

MoCA : Tool for Motif Conservation Analysis

Saket Choudhary, Anton Valouev

{skchoudh}@usc.edu

<https://github.com/saketkc/moca>

Introduction

- Motifs are Short DNA sequences that appear recurrently and often act as sequence specific binding sites for transcription factors
- Determining the quality of a reported ChIP-Seq motif is hard
- Motif analysis tools such as MEME[1] can often report ‘false motifs’ and still have significant p-value(or E-values)
- ‘Distance from center’ approach fails to identify new co-transcription factor motifs

Materials and Methods

Hypothesis

- ‘True’ transcription factor(TF) motifs have significant correlation between motif and conservation
- ‘True’ TF motifs have higher conservation scores at motif bases as compared to flanking bases

MoCA

- Any metric to assess the quality of motifs, should also rely on biological relevance besides the statistical analysis
- Since the motif acts as a specific binding sequence, it can be expected to be conserved evolutionarily
- MoCA makes use of the PhyloP and Gerp scores to assess the conservation profile
- Automated analysis of ENCODE datasets, with a RESTful api

Results

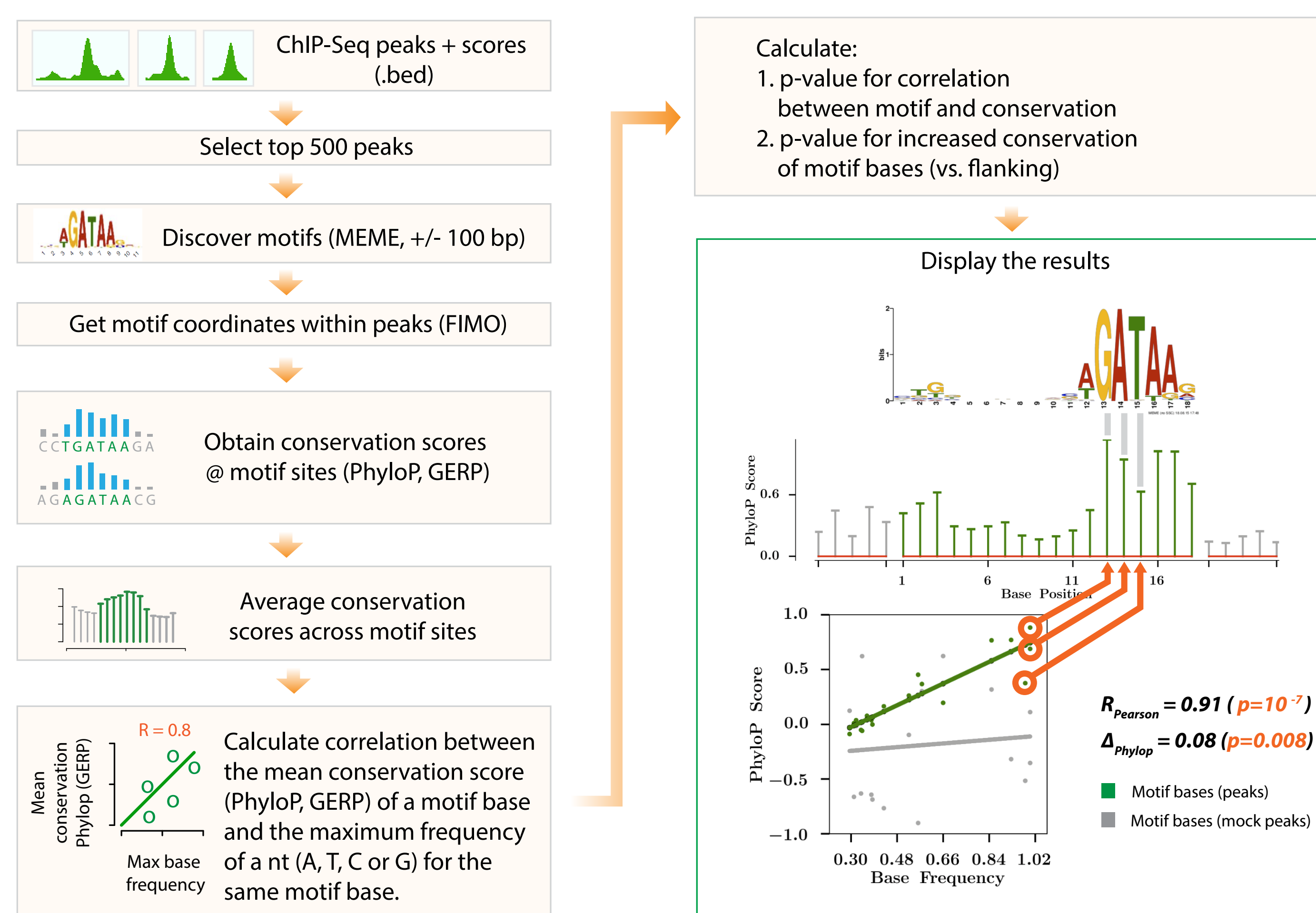


Figure 1: MoCA Workflow

Conclusions

- MoCA is a helpful tool for identifying ‘true’ motifs
- MoCA’s RESTful api allows automated analysis of ENCODE ChIP-seq datasets

References

- [1] Timothy L Bailey, James Johnson, Charles E Grant, and William S Noble. The meme suite. *Nucleic acids research*, page gkv416, 2015.
- [2] Adam Siepel, Katherine S Pollard, and David Haussler. New methods for detecting lineage-specific selection. In *Research in Computational Molecular Biology*, pages 190–205. Springer, 2006.

