Supplementary Materials

Supplementary Text

1. Parameter tuning of DeepSignal

We choose the k-mer window size and the number of neural network layers of our model based on the experiments of CpG methylation detection with *E.coli* R9 2D data. We use randomly selected 20 million samples as training set and another 1 million samples as validation set. We tested 11, 13, 15, 17, 19 and 21 as the k-mer window size on a 3-BRNN-layer and 11-inception-layer DeepSignal model, and select 17 based on the performances of models on the validation set (Fig. S11A). In the validation set, there are 719,442 singleton CpGs (of which there are no other CpGs in the up and down 10bp region). And there are 280,558 non-singleton CpGs. As shown in Fig. S10A, the accuracy for non-singleton CpGs keep increasing as the k-mer window size increases. However, the accuracy for singleton CpGs decreases when the k-mer window size is 19. We tested 1, 2, 3, 4 and 5 as the number of bidirectional recurrent neural network (BRNN) layers (Fig. S11B). And we tested 5, 11 and 17 as the number of inception layers (Fig. S11C). DeepSignal achieves the highest accuracy when there are 3 BRNN layers and 11 inception layers. We also tested the influence of number of samples for training on our model (Fig. S11D). The result shows that 20 million samples for training is sufficient for our model. Thus in this study, if there are less than 20 million samples in the training dataset, we use the whole samples to train DeepSignal. Otherwise we use randomly selected 20 million samples to train DeepSignal.

2. Time cost and memory usage

There are three steps of DeepSignal: feature extraction, training, testing. We performed the feature extraction step on a server with 48 CPU processors (Intel(R) Xeon(R) Gold 6126 CPU @ 2.60GHz). And we performed the training and testing steps on a server with 4*12GB TITAN V GPUs. Based on the experiments, we use 20 million samples (half positive, half negative) at most for training. The time cost and memory usage of DeepSignal are listed in Table S3. For the feature extraction step, the peak memory may vary depending on the number of processes used and the IO speed of the disk. For the testing step, besides performing the testing step on GPU, we also performed the testing step using 48 processors on the CPU server in the *E.coli/H.sapiens* experiment. The time cost and peak memory are 115.6min/18.5GB and 84.5min/18.3GB, for template and complement strand respectively. The results show that testing with CPU will cost at least ~2.5 times more time with ~33% less peak memory than using GPU.

Supplementary Figures

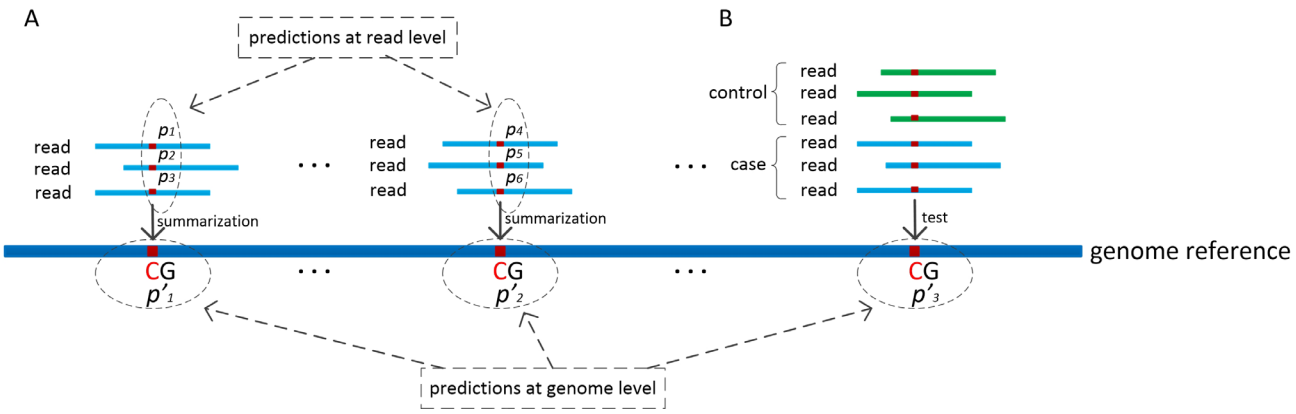

Fig. S1. Predicting methylation states of targeted DNA bases by model based methods (**A**) and statistics based methods (**B**). (Here we take CpG methylation state prediction as a example. p_1-p_6 are predictions at read level, p'_1-p'_3 are predictions at genome level.)

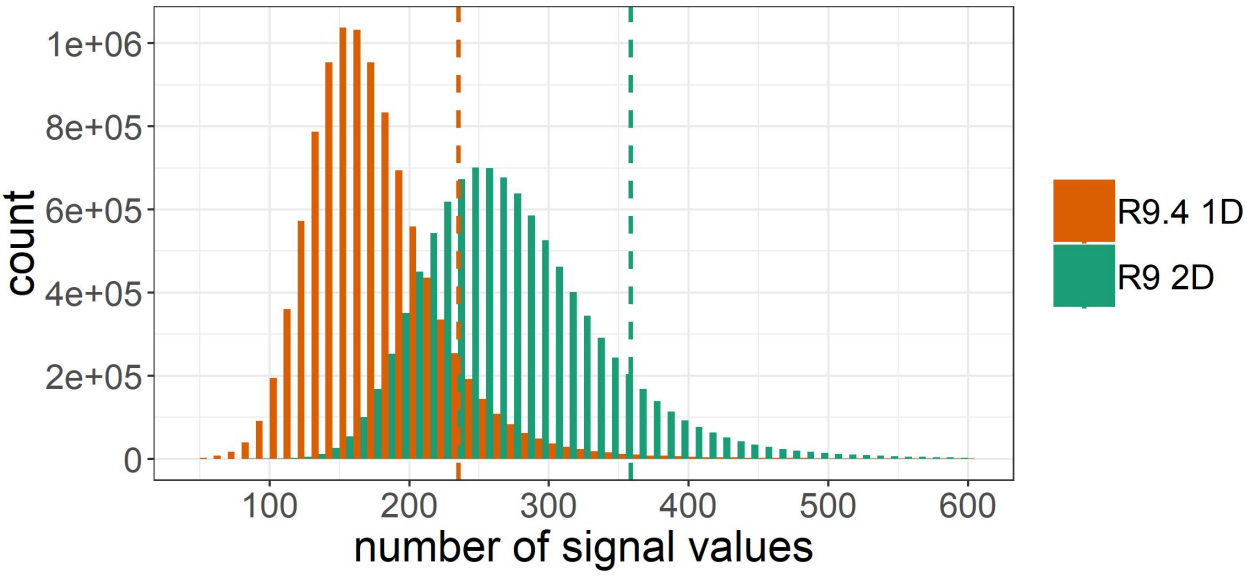

Fig. S2. The number of signal values of 10 million randomly selected 17-mers in R9.4 1D data and R9 2D data (Suppose the number of signal values obeys a Gaussian distribution (u, σ^2), the dash lines indicate approximately $u+\sigma$ signal values: 235 for R9.4 1D and 360 for R9 2D data, respectively.)

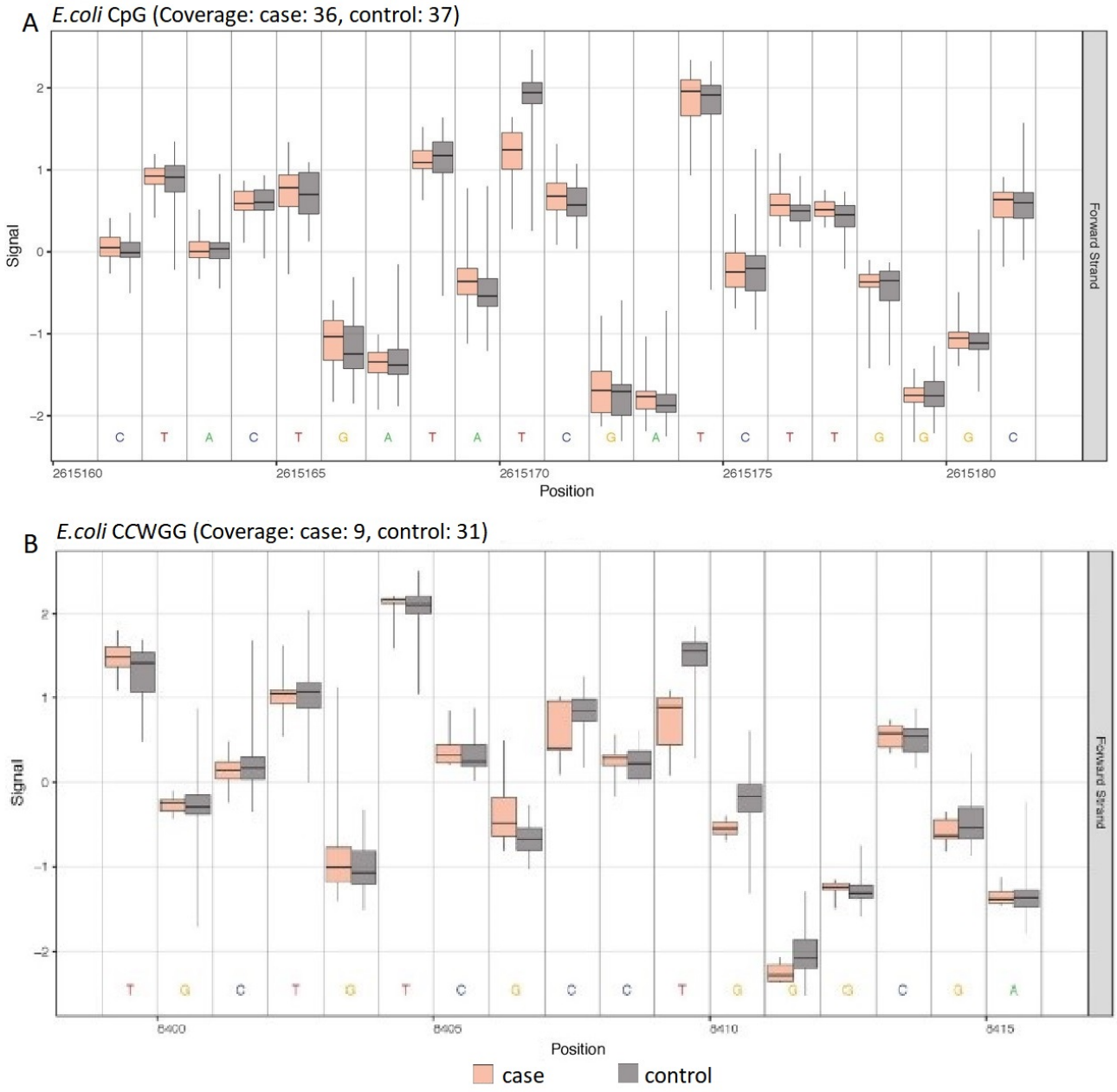

Fig. S3. Boxplots of signal distribution of each base in methylated and unmethylated CpG and CCWGG (x-axis is the position of bases in the genome reference, y-axis is the normalized raw signals. Case represents signals from methylated reads, control represents signals from unmethylated reads). **A:** CpG, signals are extracted from 36 methylated reads and 37 unmethylated reads of *E.coli* data. **B:** C*C*WGG, signals are extracted from 9 methylated reads and 31 unmethylated reads of *E.coli* data.

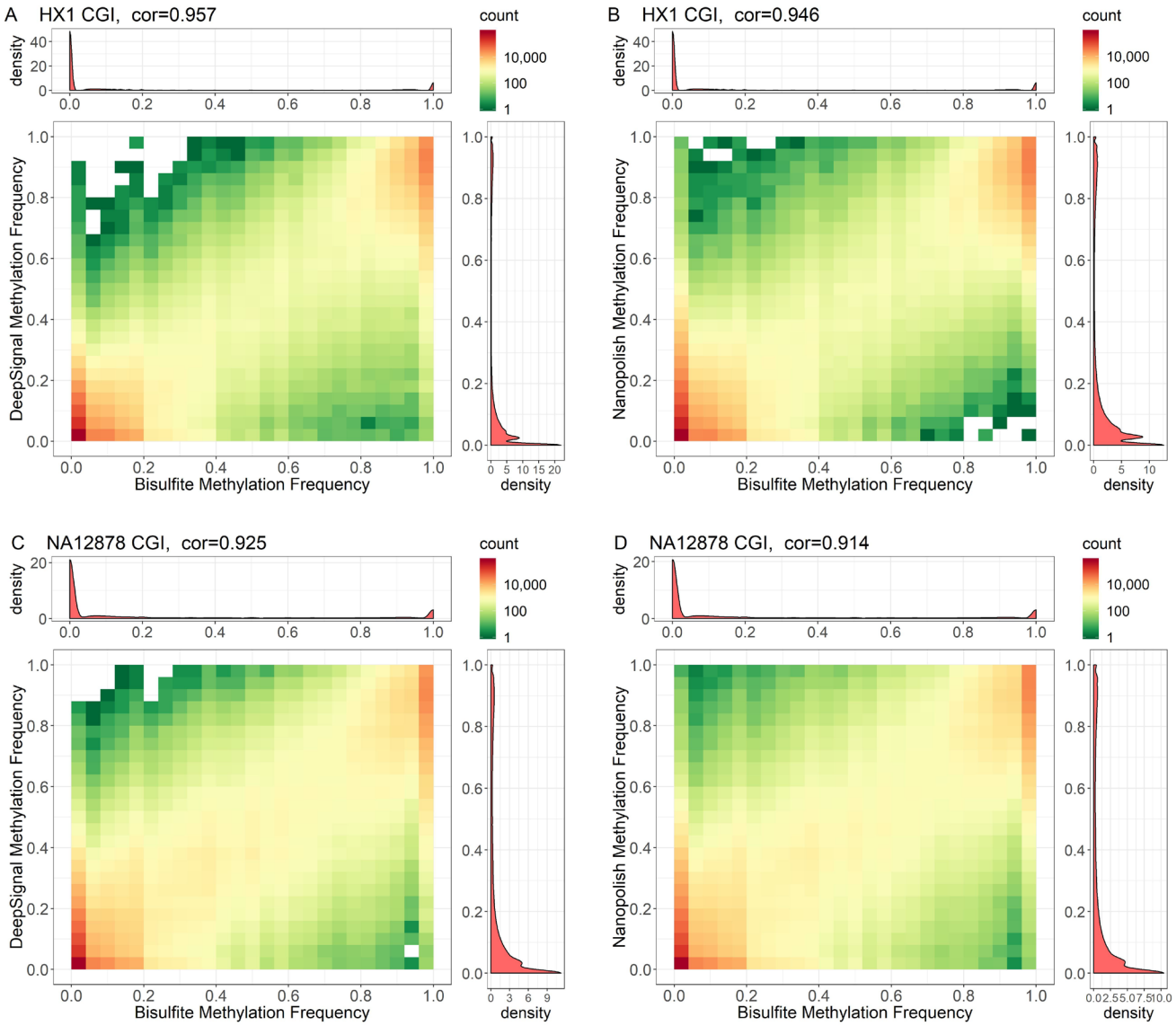

Fig. S4 Comparison of methylation frequencies of CpGs in CpG islands calculated by DeepSignal/nanopolish with those from bisulfite sequencing (cor is Pearson correlation). **A-B:** HX1; **C-D:** NA12878.

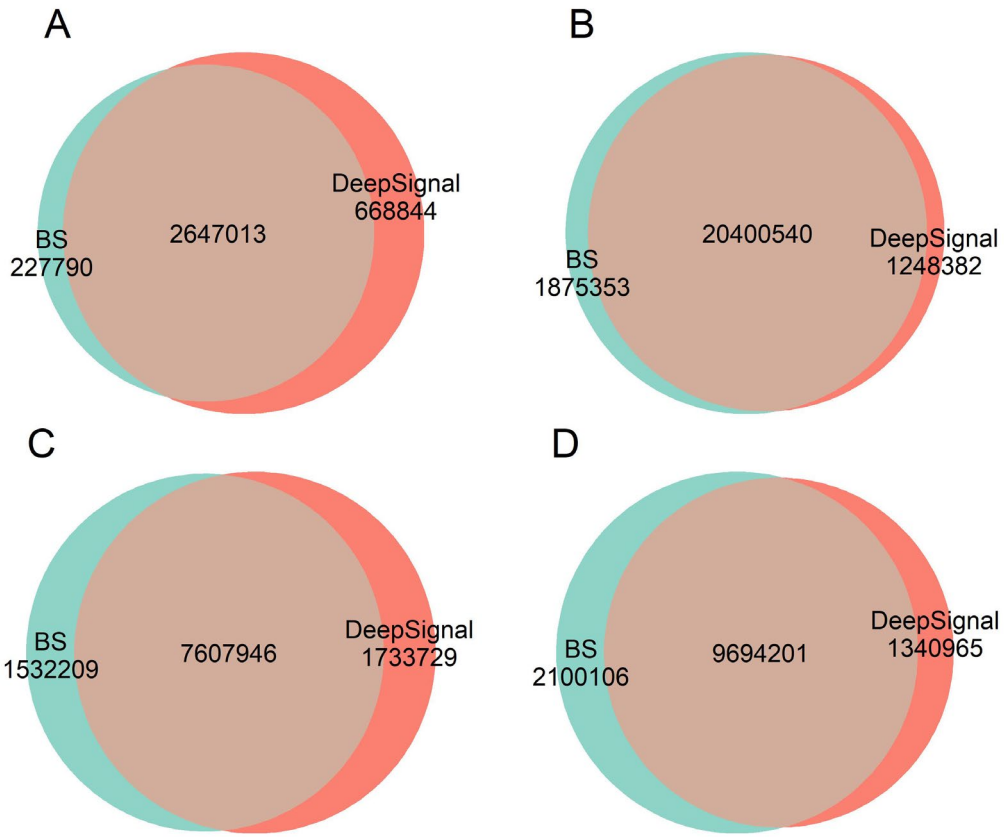

Fig. S5 Comparison of lowly methylated (methylation frequency <= 0.3) and highly methylated (methylation frequency >= 0.7) CpGs predicted by DeepSignal and bisulfite sequencing. Only CpGs covered with at least five reads are considered. **A:** number of lowly methylated CpGs in HX1; **B:** number of highly methylated CpGs in HX1;**C:** number of lowly methylated CpGs in NA12878; **D:** number of highly methylated CpGs in NA12878.

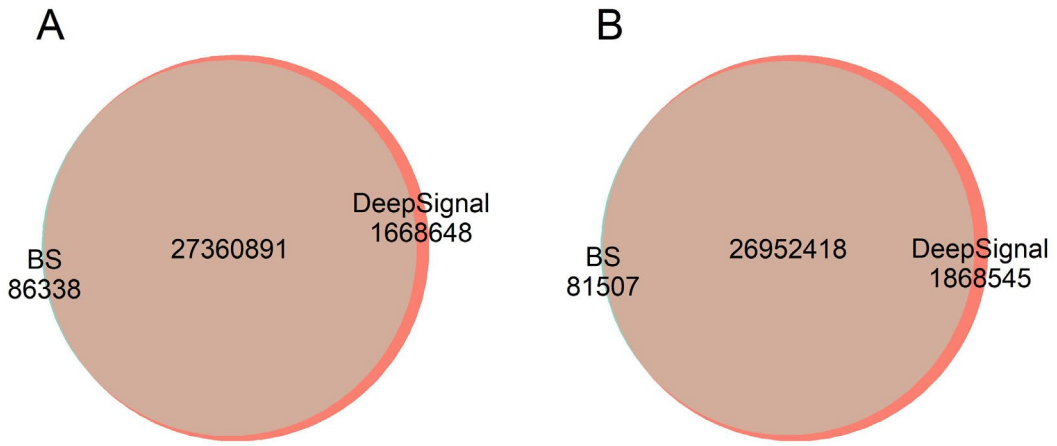

Fig. S6 Comparison of all CpGs in HX1 and NA12878 predicted by DeepSignal and bisulfite sequencing. **A:** HX1; **B:** NA12878. Only CpGs covered with at least five reads are considered.

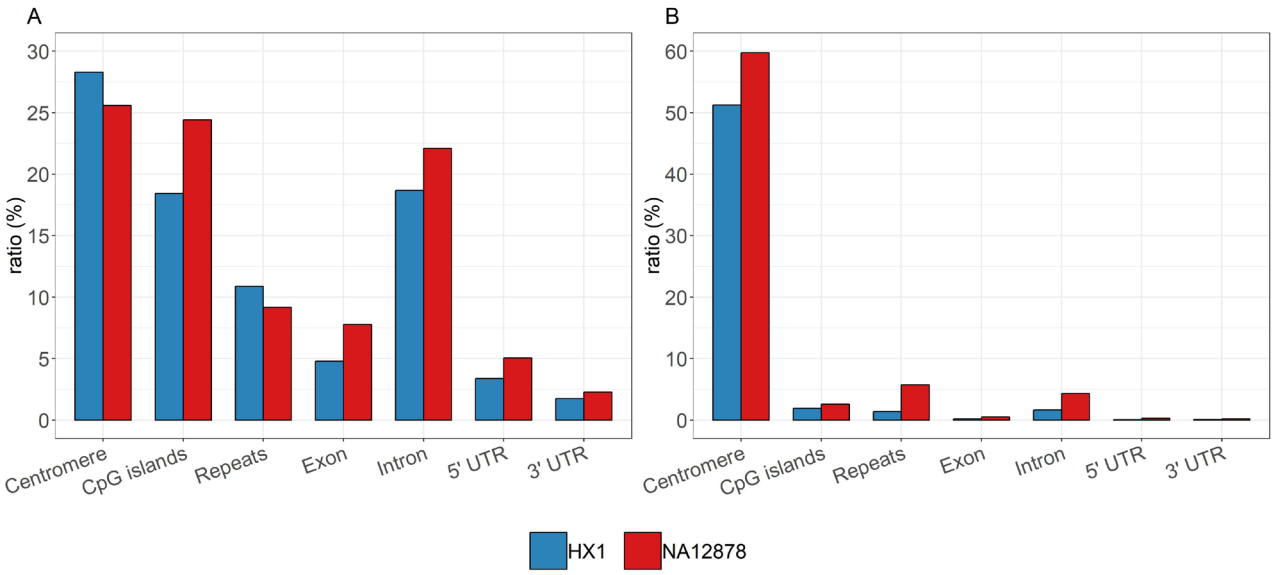

Fig. S7 Proportion of CpGs detected by DeepSignal or Bisulfite sequencing only in regions: Centromere, CpG islands, Repeats, Exon, Intron, 5' UTR and 3' UTR. (The annotations are all got from UCSC Genome Browser with *H. sapiens* GRCh38/hg38 genome version. Specifically, we get Centromere annotation, CpG islands annotation, Repeats annotation by choosing 'Centromere Locations', 'CpG Islands' and 'RepeatMasker' as track, respectively. We get Exon, Intron, 5' UTR and 3' UTR annotation by choosing 'GENCODE v29' as track.) **A:** CpGs (1,668,648 for HX1 and 1,868,545 for NA12878) detected by DeepSignal only; **B:** CpGs (86,338 for HX1 and 81,507 for NA12878) detected by Bifulfite sequencing only.

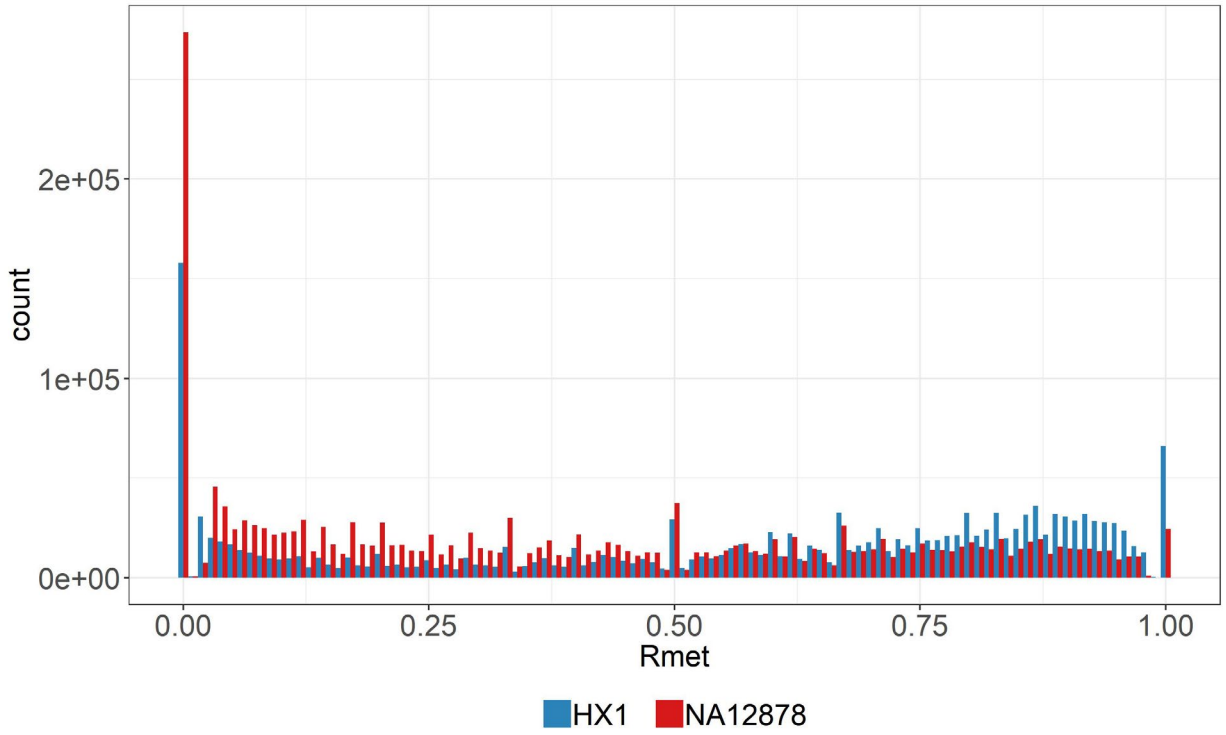

Fig. S8 Distribution of methylation frequency of the CpG sites can be detected by DeepSignal only.

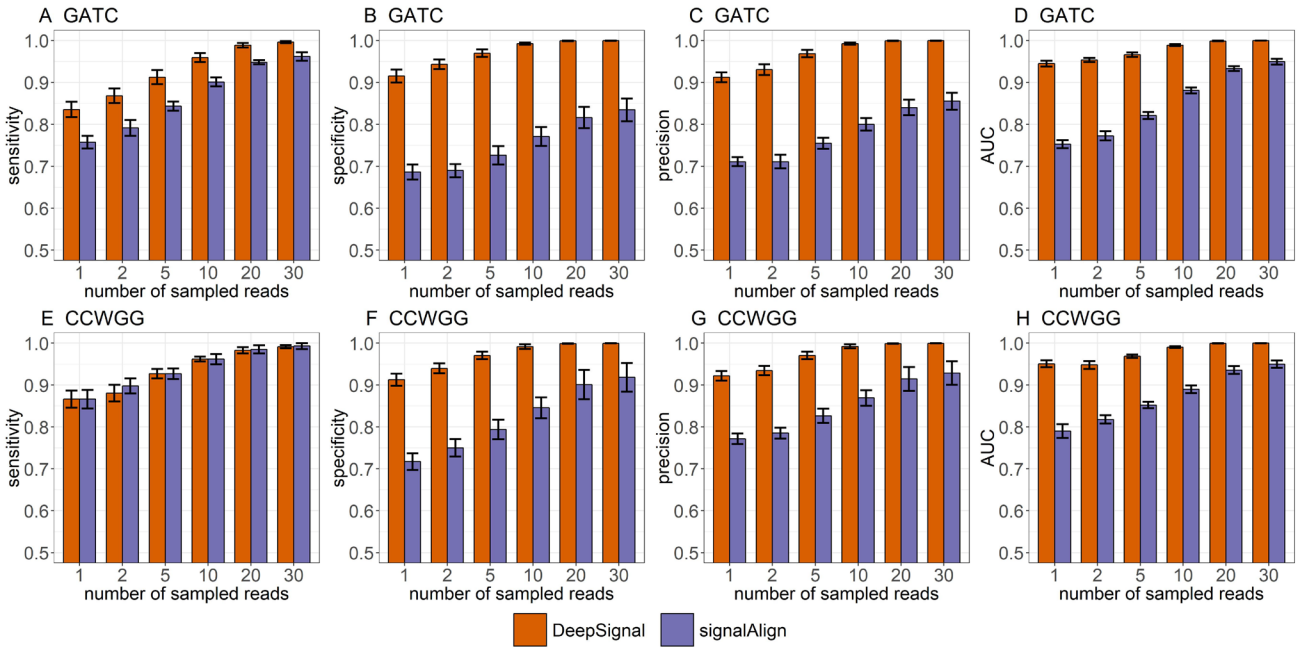

Fig. S9 Comparison of DeepSignal and signalAlign for predicting methylation states of 6mA in G*A*TC motifs and 5mC in C*C*WGG motifs of pUC19 DNA under different number of sampled reads. Values are average and standard deviation of 10 replicated train-test splits (To calculate the values of accuracy metrics, we first calculate the average values of 100 replicated tests in each train-test split. Then we calculate the average and standard deviation values of 10 train-test splits.). **A-D:** performances on predicting methylation states of 6mA in G*A*TC motifs; **E-H:** performances on predicting methylation states of 5mC in C*C*WGG motifs.

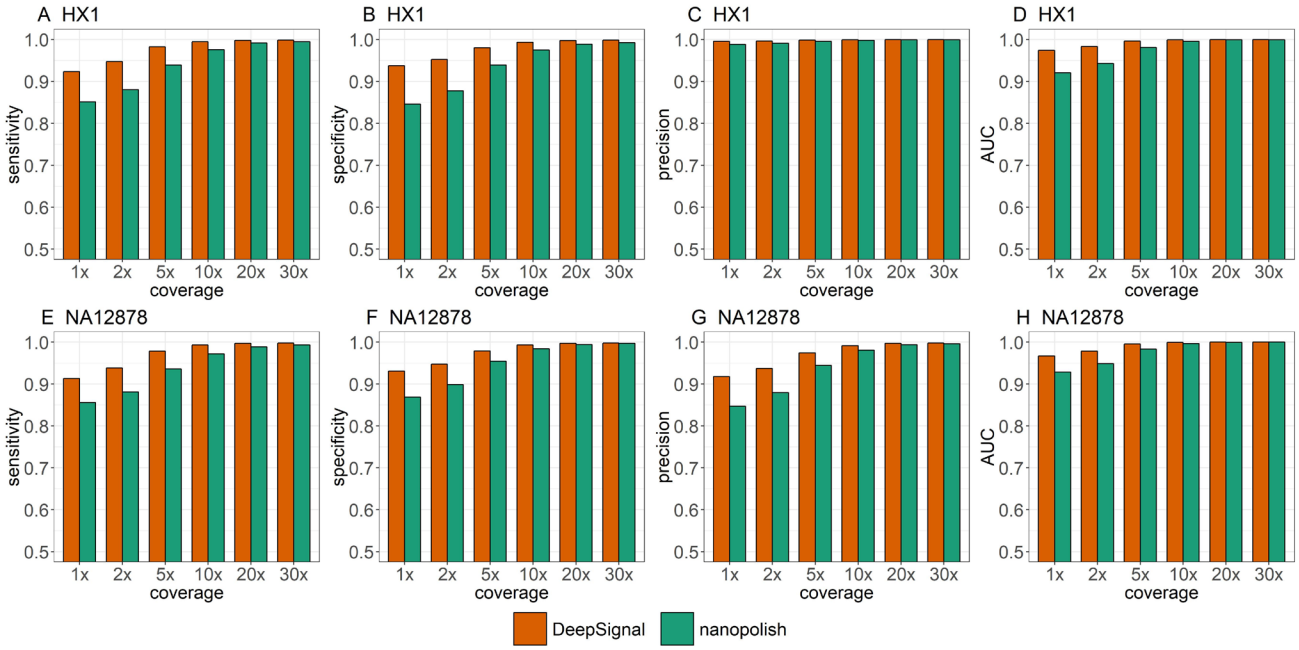

Fig. S10 Comparison of DeepSignal and nanopolish for predicting methylation states of CpGs at genome level under different coverages of H.sapiens R9.4 1D data (For each coverage, the reads are randomly shuffled and selected. Values are average of 10 replicated tests.). **A-D:** HX1; **E-H:** NA12878.

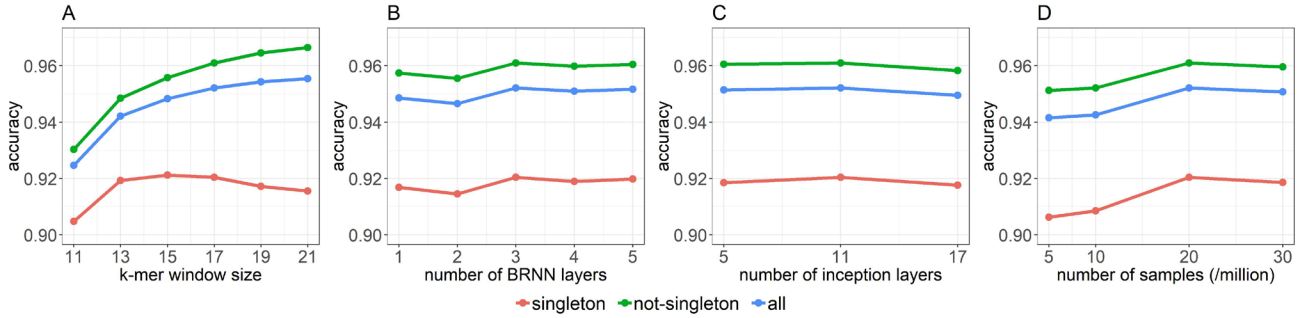

Fig S11. Parameter tuning of DeepSignal model. **A:** k-mer window size tuning (on a model of 3 BRNN layers and 11 inception layers) ; **B:** number of BRNN layers tuning (on a model of 11 inception layers, with 17 as k-mer window size); **C:** number of inception layers tuning (on a model of 3 BRNN layers, with 17 as k-mer window size); **D:** number of samples tuning (on a model of 3 BRNN layers and 11 inception layers, with 17 as k-mer window size)

Supplementary Tables

Table S1. Accuracies of DeepSignal and signalAlign for predicting methylation states of 6mA in G*A*TC motifs and 5mC in C*C*WGG motifs of pUC19 DNA under different number of sampled reads. Values are average and standard deviation of 10 replicated train-test splits (To calculate the values of accuracy, we first calculate the average values of 100 replicated tests in each train-test split. Then we calculate the average and standard deviation values of 10 train-test splits.).

method	6mA in G*A*TC motif prediction (Number of sampled reads)					
	1	2	5	10	20	30
DeepSignal	0.876 (+/-0.011)	0.901 (+/-0.009)	0.940 (+/-0.006)	0.976 (+/-0.005)	0.994 (+/-0.003)	0.998 (+/-0.001)
signalAlign	0.721 (+/-0.008)	0.736 (+/-0.010)	0.783 (+/-0.011)	0.836 (+/-0.009)	0.882 (+/-0.011)	0.898 (+/-0.013)
method	5mC in C*C*WGG motif prediction (Number of sampled reads)					
	1	2	5	10	20	30
DeepSignal	0.890 (+/-0.013)	0.905 (+/-0.011)	0.948 (+/-0.007)	0.977 (+/-0.003)	0.991 (+/-0.004)	0.996 (+/-0.002)
signalAlign	0.792 (+/-0.013)	0.819 (+/-0.009)	0.860 (+/-0.010)	0.904 (+/-0.011)	0.943 (+/-0.018)	0.956 (+/-0.017)

Table S2. Number of samples used for training, validation and testing by DeepSignal. (For R9 2D data, we train and test the data from template and complement strand of reads separately. For the *E.coli/E.coli*, *H.sapiens/H.sapiens* and pUC19/pUC19 experiments, numbers are average of 10 replicated train-test splits.)

type	data (train/test)	motif	read strand	number of training samples	number of validation samples	number of testing samples
R9 2D	*E.coli/H.sapiens*	CpG	template	20,000,000	1,000,000	4,804,942
			complement	20,000,000	1,000,000	3,608,191
	E.coli/E.coli	CpG	template	20,000,000	1,000,000	30,520,729
			complement	20,000,000	1,000,000	23,609,188
	H.sapiens/H.sapiens	CpG	template	2,284,511	100,000	2,380,892
			complement	1,646,495	100,000	1,748,208
	pUC19/pUC19	G*A*TC	template	177,908	10,000	281,982
			complement	160,274	10,000	255,810
	pUC19/pUC19	C*C*WGG	template	59,865	3,000	94,263
			complement	53,382	3,000	84,422
R9.4 1D	*H.sapiens* HX1/HX1	CpG	template	20,000,000	1,000,000	1,288,636,329
	H.sapiens HX1/NA12878	CpG	template	20,000,000	1,000,000	902,181,283

Table S3. Time cost and memory usage of DeepSignal (The feature extraction step is performed using 48 CPU processors each time. The training and testing step is performed using one GPU each time. For R9 2D data, we train and test the data from template and complement strand of reads separately.).

type	data (train/test)	motif	read strand	feature extraction (time/peak memory)	training (time/peak memory)	testing (time/peak memory)
R9 2D	*E.coli/H.sapiens*	CpG	template	216.4min/143GB	1714.9min/28.3GB	32.6min/27.7GB
			complement	142.5min/137GB	1769.5min/28.3GB	24.6min/27.7GB
	E.coli/E.coli	CpG	template	185.8min/143GB	1714.9min/28.3GB	254.4min/29.8GB
			complement	131.3min/137GB	1769.5min/28.3GB	232.2min/29.6GB
	H.sapiens/H.sapiens	CpG	template	30.6min/141GB	444.9min/28.4GB	16.3min/27.7GB
			complement	11.2min/125GB	345.4min/28.3GB	12.3min/27.7GB
	pUC19/pUC19	G*A*TC	template	3.3min/ 95.9GB	282.1min/ 28.3GB	2.8min/ 28.5GB
			complement	3.1min/95.7GB	244.0min/28.3GB	2.8min/ 28.5GB
	pUC19/pUC19	C*C*WGG	template	2.9min/97.7GB	90.0min/28.4GB	1.8min/27.7GB
			complement	2.7min/92.3GB	80.9min/28.4GB	1.6min/27.7GB
R9.4 1D	*H.sapiens* HX1/HX1	CpG	template	2374.8min/215.6GB	1769.5min/28.3GB	10533.4min/29.8GB
	H.sapiens HX1/NA12878	CpG	template	1360.6min/215.6GB	-	7366.1min/29.8GB