


MotifSeeker

A Motif Finding and Enrichment Tool

Justin Hii
Tusha Karnani
Numa Yazadi

A dark blue diagonal gradient bar that starts from the bottom left and extends towards the top right, covering the lower half of the slide.

Overview

- Tool used to identify and validate motifs in genomic sequences
- **Extracts and analyzes sequences** from genomic regions.
- **Computes and utilizes PWMs** to find and validate motifs.
- **Generates background sequences** and calculates statistical significance.
- **Helps identify biologically relevant motifs** and their roles in genomic functions.

Benchmark

MotifSeeker vs HOMER

- User Interface
 - HOMER has more friendly UI
- Computational Efficiency
 - Longer runtime

Benchmark

Runtime

- On public dataset from ENCODE, on embryo brain tissue
- <https://www.encodeproject.org/experiments/ENCSR817LUF/>
- Used the ENCFF803UAK.bed and GRCh38.fa genome.

Results

1. Sequences from BED Files

- A list of sequences that correspond to the regions specified in the BED file

2. PWM Computations

- A PWM that quantifies the likelihood of each nucleotide (A, C, G, T) at each position within the motif

3. Motif Enrichment Analysis

- Uses PWMs to scan the extracted sequences for motif occurrences: Number of Peaks Passing, Number of Background Sequences Passing, P-value

4. Threshold Calculation

- determines the PWM score threshold that corresponds to a desired p-value

Challenges

- Computational Load
 - Large genomic datasets mean long processing time
- Motif Length
 - Different motifs can vary significantly in length and composition
- Quality of Data
 - Accuracy of results heavily depends on the quality and completeness of input data (BED Files)