Stage 1: scan the genome to construct candidate regions (CRs)

For each CpG j, compute variance across cells.

$$\sigma_j^2 = \frac{\sum(\widehat{x_{ij} - \bar{x}_j})^2}{n_j - 1}$$

To alleviate noisiness: relative methylation levels are applied with kernel smoothing (borrow strength from nearby sites).

Identify contiguous CpG sites with high variance as candidate regions (CRs).

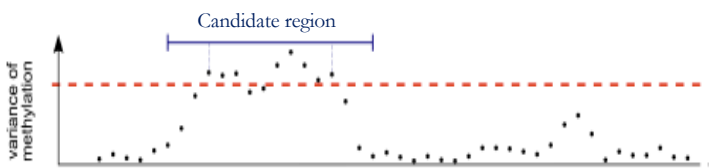
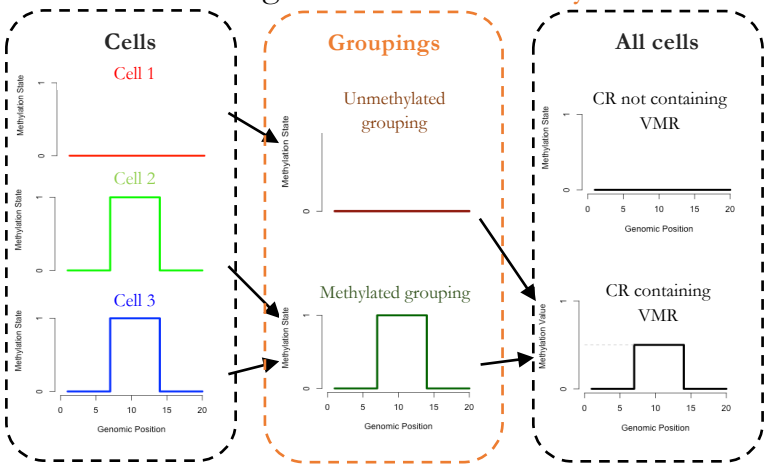


Image credit: Kremer et al., 2022

Stage 2: detect variably methylated regions (VMRs) within CRs

Hypothesize each CR containing at most one VMR and take advantage of the data's binary nature.



Decode hidden states in hidden Markov model and determine number of groupings.

$$p(m|n, N_{\text{Grouping}} = 2) \; ?> \; p(m|n, N_{\text{Grouping}} = 1)$$