Supplementary Figures

**Supplementary Figure 1. Alignment performance of simulated data.**

A total of 100,000 reads were simulated for each size category considered, in the presence of typical Illumina sequencing errors as well as nucleotide mis-incorporations remaining following USER-treatment. MQ refers to the mapping quality scores of the read alignments. This figure is the same as Figure 2, except that the respective contributions of each individual mapping quality scores are indicated. **(A) Fractions of true positive, false positive and false negative alignments.** **(B) Mapping quality scores of false positive alignments. (C) Mapping quality scores of true positive alignments.**

****

**Supplementary Figure 2. Mapping quality scores of false positive alignments.**

This figure zooms the fraction of false positive alignments with mapping quality scores between 20 and 30. It corresponds to the same information as shown in Figure S1B, however, the scale of the Y-axis now allows to identify the relevant classes of sequencing reads.



**Supplementary Figure 3. Read alignment sensitivity and positive predictive values.**

The performance metrics were calculated using simulated sequence data in the presence of typical Illumina sequencing errors as well as nucleotide mis-incorporations remaining following USER treatment. Alignments were filtered for minimal mapping quality scores of 30 and PCR duplicates. **(A) Alignment sensitivity** (ie true positives / (true positives + false negatives)**. (B) Alignment positive predictive value** (ie true positives / (true positives + false positives)).



**Supplementary Figure 4. Average depth-of-coverage in four dinucleotide contexts with soft-clipped bases (real data).**

The average depth-of-coverage was estimated filtering alignments for minimal mapping quality scores of 30 (MQ≥30) and removing PCR duplicates. Coverage values are calculated in the dinucleotide sequence context most affected by DNA methylation (CpG), as well as the three other dinucleotides potentially affected by post-mortem Cytosine deamination at the same position (ie CpA, CpC and CpT). The differences observed are not due to soft-clipped bases as the values returned in the presence or not of soft-clipping masking are identical.



**Supplementary Figure 5. Computational running times.**

The times provided represent the running time that was necessary for Paleomix to process the sequence data of each specimen or individual indicated, using the same number of CPUs (E5-2683 v4 at 2.10GHz). Total running times were added for those individuals showing sequencing data generated both in the absence (USER-) and in the presence of USER treatment (USER+) in order to improve the figure readability.

