Why There Are So Many Peaks in ChIP-Seq

Renqiang Min
to be joint work with Chao Cheng, Arif Harmanci, and Koon-Kiu Yan

Mar 2, 2011

ChIP-Seq versus ChIP-chip

- Chromatin ImmunoPrecipitation
- Chip signals are noisy due to cross hybridization
  - Many bindings are undetectable due to low signal to noise ratio under stringent statistical test criteria
- Sequencing is very sensitive. Based on robust and reliable peak calling methods, we believe that:

  - If there's a significant peak (different from background peak), there must be a "binding" there: either functional or nonfunctional, either direct or indirect, either stable or non-stable

A quick paper on Figuring Out (In) Direct, Stable, and Functional Bindings

- Utilizing all types of sources of high-throughput data
  - PPI
  - Nucleosome Occupancy
  - Binding Affinity based on PWMs
  - Expression data
- Try our best to explain the bindings we see from ChIP-Seq

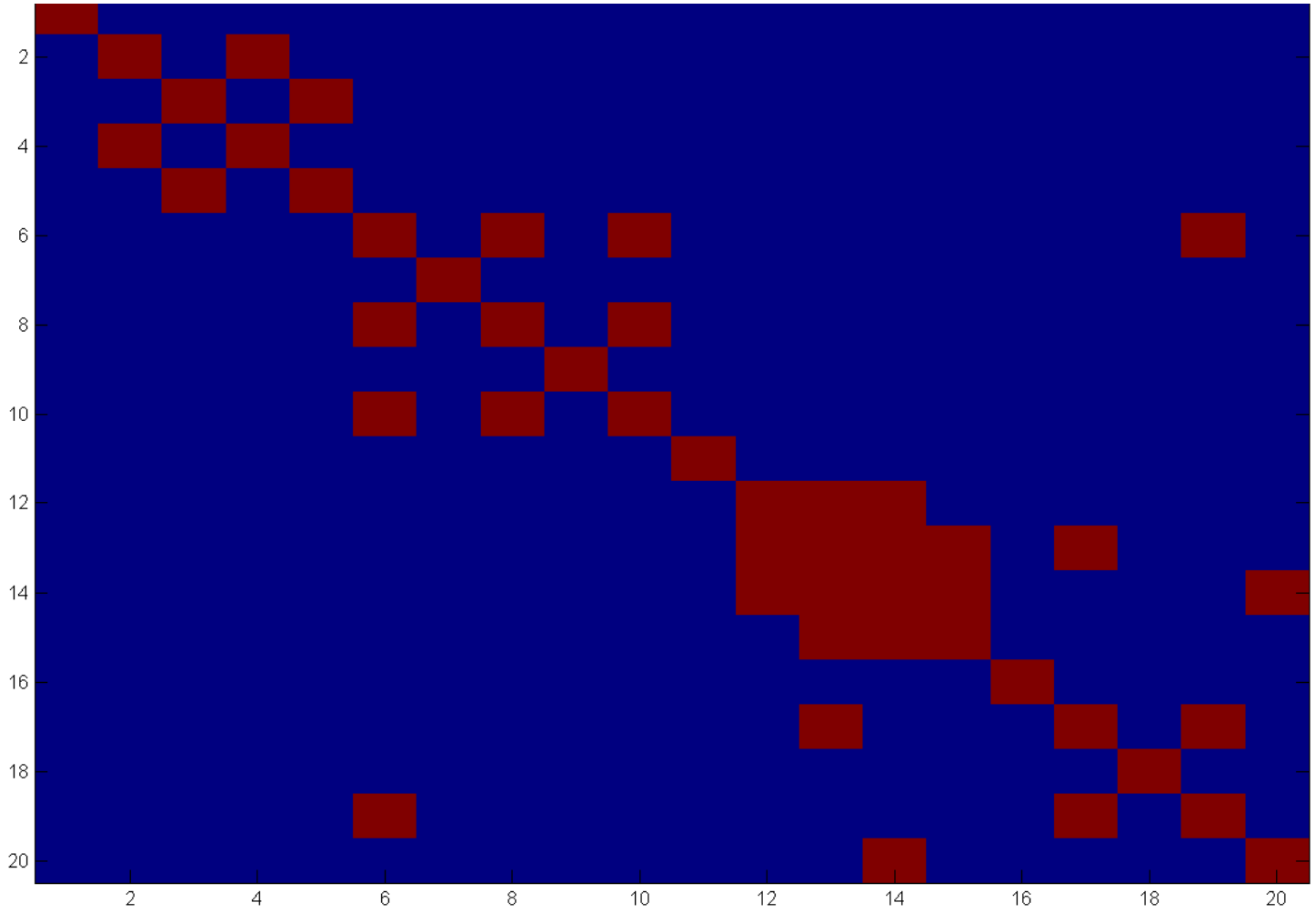
Models

- Assumptions:
  - For a stable direct binding: the binding site is accessible based on nucleosome occupancy data (prior term), and the binding affinity is reasonably high based on PWM calculations (likelihood term)
  - If there's an indirect stable binding, there should be a random-walk path with length smaller than k in the PPI network from the TF under consideration to another TF with direct stable binding
  - If stable bindings are highly confidently functional, genes sharing common binding patterns in their promoter regions should tend to have similar expressions or functions

Models (Continued)

- Choose some confident binding sites and negative sites, parameterize the prior distribution and the likelihood term of positive binding, and learn these paramters (learning could be skipped)
- Rank the Peak sites using the learned model and filter out stable direct binding
- Calculate k-step transition probability matrix between genes with self-allowed transitions
- Identify indirect stable bindings
- Based on gene expression clusters or functional categories, identify common binding patterns (feature selection)

Sample PPI



Sample Binding Propagation (Matrix Multiplication)