

Motifs and Position Weight Matrices

Martin Morgan¹

June 23 – 28, 2013

¹mtmorgan@fhcrc.org

Overview

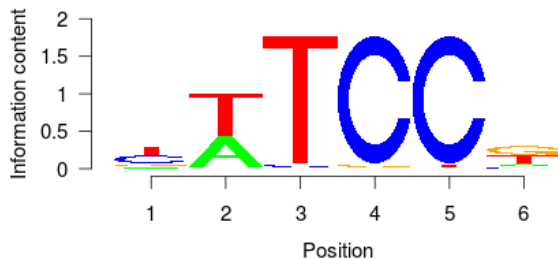
- ▶ What is a Position Weight Matrix?
- ▶ When are PWMs useful?
- ▶ How can we use PWMs in *Bioconductor*?
- ▶ A quick case study – Huang et al., 2013, Highly Recurrent TERT Promoter Mutations in Human Melanoma, Science 339:957

What?

PWM for transcription factor ETS1

	1	2	3	4	5	6
A	0.1	0.425	0.000	0.000	0.000	0.125
C	0.4	0.000	0.025	0.975	0.975	0.075
G	0.1	0.000	0.000	0.025	0.000	0.425
T	0.4	0.575	0.975	0.000	0.025	0.375

seqLogo:



When?

- ▶ Motif discovery, e.g., identify shared sequence under ChIP peaks
- ▶ Annotation: identify parts of genome with high similarity to existing motif

How?

Motif / PWM discovery

- ▶ Naive, e.g., `consensusMatrix` of candidate sequences
- ▶ Fancy

PWM scoring: *Biostrings*

- ▶ `matchPWM`, `scorePWM`
`library(Biostrings)`

Catalogs and visualization

- ▶ *MotifDb* for a catalog of known motifs
- ▶ *seqLogo* for visualization

Case study

Background

- ▶ Huang et al., Science 339:957:

Task

- ▶ Score catalog of known PWMs for match to reference versus variant sequence