

Figure S2. Pseudo code for the semi-supervised CoSBI algorithm

Three dimensions of the input GCP matrix are Genomic loci, Chromatin modifications, and ChIP-seq signal Positions. Output is the set of biclusters.

Input: 3D GCP data, prior co-occurrence probability matrix P
 $\alpha, \beta, \min_g, \min_s$

Output: set of coherent biclusters $\{Bi_{ij}\}$

Begin

for each genomic locus{

generate histone mark similarity C ;

$D_{ij} = 1 - C_{ij}$; //convert similarity matrix to distance matrix

$D'_{ij} = D_{ij} \times (1 - P_{ij})$; //incorporate prior knowledge to the biclustering process

$D' = \text{all-pair-shortest-path}(D')$; // update the distance matrix

$C'_{ij} = 1 - D'_{ij}$; // convert distance matrix back to similarity matrix

$\text{maximal_sample_sets} = \text{depth-first-search}(C', \beta)$; //identify all maximal sample sets

}

$\{Bi_{ij}\} = \text{tree_enumeration}(\{\text{maximal_sample_sets}\}, \alpha, \min_g, \min_s)$

$\{Bi_{ij}\} = \text{merging}\{\{Bi_{ij}\}\}$;

End