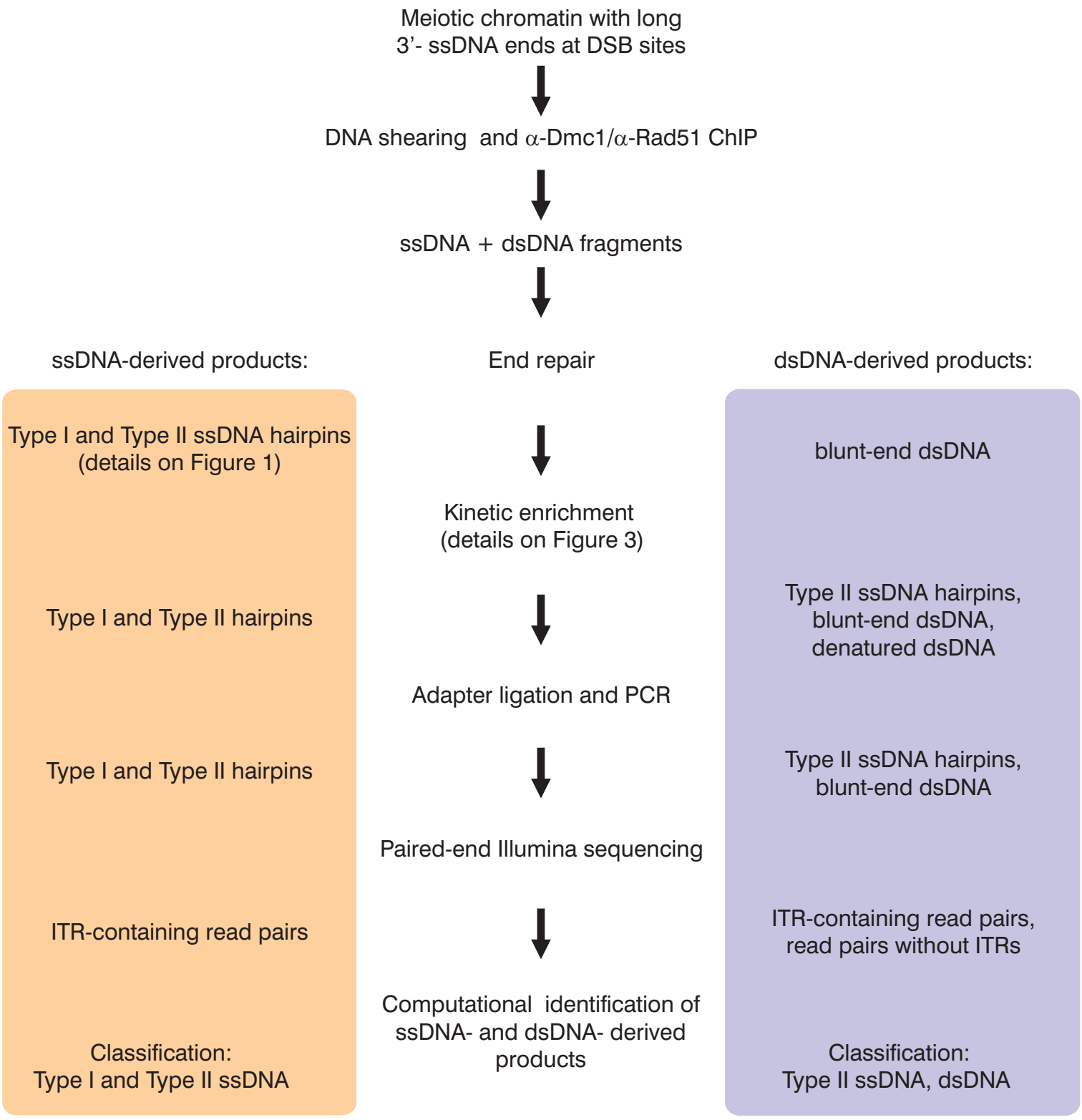
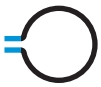
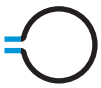
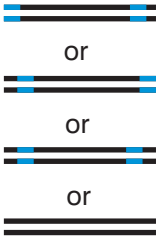
Supplemental Figure 1



Supplemental Figure 1. Workflow of SSDS. Intermediate products are listed separately for ssDNA (left) and dsDNA (right) derived fragments. Because Type II ssDNAs can be either dsDNA-derived or ssDNA-derived, they were not used for hotspot detection.

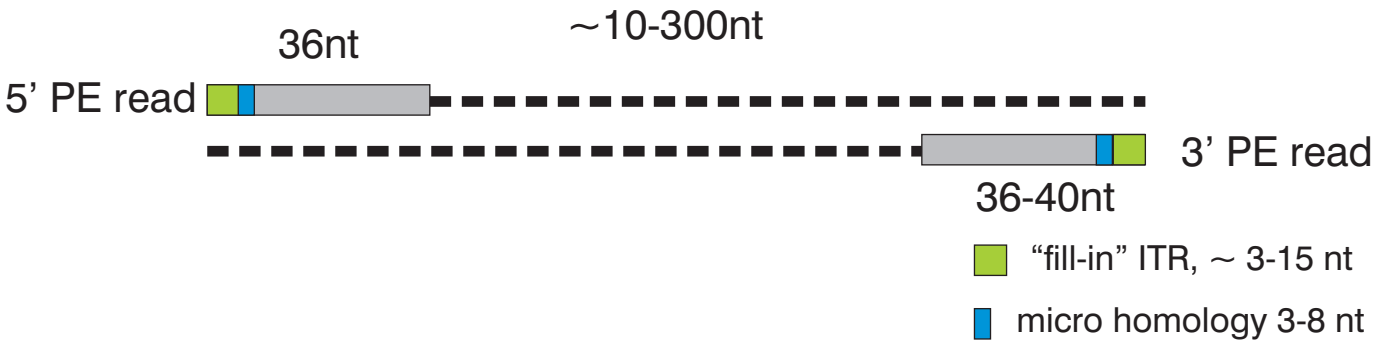
Supplemental Figure 2

Substrate		ITR	Fill-in	ssDNA hairpin			Genomic alignment	Classifi-cation
				original	after exo	after fill-in		
ssDNA		>5 nt	Yes					Type I
		>5 nt	Yes					Type I
		>5 nt	No					Type II
		>5 nt	No					Type II
dsDNA		>5 nt	N/A	dsDNA				Type II
		3-5 nt	N/A	dsDNA				N/A
		<3 nt	N/A	dsDNA				dsDNA
		N/A	N/A	dsDNA				dsDNA

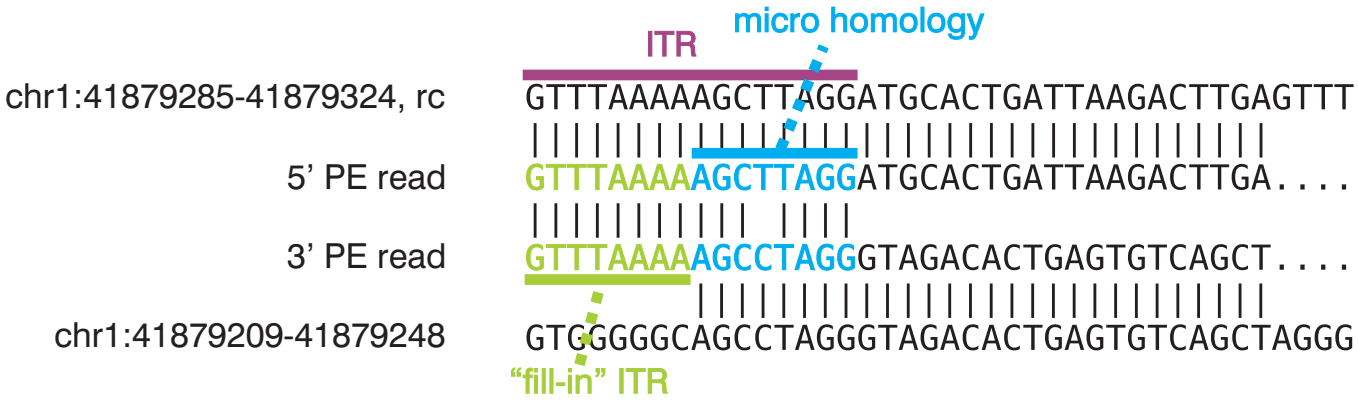
Supplemental Figure 2. Classification of ssDNA and dsDNA fragments into Type I, Type II ssDNA and dsDNA. Initial ds/ssDNA substrates are processed differentially to yield products that can be classified based on their genomic alignment and primary structure of the ends. Schematic structure of the intermediate ssDNA hairpin substrates is shown for different stages: original, following exonuclease digestion (after exo) and following polymerase fill-in (after fill-in). Blue bars represent micro-homology regions, green bars represent fill-in ITR regions, orange - 3' overhangs digested by exonuclease activity. ssDNAs are shown in a 5' (left) to 3' (right) orientation. As we perform paired-end sequencing, both reads are shown aligning to the genome. The first read is shown on the left, the second read on the right. The genomic DNA is illustrated in grey. Intermediate structures of dsDNA substrates are not shown.

Supplemental Figure 3

A



B

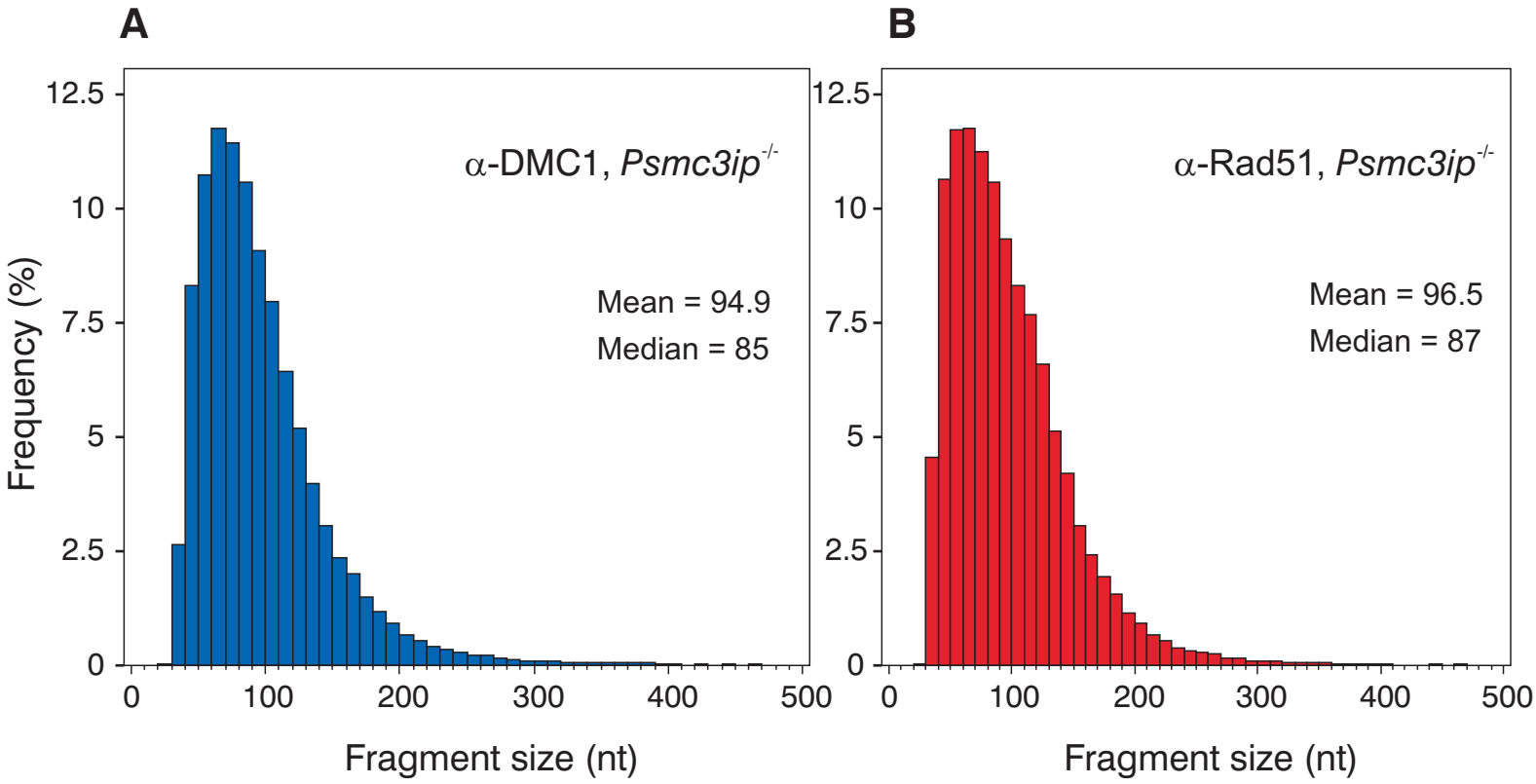


Supplemental Figure 3. Detailed view of an individual ssDNA-originated library fragment.

(A) Schematic structure of individual ssDNA-derived library fragments.

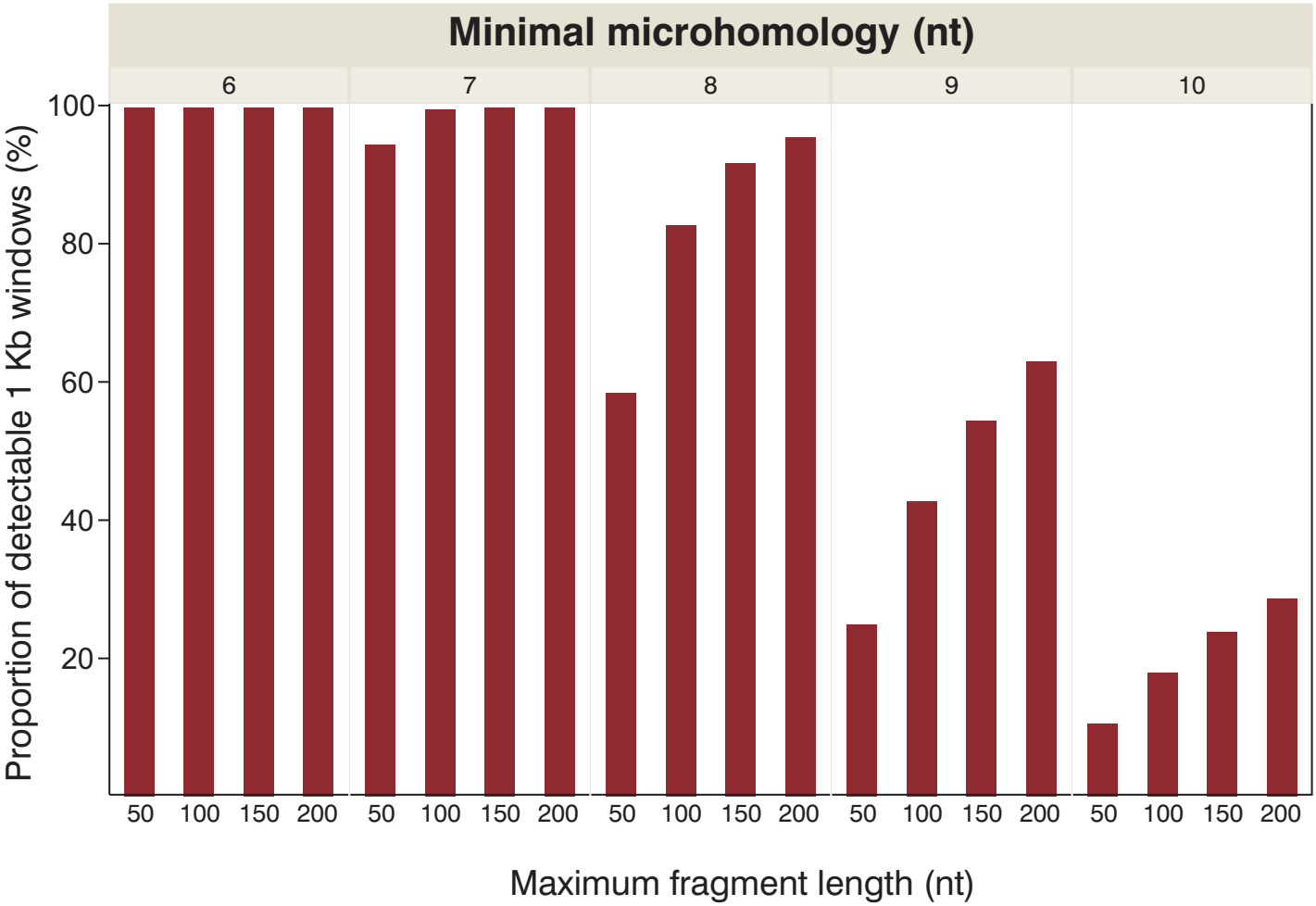
(B) An example of a ssDNA-derived read pair and its alignment to the genome.

Supplemental Figure 4



Supplemental Figure 4. Fragment length distribution of type I ssDNAs in α–RAD51 and α–DMC1 *Psmc3ip*$^{-/-}$ libraries.

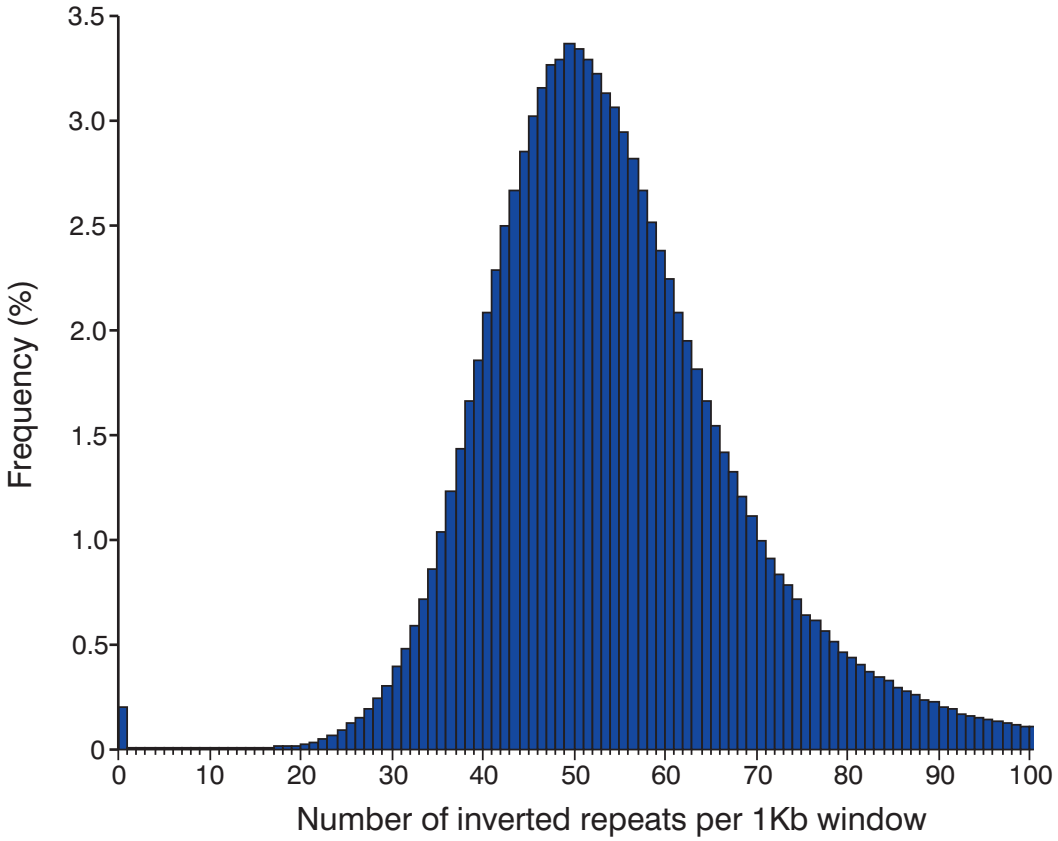
Supplemental Figure 5



Supplemental Figure 5. Hotspots in the most genomic regions can be detected through the formation of the short ssDNA hairpins. Proportion of 1 Kb windows containing at least one inverted repeat of a given length (6-10 nt microhomology) within an indicated distance (x-axis). These potential hairpins represent the putative targets of our method. We consider a given window to be "detected" if it contains at least one properly positioned inverted repeat (potential hairpin). Calculations were performed for mouse chromosome 2.

Supplemental Figure 6

A



B

Quantiles

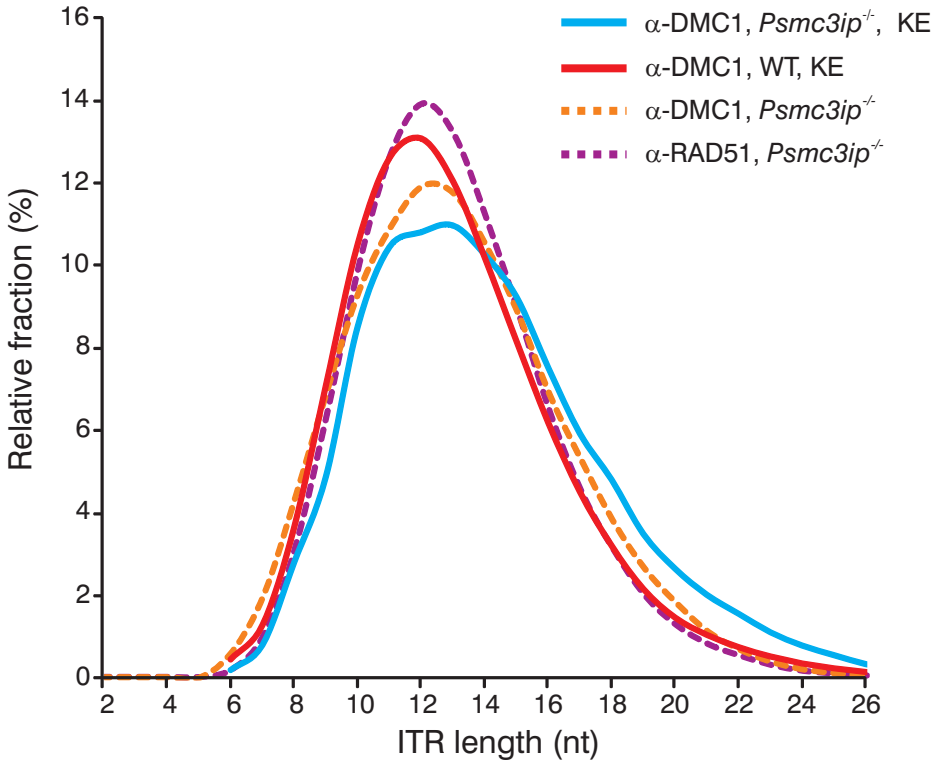
100.0%	maximum	1211
99.5%		158
97.5%		101
90.0%		74
75.0%	quartile	62
50.0%	median	52
25.0%	quartile	45
10.0%		39
2.5%		32
0.5%		25
0.0%	minimum	0

Moments

Mean	55.8
Std Dev	20.2
Std Err Mean	0.015
N	1787471

Supplemental Figure 6. The number of potential hairpins (inverted repeats) that can be detected by our method in 1 Kb regions of mouse genome. Histogram (A) and statistical moments (B) of closely spaced inverted repeats. We calculated the number of palindromes with stems of 6 nt or longer within 100 nt from each other per 1 Kb window. Calculations were performed for mouse chromosome 2 in 100 bp increments.

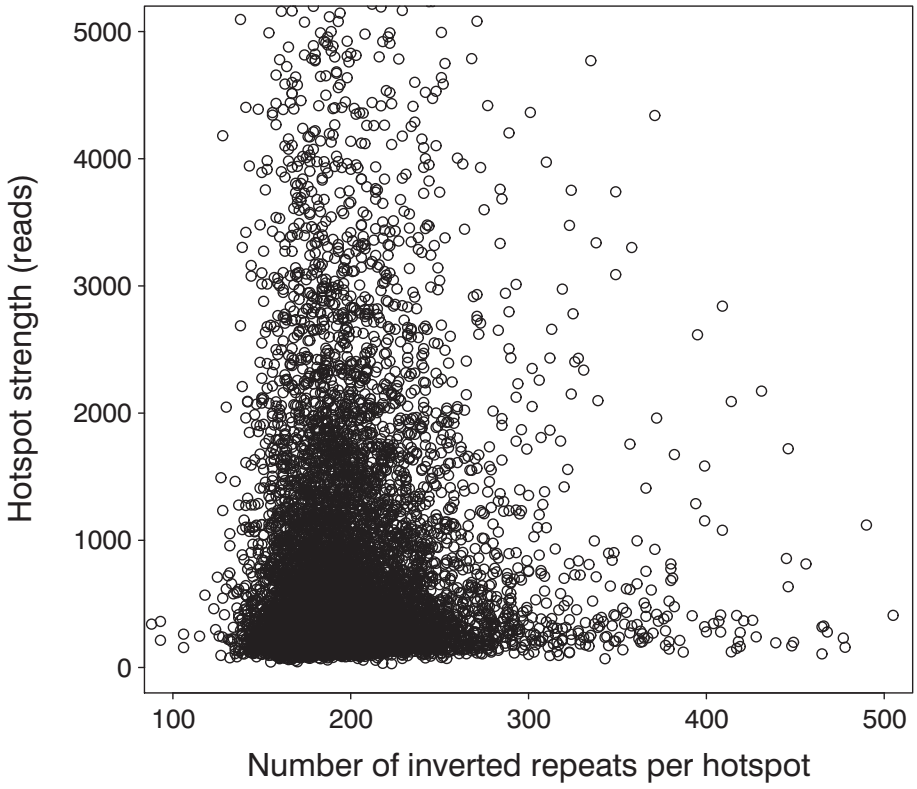
Supplemental Figure 7



Supplemental Figure 7. ITR length distributions of type I ssDNAs in KE and non-KE libraries.

Supplemental Figure 8

A



B

Minimum ITR length (nt)

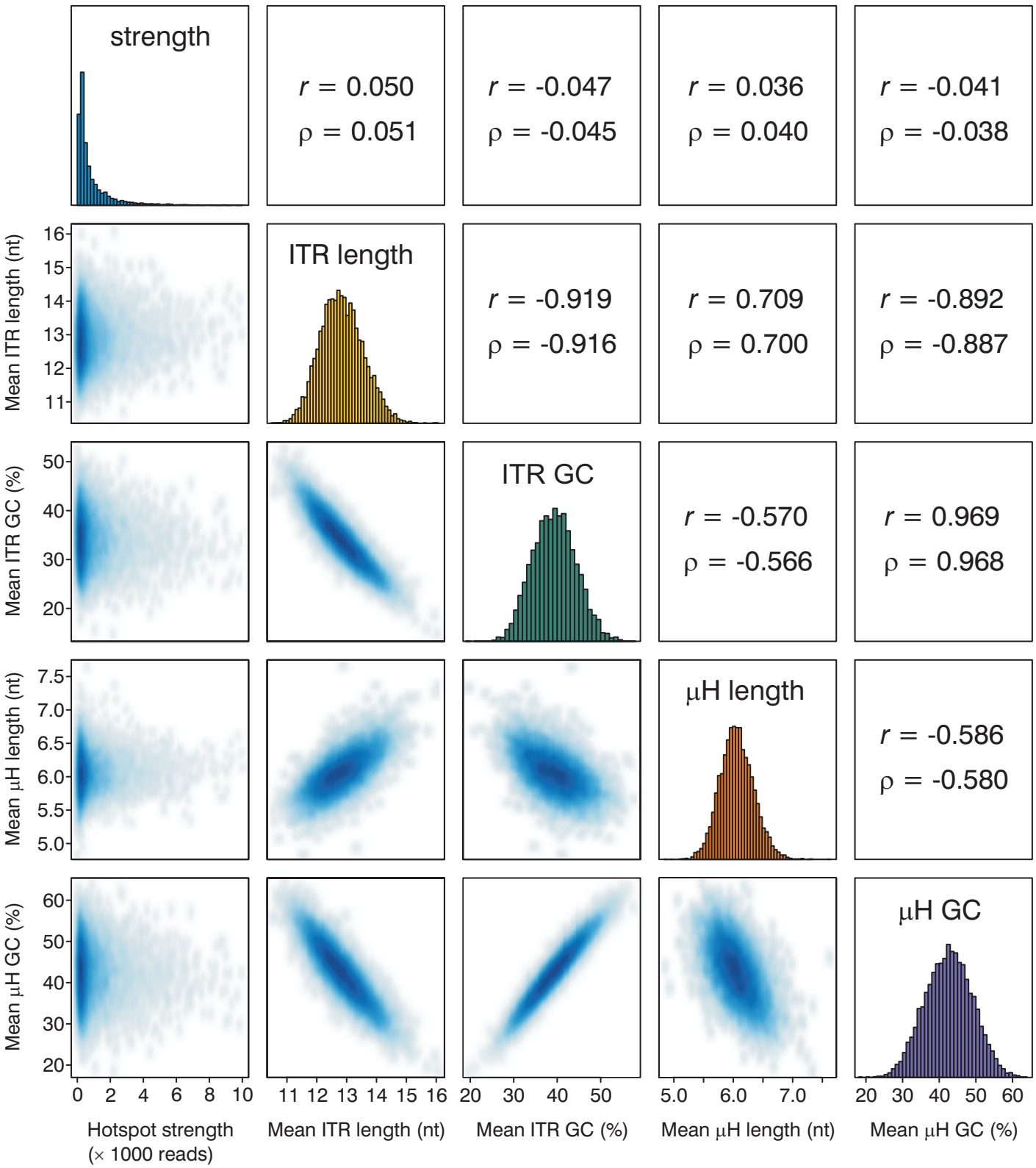
Maximum hairpin length (nt)	6	7	8	9	10
50	-0.006	-0.013	-0.014	-0.010	-0.007
100	0.003	-0.006	-0.012	-0.008	-0.008
150	0.005	-0.008	-0.013	-0.010	-0.009
200	0.005	-0.008	-0.012	-0.012	-0.011

C

Spearman	6	7	8	9	10
50	0.025	0.003	0.005	-0.003	0.004
100	0.038	0.007	0.006	0.000	-0.003
150	0.045	0.015	0.009	-0.001	-0.002
200	0.049	0.020	0.015	0.001	0.002

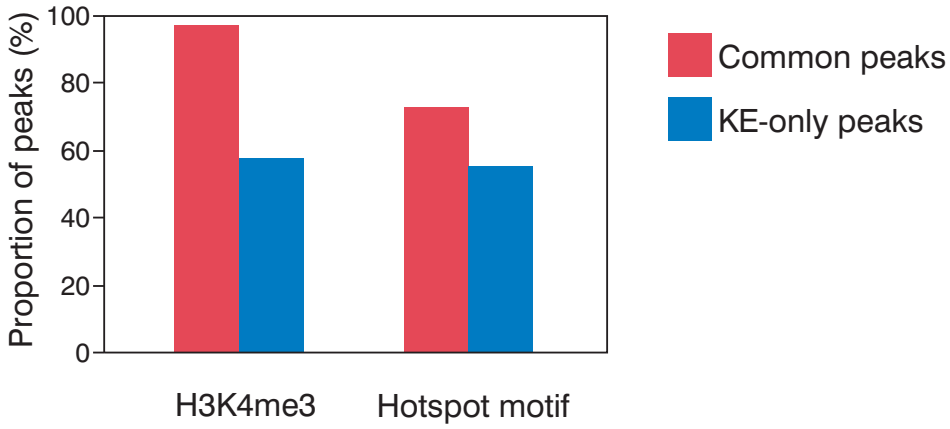
Supplemental Figure 8. The correlation of the hotspot strength with the number of inverted repeats per hotspot is very weak. (A) Scatterplot of number of reads versus number of inverted repeats (maximum fragment length = 150 bp, minimum ITR = 6 bp) in the hotspot window. Table of Pearson (B) and Spearman (C) correlation coefficients between hotspot strength and inverted repeat content for different ITR and hairpin lengths thresholds.

Supplemental Figure 9



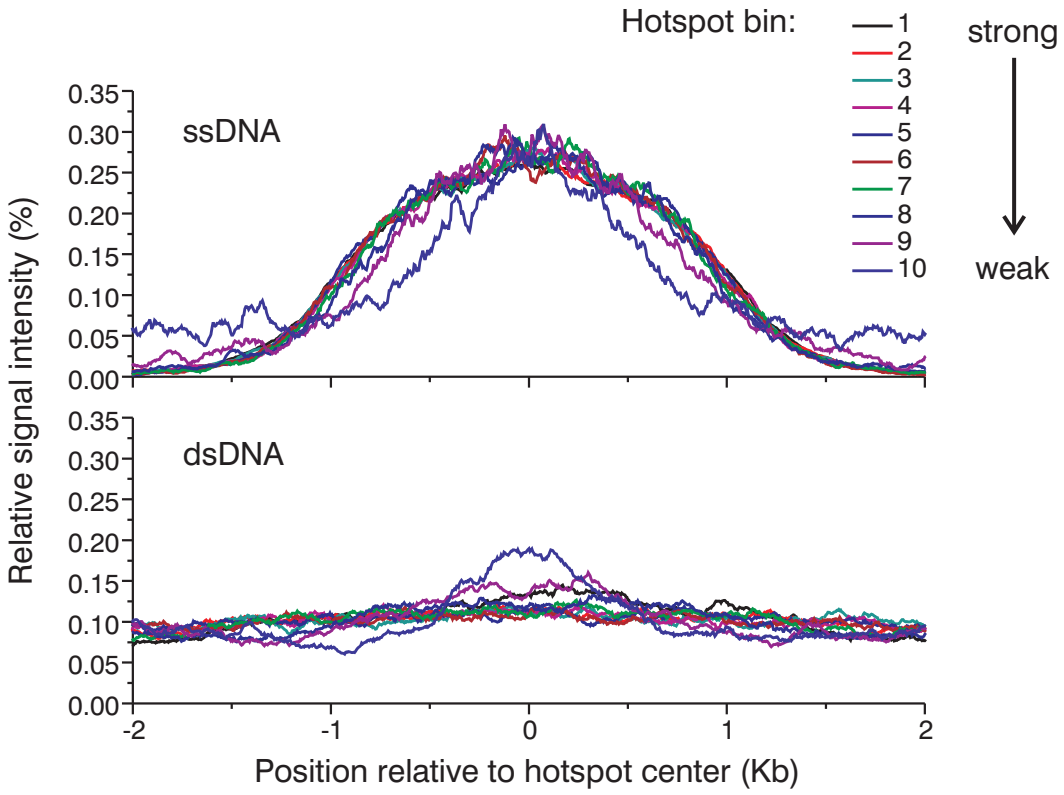
Supplemental Figure 9. Scatterplot matrix of correlations between mean ITR length, mean ITR GC content, mean microhomology (μH) length, mean microhomology GC content and hotspot strength. Above diagonal Pearson (*r*) and Spearman (ρ) correlation coefficients for corresponding comparisons are shown.

Supplemental Figure 10



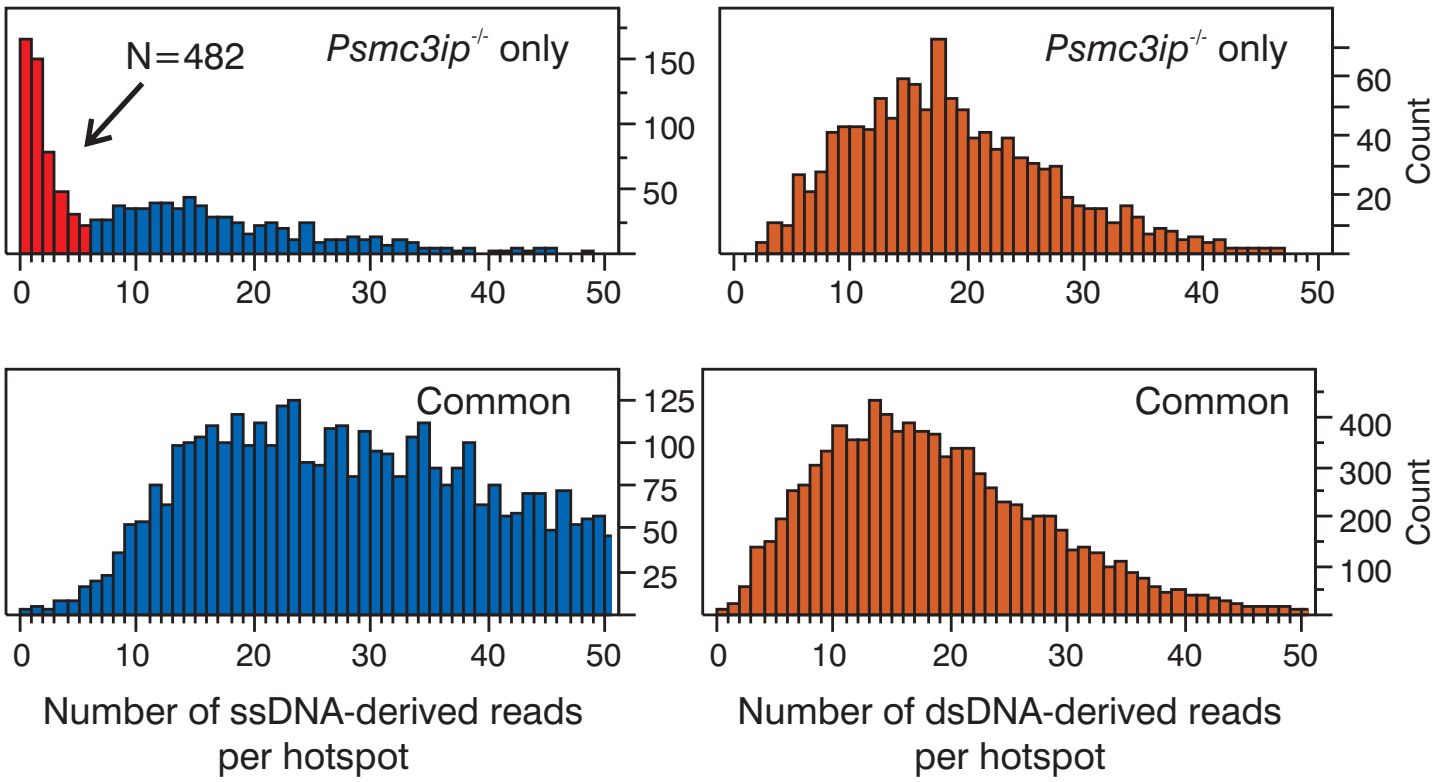
Supplemental Figure 10. Proportion of common and KE-only peaks overlapping H3K4me3 marks and the previously detected hotspot motif. Positions of H3K4me3 peaks and consensus motif were defined previously in (Smagulova et al. 2011). The hotspot strength correlates with the strength of the H3K4me3 signal and with the quality of the match to the hotspot motif (Smagulova et al, 2011). Therefore, the decrease in the overlap between KE-only hotspots and H3K4me3 (or hotspot motif) is likely the reflection of the fact that these additional hotspots are very weak.

Supplemental Figure 11



Supplemental Figure 11. ssDNA and dsDNA coverage in strong and weak hotspots. All hotspots were ordered by strength and divided into ten bins, bin 1 being the strongest. ssDNA and dsDNA coverage profiles were then calculated for all bins and plotted on the graph. The relative signal intensity was calculated for each profile by dividing the magnitude of the specific signal at each position by the total signal in the 5 Kb window centered at the hotspots. Profiles were calculated in 5 bp increments.

Supplemental Figure 12



Supplemental Figure 12. The distribution of ssDNA-derived and dsDNA-derived sequencing read counts in *Psmc3ip*$^{-/-}$ only and common peaks.