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

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

Supplementary Algorithm 1. Seed detection

Input: $\delta, d$

1. for $i \leftarrow 1$ to $N$
2. if ($\beta_i \leq \delta$) then
3.  startPosition $\leftarrow GCH_i$
4.  while ($\beta_i > \delta$ or distance($GCH_i, GCH_{i+1}$) $> d$)
5.    $i \leftarrow i + 1$
6.  endwhile
7.  endPosition $\leftarrow GCH_i$
8.  average $\leftarrow$ calculate(startPosition, endPosition)
9.  insert(startPosition, endPosition, average) to $S'$
10. endif
11. endfor
12. return $S'$

Supplementary Algorithm 2. Seed extension

Input: $S', \mu, \epsilon, d, \Delta$

1. while ($\forall S$ is considered, $S \in S'$)
2.  $S_i \leftarrow$ selectStartingPoint($S'$)
3.  while (!stopCondition)
4.    (peak, valley) $\leftarrow$ searchNextPeakValley($S_i$)
5.    tempAverage1 $\leftarrow$ calculate($S_i$, valley)
6.    tempAverage2 $\leftarrow$ calculate($S_{i+1}$, valley)
7.    if ($\beta_{\text{peak}} < \Delta$ and tempAverage1 $\leq \mu$ and tempAverage2 $\leq \epsilon$)
8.      $S_i \leftarrow$ extendSeed($S_i$, valley)
9.    else
10.       (ave, std) $\leftarrow$ npMixture()
11.       for $j \leftarrow S_i$ to valley do
12.         if ($\beta_{j+i} > \text{ave} + \text{std}$) then stop
13.       endfor
14.       $S_i \leftarrow$ extendSeed($S_i$, $j$)
15.       stopCondition $\leftarrow$ true
16.    endif
17.  endwhile
18.  insert($S_j$) to $R'$
19. endwhile
20. return $R'$
**Supplementary Table. 1.** Performance change of CAME by CCR and OCR detection methods. For the OCR detection method, parameters $d=150$, $\Delta=0.6$, and $(\delta, \varepsilon) = (0.9, 0.6), (0.8, 0.5), (0.7, 0.5), \text{and} (0.6, 0.5)$ were used.

<table>
<thead>
<tr>
<th></th>
<th>CCR detection</th>
<th>OCR detection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>D1</td>
<td>0.9954</td>
<td>0.9999</td>
</tr>
<tr>
<td>D2</td>
<td>0.9964</td>
<td>0.9997</td>
</tr>
<tr>
<td>D3</td>
<td>0.9941</td>
<td>0.9228</td>
</tr>
<tr>
<td>D4</td>
<td>0.9470</td>
<td>0.9271</td>
</tr>
</tbody>
</table>

**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 $\delta$: 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 $\Delta$: 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).