**HTGTS**

Libraries were prepared as described previously and sequenced (Illumina MiSeq). Reads from demultiplexed FASTQ files were aligned to the genome build mm10/GRCm38 through Bowtie2, and processed through the HTGTS pipeline.

**Recurrent double-stranded break clusters calling**

**Libraries**

For RDCs detection we combined samples from published data (Wei et al. PNAS 2018, 59 APH treated and 59 DMSO control) and newly generated samples (64 APH treated and 51 DMSO control). This in total resulted in 4147476 junctions. Only APH-treated inter-chromosomal junctions where used for the final list of RDCs, other subsets (Intra-chromosomal junctions and DMSO control libraries) where used to compare the performance RDC-prediction algorithm under different conditions.

**Data clean-up**

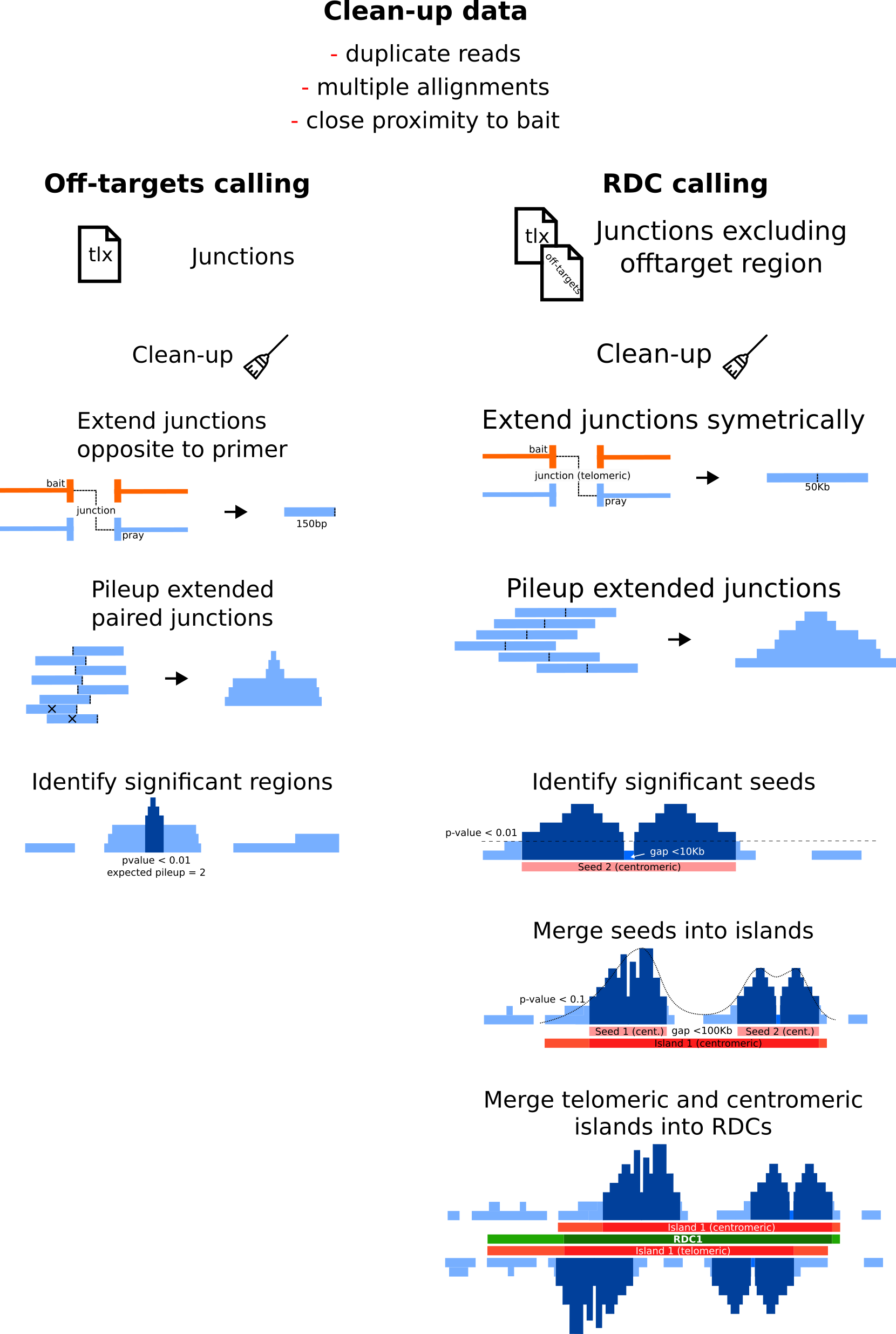
Junctions from bait region (±6Mb, 2690102 junctions) and from bait off-target regions (±50Kb, 58931 junctions) were removed. Also, duplicate reads (176774 junctions) and reads that have multiple alignments (292903 junctions) were removed. Additionally, intra-chromosomal junctions from libraries where bait chromosome was genetically modified were excluded from subset used for RDC calling (270332 junctions). The final dataset contained 1116629 reads distributed across 4 conditions: APH-Inter (517947), APH-Intra (152146), DMSO-Inter (365270) and DMSO-Intra (81266).

**Offtarget calling**

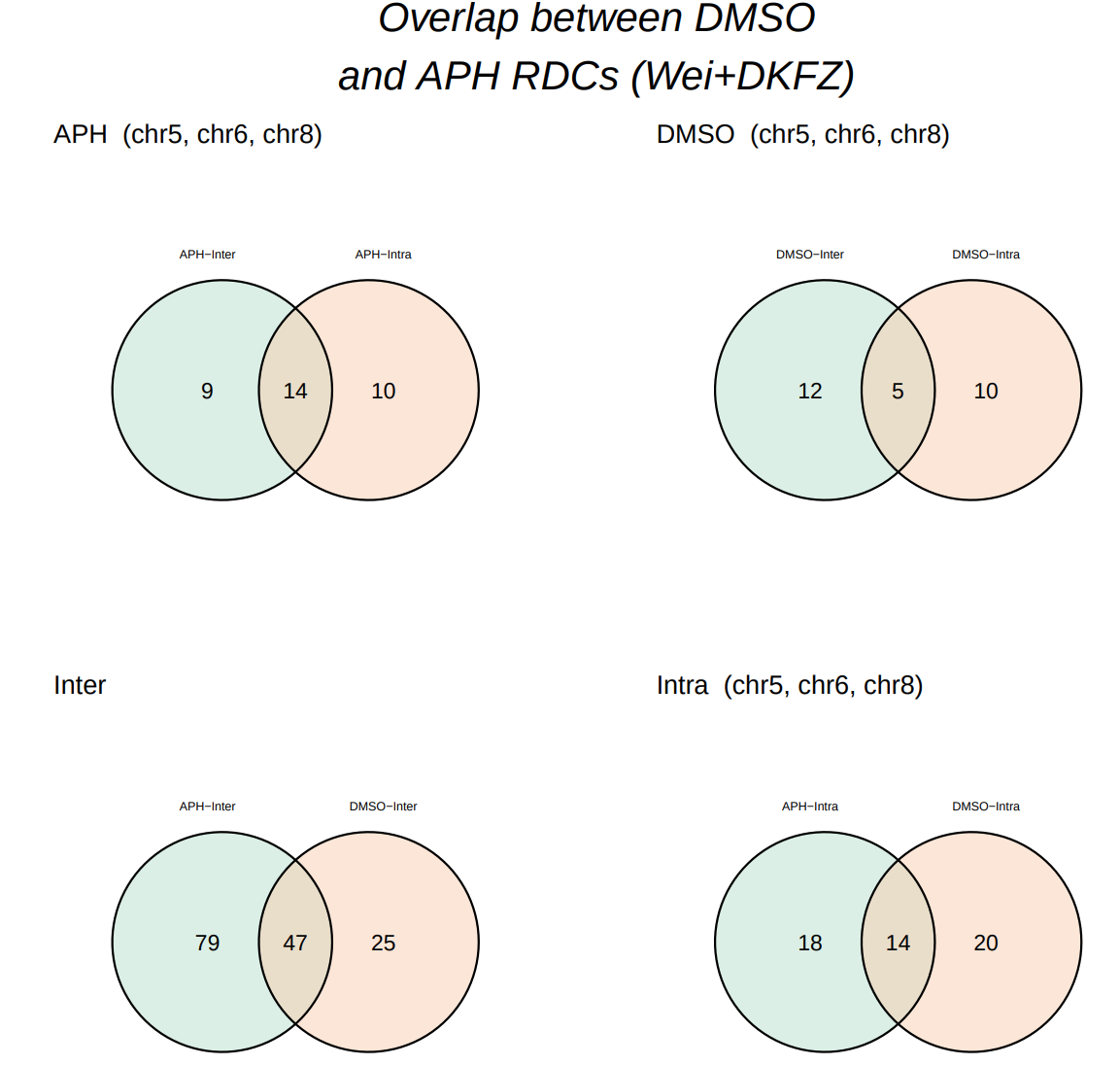
Junctions where extended 150bp in opposite direction from translocated prey and only junctions that are overlapping with opposite direction junction where kept so that there is equal amount of centromeric and telomeric oriented junctions. A pileup was calculated from the resulting junctions and a poisson distribution (mean=2) was used to calculate significance for each interval. Continous significant regions (p-value<0.01) where further filtered to contain more telomeric translocations upstream from offtarget site and more centromeric translocations downstream (Fisher exact test, p-value<0.01)

**RDC calling**

Junctions where extended 50Kb in symmetrically in both directions and pileup was calculated for telomeric-only, centromeric-only and all junction orientations. A negative binomial model for estimating the expected pileup value for each chromosome/condition/junction-orientation triplet was derived and a p-value was calculated for each pileup value in respect to model expectation. Regions where p-value was below 0.01 where joined together (maximal gap 10Kb) to create seeds. These seeds where further joined with other seeds (maximal gap 100Kb) to form islands. Islands are extended up- and down-stream to include regions below 0.1 significance. Overlapping orientation-specific islands are further joined to form initial RDC list that is further filtered to contain at least 100Kb below 0.01 p-value and be of at least 300Kb in length when considering extended regions







There is less bias in Inter-chromosomal centromeric/telomeric junctions bias

