

Bioinformática 2019/2020

Exercise Aula 9- Deterministic Motif Mining

Student 1: Orlando Melo

Number: 201604476

Student 2: Bruno Pinto

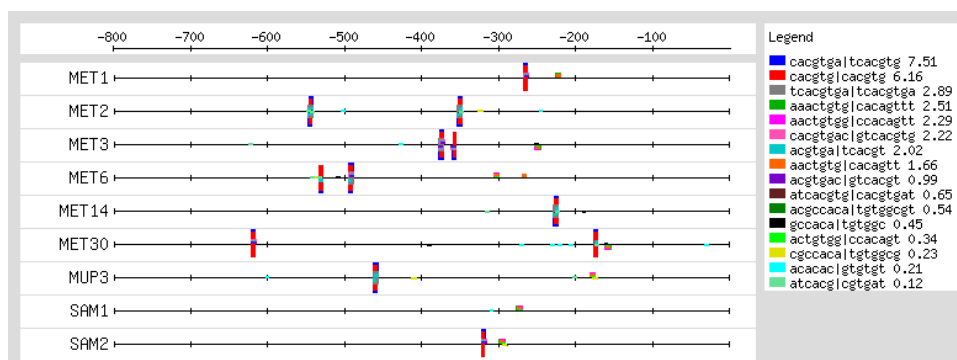
Number: 201603939

Task1

- Esta afirmação está incorreta pois a localização de motivos reguladores é uma tarefa especialmente difícil em eucariontes devido às sequências de consenso reconhecidas por fatores transcripcionais serem geralmente mais curtas que em procariontes, além disso são bastante variáveis e estão dispersas por distâncias grandes.
- É um motif curto que se encontra em toda a parte do genoma. No entanto, um único motif é insuficiente para exercer a função reguladora. A especificidade do GATA vem da sua repetição em várias cópias na região upstream dos genes controlados.
- Dada a flexibilidade dos mecanismos regulatórios, dificilmente se pode conceber um método abrangente que possa detectar sistematicamente todos os tipos de sinais regulatórios.
- Os principais passos do algoritmo são os seguintes:
 - Extração de regiões upstream das sequências
 - Detecção de oligo nucleotídeos sobre-representados
 - Procura de todas as posições equivalentes no conjunto de sequências upstream
 - Desenho de um feature map
- Consiste no agrupamento de padrões altamente significativos com outros menos significativos. Este agrupamento, em alguns casos, reflete uma tendência devido à alta frequência de um local de reconhecimento mais curto.
- Os valores mais altos para esse parâmetro correspondem aos padrões mais representados.
-
- $w=7$, $N_{pal}=0$, $D=8192$ | $w=8$, $N_{pal}=256$, $D=32896$

Task2

1.



2.

Motif | Score | Number of occurrences | Z-score | Frequency | Obs/exp ratio

cacgtga/tcacgtg | 7.51 | 12 | 14.46 | 0.0016925246827 | 19.3634 |

cacgtg/cacgtg | 6.16 | 12 | 11.53 | 0.0016870518768 | 13.0094 |

Tcacgtga/tcacgtga | 2.89 | 6 | 13.76 | 0.0008463817182 | 33.5442 |

3.

Ocorrências correspondem ao número de vezes que um padrão ocorre nas sequências, enquanto a frequência é uma frequência relativa entre o número de ocorrências e o total do padrão.

4.

Ambas estimam o número de ocorrências de um padrão num conjunto de sequências, mas o z-score melhora a acurácia do resultado pois elimina do cálculo os padrões sobrepostos.