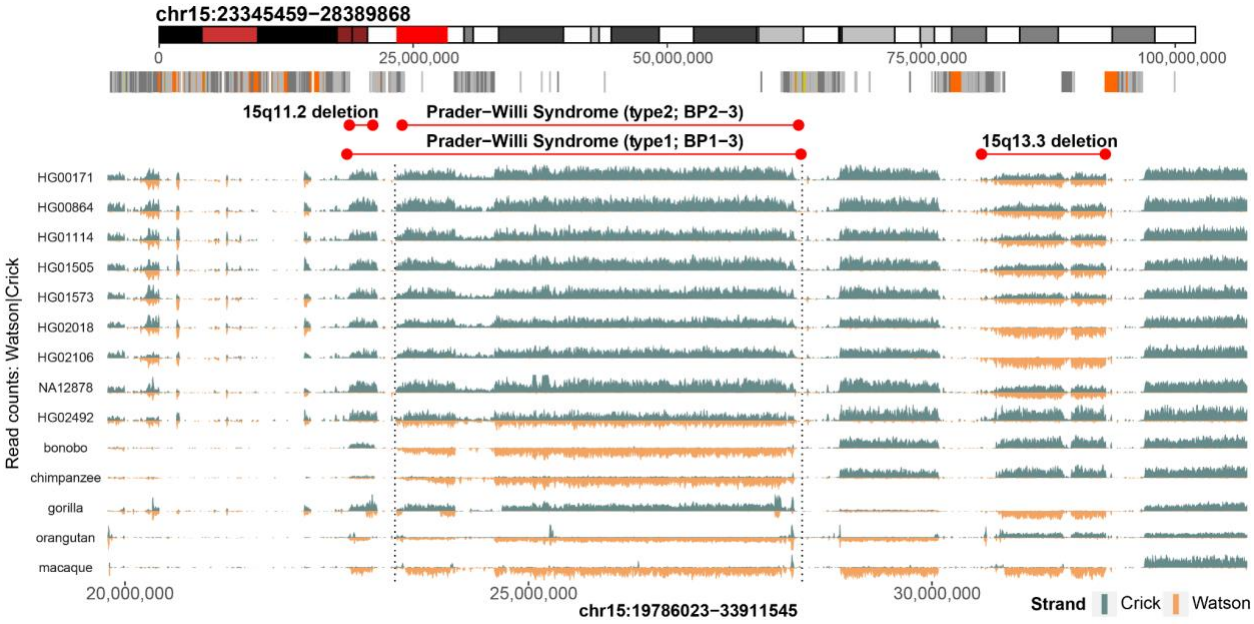
From top to bottom: A chromosome 16 ideogram with inverted region highlighted in red color. Next, read coverage profiles of Strand-seq data summarized as binned (bin size: 10,000, step size: 1000) read counts and represented as bars above (teal; Crick read counts) and below (orange; Watson read counts) midline. The region with roughly equal coverage of Watson and Crick reads represents a heterozygous inversion as only one homologue is inverted in respect to the reference. Vertical dotted lines highlight the inverted region of interest. Last, a copy number (CN) heatmap based on high-coverage Illumina read-depth profiles around the inversion breakpoints (black arrowheads at the bottom) separately for each sample used in this study (n=44, rows). Note that SD blocks at the predicted inversion breakpoint exhibit an increased copy number in this individual in comparison to other samples in our cohort; however, this copy number increase was also seen in a sample (HG02587) lacking the inversion. Regional increase in CN for sample HG02587 and HG04217 is highlighted by black horizontal lines at the top of the heatmap.



Supporting data figure for Figure 6B.

Figure xi: Full set of 1p36.13 structural haplotypes detected.

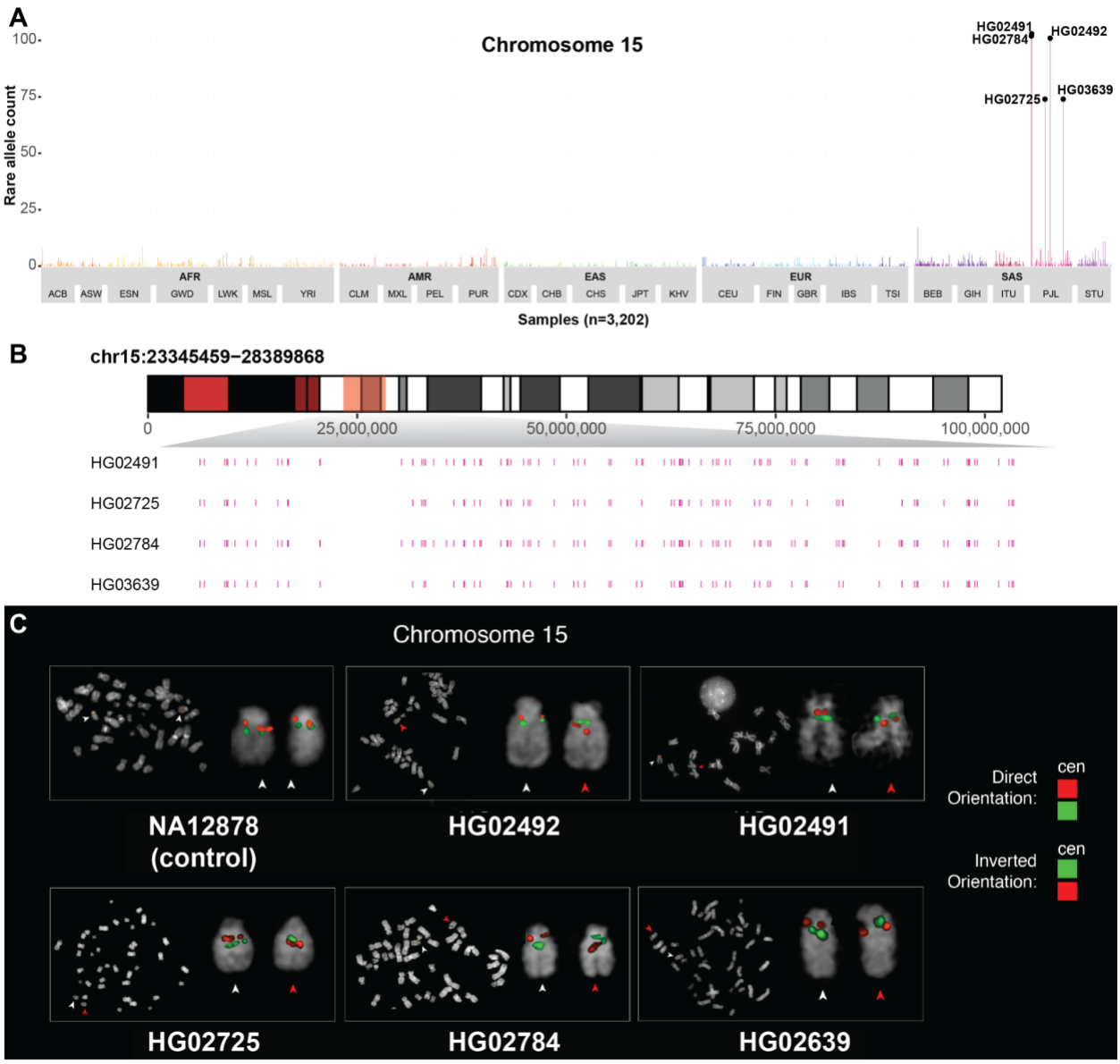
1p36.13 haplotypes were identified in all samples (n=35) using Bionano optical mapping data. Colored arrows represent 1p36.13 segments.



Supporting data figure for Figure 6D.

Figure xii: Large inversion overlapping the PWAS type II region on chromosome 15.

From top to bottom: A chromosome 15 ideogram with inverted region highlighted in transparent red color. Below is the SD annotation for a given region represented as a set of rectangles colored by sequence identity of each SD. Lastly, underneath are the read coverage profiles of Strand-seq data summarized as binned (bin size: 50,000, step size: 10,000) read counts and represented as bars above (teal; Crick read counts) and below (orange; Watson read counts) midline. The region with roughly equal coverage of Watson and Crick reads represents a heterozygous inversion as only one homologue is inverted in respect to the reference. Region with reads aligned only in Watson orientation represents a homozygous inversion as both homologs are inverted with respect to the reference. Vertical dotted lines highlight the inverted region present in sample HG02492 as a heterozygous inversion.

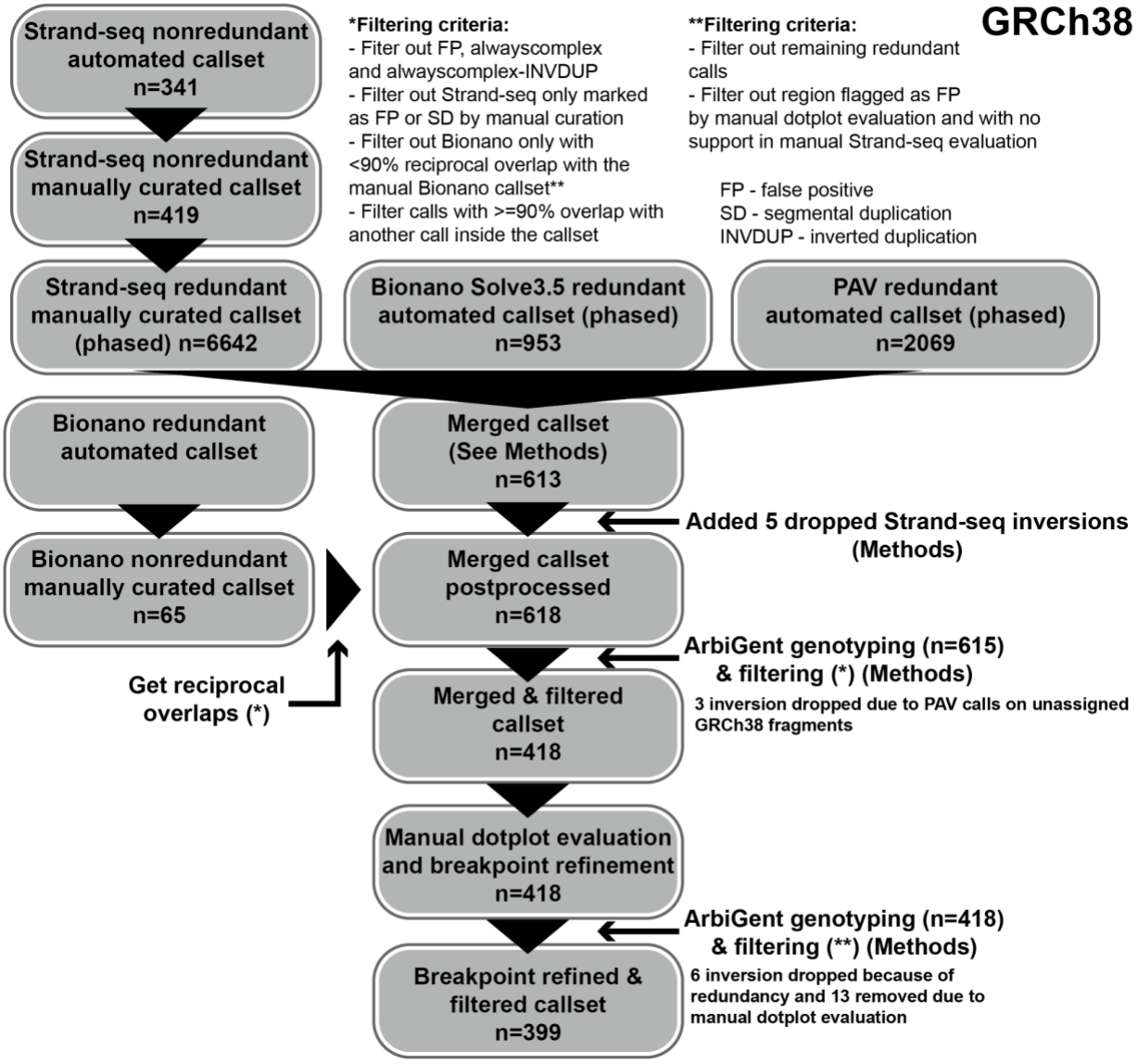


Supporting data figure for Figure 6E; 6F.

Figure xiii: Inference in short-read data and validation of the large balanced inversion on chromosome 15.

A) A barplot showing the number of rare alleles shared between an index individual (HG02492) and all 1KG samples (n=3,202) stratified by superpopulation (AFR, AMR, EAS, EUR and SAS) and population (same abbreviations used as in 1KG (1000 Genomes Project Consortium et al., 2015)). Individuals with the largest number of shared rare alleles are highlighted by a black dot and a sample-specific identifier. **B)** A distribution of detected shared rare alleles along the inverted region. Inverted region is highlighted by a transparent red rectangle. Note: Rare alleles for chromosome 15-specific inversion are evenly distributed along the whole inversion for all predicted inversion carriers. **C)** FISH validation. Both chromosome 15 homologs have a direct orientation in the control individual (NA12878) while HG02491, HG03639, HG02725 and HG02784 are all carriers of the inversion in heterozygous state. White arrowheads indicate chromosomes in direct orientation while red arrowheads indicate chromosomes with the inversion (ABC8-41788900G7 in red mapping at chr15:23751929-23796236; RP11-640H21 in green mapping at chr15:27894428-28091240). cen - centromere.

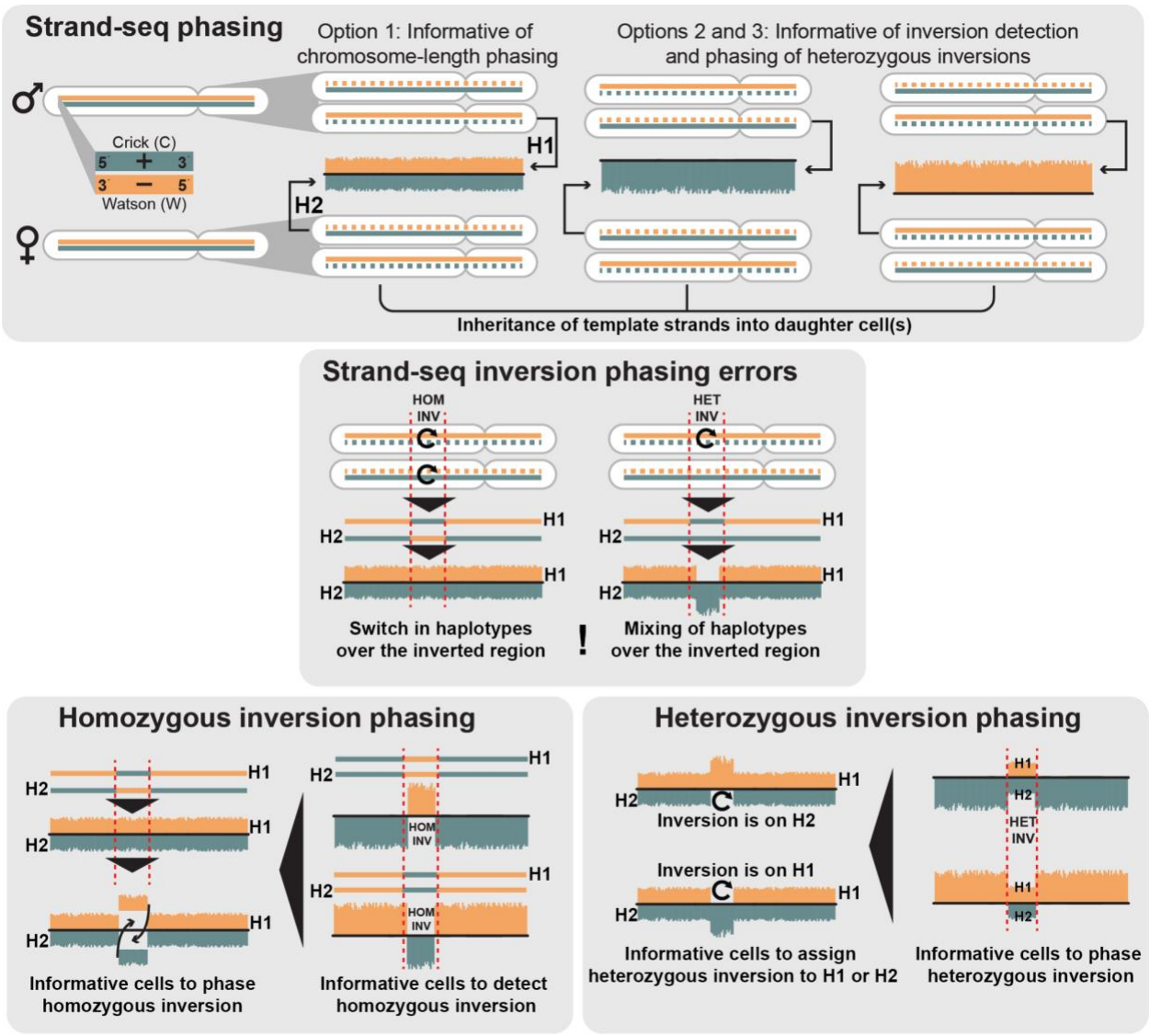
Supporting data for STAR Methods



Supporting data figure for the STAR Methods sections: Strand-seq-based inversion discovery, Inversion merging into a provisional integrated callset, and Inversion genotyping and phasing with ArbiGent.

Figure xiv: Data flowchart to generate unified inversion callset in respect to GRCh38, for inversions outside of L1-internal sequence.

Inversions forming in L1-internal sequences were discovered as described in detail in the Methods section.

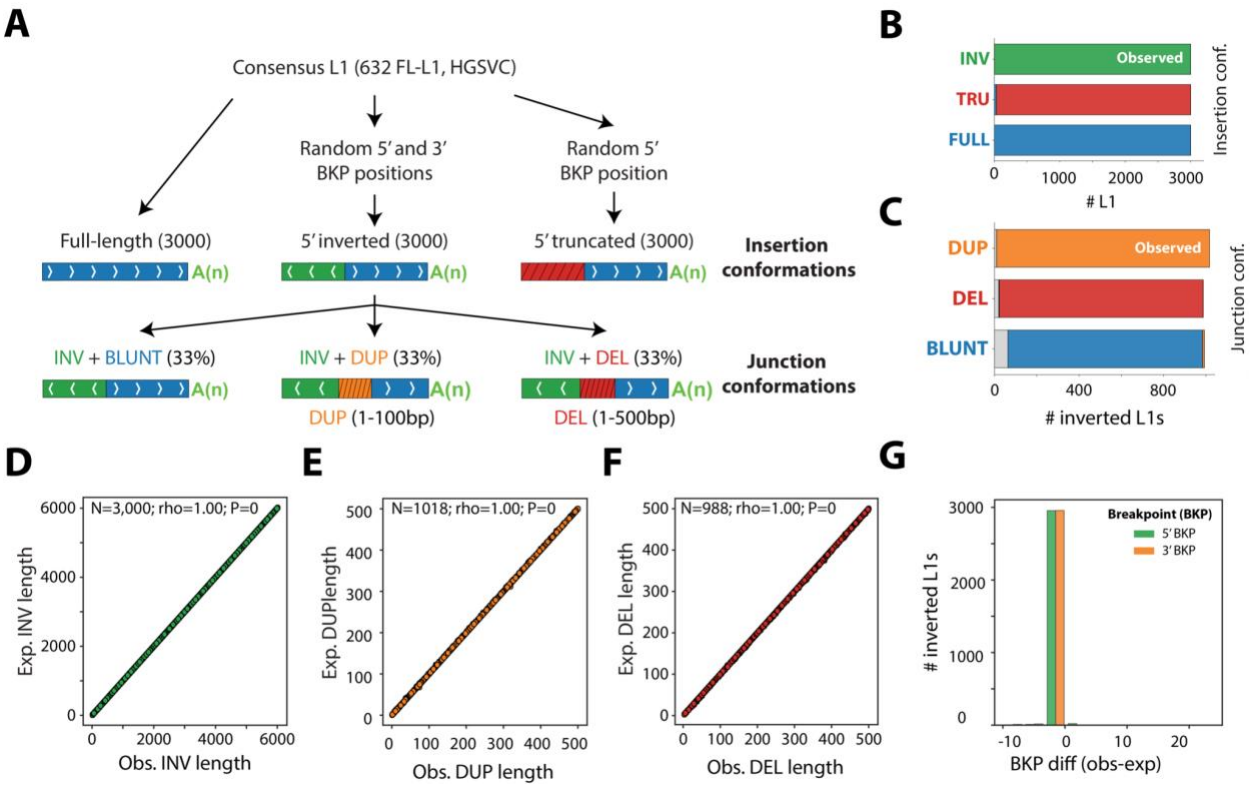


Supporting data figure for the STAR Methods section: Phasing and correction of chromosome-length inversion haplotypes.

Figure xv: Correction of inversion phasing in Strand-seq datasets.

From top to bottom: Using strand-specific sequencing reads (Strand-seq reads), each maternal and paternal homolog is composed of a positive (Crick (C); teal) and a negative (Watson (W); orange) DNA strand. The key to Strand-seq (Falconer et al., 2012) is to let single cells replicate in the presence of bromodeoxyuridine (BrdU), which is a thymidine analogue. During DNA replication a cell incorporates BrdU into the newly synthesized DNA strands. This results in sister chromatids that contain one original template strand (solid line) and one newly synthesized, BrdU-incorporated strand (dashed line). One single-cell division leads to a random assortment of paternal and maternal sister chromatids to daughter cells. Note that newly formed DNA strands containing BrdU are selectively removed in daughter cells during library preparation, such that only the original template DNA strands are being sequenced. There are three possible combinations of template strands: Option 1 - when the daughter cell inherits one Watson and one Crick strand (WC state) from each parent. Such a combination of parental template strands provides chromosome-length phasing information. Options 2 and 3 are a result of inheriting only Watson (WW state) or only Crick (CC state) template strands from both parents. Such a combination of parental template strands provides information to detect inversion as well as phasing of heterozygous inversions. Inversion phasing errors differ between homozygous (HOM INV) and heterozygous (HET INV) inversions. Homozygous inversions appear as a complete switch of haplotypes in comparison to the haplotypes from uninverted regions. Such switches in haplotypes are not visible in Strand-seq cells with a WC state. This is caused by a switch in directionality of Crick (plus) and Watson (minus) reads with respect to the uninverted regions. Homozygous inversions can easily be detected in Strand-seq cells with either WW or CC state. After detection, a switch in haplotypes inside a homozygous inversion is easily corrected by flipping such haplotypes. In contrast, heterozygous inversions are more difficult to correct as only one strand is

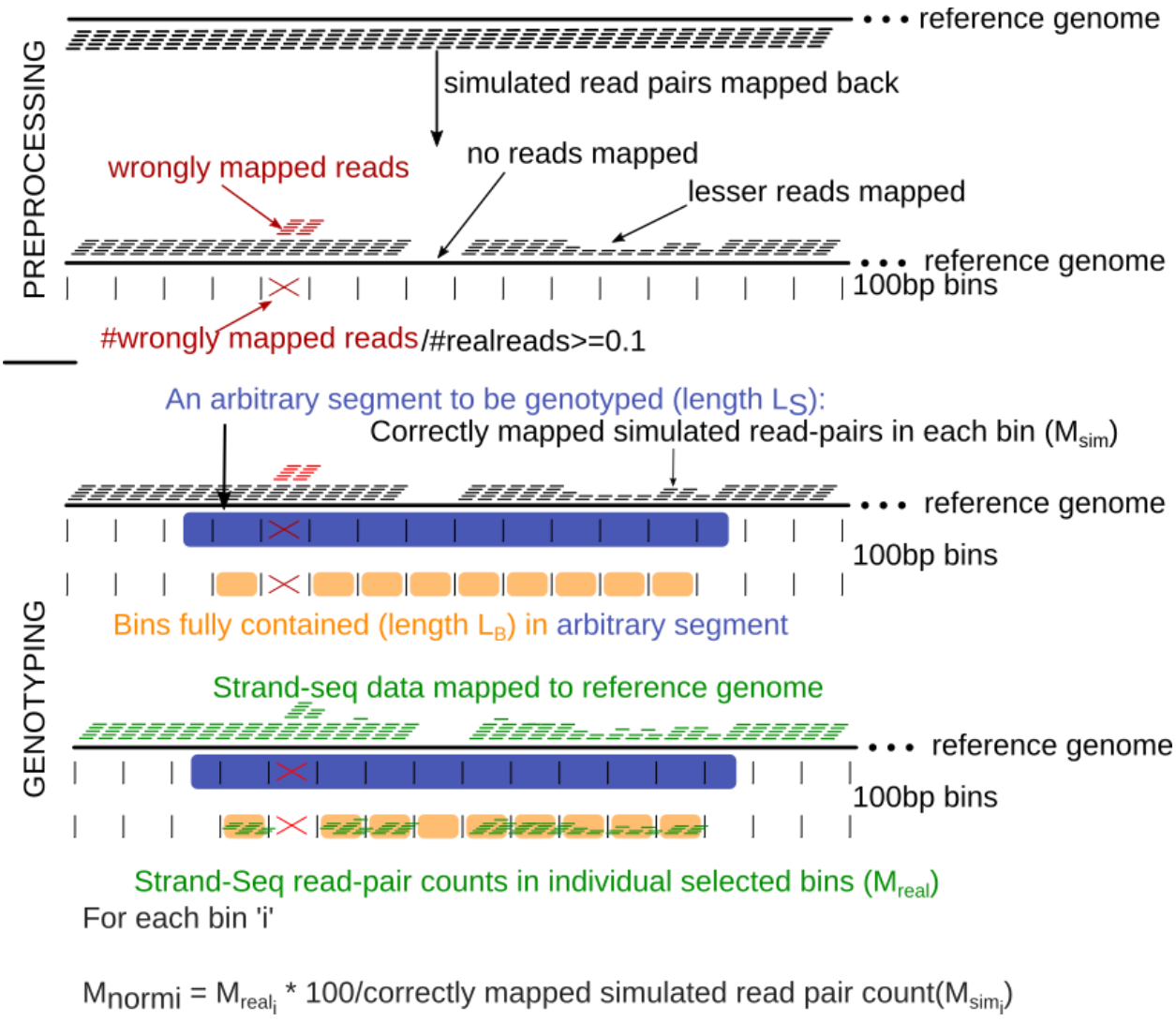
inverted and thus either the Watson or Crick strand changes its directionality over the inverted region. This creates a mixing of alleles in heterozygous state and thus heterozygous inversions need to be phased *de novo* using Strand-seq cells with WW or CC state for a given chromosome. In such cells, heterozygous inversions appear as an equal mixture of Crick and Watson reads (WC state) as only one strand (one haplotype) is inverted. Such cells are informative and can be used for unambiguous phasing for a given inverted region, since Crick and Watson reads are coming from different parental homologs. The inverted haplotype is assigned to a respective parental homolog based on the phasing of reads from Strand-seq cells that inherited either Watson or Crick strands (WC state) from each parent.



Supporting data figure for the STAR Methods section pertaining to twin-priming.

Figure xvi: Evaluation of L1 annotation pipeline using simulations.

A) Workflow for simulating L1s (n=9,000) with diverse insertion conformations from a consensus sequence derived from a collection of 632 full-length L1s (Ebert et al., 2021). Simulated L1s were used to assess the accuracy of the L1 annotation pipeline. (BKP - breakpoint position) **B)** Consistency between pipeline annotation and expectations for the configuration of L1 insertions (INV: inverted; TRU: truncated; FULL: full-length). **C)** Consistency between pipeline annotation and expectations for the configuration of the inversion junctions at L1 insertions (DUP: duplicated; DEL: deleted; BLUNT: blunt joints). **D-F)** Correlation between observed and expected inversion (D), duplication (E), and deletion (F) length at the inversion junction. Spearman's rho and p-value provided for each variable. **G)** Distribution of deviations between observed and expected breakpoints for both inversion ends (5' and 3').

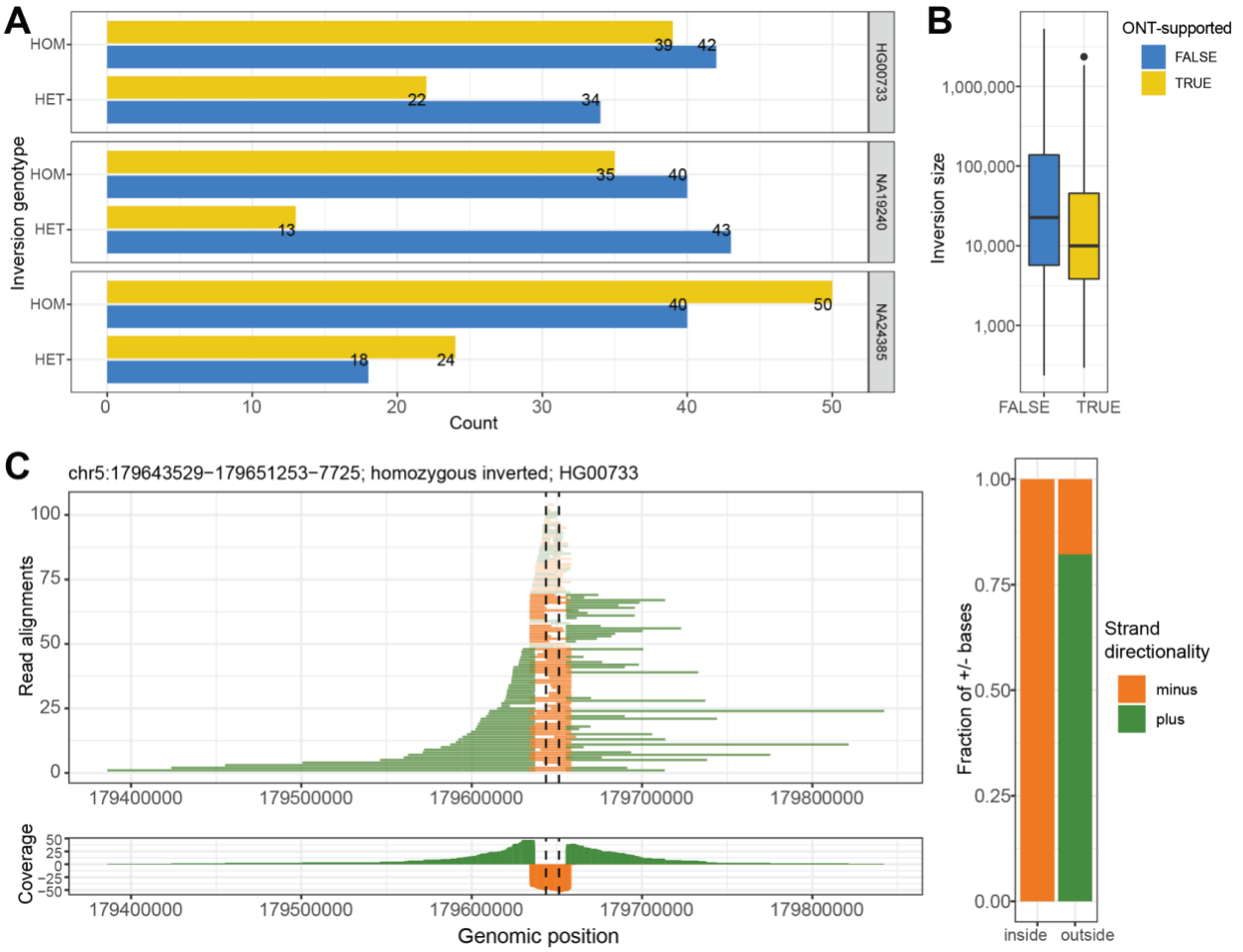


Supporting data figure for the STAR Methods section: Inversion genotyping and phasing with ArbiGent.

Figure xvii: Leveraging mappability of a region to normalize Strand-seq read counts for ArbiGent genotyping.

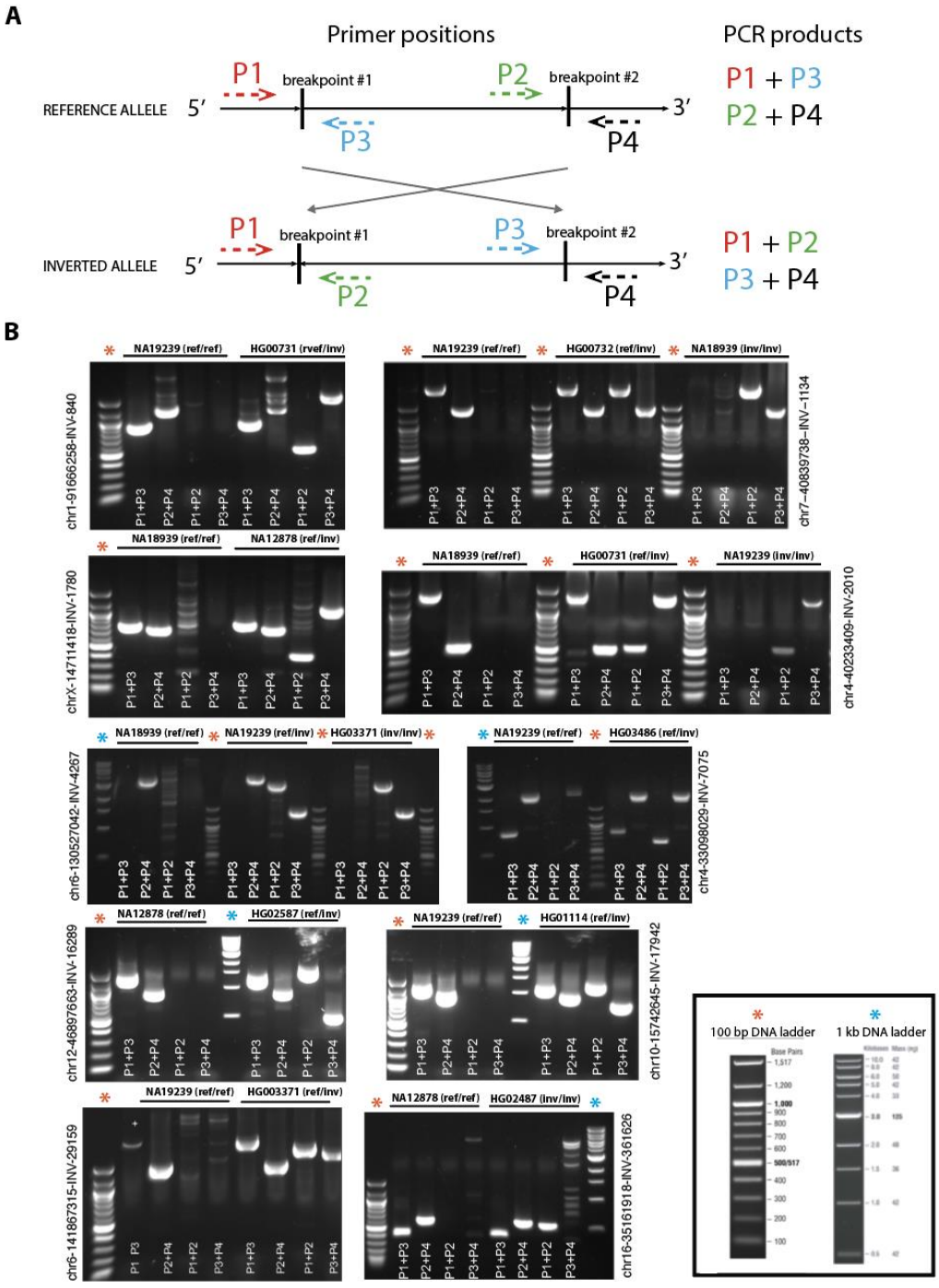
A schematic representing a read-simulation-based approach devised to determine the mappability of a genomic region, which was used for normalizing Strand-seq read counts before making a variant call by ArbiGent. Simulated paired end reads are created for each position across the reference genome. These read pairs are then mapped onto the reference genome in the same way as done for the short-read Strand-seq data. Resulting alignments are evaluated to quantify the mappability of each 100 bp region as the fraction of correctly mapped reads. Bins where the number of spurious simulated reads exceeds the correctly mapped ones by a

factor of 0.1 or greater are discarded. For each arbitrary segment to be genotyped, the Strand-seq reads coming from each 100 bp bin contained in the segment are normalized individually using the mappability-based normalization factor and aggregated.



Supporting data figure for the STAR Methods section: Inversion site verification using ONT reads. Figure xviii: ONT long-read support.

A) A horizontal barplot showing for each sample the number of inversion calls (HG00733: n=137, NA19240: n= 131 and NA24385: n=132) supported by Oxford Nanopore (ONT) reads (TRUE - yellow) and those that are not (FALSE - blue) stratified by inversion genotype (HOM - homozygous inversion, HET - heterozygous inversion). (Inversions internal to L1 sequences were not analyzed using ONT reads.) **B)** A boxplot showing the size distribution of inversions supported by ONT data (TRUE - yellow, median: 9,965 bp) and those not supported (FALSE - blue, median: 22,616 bp). **C)** An example of an inversion supported by ONT reads. Left: Visualization of long ONT read alignments of the reported inversion region (dashed lines). ONT reads mapped as split alignments are shown in full color while those mapped continuously without a change in directionality are shown in faded color. Below is the summary of bases aligned either in plus ('+') or minus ('-') orientation to the reference genome (GRCh38). Right: A barplot reporting the fraction of bases aligned in plus and minus orientation over the inverted region (inside of dashed lines) and those outside of the inverted region (outside of dashed lines).

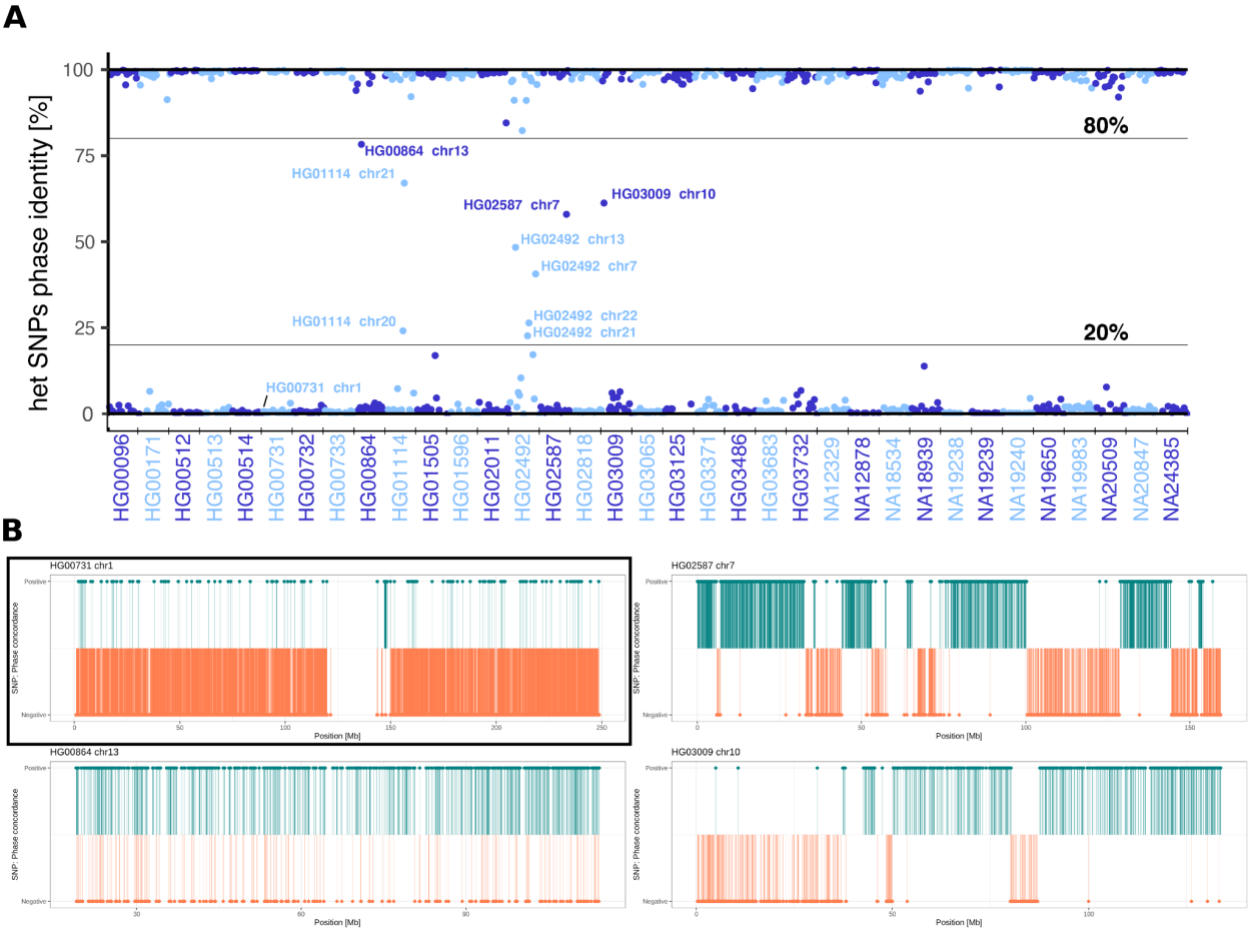


Supporting data figure for the STAR Methods section: PCR validation.

Figure xix: PCR validation.

A) Schematic view of the expected primer positions and orientation in reference and inverted alleles. Each inversion breakpoint is spanned by a pair of primers on opposing ends, leading to expected PCR products for primer pairs "P1/P3" and "P2/P4" in reference haplotypes, and "P1+P2" and "P3+P4" in inverted haplotypes. **B)** PCRs testing genotypes of 10 randomly selected sequence-resolved balanced inversions. Molecular weight markers are indicated with a blue or orange asterisk, respectively. We successfully validated both breakpoints for 9/10 inversions and one inversion breakpoint for the tenth event. One primer pair designed to test an

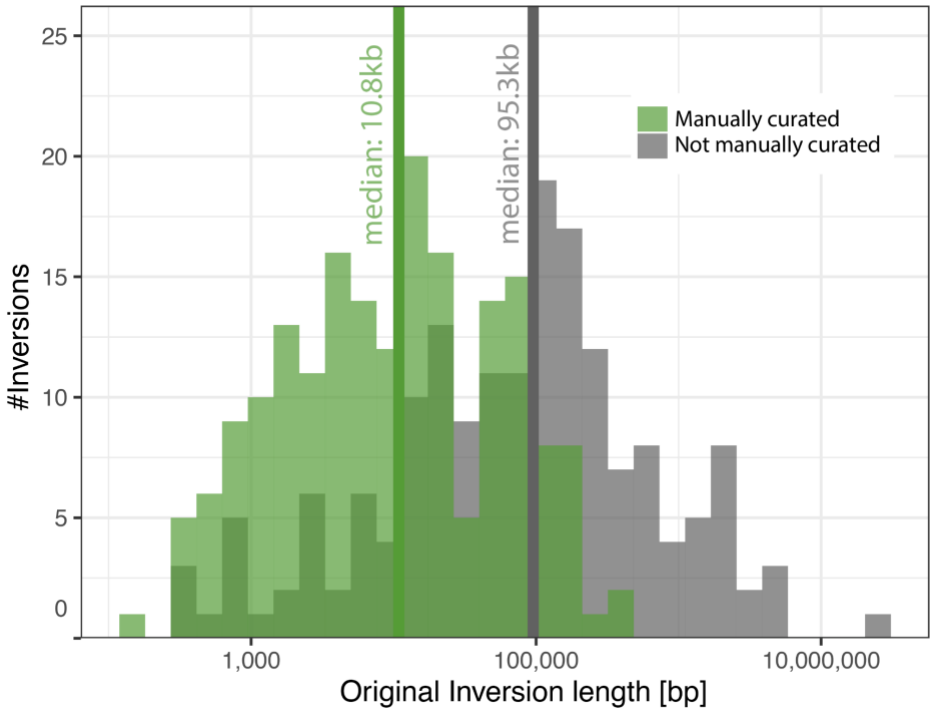
inversion (“P1+P3”, chr6-130527042-INV-4267) failed to produce a PCR product, perhaps owing to additional sequence complexities present within the respective inversion locus.



Supporting data figure for the STAR Methods section: Inversion genotyping and phasing with ArbiGent.

Figure xx: Comparison between Strand-seq-based and assembly-based (PAV) inversion phasing.

A) Comparison of phased heterozygous (het) SNP identity between calls made by Strand-seq (StrandPhaseR) and PAV. With the two phasing approaches conducted independent of each other, we expect het SNPs per chromosome to be either close to 100% ($h1_{StrandSeq} = h1_{PAV}$) or 0% ($h1_{StrandSeq} = h2_{PAV}$), which suggests concordance or discordance with respect to phasing. Inversion genotypes derived from Strand-Seq data on discordant chromosomes were subsequently flipped in phase to match the haplotype assignment used for PAV. Chromosomes with a het-SNP identity between 20-80% were considered outliers and were not phase-adjusted. **B)** Visualization of phasing from one successful (top left) and three outlier chromosomes. Each lollipop represents one SNP; color and direction of the lollipop indicate phase agreement (yes/no).



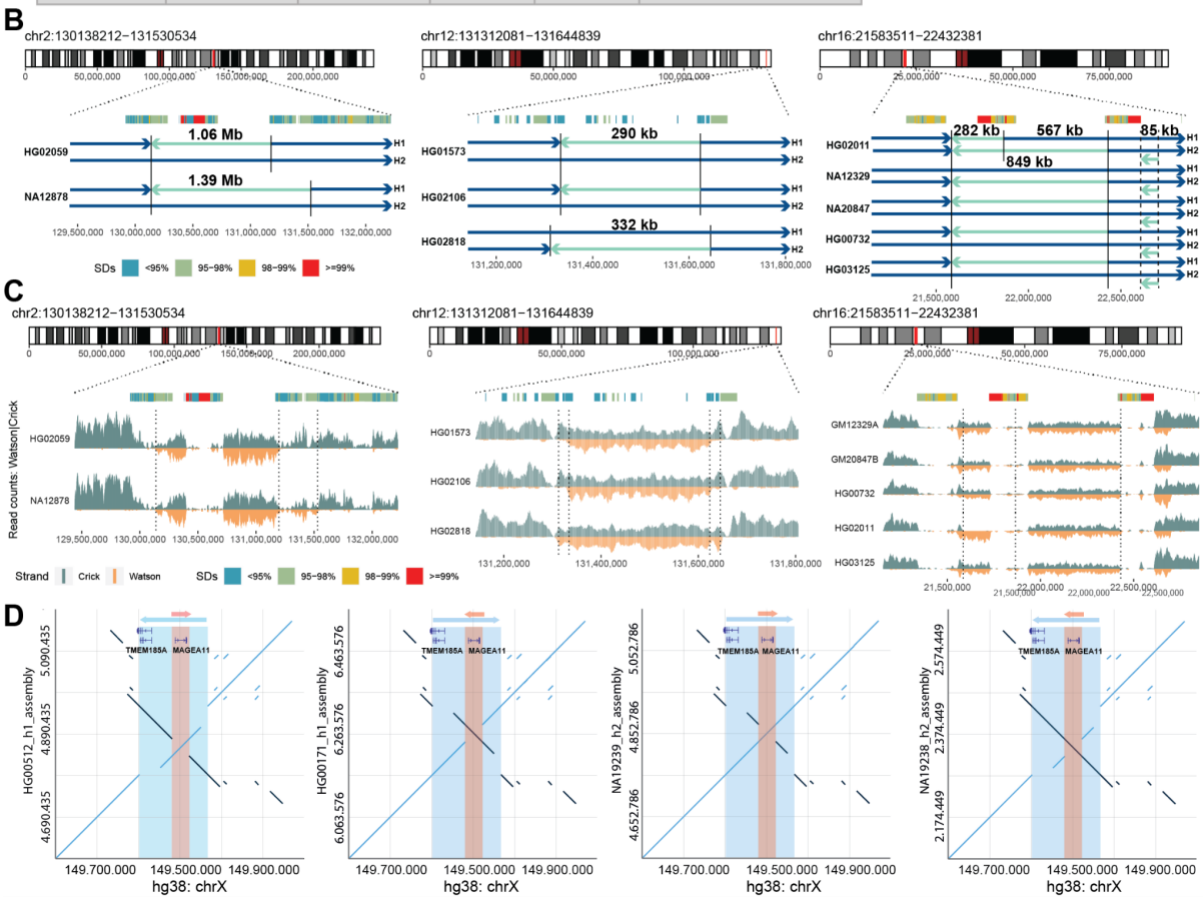
Supporting data figure for the STAR Methods section: Inversion refinement with dot plots and global genome alignments.

Figure xxi: Manual inversion curation using phased genome assemblies.

Size distribution of inversions stratified by whether or not they could be manually curated using genome assemblies from (Ebert et al., 2021), for inversions outside of L1-internal sequences. Long inversions were typically not accessible in the phased genome assemblies, since they fell into assembly breaks primarily caused by SDs.

A

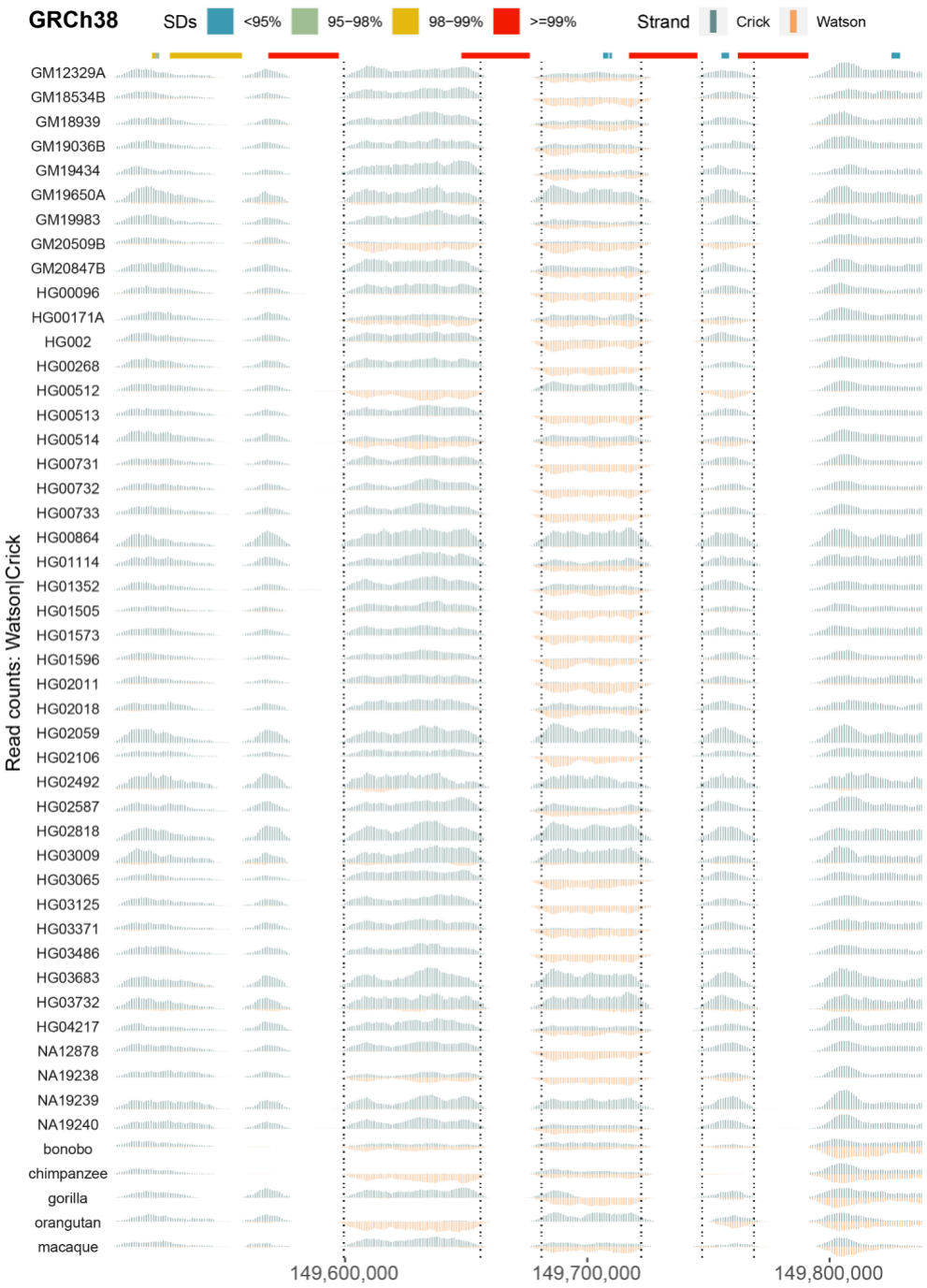
seqnames	start	end	width	position	id
chr2	130138212	131200602	1062391	nested	breakpoint_reuse_1
chr2	130138212	131530534	1392323	outer	breakpoint_reuse_1
chr12	131333943	131623807	289865	nested	breakpoint_reuse_2
chr12	131312081	131644839	332759	outer	breakpoint_reuse_2
chr16	21864964	22432381	567418	nested	breakpoint_reuse_3
chr16	21583511	22432381	848871	outer	breakpoint_reuse_3
chrX	149681035	149722249	41215	nested	nested inversion
chrX	149599490	149768782	169293	outer	nested inversion



Supporting data figure for the STAR Methods section: Detection of nested inversions and inversions showing signs of breakpoint reuse.

Figure xxii: Inversion breakpoint reuse and nested inverions.

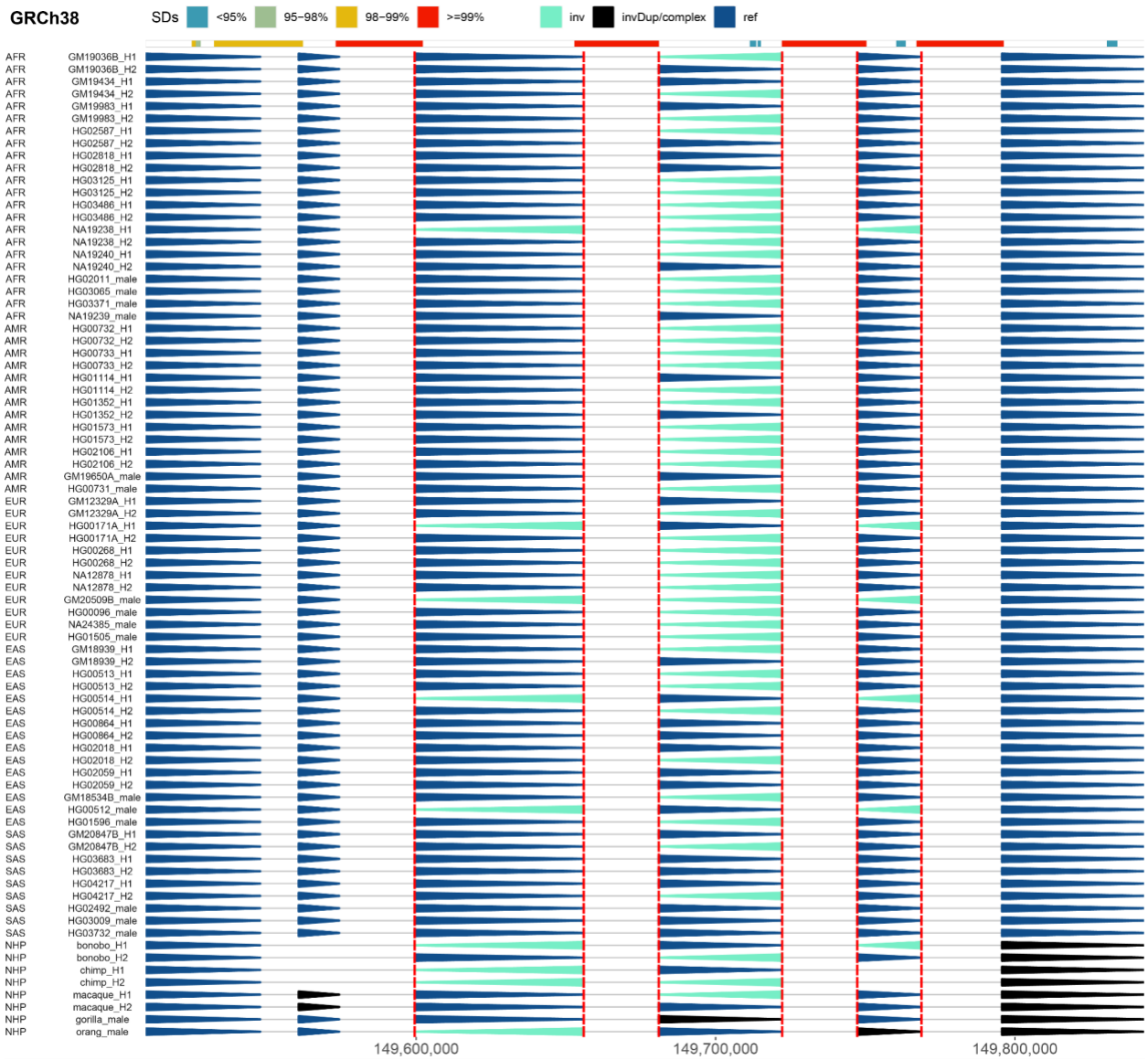
A) GRCh38 coordinates of inversion with breakpoint reuse or defined as nested. Field 'Position' denotes if an inversion is 'nested' within a larger inversion range, denoted as 'outer'. **B)** Regions (n=3) with disparate inversion breakpoint reuse are highlighted on a chromosome-specific ideogram by a red rectangle. The panels below zoom into the respective regions: Top annotated SDs, colored based on sequence identity. Samples with detected inversions are shown as sample-specific haplotypes (H1 and H2) with direct (dark blue) and inverted (bright blue) regions depicted as arrows. Vertical solid lines highlight detected inversion breakpoints. **C)** Read coverage profiles of Strand-seq data for three different regions with signs of balanced inversion breakpoints shifting. Strand-seq reads are summarized as binned (bin size: 10 kbp, step size: 1 kbp) read counts represented as bars above (teal; Crick read counts) and below (orange; Watson read counts) midline. Region with roughly equal coverage of Watson and Crick count represents a heterozygous inversion as only one homologue is inverted with respect to the reference while region with reads aligned only in Watson orientation represents an homozygous inversion. Vertical dotted lines highlight the inversion breakpoints. Each inverted region is highlighted on chromosome-specific ideogram by a red rectangle. **D)** Dot plots visualizing sequence alignments between GRCh38 and *de novo* phased assemblies in the region of a nested inversion on chromosome X. The outer inversion locus is highlighted in blue, the inner in red. All four possible combinations of nested inversion (inv) states (ref/ref, ref/inv, inv/ref, inv/inv; with ref. for reference orientation) were seen, suggesting inversion recurrence.



Supporting data figure for the STAR Methods section: Detection of nested inversions and inversions showing signs of breakpoint reuse.

Figure xxiii: Strand-seq read distribution at a structurally complex region on chromosome X.

Top: Reference genome-specific SD annotation for a given region represented as a set of rectangles colored by sequence identity of each SD. Underneath are the read coverage profiles of Strand-seq data over a given region summarized as binned (bin size: 10 kbp, step size: 1 kbp) read counts represented as bars above (teal; Crick read counts) and below (orange; Watson read counts) midline. Region with roughly equal coverage of Watson and Crick reads represents a heterozygous inversion as only one homologue is inverted in respect to the reference. Region with reads aligned only in Watson orientation represents a homozygous inversion as both homologs are inverted in respect to the reference while region with purely Crick reads is represented by both homologs being in direct orientation in respect to the reference. Vertical dotted lines highlight the inverted regions.



Supporting data figure for the STAR Methods section: Detection of nested inversions and inversions showing signs of breakpoint reuse.

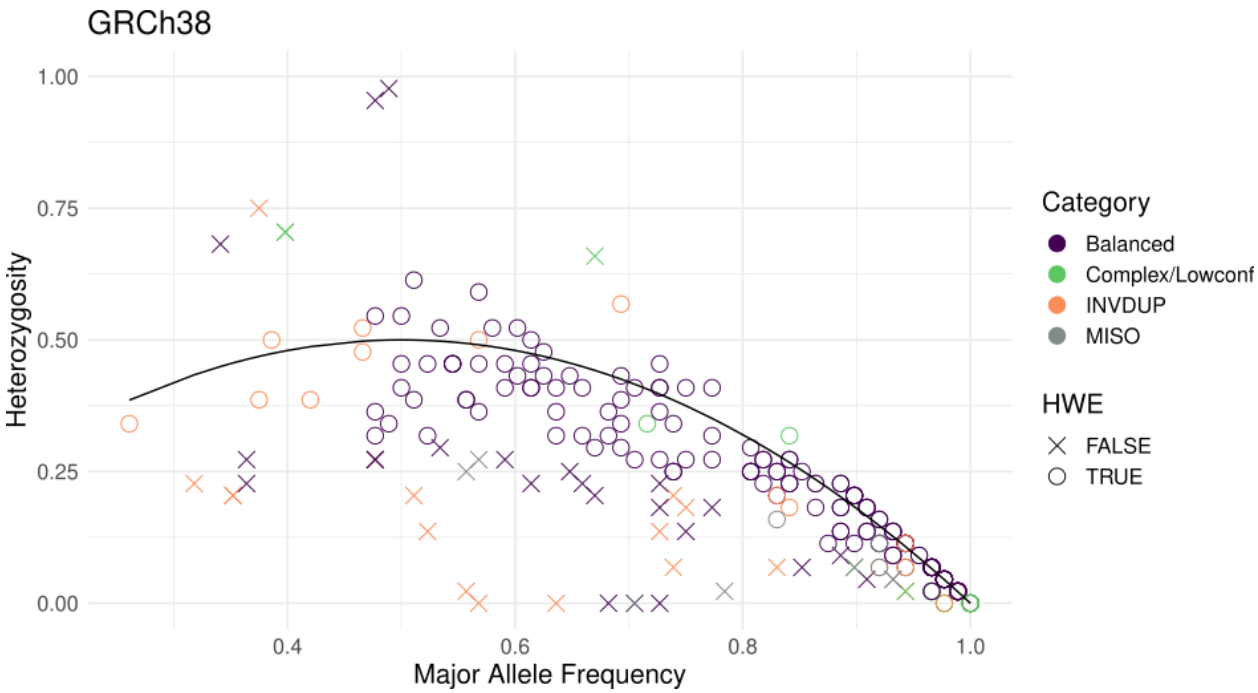
Figure xxiv: Inversion phasing at a structurally complex region on chromosome X.

Two arrowhead plots showing the inverted status of each defined region reported as colored arrowheads (dark blue - direct (ref), bright blue - inverted (inv), black - inverted duplication/complex genotype, see the legend) for the region chrX:149509687-149843091. Top: Reference genome-specific SD annotation colored by sequence identity. Inverted regions are highlighted by vertical red lines. Deleted chunks of the genome in bonobo and chimpanzee are visible as empty spots where other samples show phased arrowheads. Note: Phase between haplotype 1 (H1) and haplotype 2 (H2) in this plot might not necessarily match ArbiGent genotypes, as inverted regions in this plot were genotyped *de novo* using Strand-seq data presented as raw read counts in the previous figure.

Supporting data figure for the STAR Methods section: Quality control of recurrent inversions detected by the tiSNP-based method.

Figure xxv: Flanking inverted repeat length versus fraction of tiSNPs.

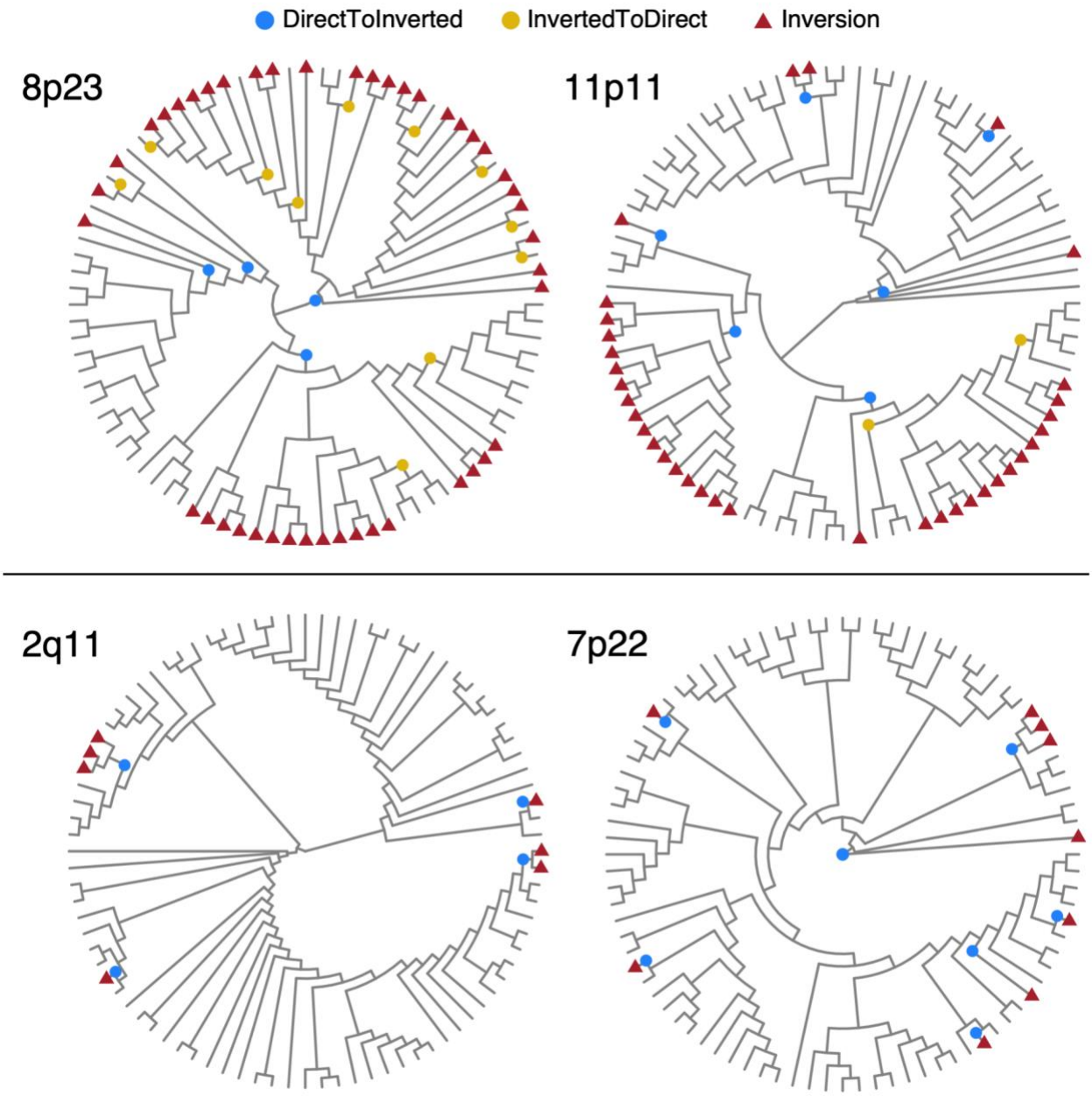
Plot showing relationship between the fraction of tiSNPs observed at each locus and length of the inverted repeat at its flanks. All 252 balanced inversions tested for recurrence using the tiSNP-based approach are shown. With increasing length of the flanking inverted repeats, the fraction of tiSNPs clearly appears to increase. Moreover, for (49) inversions where the tiSNPs-based approach detects evidence of recurrence (TRUE), an enrichment for longer flanking inverted repeats is observed.



Supporting data figure for the STAR Methods section: Hardy-Weinberg equilibrium test and multi-allelic sites.

Figure xxvi: Hardy-Weinberg equilibrium (HWE) evaluation.

GRCh38-based inversion callset of inversions outside of L1-internal sequences (n=399). For generating this plot, we removed inversions on the sex chromosomes (60), followed by those with missing sample genotypes (64). As a consequence, 275 inversions were tested and 224 (81.45%) of them were in Hardy-Weinberg equilibrium. Balanced - simple balanced inversion; Complex/Lowconf - Complex or low confidence inversion; INVDUP, Inverted duplication; MISO - Likely misorientation.

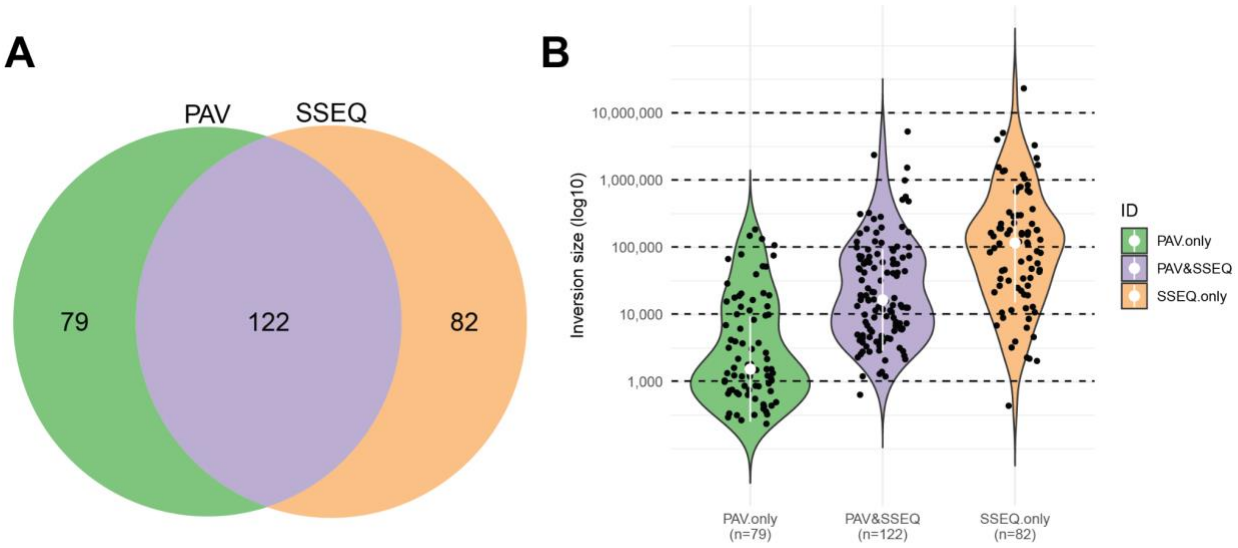


Supporting data figure for the STAR Methods section:Analysis of the genetic architecture of recurrent inversion loci

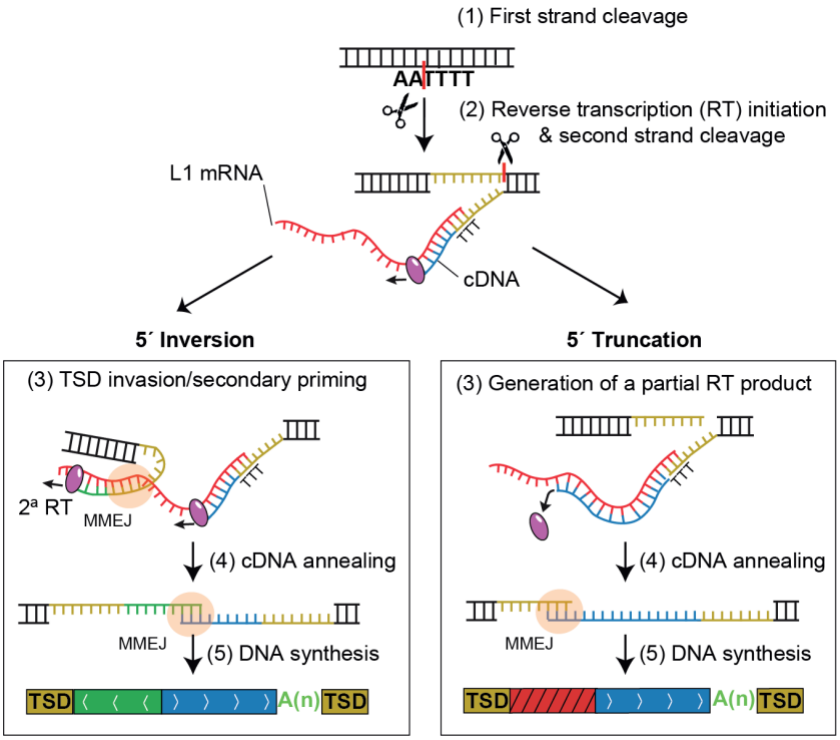
Figure xxvii: Inferred trees of recurrent inversions with evidence for serial (8p23 and 11p11) and non-serial (2q11 and 7q22) toggling.

On each tree, an inversion event can either change the sequence from direct to inverted orientation with respect to the reference genome (blue) or vice versa (yellow). Serial toggling is thus defined as recurrent inversion events affecting the same segment along the same lineage multiple times leading to a segment toggling back and forth in orientation on a given tree. By comparison, a recurrent inversion affecting the same segment in different parts of the tree could be defined as 'parallel' or 'non-serial' toggling. We find that 23 out of the 32 recurrent inversions on the autosomes and chromosome X show evidence for serial toggling (similar to the inversions at 8p23 and 11p11; top panels), while the remaining 9 loci exclusively harbor non-serial toggling events (similar to the inversions at 2q11 and 7q22; bottom panels). We caution that "incomplete sampling" of our diversity panel (consisting of 82 haplotypes) in the trees generated means that several cases now appearing as non-serial toggling are likely to eventually show evidence of serial toggling once inversion surveys include more samples in the future.

Supporting data for claims made in discussion



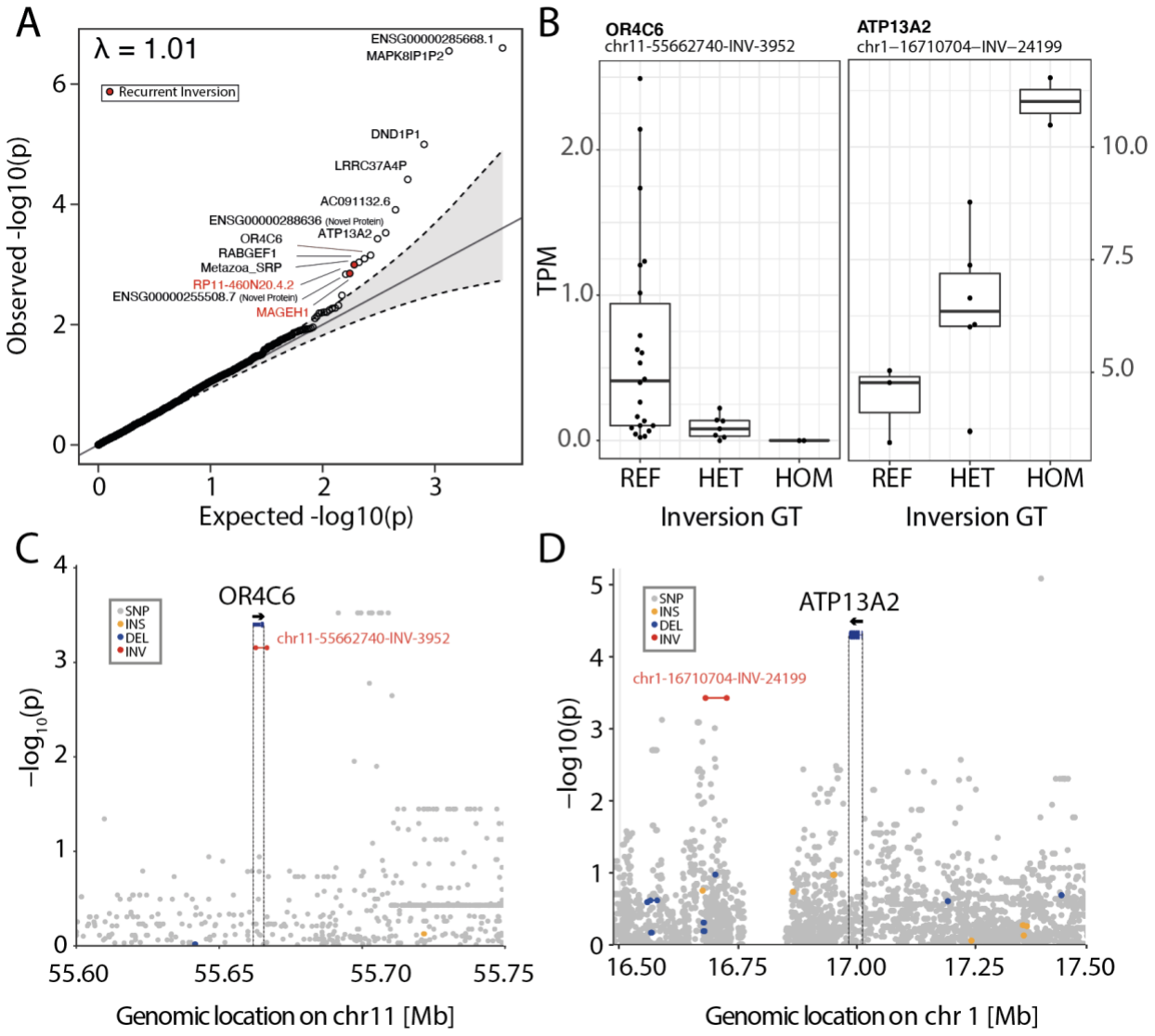
Supporting data figure for Figure 1E and discussion point about the Limitations of the study.
Figure xxviii: Detection limits of assembly-based and Strand-seq-based inversions.
A) A Venn diagram showing the set of inversions detected by both assembly-based approach (PAV) and Strand-seq (SSEQ) as well as inversions detected by each method only, for inversions outside of L1-internal sequences. (Bionano-based inversion calls not shown in this figure.) **B)** Size distribution of inversions from each intersection of the Venn diagram from (A). White dots show the median of each distribution.



Supporting data figure for discussion point about inversions internal to L1s occurring via twin-priming

Figure xxix: Mechanistic models for inversion and truncation of L1 sequences during retrotransposition.

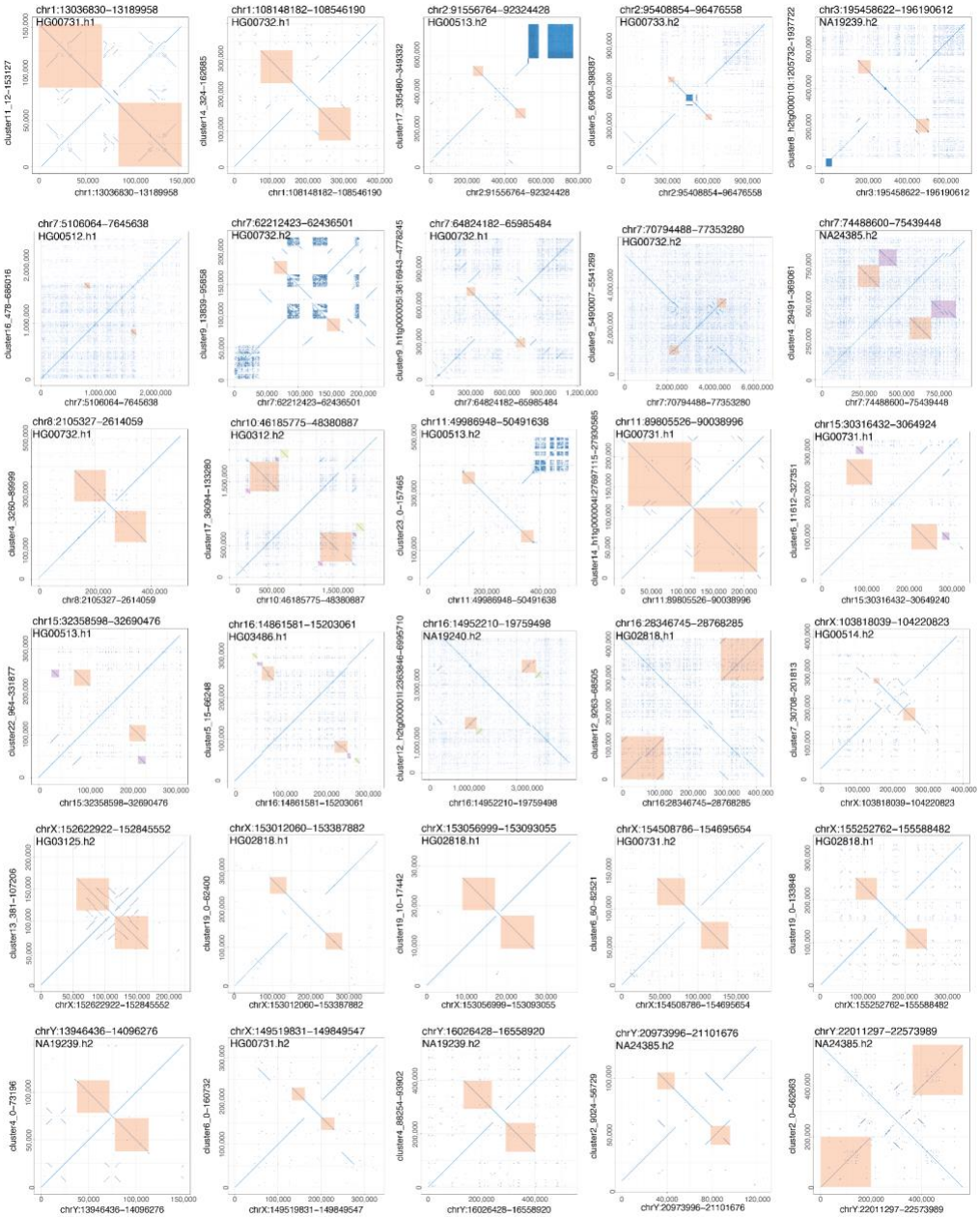
L1 retrotransposition is initiated by the cleavage of the first DNA strand by the L1-encoded endonuclease at the target degenerated motif of 3'-AA|TTTT-5' (Cost and Boeke, 1998; Cost et al., 2002; Luan et al., 1993) (1). The resulting T-rich single-stranded DNA serves for the annealing of the L1 mRNA poly(A) (Monot et al., 2013) and provides a 3 '-hydroxyl that primes reverse transcription (RT) of the L1 transcript (2). **Left:** 5' inversion model. After the cleavage of the second DNA strand, by a currently unknown molecular mechanism, the derived single-stranded overhang at the target site anneals internally to the L1 transcript likely through microhomology-mediated end joining (MMEJ) (Kojima, 2010; Ostertag and Kazazian, 2001; Zingler et al., 2005), serving as a primer for a secondary RT reaction (3). Then, the inverted and non-inverted cDNA products are joined through MMEJ (4) and retrotransposition finalizes after the remaining DNA synthesis is completed and the second L1 DNA strand is joined to the target site by the action of a cellular ligase (5). **Right:** 5' truncation model. Primarily due to the action of cellular DNA damage repair pathways (Coufal et al., 2011; Suzuki et al., 2009), L1 reverse transcription prematurely finishes before reaching the 5' end of L1 mRNA (3). Then, the nascent partial RT products are ligated via MMEJ to the 3' overhang at the target site originating from the second strand nick (Kojima, 2010; Zingler et al., 2005) (4) and retrotransposition finalizes as described for the 5' inversion model. In both models, the engagement of MMEJ repair machinery (highlighted as red circles) is suggested by the identification of microhomology patches at the majority of breakpoint junctions. TSD, target site duplication.



Supporting data figure for discussion point about functional impact of inversion polymorphisms on gene expression.

Figure xxx: eQTL analysis of inversions outside of L1-internal sequences and gene expression in 33 samples.

A) A qq-plot showing the most observed vs. expected p-values and the genes most strongly affected by an inversion. **B)** Expression values for two previously described inversion eQTLs, *OR4C6* and *ATP13A2*. **C)** Manhattan plot for the *OR4C6* locus. The gene is overlapped by an inversion (red). **D)** Manhattan plot for ATPase Cation Transporting protein *ATP13A2*, located ~250 kbp downstream of inversion chr1-16710704-INV-24199. The inversion overlaps an enhancer-rich region with further high-scoring SNP variants. REF, reference (direct) orientation; HET, heterozygous inverted; HOM, homozygous inverted.



Supporting data figure for discussion points about inversion breakpoint resolution.

Figure xxxi: Dotplots of 30 sequence-resolved recurrent inversion loci.

Hifiasm assemblies allowed us to visualize putative sequence structures of 30 out of 40 recurrent inversions, including their flanks (shown above). We identified assemblies of at least one direct and one inverted haplotype for 23/40 recurrent inversion loci. For 7/40 recurrent inversion loci, only one state was sequence assembled (reference state: 6 loci; inverted state: 1 locus); whereas for the remaining 10/40 recurrent inversions, we were neither able to assemble the direct nor the inverted state. Each panel depicts alignments between the GRCh38 reference genome (x-axis) and a selected *de novo* assembly from a haplotype carrying the respective inversion (y-axis, sample name indicated in the header). Flanking SDs are highlighted with colored boxes and, in each case, correspond to the putative breakpoint interval. For each locus, we generated dotplots from 28 hifiasm assemblies, which we have made accessible as part of this resource (**Data and Code Availability**). As a note, haplotype assignments depicted (h1/h2) to visualize the assemblies are based on the hifiasm algorithm and not synchronized to haplotype assignments used elsewhere in this study.