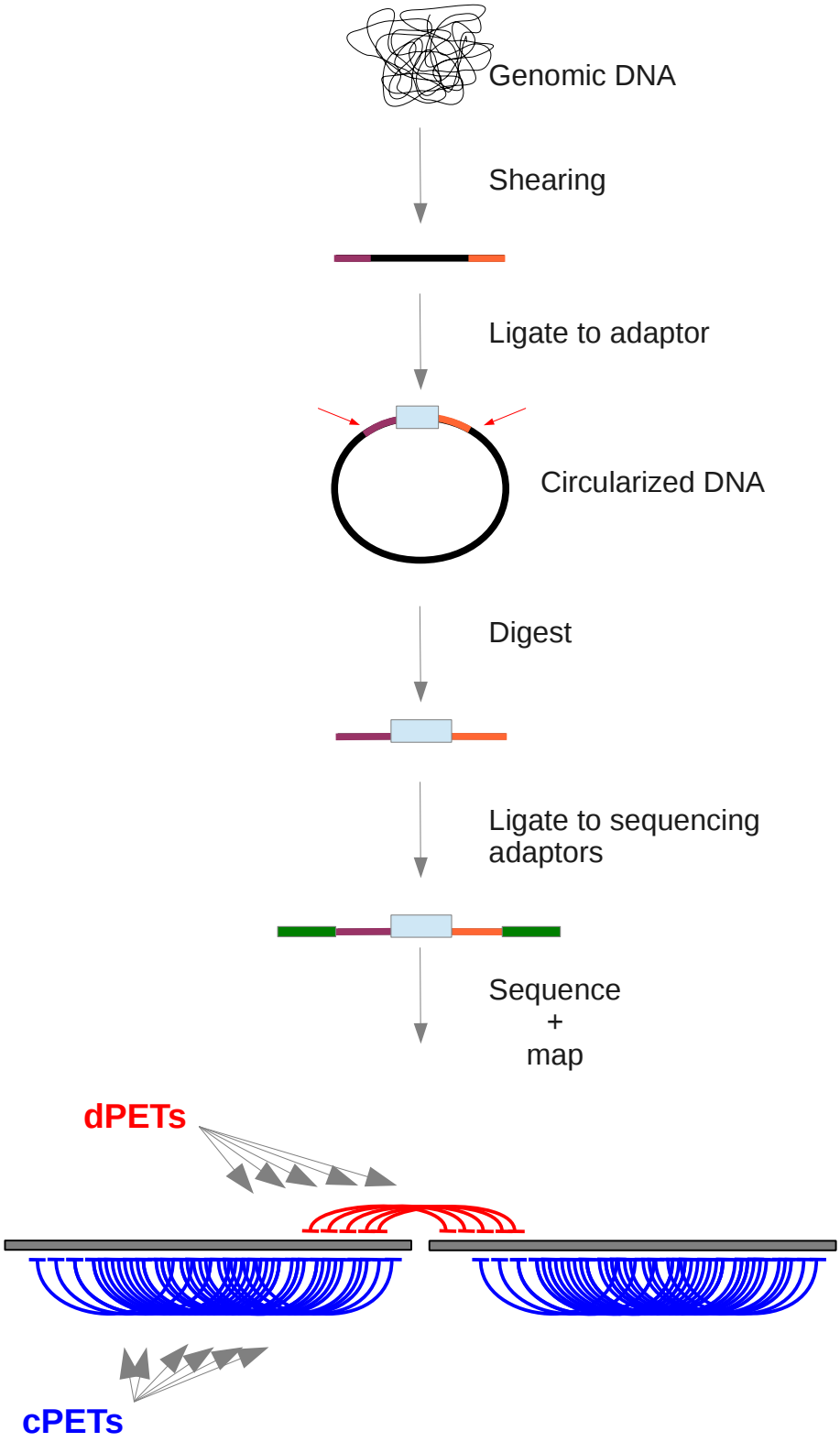
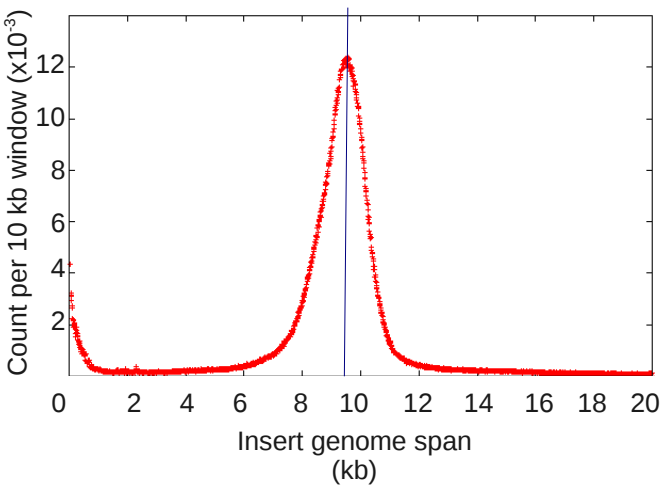
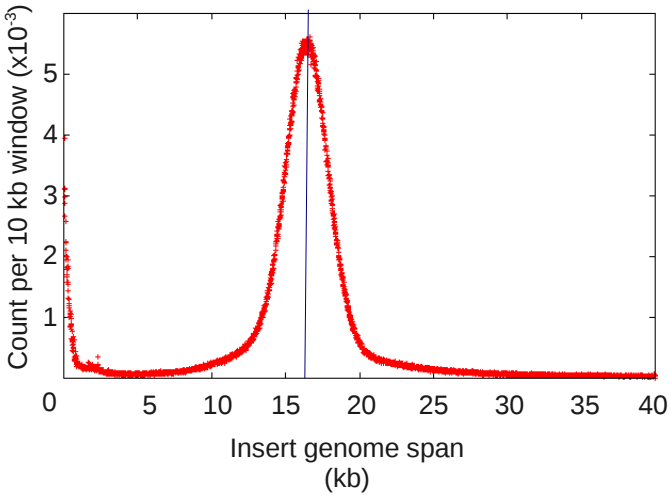


Supplementary Figure 1: Principle of DNA-PET.
Genomic DNA is fragmented in 10 or 17kb fragments and circularized with adaptors containing a EcoP15I restriction site (pale blue). After restriction with EcoP15I, released DNA fragments are ligated to sequencing adaptors (green), sequenced and mapped on the reference genome. cPETs (deep blue) correspond to PETs where the two tags map on the same scaffold whereas dPETs (red) correspond to PETs where the two ends map on different scaffolds. dPETs can be used to improve scaffolding. The 5' and 3' ends of sheared genomic DNA fragments are colored in purple and orange, respectively.

A

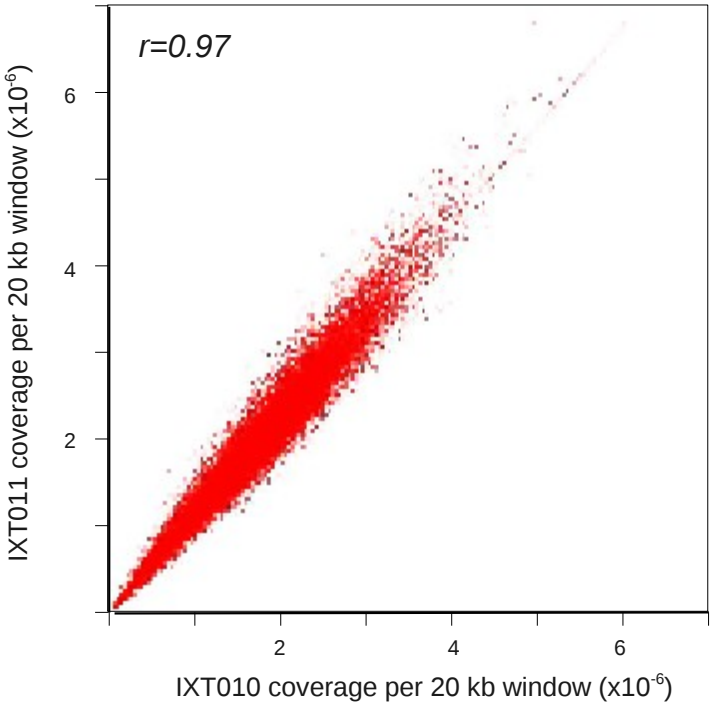


Average insert size IXT010 : 9.6 kb

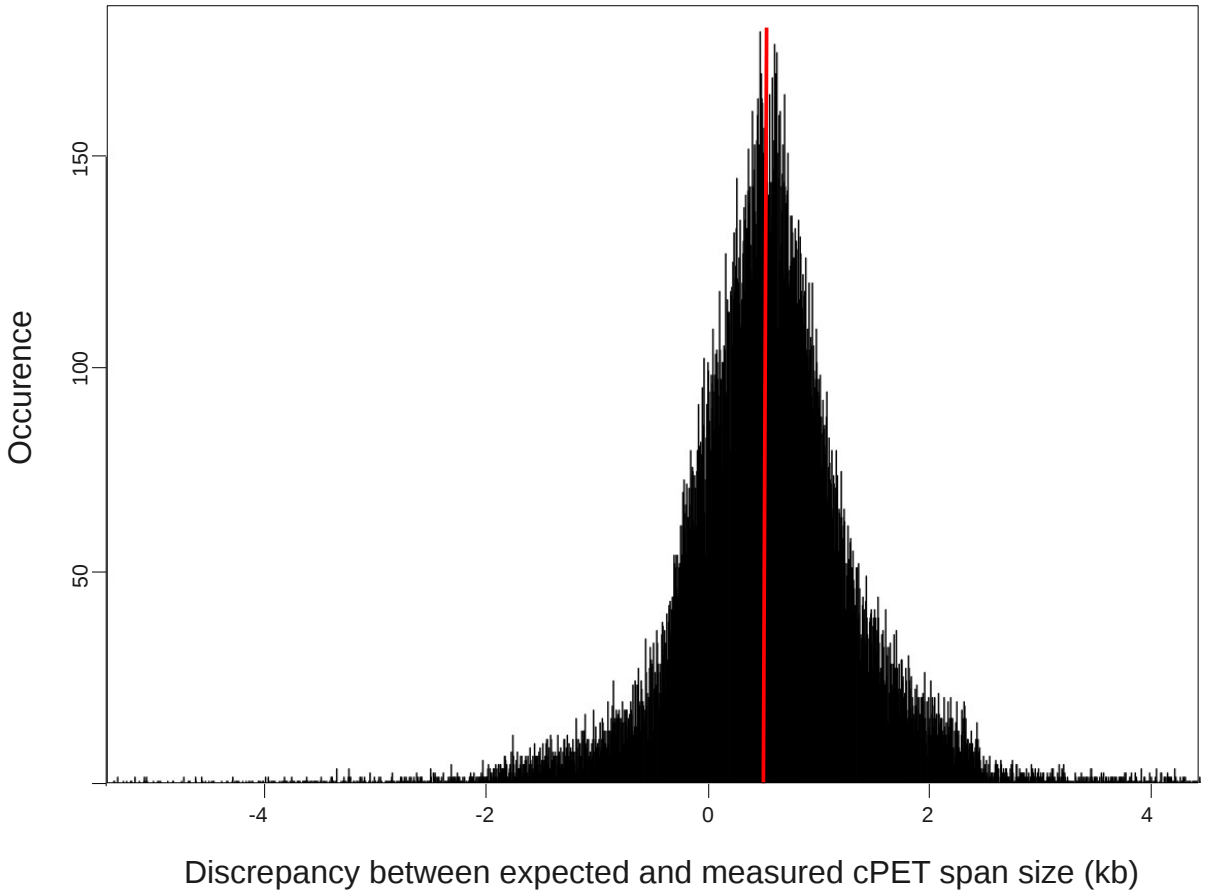


Average insert size IXT011 : 17.5 kb

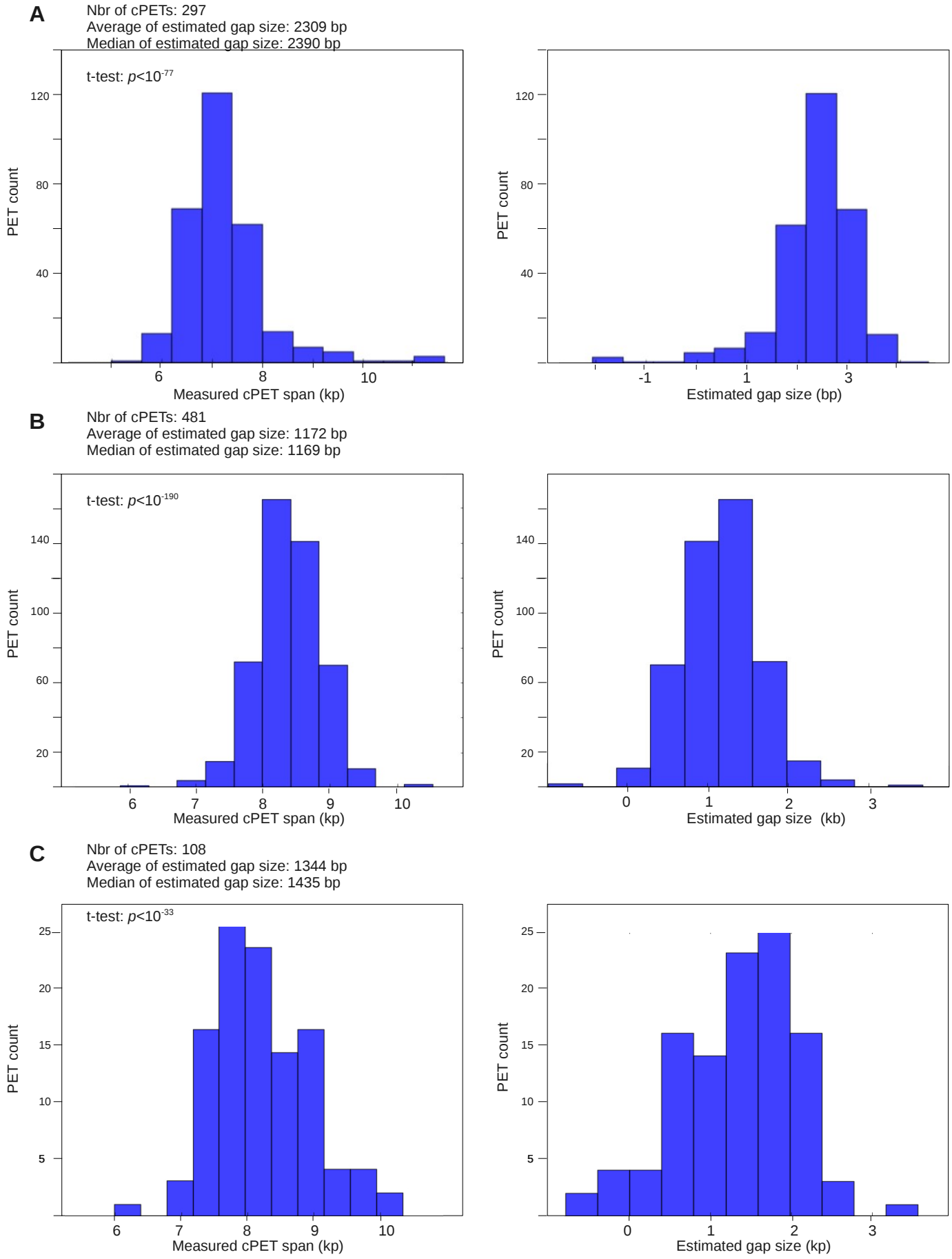
B



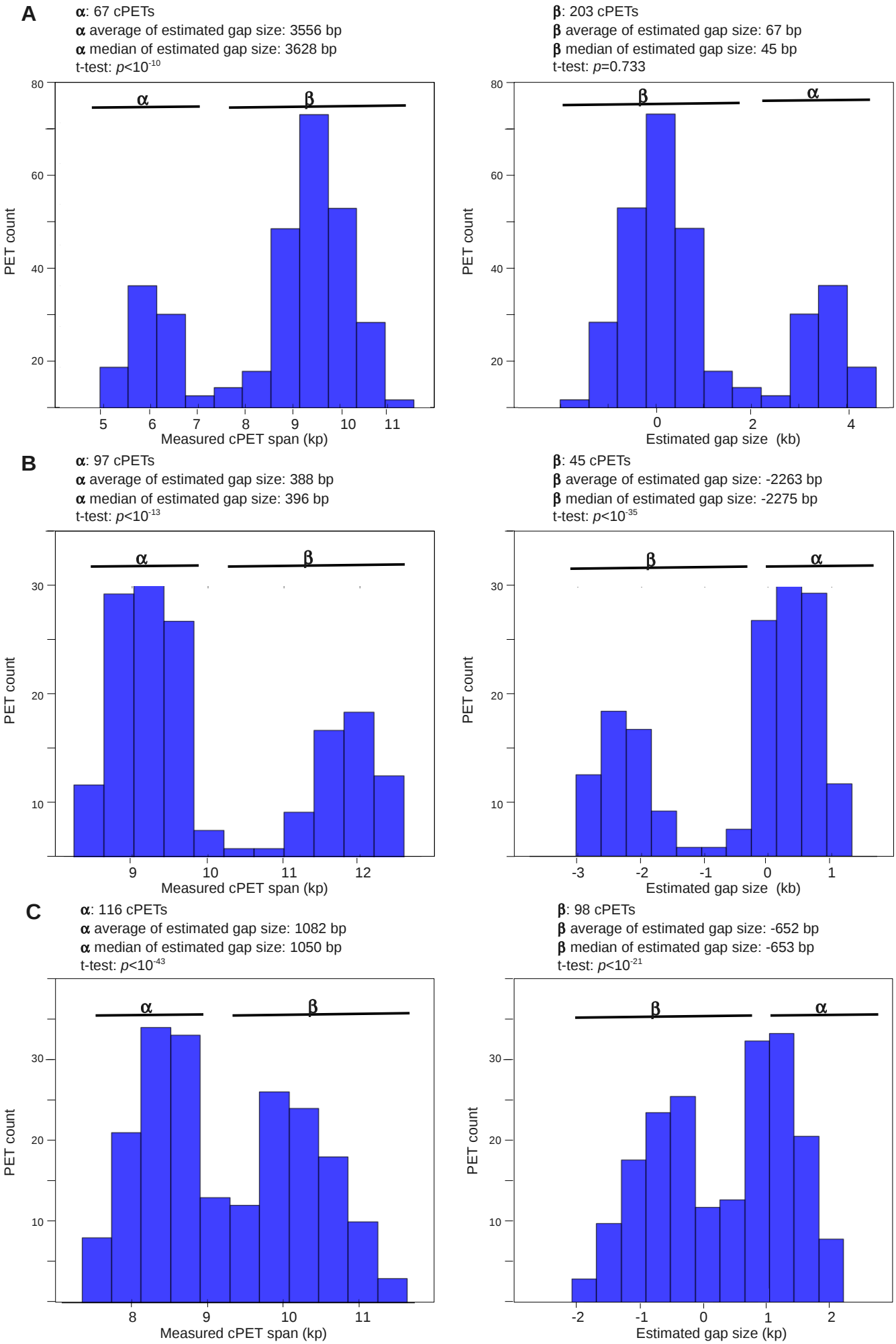
Supplementary Figure 2: Quality control of the DNA-PET datasets.
A. Insert size distribution of DNA-PET libraries of 9.6 (IXT010) and 17 kb (IXT011), estimated as the average genomic span of cPETs (blue line).
B. Correlation of coverage density between DNA-PET libraries IXT010 and IXT011, computed by scoring the number of cPETs overlapping successive 20 kb windows. The tight distribution indicates that the genome coverage derived with two independent libraries is highly correlated.



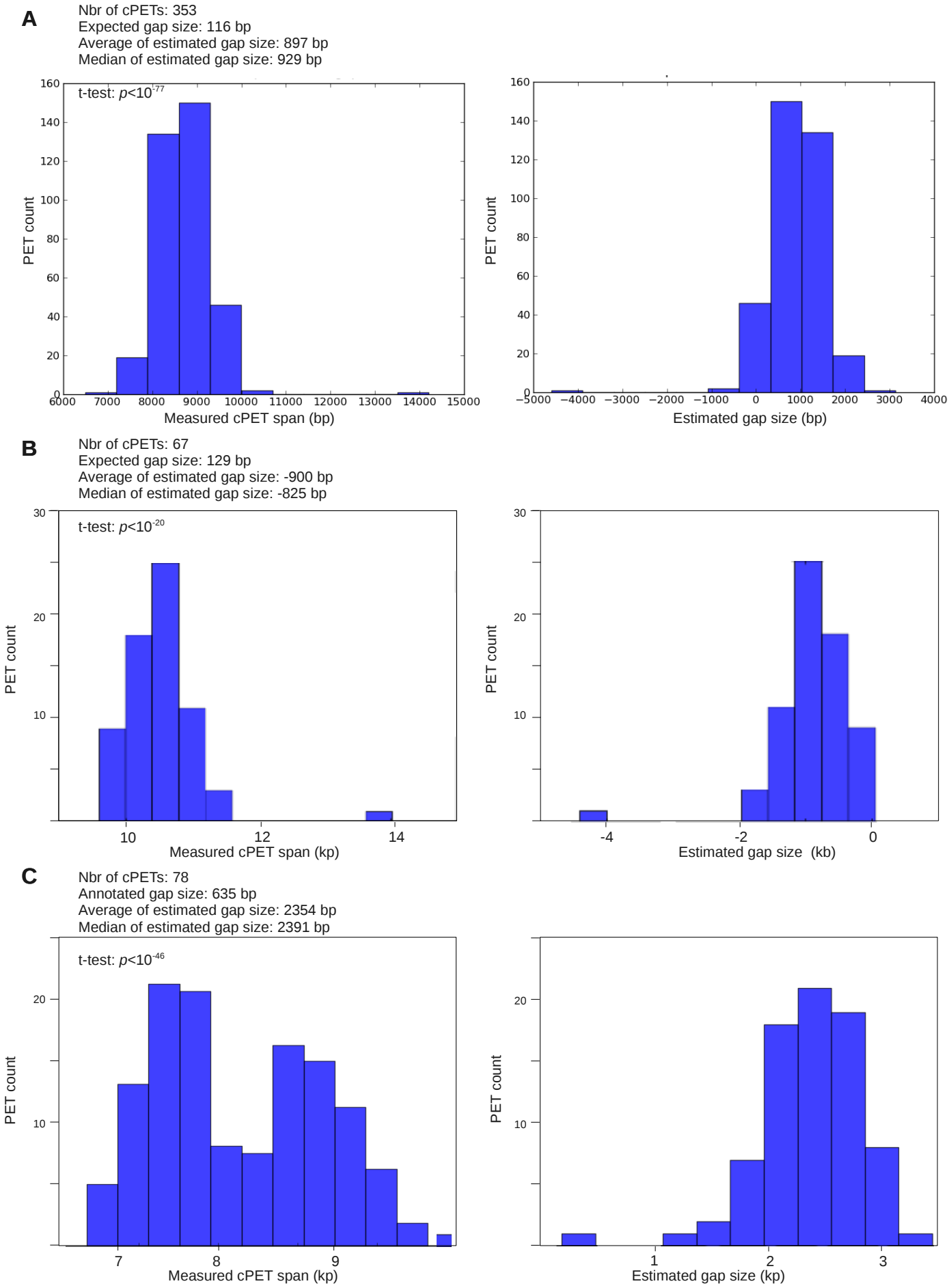
Supplementary Figure 3: Discrepancy between expected and annotated gap size, for '50 bp gaps'.
The difference between the expected and the measured span of individual cPETs has been used to estimate the length of '50 bp gaps'. This measure of discrepancy is positive for cPETs overlaping a region that is actually larger than 50 bp and conversely for negative values. The red line represent the average value.



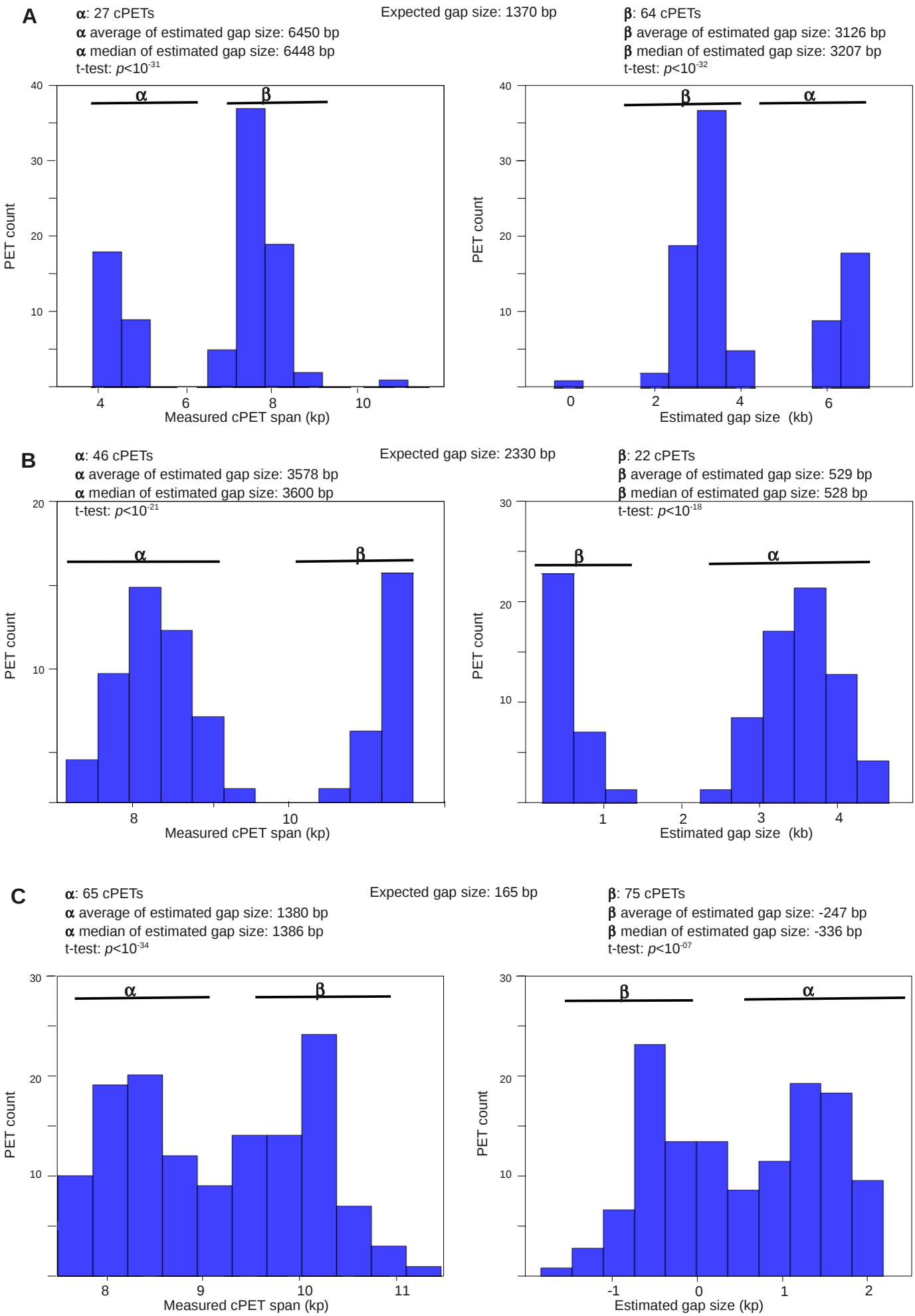
Supplementary Figure 4: Examples of discrepancy between expected and measured assembly gap length, for 50 bp gaps.
Individual cPETs spanning a single assembly gap of 50 bp have been used to estimate their actual length. For each cPET, the measured span size (left panel) was compared to the expected span size. Disagreements between expected and measured span sizes were used to derive independent estimate of the gap actual size (right panel). The number of cPETs, the average and median of the estimates are shown. A, B, C are three illustrative examples.



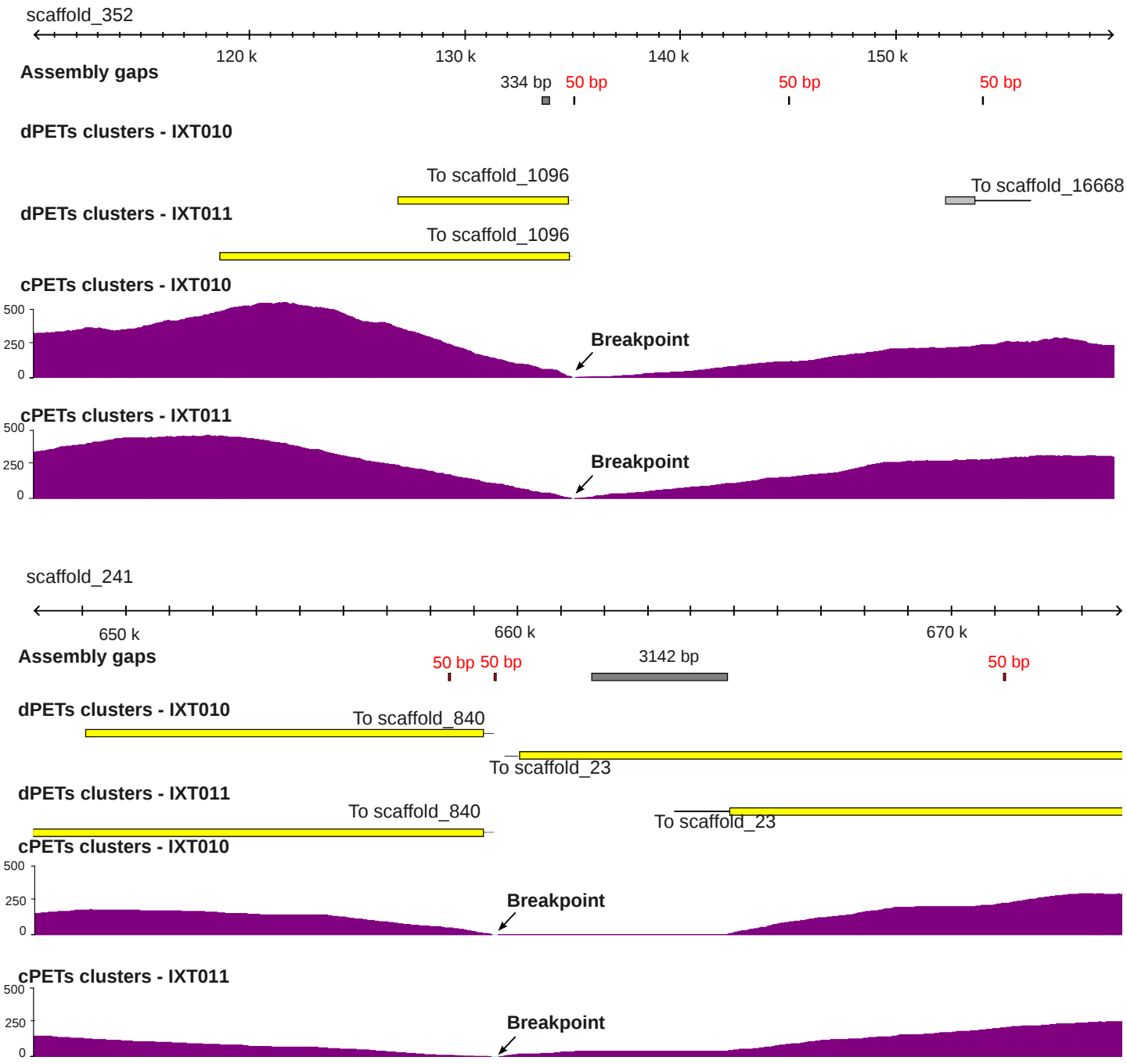
Supplementary Figure 5: Example of structural polymorphism at assembly gaps of 50 bp. In a number of cases, the measure of the actual gaps length follows a bimodal distribution, with two populations labeled α and β. See Supplementary Figure 4 for details. A, B and C are illustrative examples.



Supplementary Figure 6: Example of discrepancy between expected and measured assembly gap length, for assembly gaps other than 50 bp-long.
Individual cPETs spanning a single assembly gap (≠50 bp) have been used to estimate their acutal size. See Supplementary Figure 4 for details. The number of cPETs, the expected size, the average and median of the estimates are shown. A, B, C are three illustrative examples.

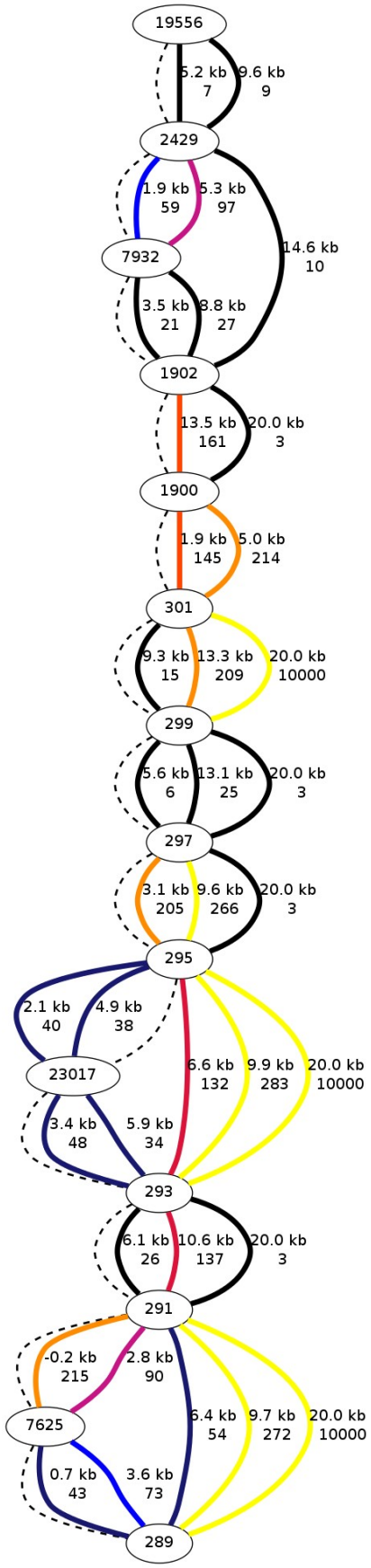


Supplementary Figure 7: Example of structural polymorphism at assembly gaps ≠ 50 bp. In a number of cases, the measure of the actual gaps length follows a bimodal distribution, indicated α and β. See Supplementary Figure 4 for details. A, B and C are illustrative examples.

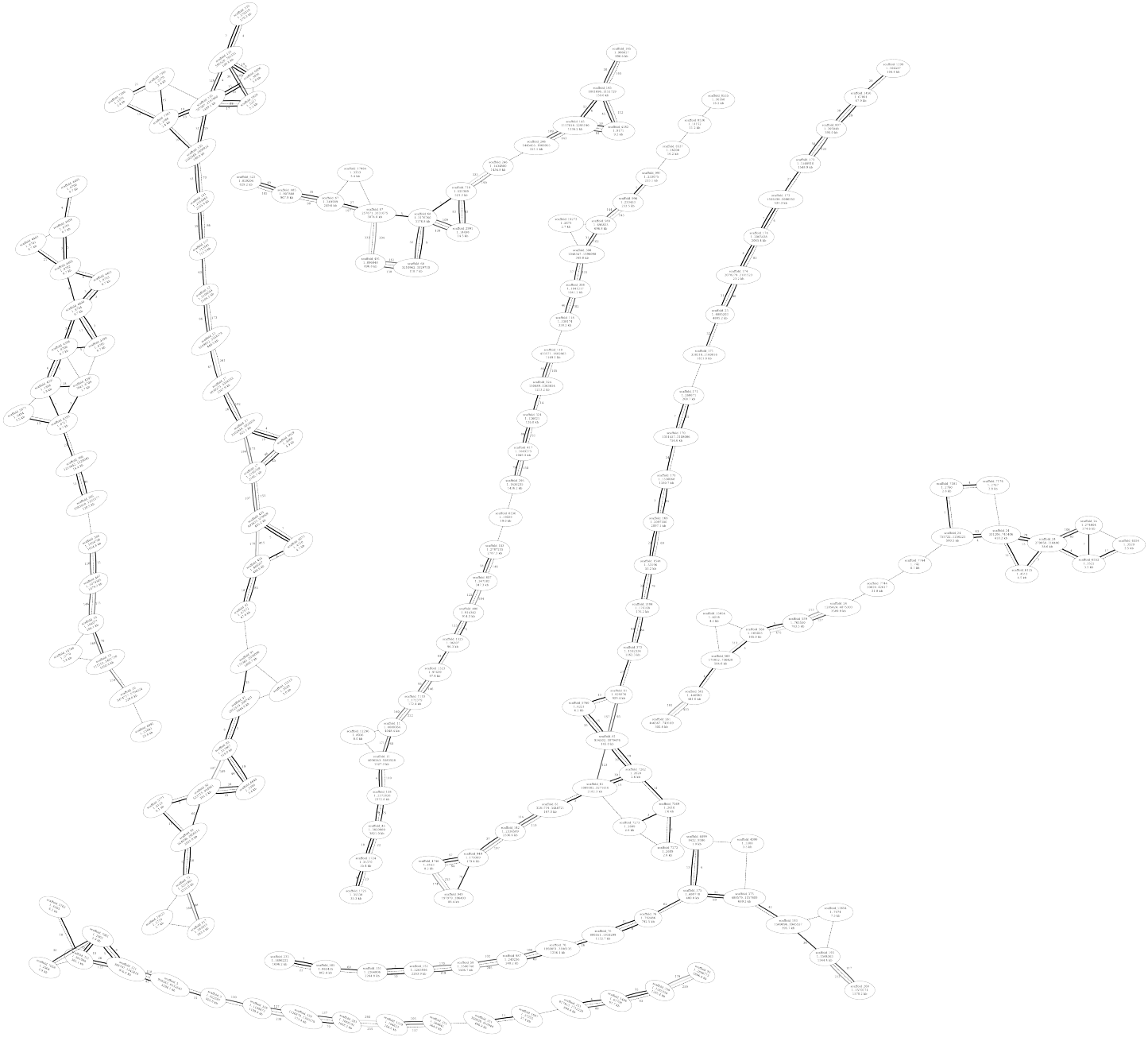


Supplementary Figure 8: Example of assembly breakpoints.

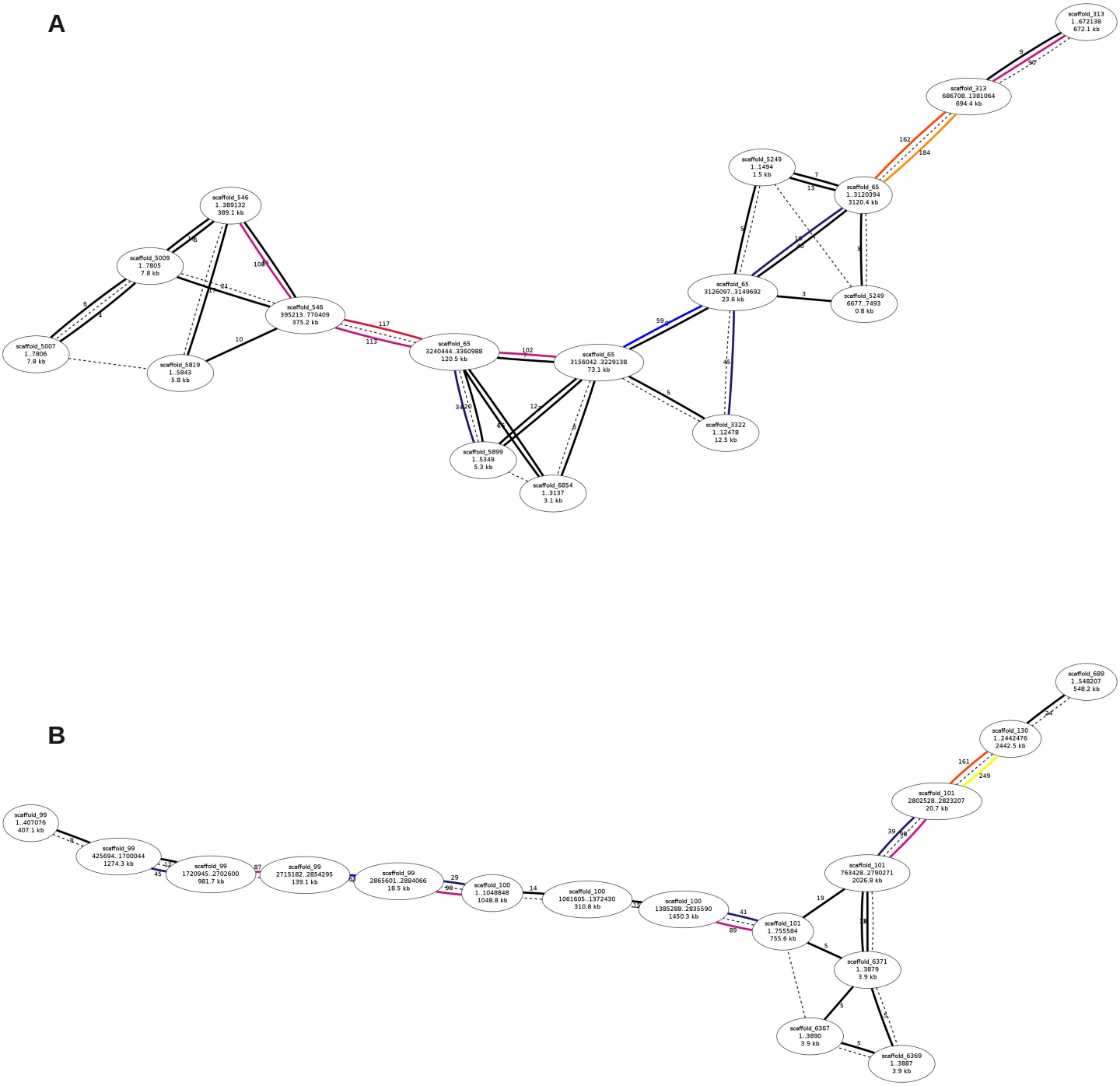
Genomic views two assembly break points. Track order: Assembly gaps, clusters of dPETs for DNA-PET libraries IXT010 and IXT011, cPETs coverage profile computed from libraries IXT010 and IXT011. The average insert size for these libraries is 9.6 kb and 17 kb, respectively. Numbers above assembly gaps indicate their length. Values in brackets correspond to their estimated length based on cPETs datas, when available. Assembly gaps of 50 bp are shown in red. Assembly breakpoints potentially occur when cPET coverage drops to zero, meaning that the two sides are not connected together, at the resolution of the DNA-PET libraries.



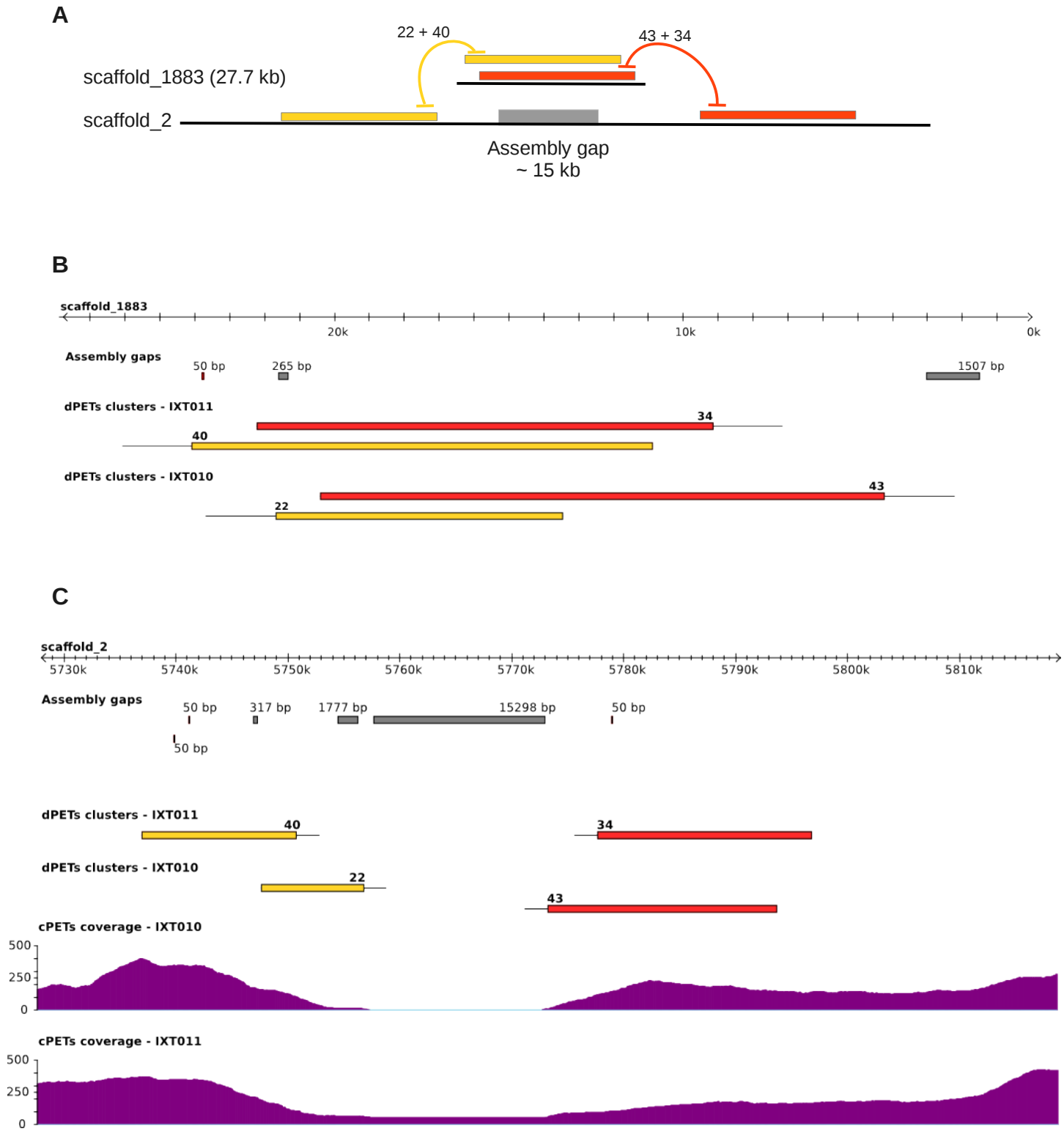
Supplementary Figure 9: Example of long range DNA-PET connectivity and re-scaffolding. Scaffolds are shown as ovals, indicating their number (*e.g.* 291 for scaffold_291). The dashed line corresponds to the path through the graph and represent the final ordering of the scaffolds relative to each other. Colored lines correspond to the connection and numbers indicate the estimated distance between scaffolds and well as dPETs support. Colors are heat map coded. In this example, a few small scaffolds (scaffold_7625, scaffold_23017, scaffold_7932) are nested between longer ones.



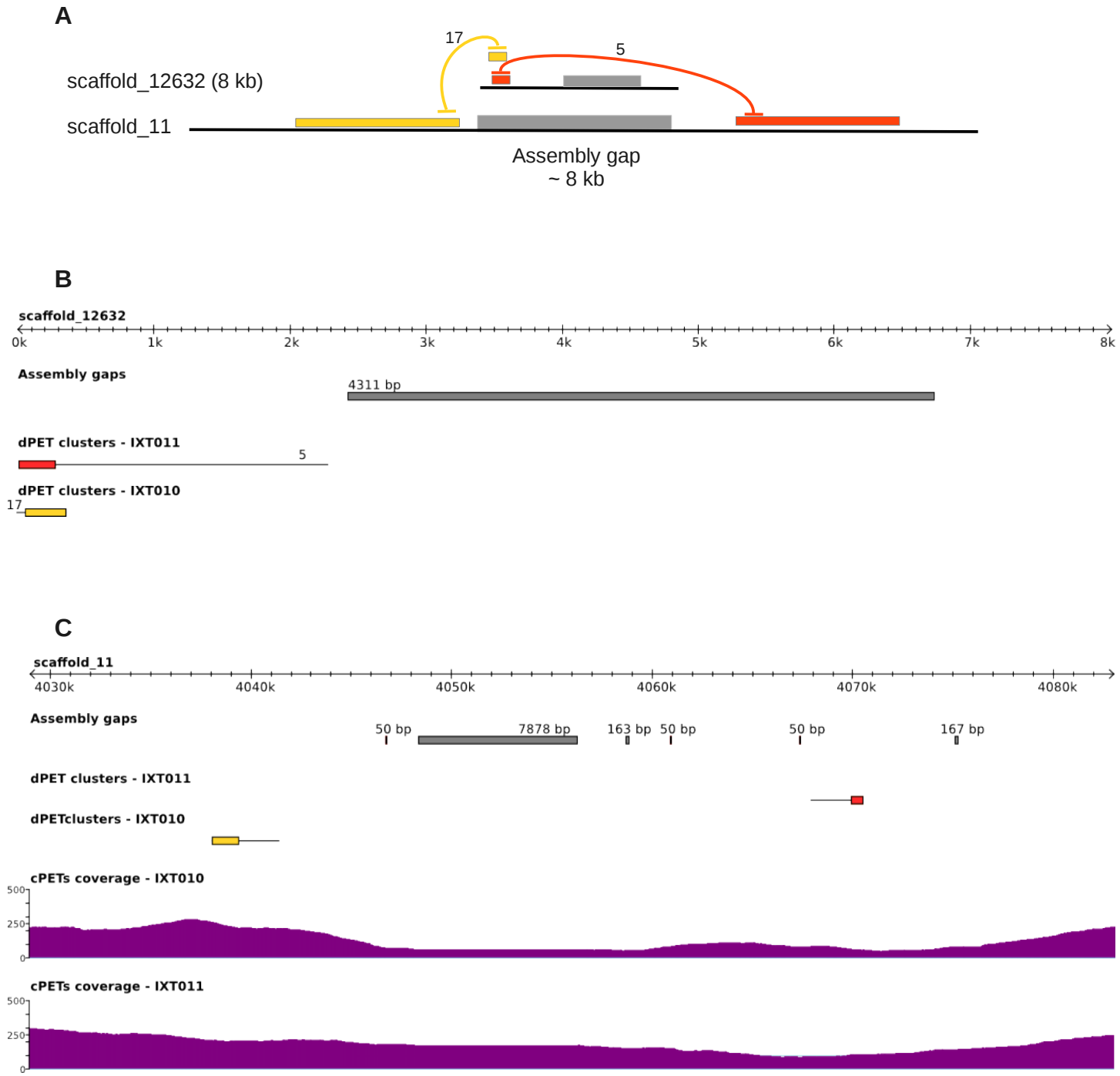
Supplementary Figure 10: Few examples of re-scaffolding.
Scaffolds are shown as ovals, indicating their name. Colored lines correspond to the connection and numbers indicate the dPETs support. Colors are heat map coded. Bubble-like structure correspond to small scaffolds inserted into larger ones or result from the complex connectivity between tandemly arranged short scaffolds.



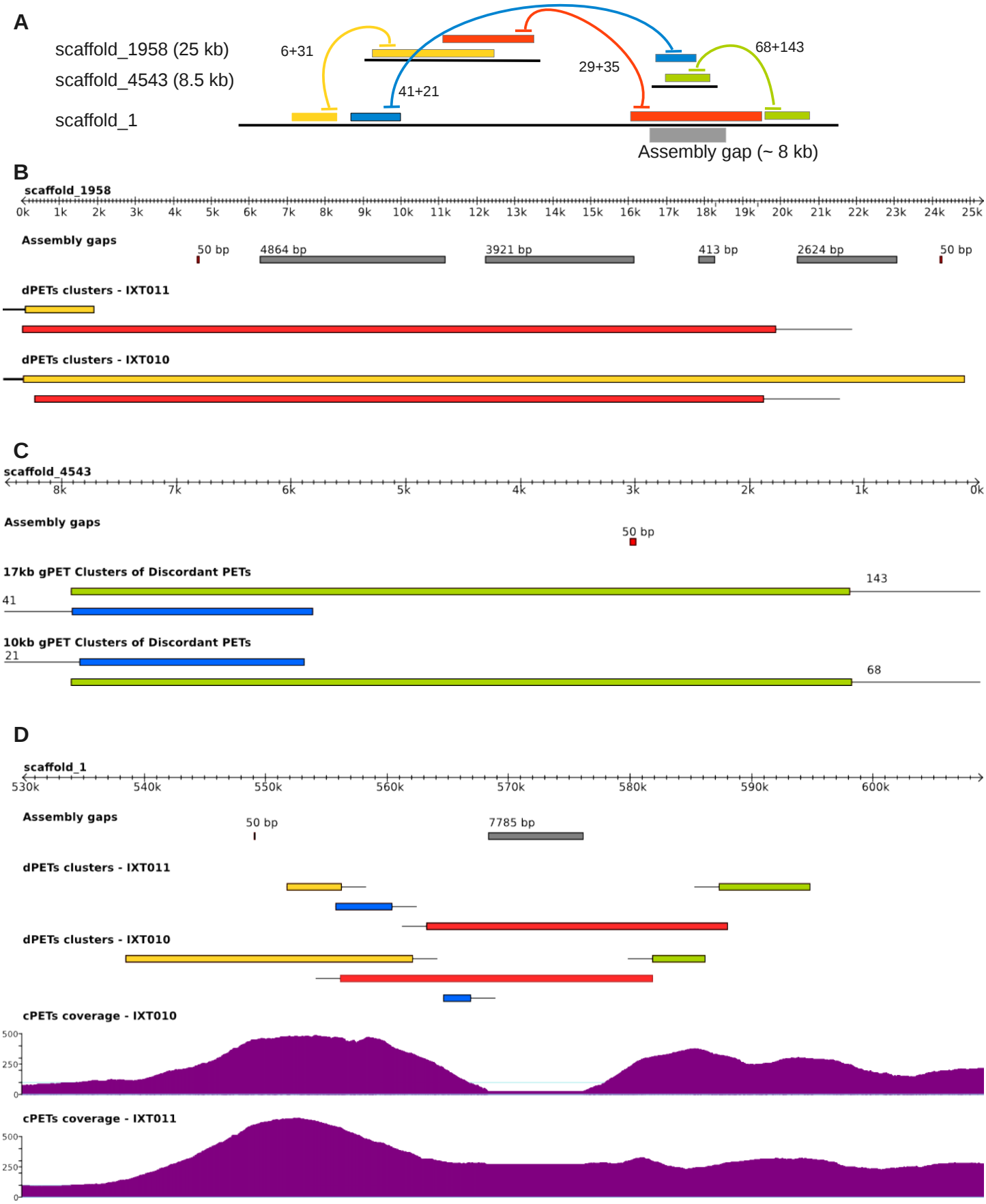
Supplementary Figure 11: Example of small scaffolds interted into assembly gaps of larger ones.
Scaffolds' fragments are shown as ovals, indicating their name (*e.g.* 291 for scaffold_291), which part of the original scaffold they correspond to, and their size. The dashed line corresponds to the path through the graph and represent the final ordering of the scaffolds relative to each other. Colored lines correspond to the connection and numbers indicate the dPETs support. Colors are heat map coded. A and B represent two examples of small scaffolds being inserted into assembly gaps present in larger ones.



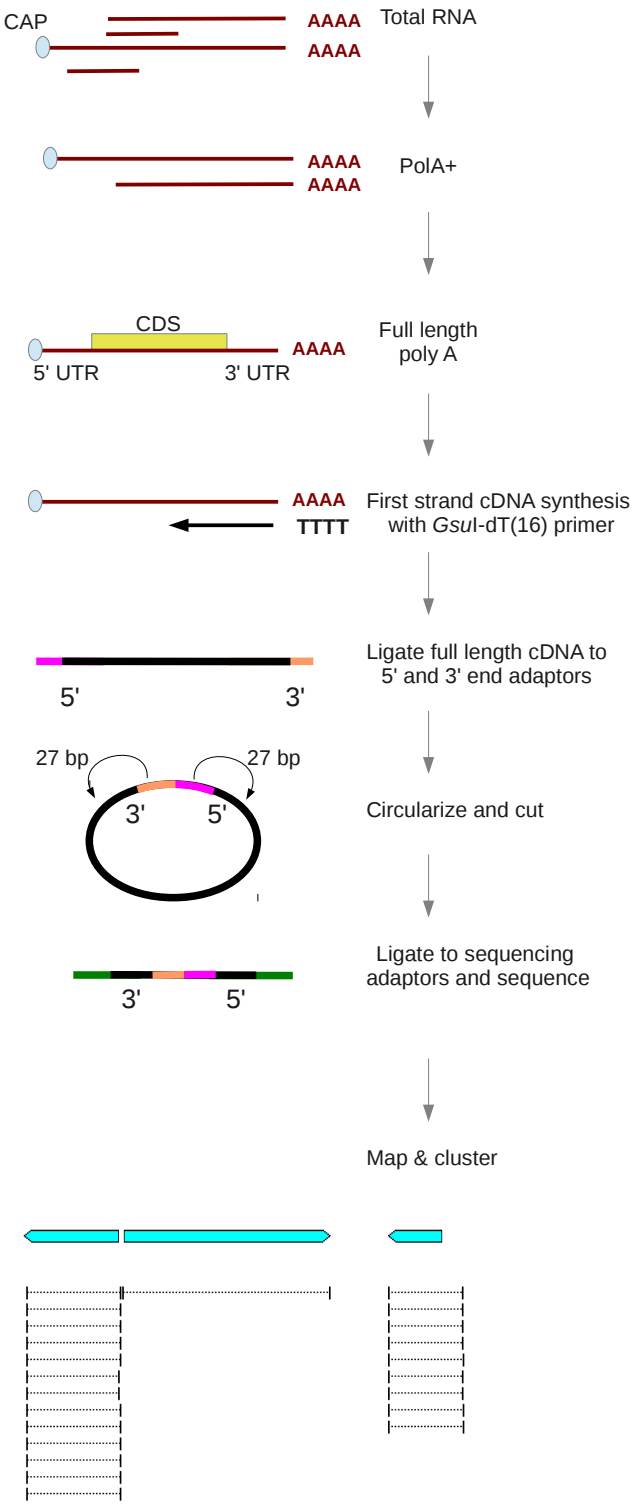
Supplementary Figure 12: Example of assembly gap fill-in by a smaller scaffold.
A. Conceptual representation of the connectivity between scaffolds_1883 and scaffold_2. Numbers above the links indicate the cPET count for each library. Clusters of cPETs connecting the left and the right side are drawn in yellow and orange, respectively. B, C. Detailed view of the entire scaffold_1883 and a sub-region of scaffold_2. The last two tracks correspond to the coverage density of cPET, for each DNA-PET library.



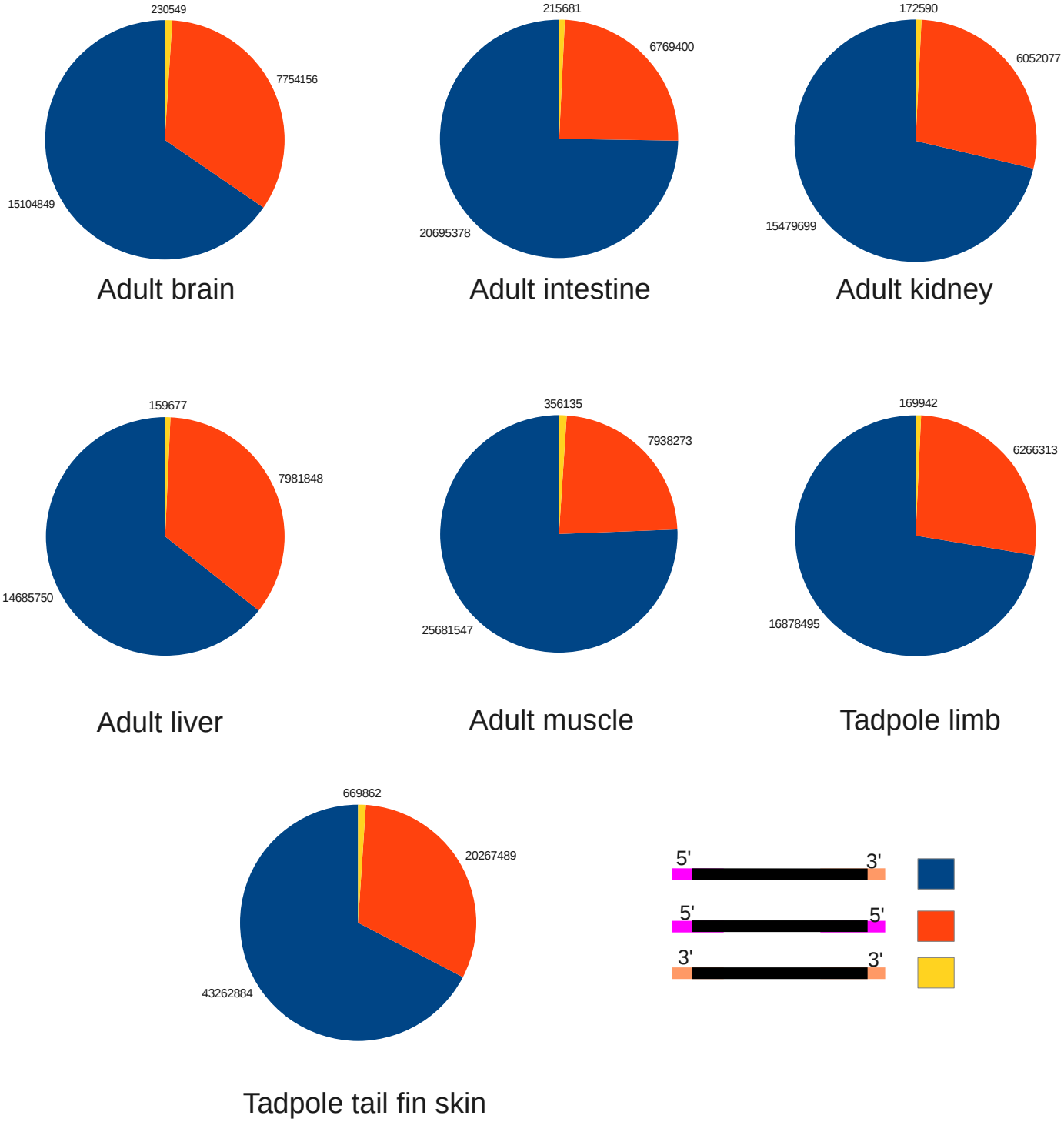
Supplementary Figure 13: Example of assembly gap fill-in by a smaller scaffold.
A. Conceptual representation of the connectivity between scaffolds_12632 and scaffold_11. Numbers above the links indicate the cPET count for each DNA-PET library. Clusters of cPETs connecting the left and the right side are drawn in yellow and orange, respectively. B, C. Detailed view of the entire scaffold_12632 and a subregion of scaffold_11. The last two tracks correspond to the coverage density of cPET, for each DNA-PET library. Note that the ~8 kb assembly gap on scaffold_11 is small compared to the average insert size of the two libraries. As a result, cPET coverage does not drop to zero.



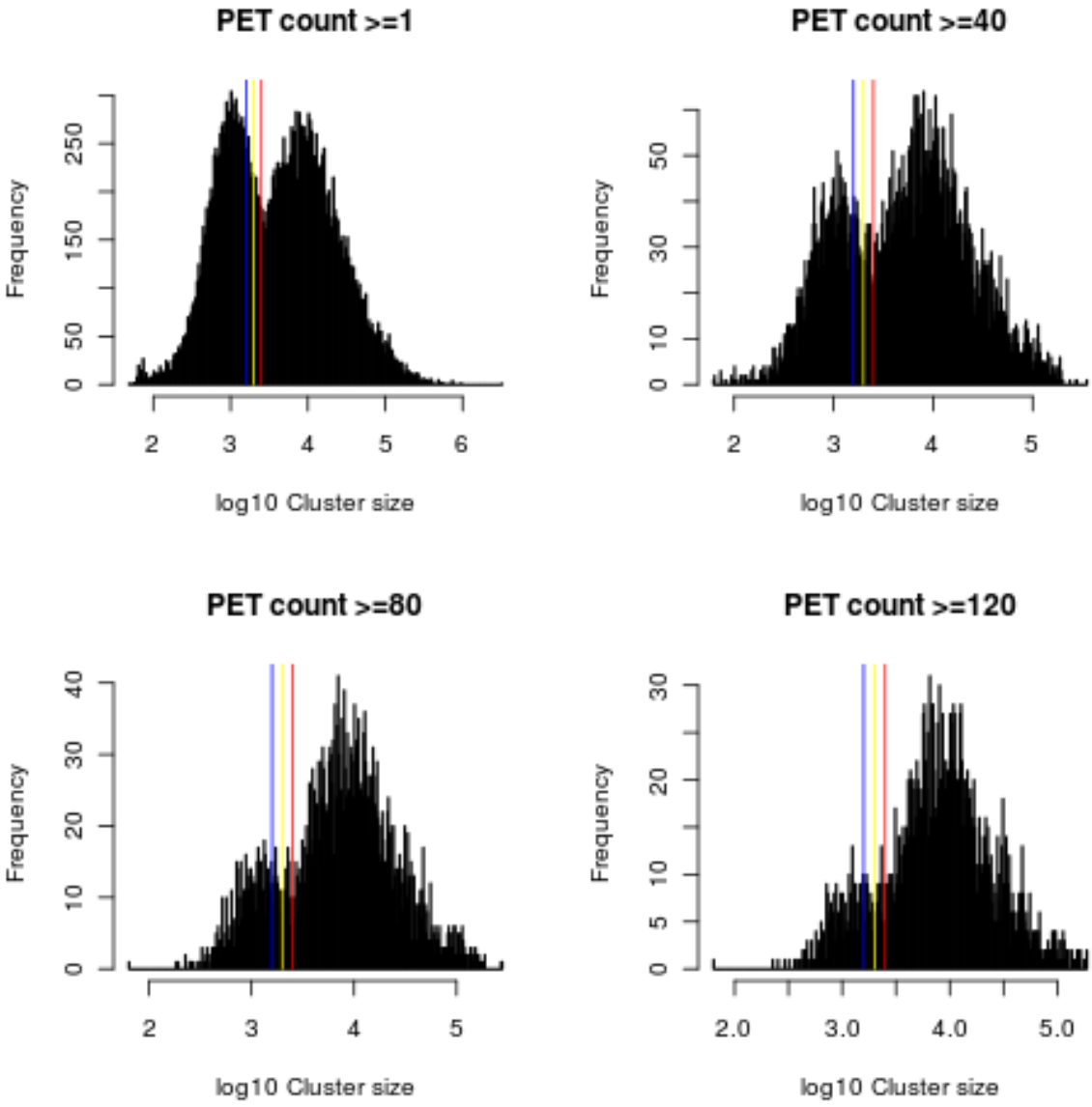
Supplementary Figure 14: Example of complex gap fill-in by smaller scaffolds.
A. Conceptual representation of the connectivity between scaffolds. Numbers above the links indicate the cPET count for each DNA-PET library. Clusters of cPETs connecting the left and the right side of the nested scaffolds are drawn in yellow/blue and orange/green. B, C, D. Detailed view of the nested scaffolds and a sub-region of scaffold_1. The last two tracks correspond to the coverage density of cPET, for each DNA-PET library. Note that the ~8 kb assembly gap on scaffold_11 is small compared to the average insert size of the two libraries. As a result, cPET coverage does not drop to zero.



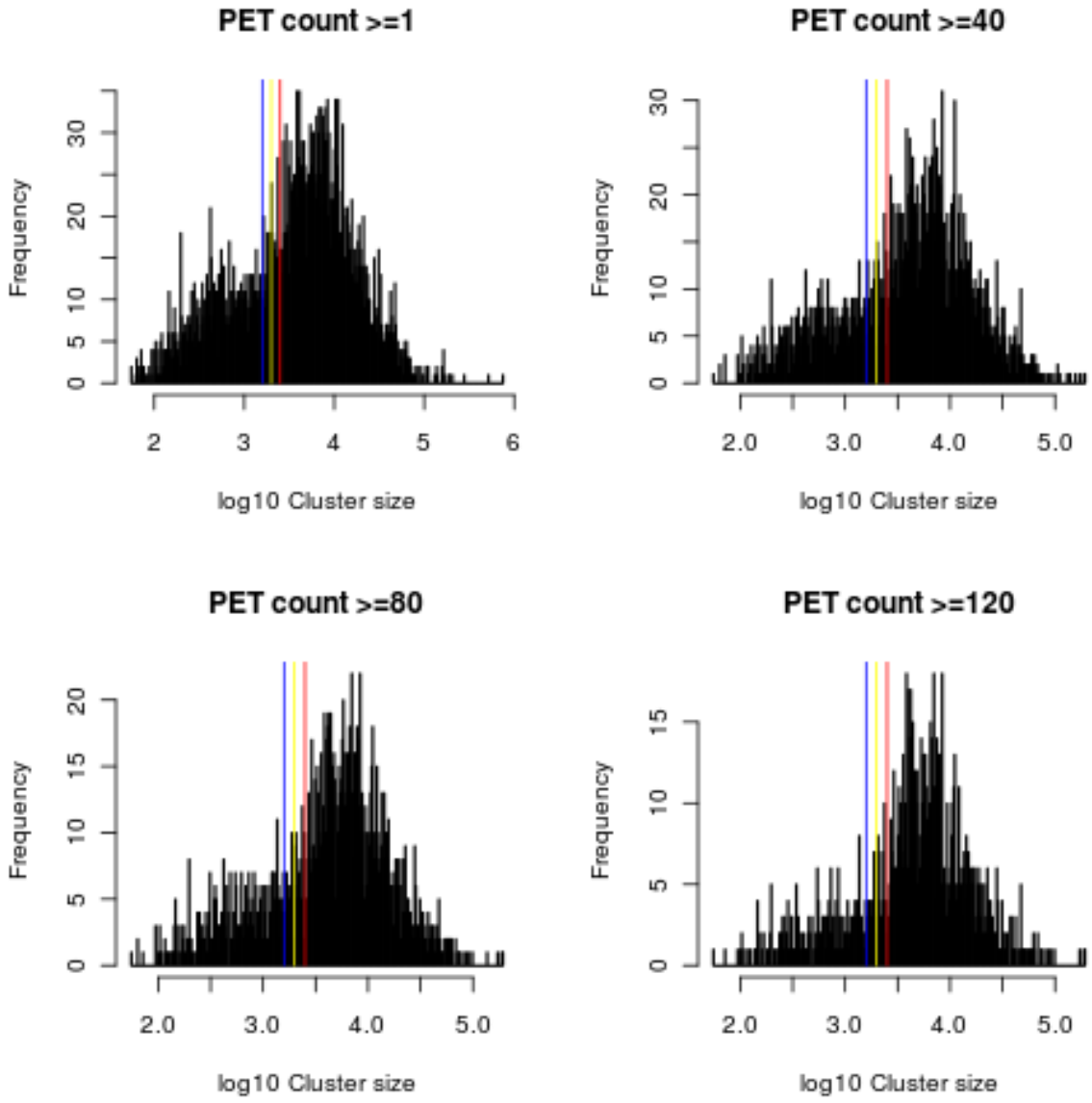
Supplementary Figure 15: Principle of RNA-PET.
Full length poly A+ RNA are used as template to generate full length cDNAs. Double stranded adaptors specific for the 5' and 3' ends (pink, orange) are ligated to the ends of the transcripts. Adaptors contain a EcoP15I restriction site, cutting 27 bp away from its binding site. After circularization and restriction with EcoP15I, sequencing adaptors are ligated to the resulting Paired-End diTag, prior to sequencing, mapping and clustering.



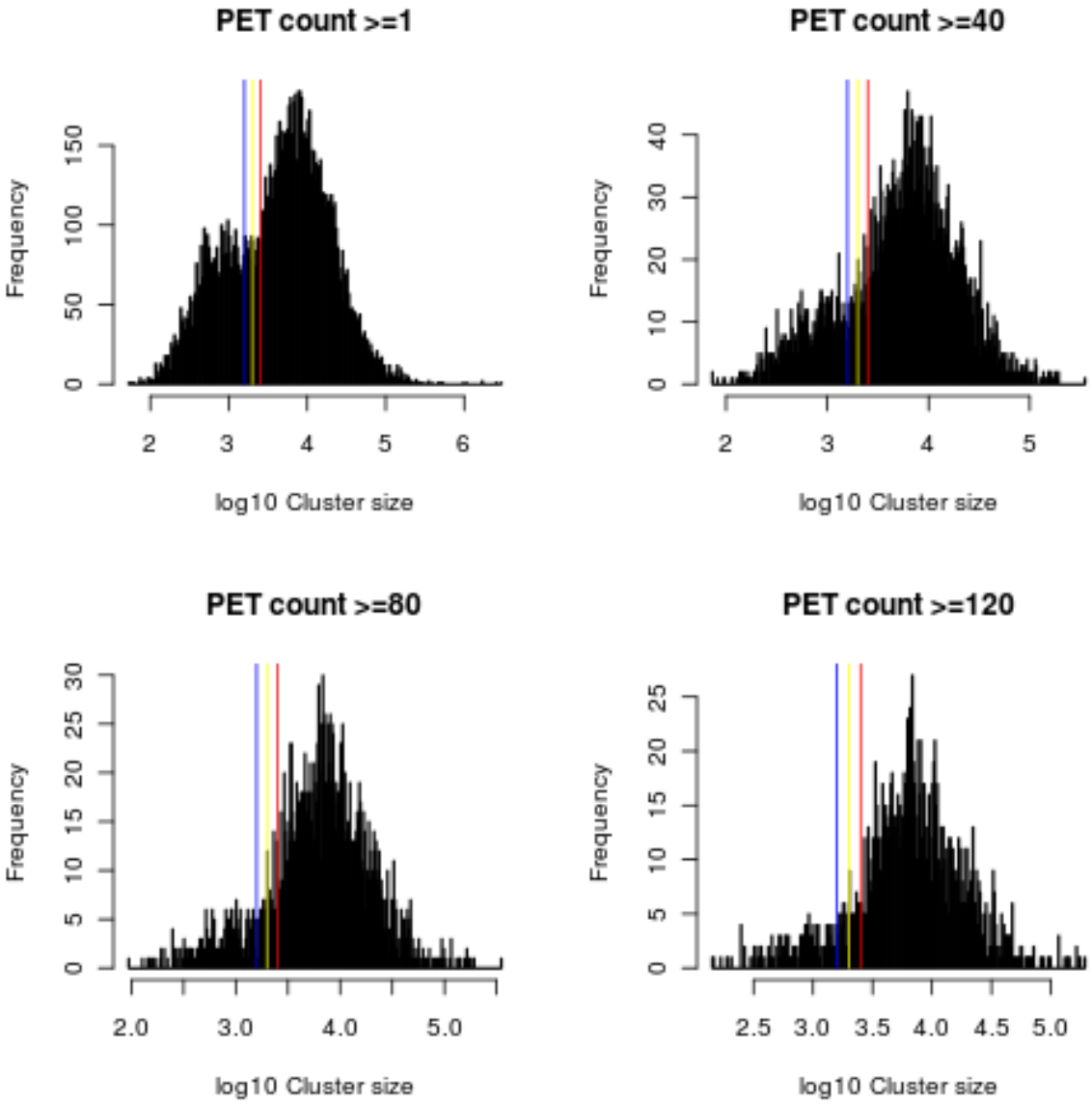
Supplementary Figure 16: Statistics of the identification of the 3'-end signature in RNA-PET di-tags.
The adaptor at the 3' end of transcripts is labeled with a specific signature (AACTGCTG). RNA-PET di-tags are labeled HT (Head and Tail), TT (Tail Tail) or HH (Head Head) on whether the signature was found at one end, both or none, respectively. Venn diagrams represent the proportion of HT, HH and TT di-tags obtained for each tissue library. Numbers indicate the total number of PETs per category.



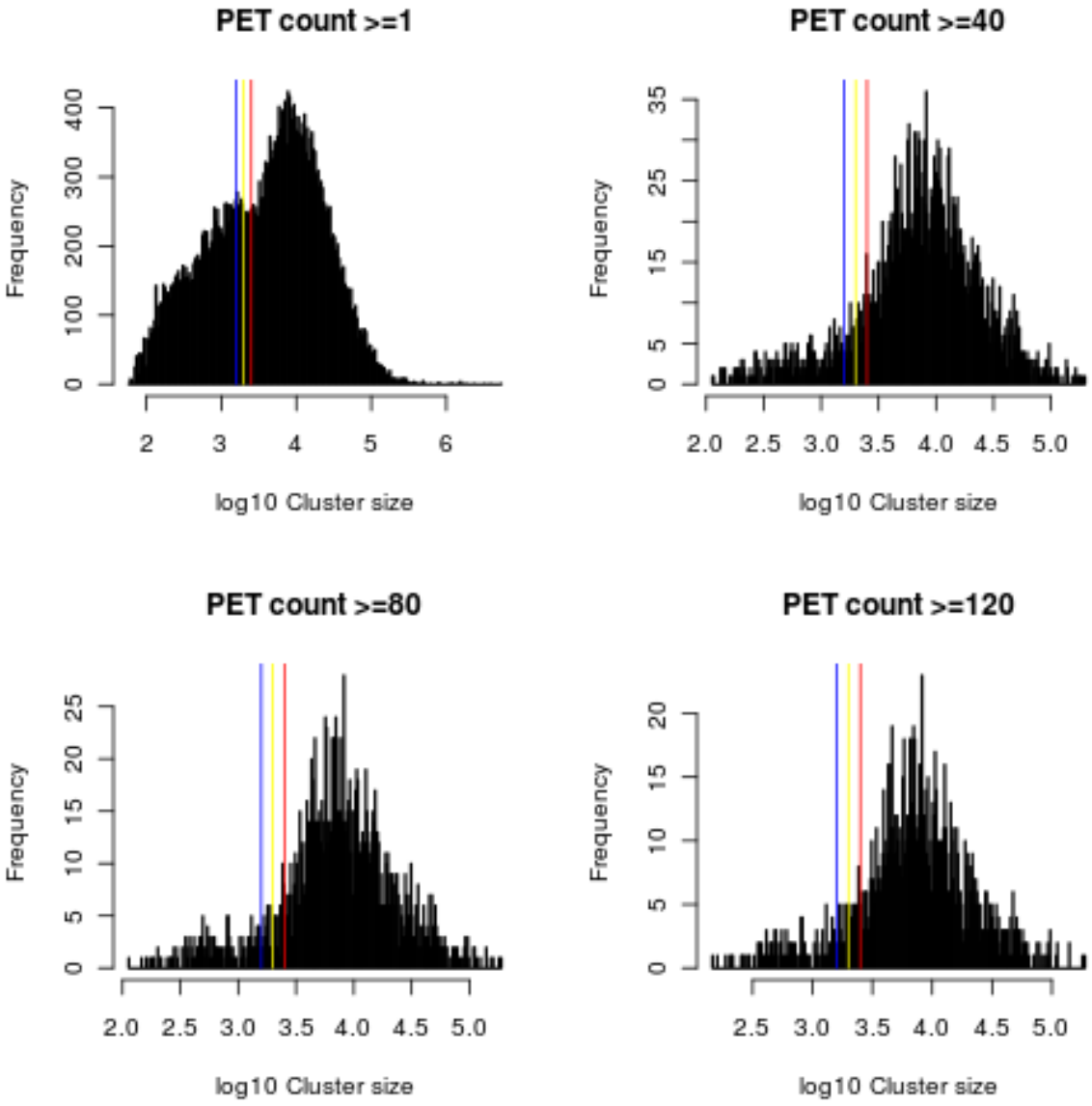
Supplementary Figure 17: Distribution of the genomic span of individual RNA-PET clusters in adult brain.
Blue, yellow and red lines denote ~1.5, ~2.0 and ~2.5 kb, respectively.



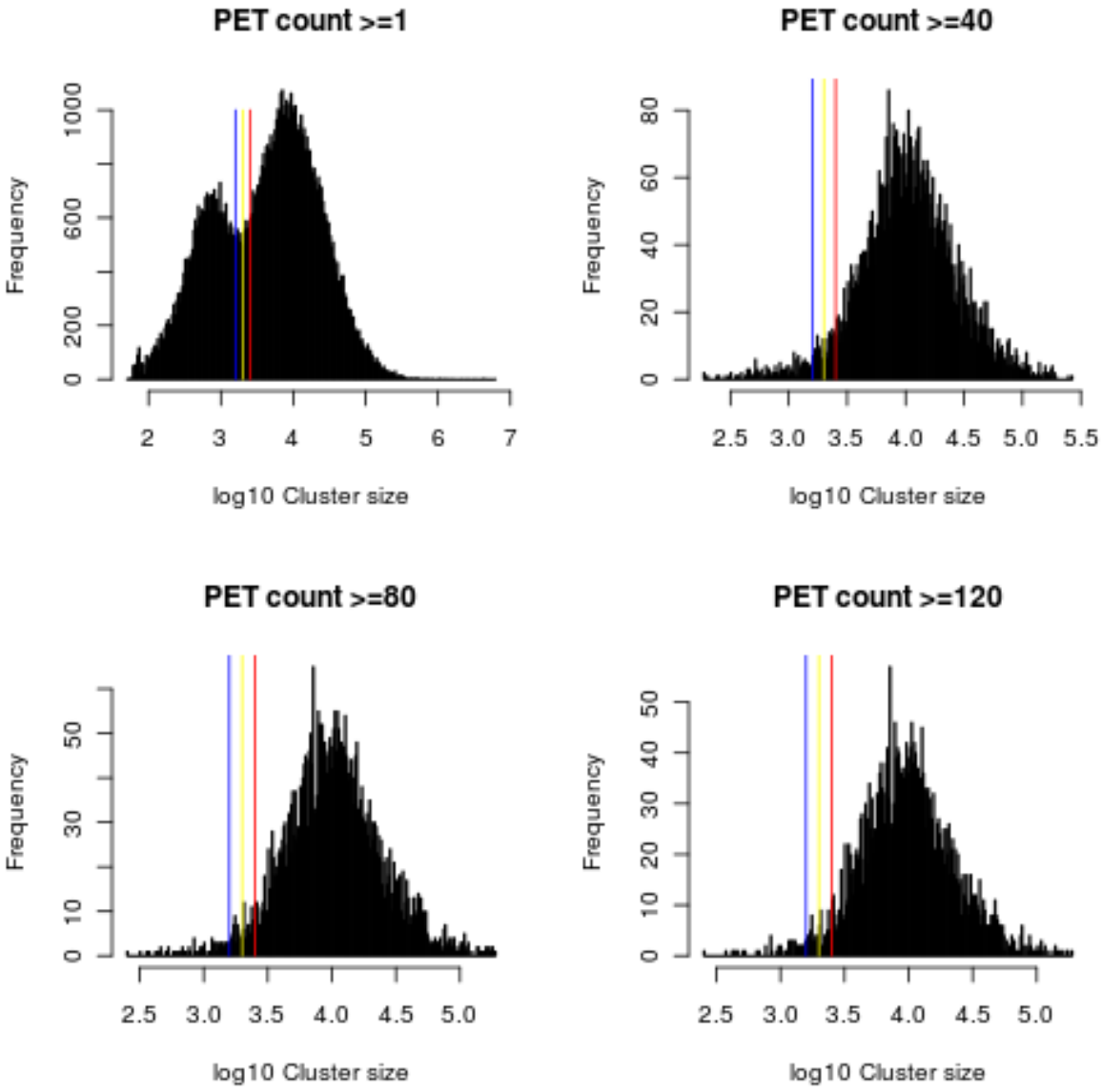
Supplementary Figure 18: Distribution of the genomic span of individual RNA-PET clusters in adult liver.
Blue, yellow and red lines denote ~1.5, ~2.0 and ~2.5 kb, respectively.



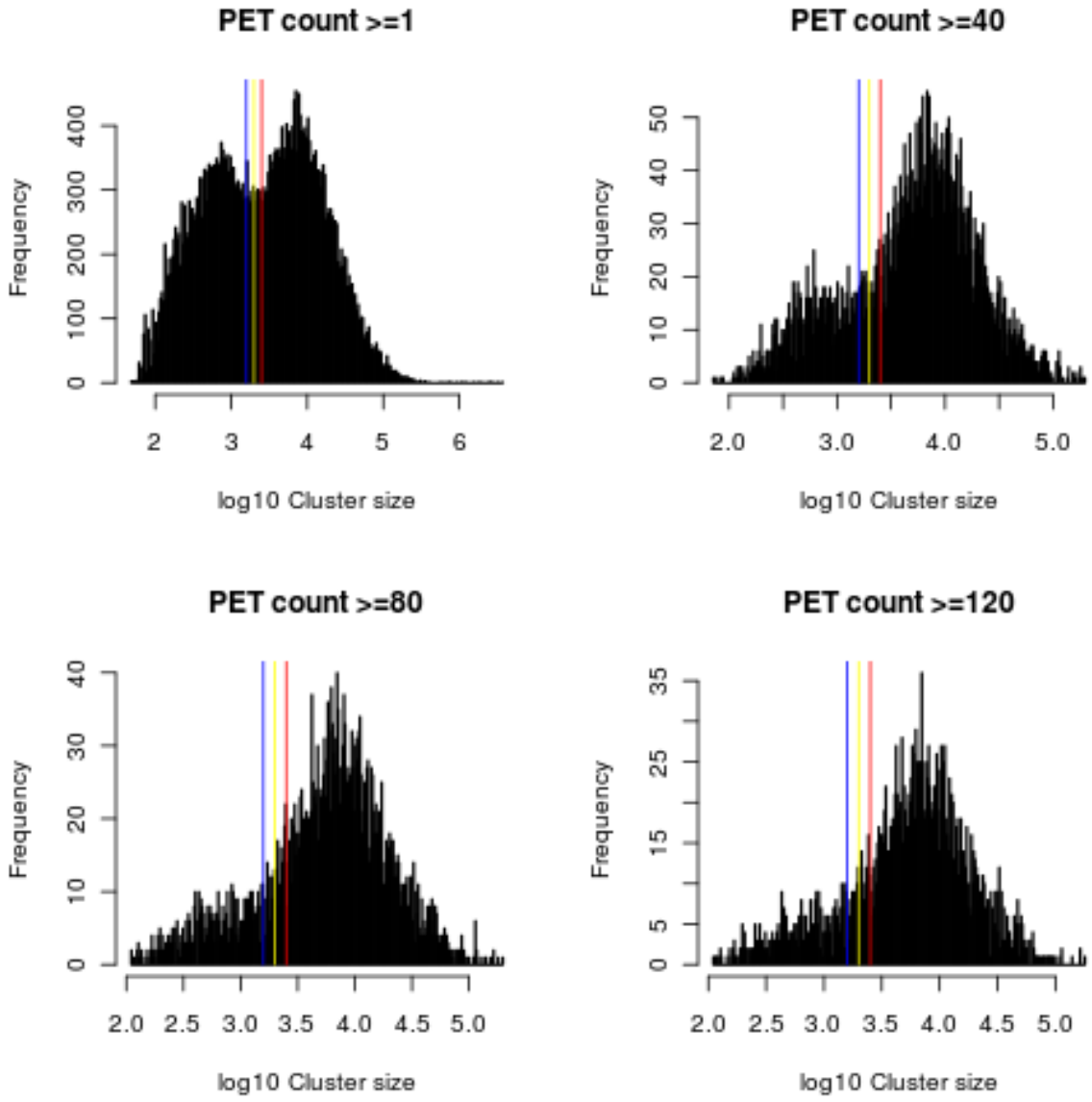
Supplementary Figure 19: Distribution of the genomic span of individual RNA-PET clusters In adult intestine.
Blue, yellow and red lines denote ~1.5, ~2.0 and ~2.5 kb, respectively.



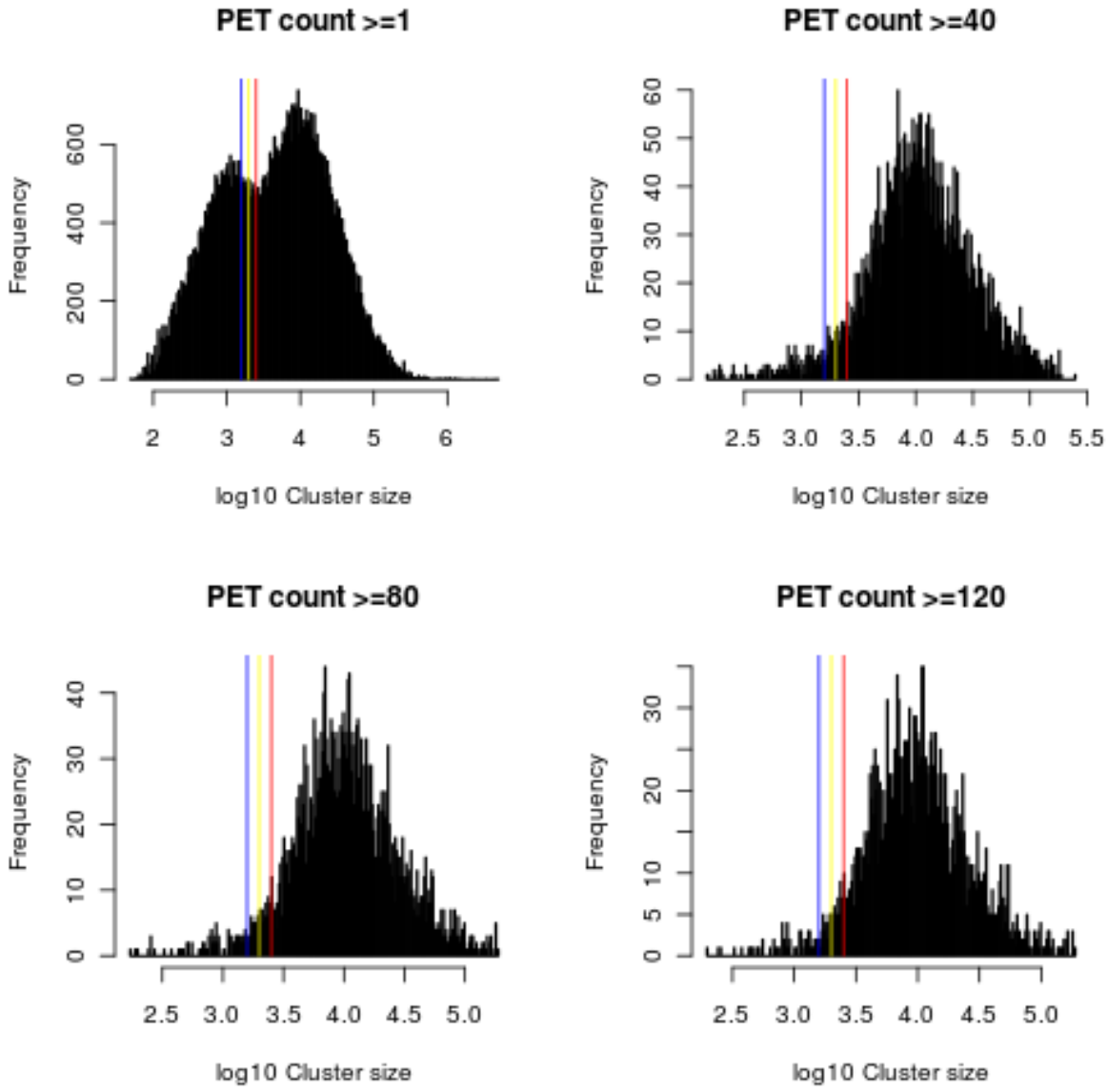
Supplementary Figure 20: Distribution of the genomic span of individual RNA-PET clusters in adult skeletal muscles.
Blue, yellow and red lines denote ~1.5, ~2.0 and ~2.5 kb, respectively.



Supplementary Figure 21: Distribution of the genomic span of individual RNA-PET clusters in adult kidney.
Blue, yellow and red lines denote ~1.5, ~2.0 and ~2.5 kb, respectively.



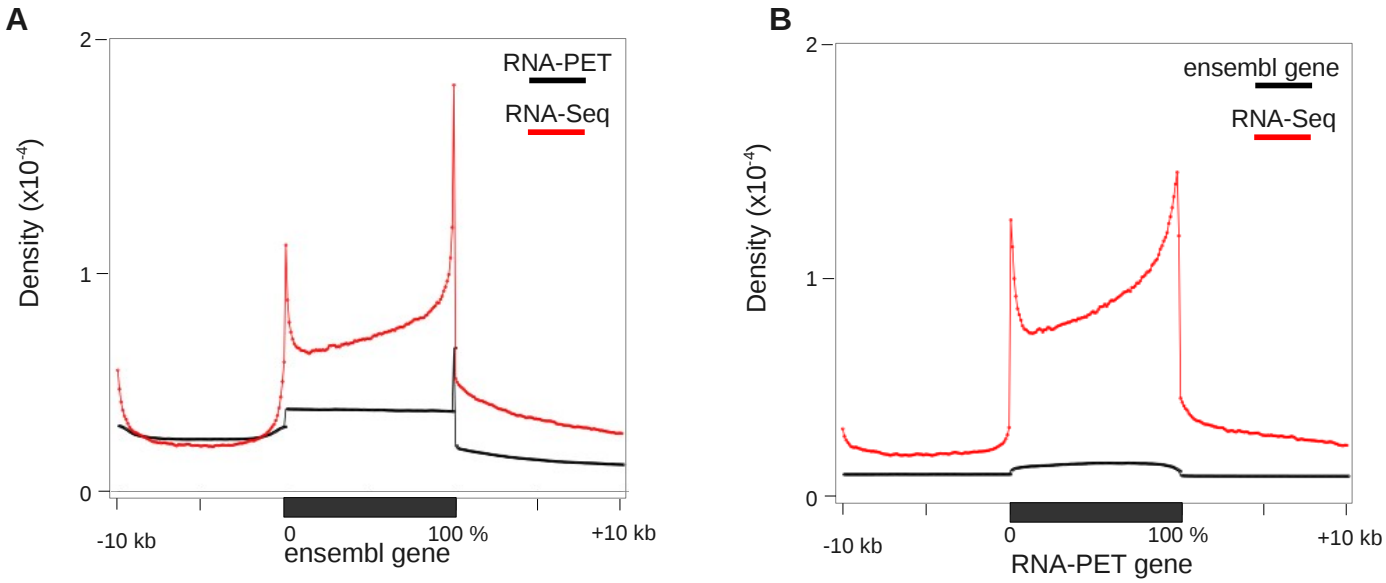
Supplementary Figure 22: Distribution of the genomic span of individual RNA-PET clusters in tadpole tail fin skin.
Blue, yellow and red lines denote ~1.5, ~2.0 and ~2.5 kb, respectively.



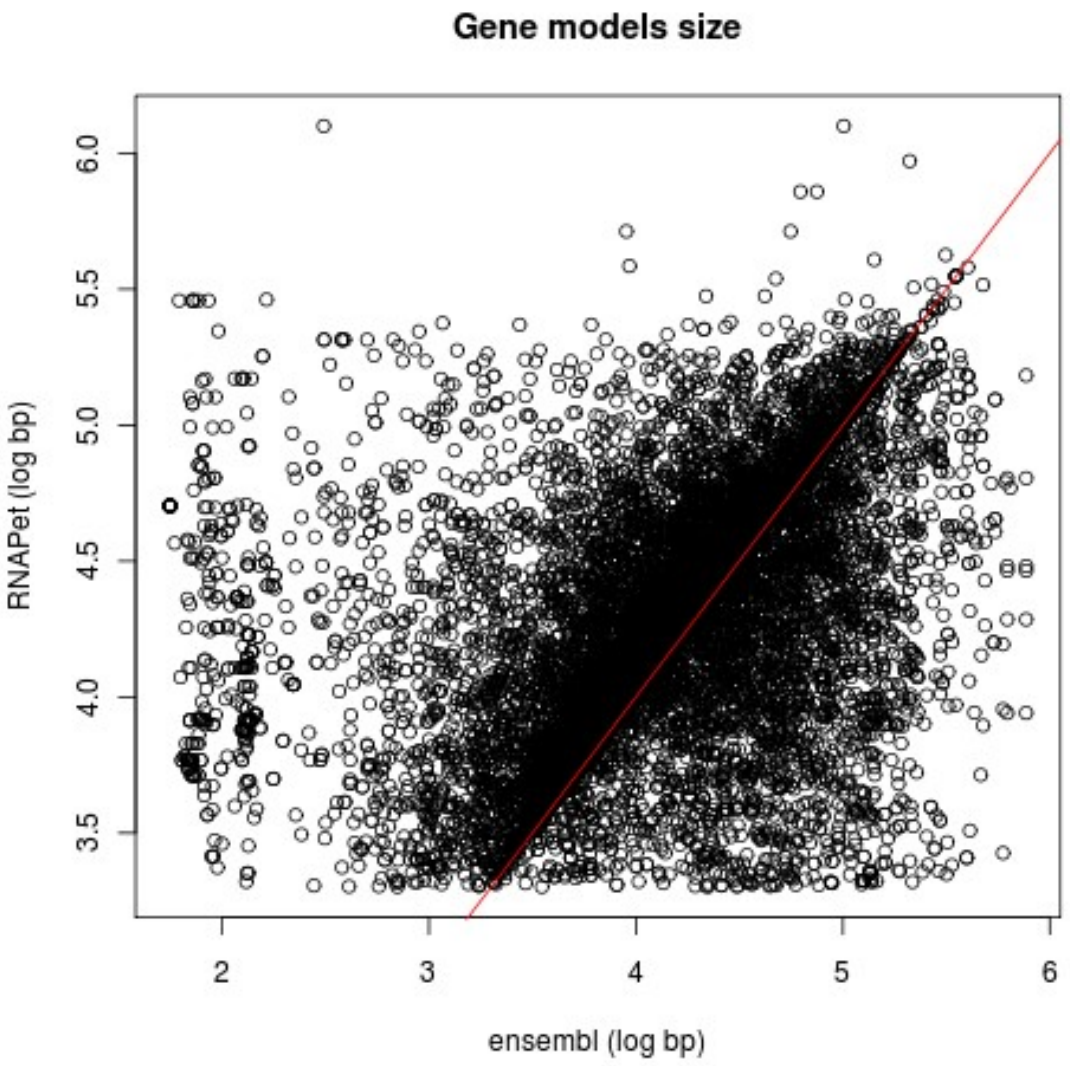
Supplementary Figure 23: Distribution of the genomic span of individual RNA-PET clusters in tadpole limbs.
Blue, yellow and red lines denote ~1.5, ~2.0 and ~2.5 kb, respectively.

Library	Total sequences	Uniquely mapped PETs	Per cent	Clusters	Filtering
Brain	20799532	7723467	37.1	366523	287331
Intestine	25918053	11569457	44.6	36310	28969
Kidney	20498589	8816384	43.0	300205	236338
Muscle	20998351	6631129	31.6	12503	9444
Liver	22834645	11763691	51.5	27311	2365
Limb	20955120	10184676	48.6	171642	133256
Tail fin skin	24058162	10934696	45.5	83702	63854

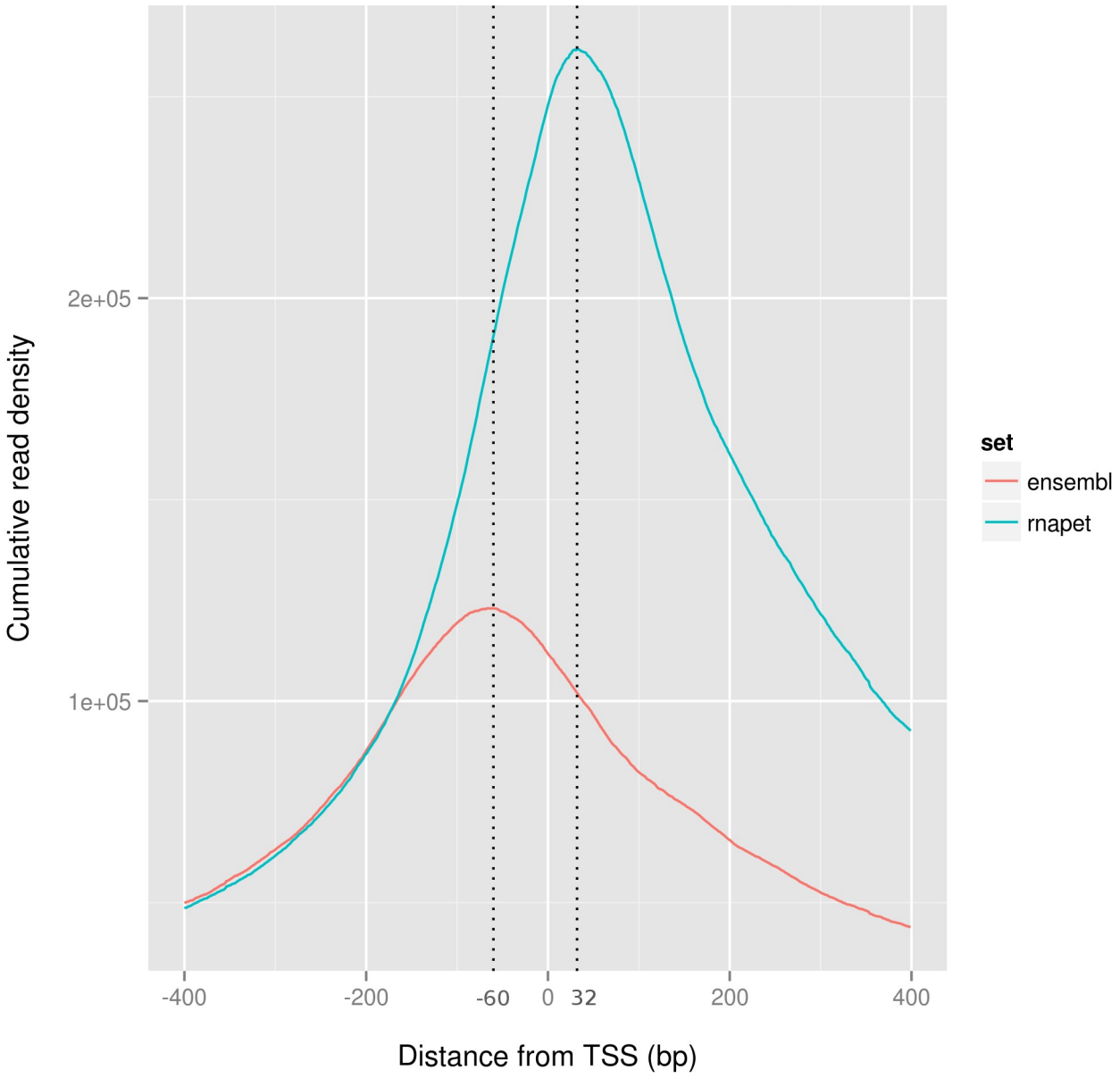
Supplementary Figure 24: RNA-PET clustering statistics.



Supplementary Figure 25: RNA-PET significantly extends gene models.
A. Density distribution of RNA-Seq reads and RNA-PET ditags overt ensembl gene models. B. Density distribution of RNA-Seq reads and ensembl genes booundaries over RNA-PET based gene models. Gene coverage is expressed in per cent of gene length.



Supplementary Figure 26: RNA-PET significantly extends gene models.
The plot corresponds to the size of individual RNA-PET-based gene models relative to their Ensembl counterpart. Gene models of the same size between the two datasets would follow the red line.



Supplementary Figure 27: Enrichment of RNA-Pol II at the 5' end of gene models improved by RNA-PET.
The RNA-Pol II density was computed at the 5' end of gene models, based on Ensembl and RNA-PET. The 5' end of Ensembl models is located 60 bp downstream of the RNA-Pol II peak (red curve), typically located ~25-45 downstream the TSS. In contrast, the 5' end of models enriched with RNA-PET data extend 32 bp upstream of RNA-Pol II peak.