Update:

Five-fold cross-validation combined with logistic regression were used to show the prediction performance for low-pass WGBS data. The detailed procedure is that DNA methylation data were divided into 5 equal parts and each of them as the test dataset and the remaining as the training dataset. In the training stage, prediction model was fitted with feature selection of AIC criteria with forward and backward selection in R. The detailed procedure is that we first starts with the full model and eliminates one predictor at a time, at each step considering whether AIC shows significant decrease by adding back in the variable removed at the previous step. Finally, we make the prediction with the prediction model built in training stage to test dataset and summarize the prediction sensitivity, specificity and accuracy.

Spe.train Sen.train Accu.train Spe.test Sen.test Acc.test

0.9869715 0.6503284 0.9418240 0.9801314 0.7126437 0.9141066

Bayesian functions for generalized linear modeling with independent normal, t, or Cauchy prior distribution for the coefficients.

The program bayesglm is the alteration of conventional logistic regression that uses an approximate EM algorithm to update the betas at each step using an augmented regression to represent the prior information.