SUPPLEMENTARY MATERIAL

Results for ISOPLOTTER 2.4

Here we provide results for ISOPLOTTER 2.4 for the simulated scenarios I and II as described in the manuscript. We used ISOPLOTTER with the default settings. Notice that the ISOPLOTTER software package also provides homogeneity tests for each segment detected by the method. In principle, such tests could also be carried for D_{JS} and B-SMUCE. But as our focus is on the accuracy of the obtained segmentation, we do not investigate this possibility further here.

Figure S-1: Sensitivity rate of B-SMUCE and ISOPLOTTER under the simulated scenario 1.

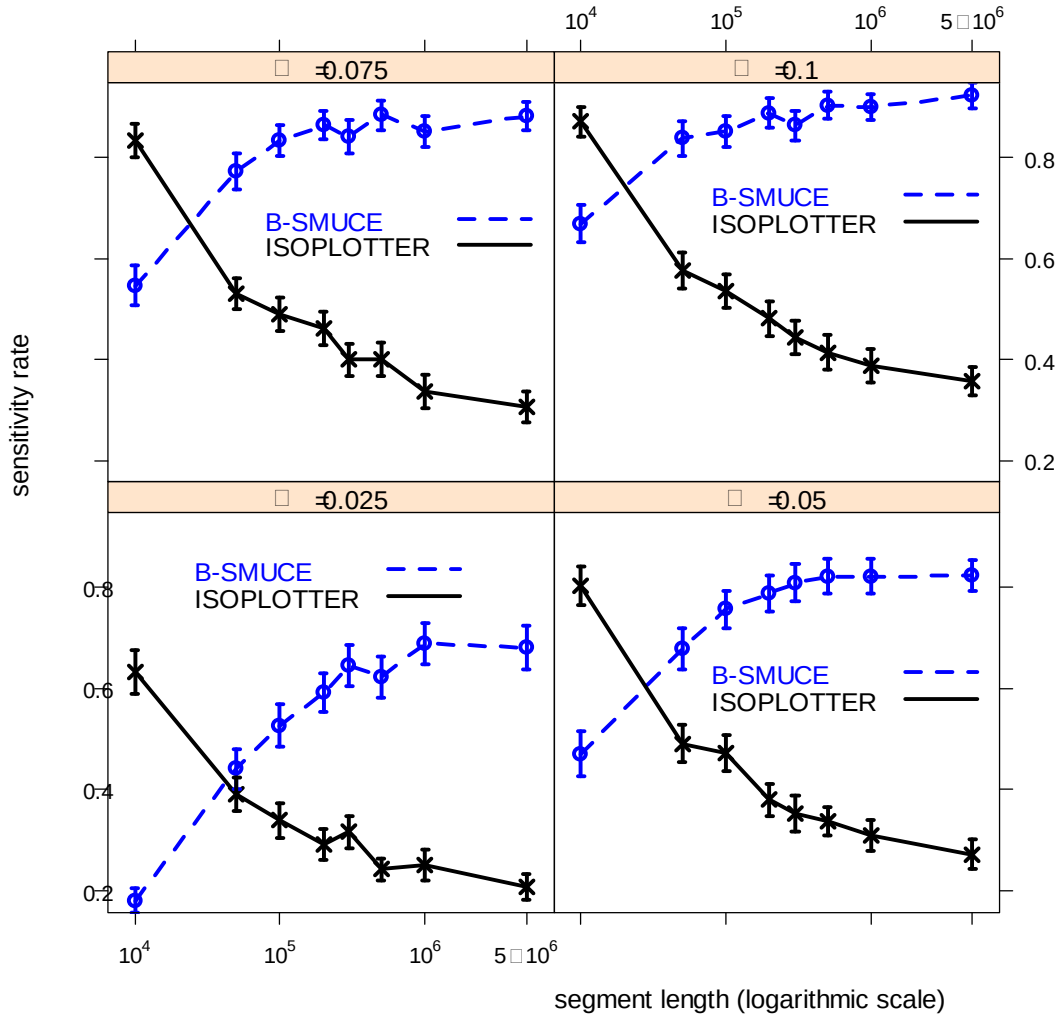
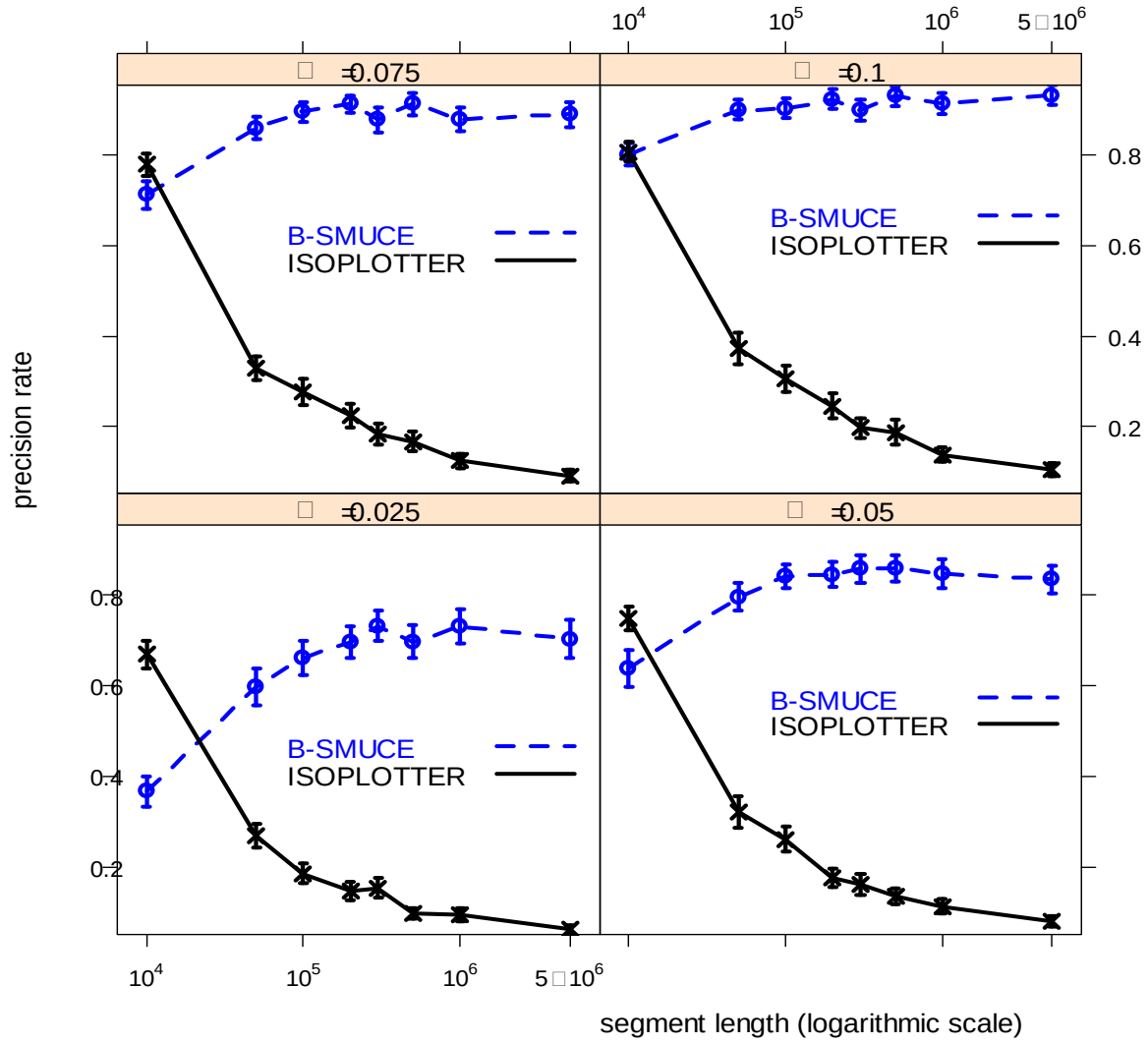


Figure S-2: Sensitivity rate of B-SMUCE and ISOPLOTTER under the simulated scenario 1.



Notice that the sensitivity and precision rate of ISOPLOTTER decreases with the sequence length. The reason is that the number of segments falsely detected by ISOPLOTTER tends to decrease as the length of the considered sequence increases.

Figure S-3: Logarithm (base 10) of FPSLE of B-SMUCE and ISOPLOTTER under the simulated scenario 1.

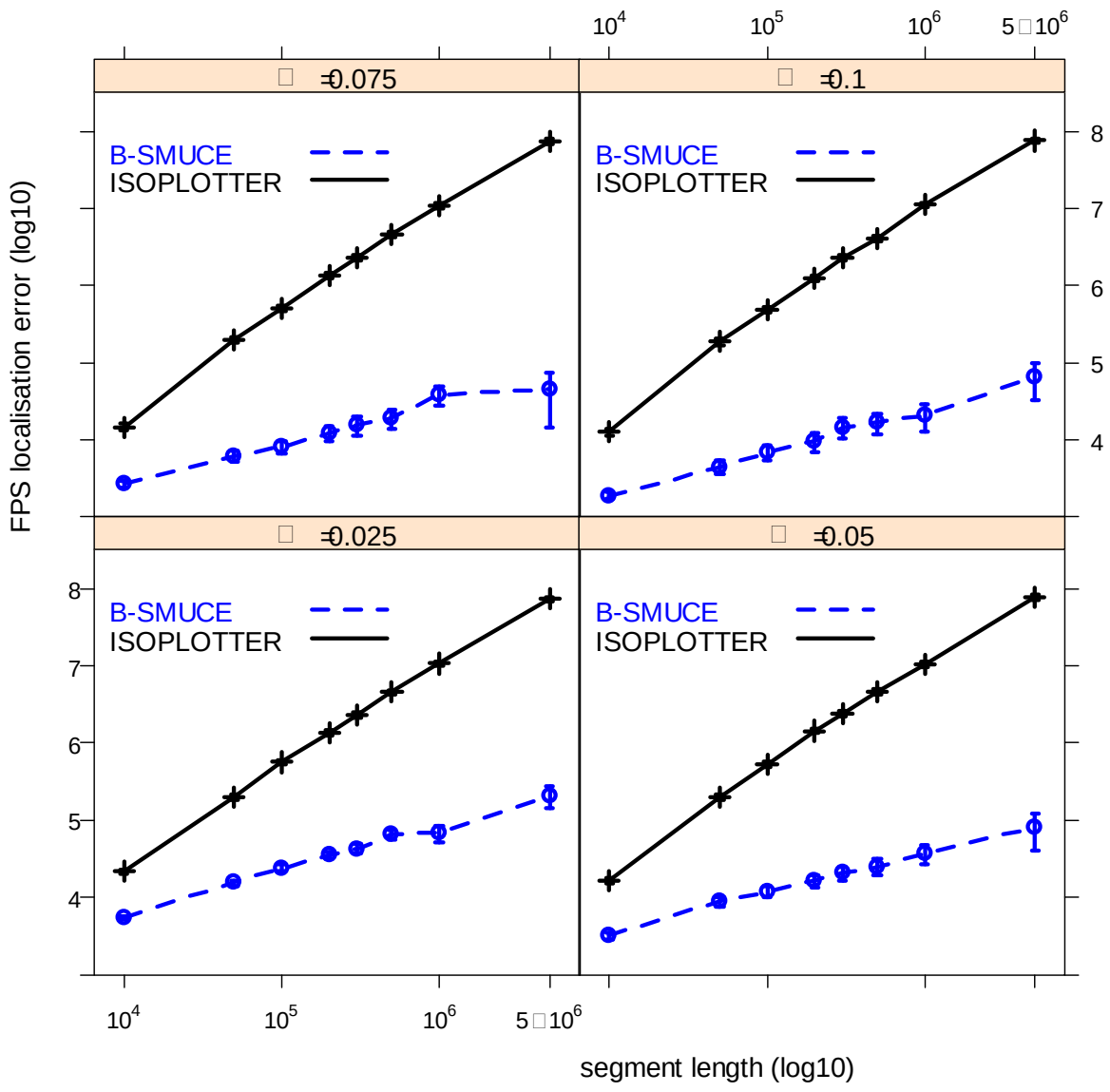


Figure S-4: Logarithm (base 10) of FNSLE of B-SMUCE and ISOPLOTTER under the simulated scenario 1.

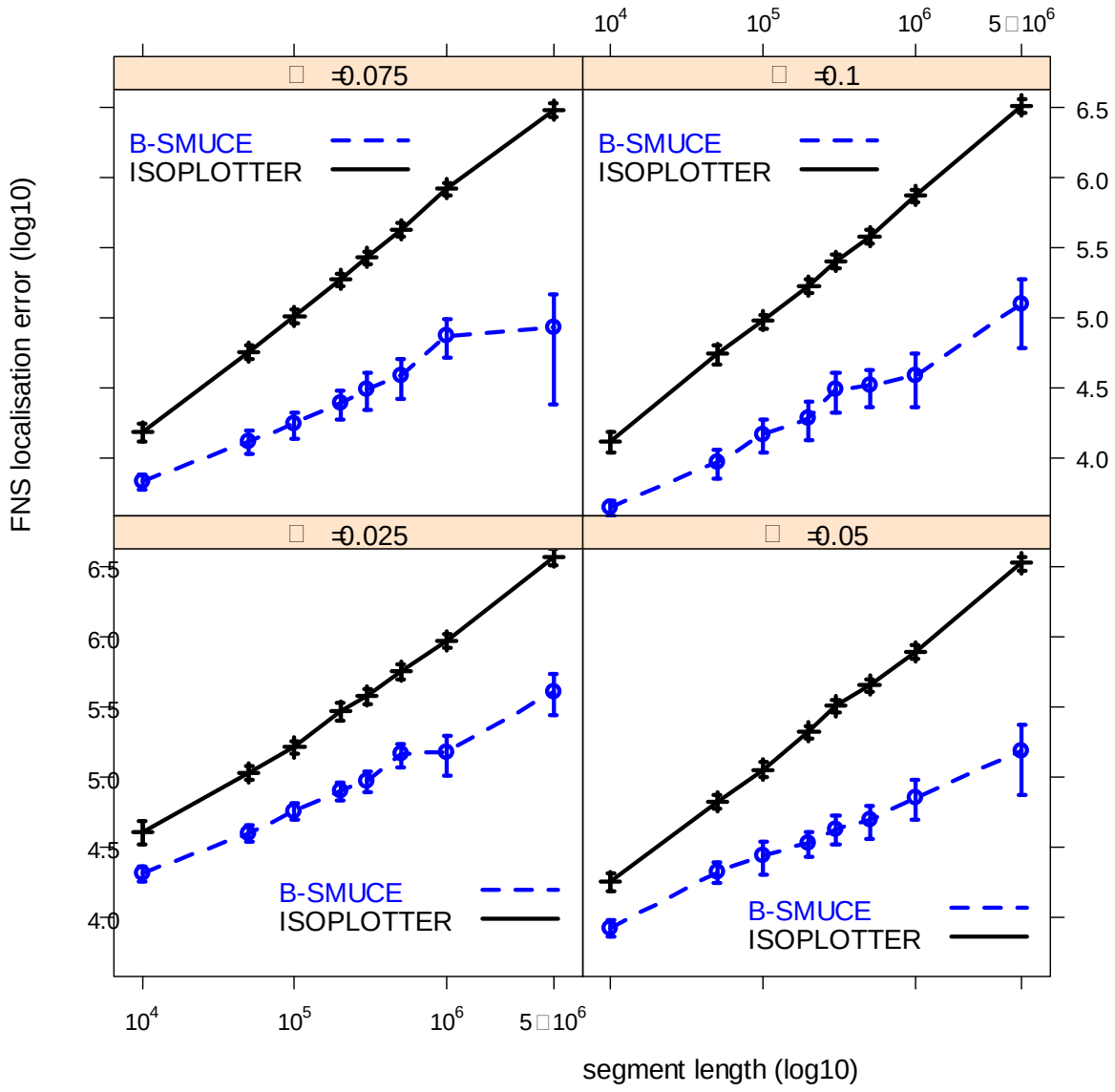


Table S-1: Average behavior of ISOPLOTTER under scenario 2 based on 100 simulation runs.

ISOPLOTTER	MEAN	STANDARD ERROR
Avg. true number of segments	27.51	1.90
Avg. number of true positives	6.63	0.57
Avg. number of false positives	4.68	0.41
Avg. number of false negatives	20.88	1.58
Avg. sensitivity	0.27	0.018
Avg. precision	0.56	0.027
FPSLE	0.55	0.018
FNSLE	3.20	1.58

HUMAN GENOME ANALYSIS

We used *bedtools 2.17 (http://bedtools.readthedocs.org/en/latest/)* to investigate the overlap between the segmentation obtained using B-SMUCE and DJS and known annotation. The human genome sequences (hg19, chr1:50000002-60000000 and hg19, chr6:29,677,952-33,289,874) we considered were obtained from the UCSC Genome Browser.

1. Overlap between segments identified on chromosome 1 and known annotation.

Using the function *intersectbed,* we filtered segments that have more than 50% mutual overlap with known genome features, such as genes, CpG islands and repeats. The required overlap ensures that the segment covers more than 50% of the feature, and the feature extends over more than 50% of the segment. In other words, the length of the intersection of the feature and the segment has to be more than 50% of both segment and feature length in order to be counted as a match.

As can be seen in the table below, D_{JS} provides more frequent overlap with genes and mRNA, whereas B-SMUCE provides a better representation for shorter sequences such as CpG islands and repeats. A plausible reason is that B-SMUCE provides segmentation at a finer scale, with segments more often also within genes, for instance at exon boundaries. This is in line with the finding that on the considered portion of the human chromosome 1, D_{JS} detected 214 segments, whereas B-SMUCE detected 1096 segments. IsoPlotter detected 27976 segments.

Table S-2: Human Chromosome 1 (hg19, chr1:50000002-60000000): Instances of more than 50% overlap between genome annotation (genes, mRNA, CpG islands, repeats) and segmentation for B-SMUCE, IsoPlotter and DJS.

Chr1	genes	mRNA	CpG islands	Simple repeats	Interrupted repeats
B-SMUCE	74	255	41	31	167
DJS	135	275	15	0	23
IsoPlotter	16	65	52	310	710
TOTAL	262	802	85	3029	2410

Since the CG content may fluctuate for various reasons, there are many segments matching none of the features detected above.

2. Overlap between segments identified on chromosome 6 and known annotation.

As in chromosome 1, we also applied a quantitative analysis of the match between segments and relevant genome annotation. Here B-SMUCE detected 640 segments (α=0.05), whereas D_{JS} found 182 and Isoplotter 227 segments. Here B-SMUCE identifies clearly many more matching features than the other considered method.

Table S-3: Human Chromosome 6 (hg18, chr6:29,677,952-33,289,874): Instances of more than 50% overlap between genome annotation (genes, mRNA, CpG islands, repeats) and segmentation for B-SMUCE, IsoPlotter and DJS.

chr6	Genes	mRNA	CpG islands	Simple repeats	Interrupted repeats
B-SMUCE	163	858	2	0	81
DJS	0	0	0	0	0
IsoPlotter	0	0	0	0	0
TOTAL	476	2923	225	812	757

As an example, we show the segmentation obtained for three sequence pieces, according to DJS (Jensen-Shannon divergence), and the B-SMUCE criterion. The shading of the segmentation is determined by the GC content of the segments: Darker segments have a lower GC content than brighter segments. As further annotation, we provide Genes, mRNA, repeats, CpG islands and microsatellites.

Figure S-5: Examples from the segmentation of the human chr. 6 (hg18, chr6:29,677,952-33,289,874) plus further annotation