GimmeMotifs:

This manuscript (permalink) was automatically generated from simonvh/gimmemotifs-manuscript@b8efae0 on August 1, 2018.

Authors

- **Simon J. van Heeringen**
  0000-0002-0411-3219 · simonvh · svheeringen
  Radboud University, Faculty of Science, Department of Molecular Developmental Biology, Radboud Institute for Molecular Life Sciences, 6500 HB Nijmegen, The Netherlands · Funded by Grant XXXXXXXX

Abstract

Introduction

The regulatory networks that determine cell and tissue identity are robust, yet remarkably flexible. Transcription factors (TFs) control the expression of genes by binding to their cognate DNA sequences, TF motifs, in cis-regulatory elements. To understand how genetic variation affects binding and to elucidate the role of TFs in regulatory networks we need to be able to accurately model binding of TFs to the DNA sequence.

The most widely adopted representation of TF binding is the position frequency matrix (PFM). This matrix, a TF motif, contains (normalized) frequencies of each nucleotide at each position in a collection of aligned binding sites. These PFMs can be derived from high-throughput experiments such as ChIP-sequencing, HT-SELEX or Protein Binding Microarrays (PBMs).

Even though the PFM is a convenient representation, it has certain limitations. A PFM cannot model inter-nucleotide dependencies, that are known to affect binding of certain TFs. Multiple different representations have been proposed [1,2,3,4,5], however, no single one of these has gained much traction.

Here, we present GimmeMotifs, a Python module and set of command-line tools for TF motif analysis. Amongst other possibilities it can be used to perfom *de novo* motif analysis, calculate enrichment statistics and identify differential motifs. We illustrate the functionality of GimmeMotifs using three different examples.

References

1. **FROM BINDING MOTIFS IN CHIP-SEQ DATA TO IMPROVED MODELS OF TRANSCRIPTION FACTOR BINDING SITES**
IVAN KULAKOVSKIY, VICTOR LEVITSKY, DMITRY OSHCHEPKOV, LEONID BRYZGALOV, ILYA VORONTSOV, VSEVOLOD MAKEEV
Journal of Bioinformatics and Computational Biology (2013-02) https://doi.org/10.1142/s0219720013400040

2. **The Next Generation of Transcription Factor Binding Site Prediction**
Anthony Mathelier, Wyeth W. Wasserman
PLoS Computational Biology (2013-09-05) https://doi.org/10.1371/journal.pcbi.1003214

3. **Varying levels of complexity in transcription factor binding motifs**
Jens Keilwagen, Jan Grau
Nucleic Acids Research (2015-06-26) https://doi.org/10.1093/nar/gkv577

4. **InMoDe: tools for learning and visualizing intra-motif dependencies of DNA binding sites**
Ralf Eggeling, Ivo Grosse, Jan Grau
Bioinformatics (2016-12-28) https://doi.org/10.1093/bioinformatics/btw689

5. **Automated incorporation of pairwise dependency in transcription factor binding site prediction using dinucleotide weight tensors**
Saeed Omidi, Mihaela Zavolan, Mikhail Pachkov, Jeremie Breda, Severin Berger, Erik van Nimwegen
PLOS Computational Biology (2017-07-28) https://doi.org/10.1371/journal.pcbi.1005176