CAME: Identification of Chromatin Accessibility from Nucleosome Occupancy and Methylome Sequencing

Yongjun Piao[3,5], Seong Keon Lee[6], Eun-Joon Lee[3], Keith D. Robertson[4], Huidong Shi[2,3], Keun Ho Ryu[5,*] and Jeong-Hyeon Choi[1,2,*]

Supplementary Algorithm 1. Seed detection

```
Input: $\delta, d$
for $i \leftarrow 1$ to $N$ d
    if ($\beta_i \le \delta$) then
        $startPosition \leftarrow GCH_i$
        while ($\beta_i > \delta$ or $distance(GCH_i, GCH_{i+1}) > d$)
            $i \leftarrow i + 1$
        endwhile
        $endPostion \leftarrow GCH_i$
        $average \leftarrow calculate(startPosition, endPosition)$
        $insert(startPosition, endposition, average)\ to\ S'$
    endif
endfor
return $S'$
```

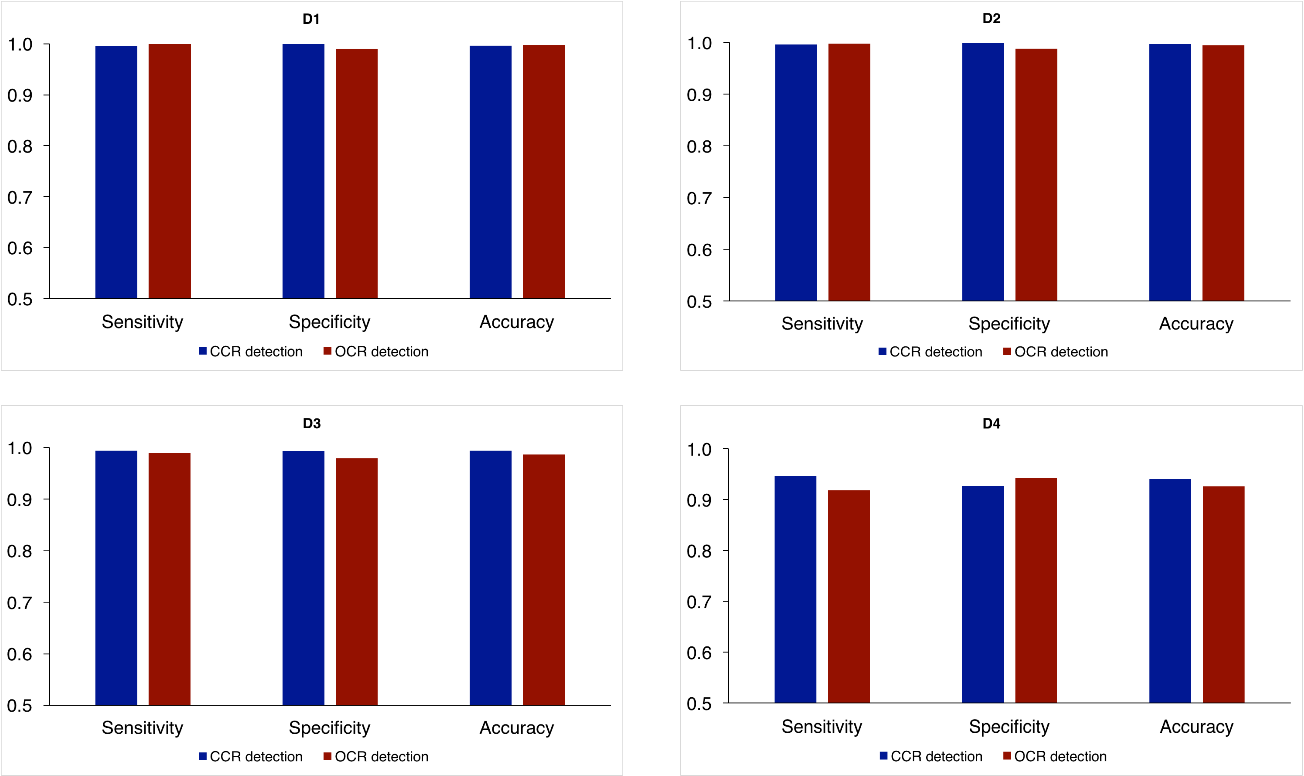
Supplementary Algorithm 2. Seed extension

```
Input: $S', \mu, \varepsilon, d, \Delta$
while ($\forall S$ is considered, $S \in S'$)
    $S_i \leftarrow selectStartingPoint(S')$
    while (! $stopCondition$)
        $(peak, valley) \leftarrow searchNextPeakValley(S_i)$
        $tempAverage1 \leftarrow calculate(S_i, valley)$
        $tempAverage2 \leftarrow calculate(S_{i+1}, valley)$
        if ($\beta_{peak} < \Delta$ and $tempAverage1 \le \mu$ and $tempAverage2 \le \varepsilon$)
            $S_i \leftarrow extendSeed(S_i, valley)$
        else
            $(ave, std) \leftarrow npMixture()$
            for $j \leftarrow S_j$ to $valley$ do
                if ($\beta_{j+1} > ave + std$) then stop
            endfor
            $S_i \leftarrow extendSeed(S_i, j)$
            $stopCondition \leftarrow true$
        endif
    endwhile
    $insert(S_i)\ to\ R'$
endwhile
return $R'$
```

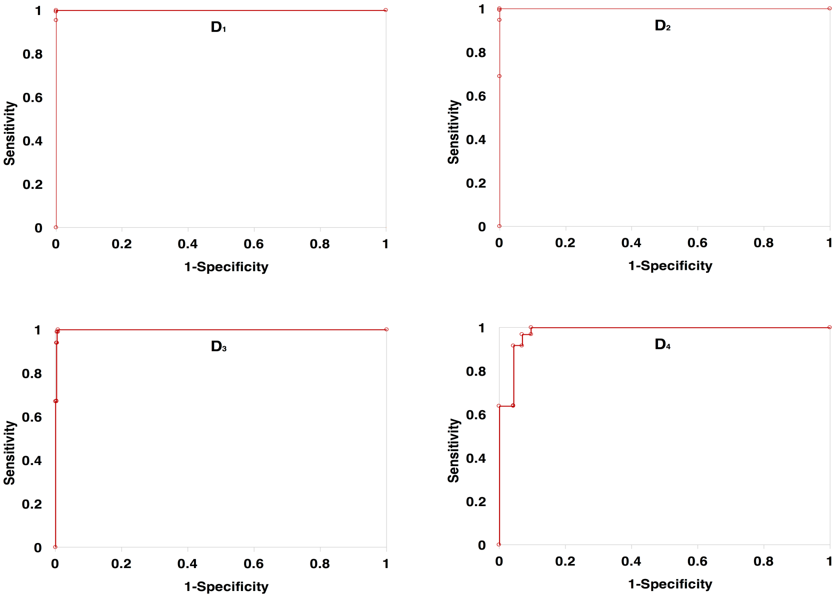
Supplementary Table. 1. Performance change of CAME by CCR and OCR detection methods. For the OCR detection method, parameters *d*=150, *Δ*=0.6, and (δ, ε) = (0.9, 0.6), (0.8, 0.5), (0.7, 0.5), and (0.6, 0.5) were used.

	CCR detection			OCR detection		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
D1	0.9954	0.9999	0.9968	0.9997	0.9907	0.9970
D2	0.9964	0.9997	0.9974	0.9976	0.9884	0.9948
D3	0.9941	0.9928	0.9937	0.9901	0.9788	0.9866
D4	0.9470	0.9271	0.9408	0.9188	0.9425	0.9261

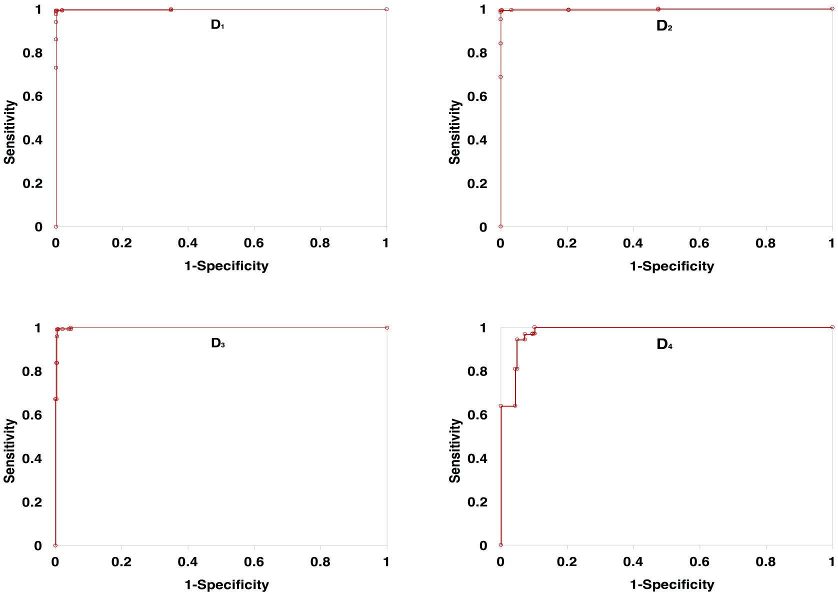
Supplementary Fig. 1. Bar charts of performance change of CAME by CCR and OCR detection methods



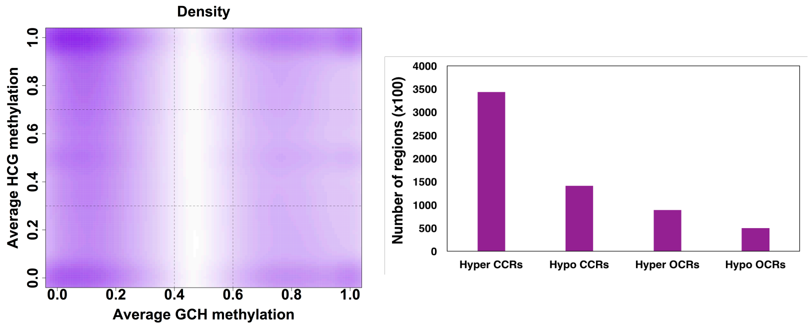
Supplementary Fig. 2. ROC curve with 10 threshold values for the average methylation score δ: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0.



Supplementary Fig. 3. ROC curve with 10 threshold values for the jump score Δ: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0.



Supplementary Fig. 4. The left plot shows a scatter plot between chromatin accessibility and DNA methylation where each x-pixel indicates average GCH methylation score and each y-pixel represents average HCG methylation score of a CCR (OCR), and each point has different degree of the color (light to dark) based on its density. The top and bottom left (right) quadrants represent hyper- and hypo-methylated CCRs (OCRs), respectively. The right plot shows histograms for the number of detected regions for each of the categories (hypo- and hyper-methylated CCRs and OCRs). CCRs (OCRs) with HCG methylation score that are > 0.7 were considered as hyper-methylated and those with HCG methylation score that are < 0.3 were defined as hypo-methylated. The histogram shows the number of detected regions for each of the categories.



Supplementary Fig. 5. DAVID analysis for hyper-methylated CCRs (A) and OCRs (B), and hypo-methylated CCRs (C) and OCRs (D).