

Chapter: Introduction

- Problem motivation: Discuss the problem and why it is important.
- Contributions of this study: What I did in a nutshell.
- Document structure: Structure of the thesis document.

Chapter: Background

- Eukaryotic regulation: Define basic terms of eukaryotic regulation, such as transcription factors
- Epigenetics: Talk about chromatin states. Open/closed chromatin and histone modifications.
- Active binding site detection: Define in more detail the main problem we are addressing. Define how it was solved before next-gen sequencing (motif matching) and the problems with that.
- Next-generation sequencing methods: Introduce DNase-seq in detail and ChIP-seq. Also formally (mathematically) describe the genomic signals. As a segway from the previous section, discuss that these methods can identify active TFBSs (TFBS grammar).
- Literature review: Discuss the state-of-the art methods that used next-gen methods in a chronological order, like telling a story (from Hon (only histones) to us).

Chapter: Method

- Signal normalization: Describe basic signal pre-processing and the normalizations we performed.
- Experimental bias correction: Explain how experimental/cleavage bias correction was performed.
- Hidden Markov models: Define the hidden markov models and how it was used to address our problem. Describe the training procedure.
- Signal processing filters: Describe the signal processing filter used and how it was used to address our problem.

Chapter: Experiments

- Data Description: Describe all data used in this study.
- ChIP-seq validation: Describe motif matching. ChIP-seq peak calling. And how they are used to assess the accuracy of a generic computational footprinting method.
- FLR-Exp validation: Describe how the expression of cells were obtained and the whole FLR-Exp evaluation procedure. In the beginning of the section describe the datasets used.
- Competing methods: Which parameters we used for all competing methods. Also include the table of resources needed by each competing method. The baseline methods would be introduced in this section.

Chapter: Results

Parameterization results

- Signal parameterization: Empirical parameterization of signal parameters such as percentile threshold, global vs local normalization, etc.

- HMM parameterization: Empirical parameterization of HMM model parameters such as topology (inclusion of UP and DOWN were better than only TOP), etc.
- Filters parameterization: Empirical parameterization of filtering parameters. Each filter type has many open parameters. We performed a grid search and this section should describe that.
- Validation parameterization: Empirical parameterization of validation parameters such as the threshold used for the motif matching, etc.

Results from 1st study

- DNase + Histones improves TFBS detection: Describe the results of the 1st study. Including:
 - DNase+histones provided the best results and which histones performed best.
 - Accuracy of best HMM method vs competing methods (comparative study) under the ChIP-seq based evaluation.
- Statistics on footprints and DHSs: In the 1st study we have interesting statistics to show.
- HMM training is cell-independent: Using data from 1st study.
- TF-oriented analysis of AUC results: In the 1st paper we performed two additional analyses:
 - correlations between our method's performance and features that describe TF binding affinity.
 - we tested our method's AUC between different TF classes (we used TFClass).

Results from 2nd study

- Footprint ranking strategy: We show that TC is the best strategy to rank footprints from all methods.
- Impact of DNase-seq experimental bias: We show all results describing the impact of DNase-seq experimental bias. This would include the clustering, the He et al correlation graphs, the line graphs with the motif logo and the experimental results from HINTBC - HINT, etc.
- Impact of CG content: We show the results of the impact of CG content on accuracy.
- TF residence time: Analysis on TF residence time and protection score.
- Comparative study with ChIP-seq based evaluation: All results from ChIP-seq evaluation methodology.
- Comparative study with FLR-Exp: All results from FLR-Exp evaluation methodology.

Case Studies

- Case study: Regulatory network of differentiation of dendritic cells
- Case study: Multimodal role of NF-kB during the intermediate-early inflammatory response

Chapter: Concluding remarks

- Discussion: What can be done with our method. Discussion of further capabilities and issues.
- Future work: Applicability of our method to other data (ATAC-seq, ChIP-exo). Maybe talk about de novo motif finding with footprints. Also talk about differential footprinting.
- Conclusion: Concluding remarks.