

Clustering Sequence Sets for Motif Discovery (T34)

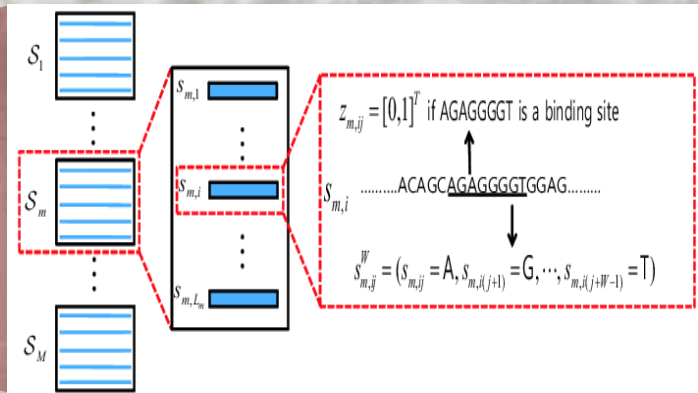
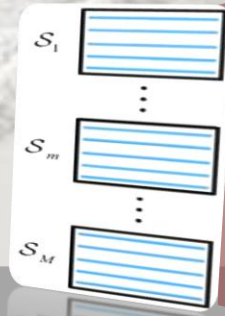
Jong Kyoung Kim, Seungjin Choi (POSTECH)

Problem Definition

Discovering sequence motifs that are enriched in a given set of DNA sequences

Key Contributions

Simultaneously cluster sequence sets and find motifs that relate the clusters



Probabilistic modeling

Single set

- single motif
- over-representation
- sensitive to noise
- applicable to limited problems

Multiple sets

- multiple motifs
- over-representation + clustering
- robust to noise
- applicable to diverse problems

