Discovering Protein-DNA Interactions
Mixture Models and EM Solutions
6.874 Lecture 9

David Gifford
March 3, 2010
mES cell Oct4 ChIP Seq
Synthetic data, EM, no prior
(events at 500 and 550 bp)
Synthetic data, EM, sparse prior
(events at 500 and 550 bp)
EM – no prior

Maximum likelihood (ML)
EM – Sparse prior
The read spatial distribution can be re-estimated using GPS predictions.
More motifs are present in GPS discovered events (High specificity)
Comparable Numbers of CTCF motifs are covered by GPS events (sensitivity)
GPS has superior spatial resolution

GPS: 12.37 bp +- 11.51
SISSRS: 17.00 bp +- 15.11
MACS: 17.21 bp +- 13.71
cisGenome: 29.90 bp +-21.52
GPS can detect joint events

**Graph Description:**
- **Y-axis:** Percentage of joint events recovered (%)
- **X-axis:** Distance between binding events (bp)
- **Legend:**
  - Blue: GPS
  - Green: SISSRS
  - Red: MACS

The graph illustrates the percentage of joint events recovered as a function of the distance between binding events, comparing GPS, SISSRS, and MACS methods. GPS shows near-perfect recovery at all distances, while SISSRS and MACS have diminishing returns, with MACS recovering fewer events at larger distances.