User Manual for ChIP-BIT Demo (Matlab)

Introduction

Bayesian Inference of Target genes using ChIP-seq profiles (ChIP-BIT) is developed to reliably detect TFBSs and their target genes by modeling binding signal intensities (of both sample and input ChIP-seq data) and binding locations of TFBSs jointly. Specifically, a unique Gaussian mixture model is used to capture both binding and background signals in sample data. As a unique feature of ChIP-BIT, background signals are modeled by a local Gaussian distribution, which can be accurately estimated from input data. The effect of TFBSs on target gene transcription is also modeled based on their binding locations for target gene identification. An expectation-maximization (EM) algorithm is then used to estimate the posterior probability of binding events for TFBSs.

Citation

Xi Chen, Jin-Gyoung Jung, Ayesha N. Shajahan, Robert Clarke, Ie-Ming Shih, Yue Wang, Luca Magnani, Tian-Li Wang, Jianhua Xuan,”ChIP-BIT: Bayesian inference of target genes using a novel joint probabilistic model of ChIP-seq profiles”.

Requirement

The Matlab package of ChIP-BIT method was tested under Windows7 64bit Matlab2012b and Ubuntu 12.04 64 bit Matlab 2012b and Matlab 2014a.

Usage

To call ChIP-BIT use:

```matlab
[Candidate_window_index, a_1, a_0, pi_1, pi_0, mean_TFBS, variance_TFBS, variance_Background, lambda_TFBS]=ChIP_BIT(gene_window, Candidate_window_index, L)
```

The inputs to the ChIP-BIT function are:

- `gene_window.gene_symbol`: gene symbol
- `gene_window.chrom`: chromosome location of gene, ‘chr1, chr2, ..., chrX’
- `gene_window.index`: index of each window at gene promoter, ‘-50, ..., -1, 1, ..., 50’
- `gene_window.start`: start location of each window
- `gene_window.end`: end location of each window
- `gene_window.chiptag`: read intensity of each window under sample ChIP-Seq data
- `gene_window.inputtag`: read intensity of each window under input ChIP-Seq data
- `Candidate_window_index`: candidate enriched window index
- `L`: number of rounds of EM iteration
The outputs are:

- $a_1$, posterior probability for binding occurrence at each candidate enriched window
- $a_0$, posterior probability for binding non-occurrence at each candidate enriched window
- $\pi_1$, prior probability for binding occurrence
- $\pi_0$, prior probability for binding non-occurrence
- $\text{mean}_{TFBS}$, mean estimation of read intensity distribution at TFBSs
- $\text{variance}_{TFBS}$, variance estimation of read intensity distribution at TFBSs
- $\text{variance}_{Background}$, variance estimation of read intensity distribution at background regions
- $\lambda_{TFBS}$, lambda estimation of distance distribution to TSS

**Demo**

‘Demo_main.m’ shows an example of ChIP-BIT on simulation data (Case 1 as described in the cited paper).

Fig. 1. Results of ChIP-BIT by using simulation Case 1.