Figure S1. Outline of the mapping algorithm details.

Algorithm BatMeth2 main Algorithms.

Input: $BatMeth2$(G,R); G: Genome; R: input sequence;
Output: Alignment hits.

1: $H_0 = BatMis(G_{c2t}|G_{g2a}, R, 0)$;
2: $Aligns = H_0$;
3: **if** $Aligns > 1$ **then**
4: report;
5: **end if**
6: $H_1 = BatMis(G_{c2t}|G_{g2a}, R, 1)$;
7: $Aligns = Aligns + H_1$;
8: **if** $H_0 > 0 \,||\, Aligns > 1$ **then**
9: report;
10: **else**
11: $Im_0 = Search_Indels(mis_0, indels_1, |R|)$;
12: $Aligns = Aligns + Im_0$;
13: **if** $Aligns > 1$ **then**
14: report;
15: **end if**
16: **end if**
17: $H_2 = BatMis(G_{c2t}|G_{g2a}, R, 2)$;
18: $Aligns = Aligns + H_2$;
19: **if** $H_1 + Im_0 > 0 \,||\, Aligns > 1$ **then**
20: report;
21: **end if**
22: $Im_1 = Search_Indels(mis_1, indels_1, |R|)$;
23: **if** $Aligns > 0$ **then**
24: $Aligns = Aligns + Im_1$;
25: report;
26: **end if**
27: $Im_2 = Search_Indels(mis_2, indels_1, |R|)$;
28: **if** $Aligns > 0$ **then**
29: $Aligns = Aligns + Im_1$;
30: report;
31: **end if**

Figure S2. The overlap of the correct methylation callings from IMR90 cell line based on 450K BeadChip data for all compared software.

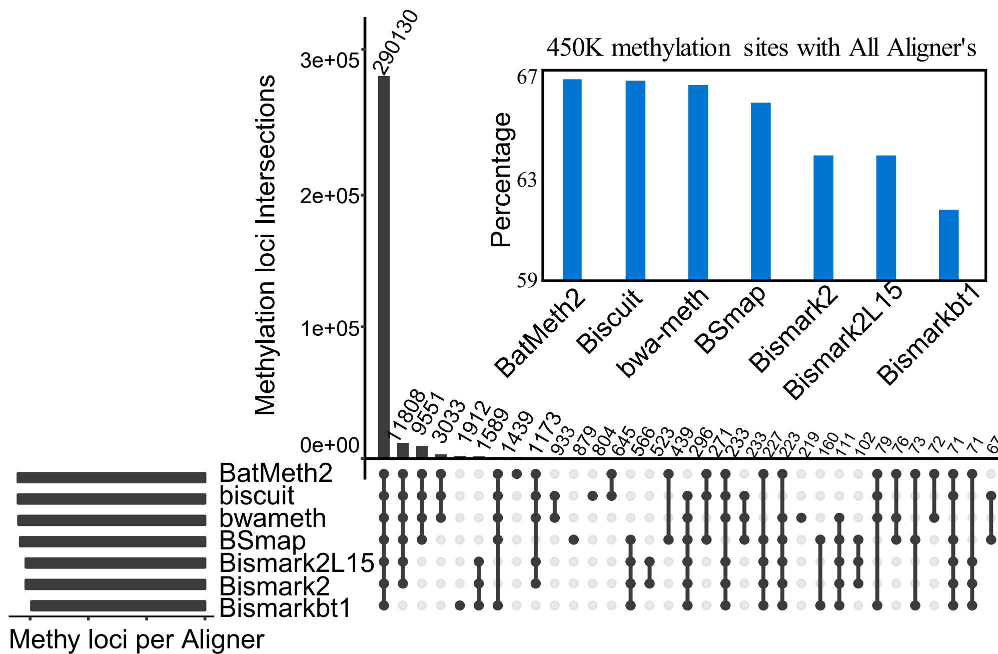

Figure S3. BatMeth2 align BS reads allowing for variable-length indels

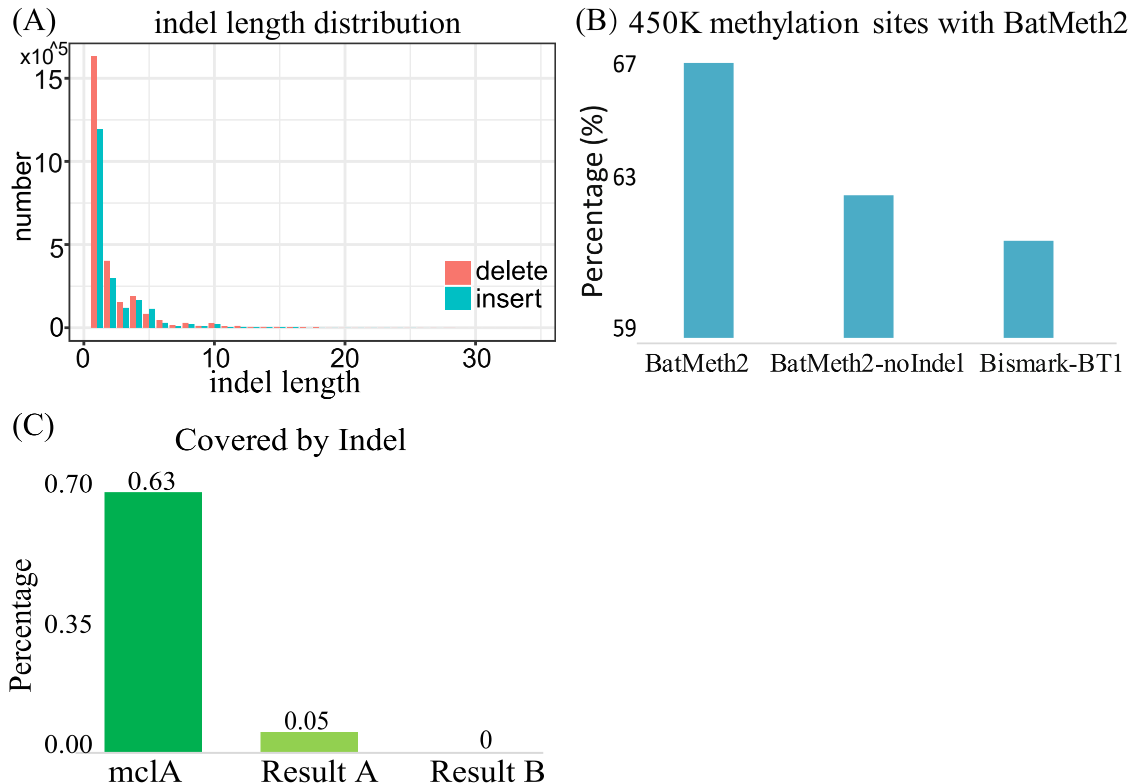

Figure S3. (A) Indel length distribution detected by BatMeth2. (B) The overlap of 450K with BatMeth2, BatMeth2 no indel detect mode and Bismark-bowtie 1(bismarkBT1). (C) More correct methylation loci in result A (mclA) covered by Indel distribution. We define the methylation sites called by BatMeth2 as Result A while the methylation sites called by BatMeth2-noIndel and Bismark as Results B. Let mclA be the methylation sites appear in Result A but not Result B.

Figure S4. The DNA methylation level distribution across exon, intron, intergenic and TEs, etc.

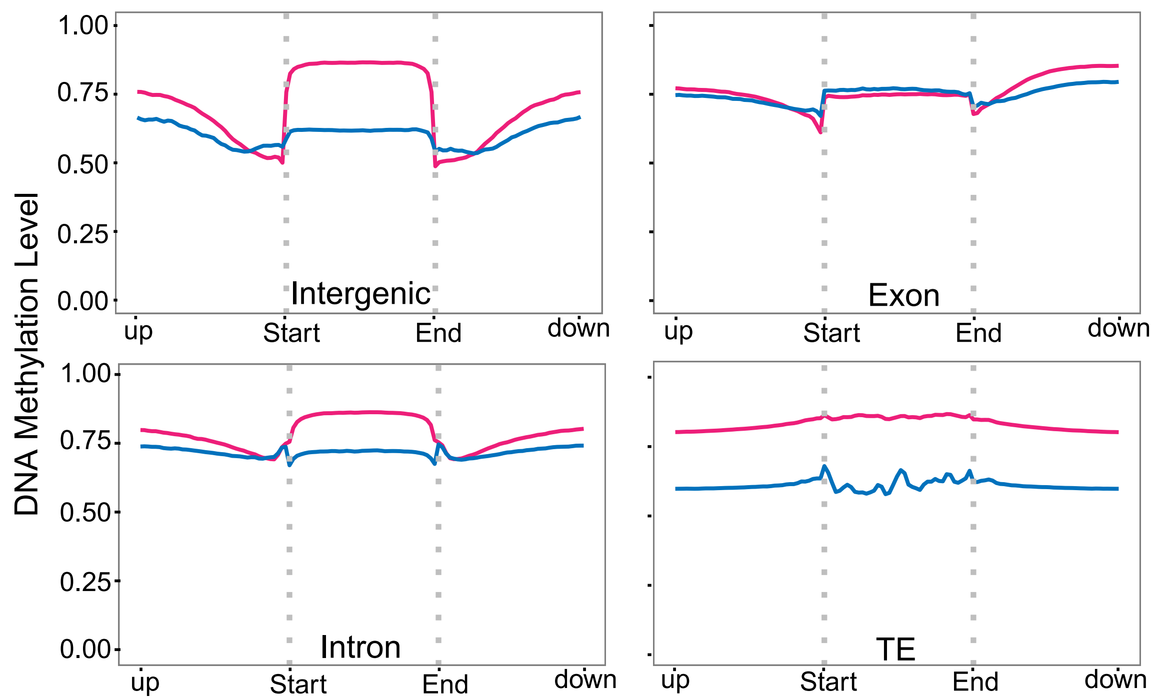

Figure S5. Methylation level under different conditions

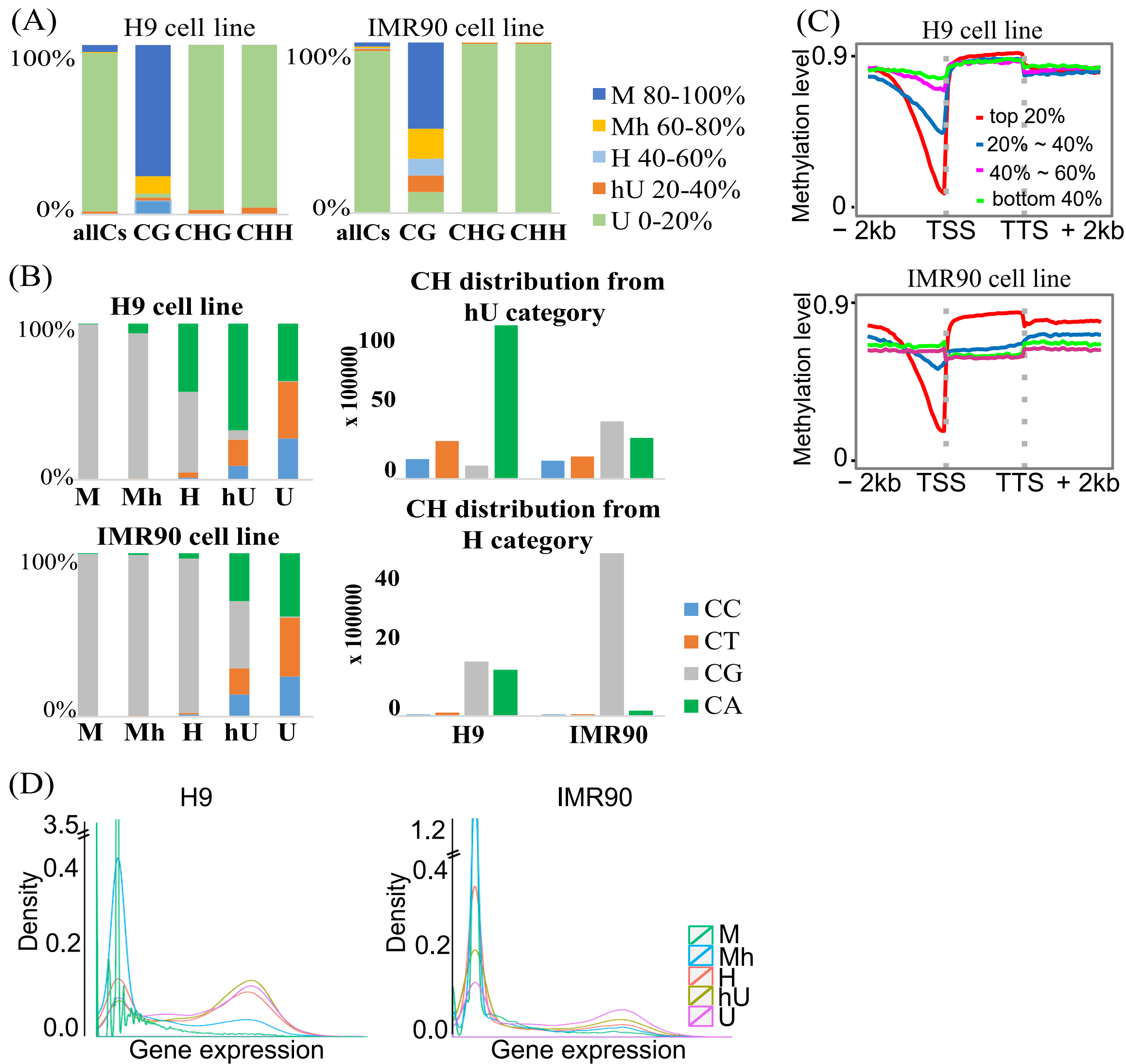

Figure S5. (A) Distribution of DNA methylation levels and the corresponding sequence context. Methylation levels were grouped into five categories: M (>80%), hM (60%<m<=80%), H 40% <m<=60%), hU (20%<m<=40%), U (<=20%). Levels of methylation were found to be highest in H9 cell line at M category than IMR90 cell line. (B) DNA methylation in various combinations of sequence contexts (CH; H = A/T/C/G) throughout the genome was examined. In the CH sequence context, CpG methylation was the predominant form, but a significant fraction of methylated cytosines were found at CpA sites while methylation level less than 40%, particularly in H9 cell line. (C) Correlation of methylation profile with expression level in H9 and IMR90 cell line. The expression levels of genes in H9 or IMR90 were divided into many categories. The higher expressed genes exhibited the lower methylation levels at the upstream from their TSS.
(D) Distribution of expression levels from genes with different promoter methylation levels in H9 and IMR90 cell lines. The methylation levels of gene promoters in H9 or IMR90 cell lines were divided into five categories. The same conclusion as C, the higher methylation levels exhibited the lower expressed level.