

Steps in a ChIP-seq analysis

Jeff Leek

@jtleek

www.jtleek.com

Background on ChIP-seq

DNA

ACTGACCTAGATCAGTGTAGCGATCGTATACGAGACCGATTTCATCGGCAT



transcription

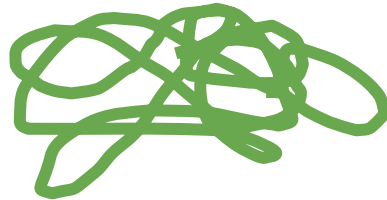
RNA

AUCAGUCGAUCACCGAU



translation

protein



DNA

ACTGACCGTTCGATCGATCGTATACGATTACAAAATCATCGGCAT



transcription

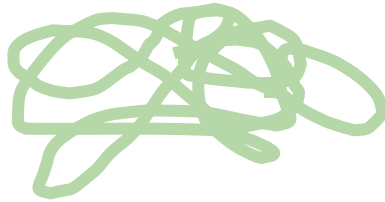
RNA

AUCAGUCGAUCACCGAU



translation

protein



GGAACCATGGGAATTCACGAATTCCTAACCATTA



Cross-link protein to DNA

GGAA

CATTA

CCATGGGAATTCACGAATTCCTAAC



Fragment DNA

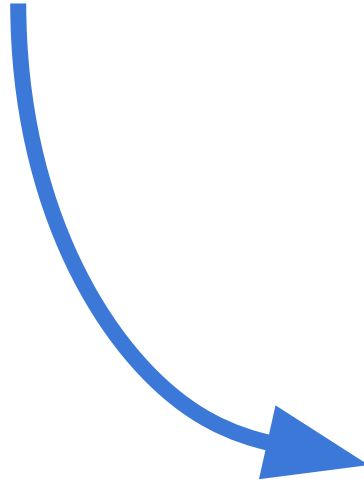
GGAA CATTAG

CCATGGGAATTCACGAATTCCTAAC



Antibody pulldown

CCATGGGAATTCACGAATTCCTAAC



Steps

1. Align
2. Peak detection
3. Counting
4. Normalization
5. Statistical tests
6. Annotation/Motif analysis

Step 1: Align

Software:

- [Bowtie2](#)
- [BWA](#)

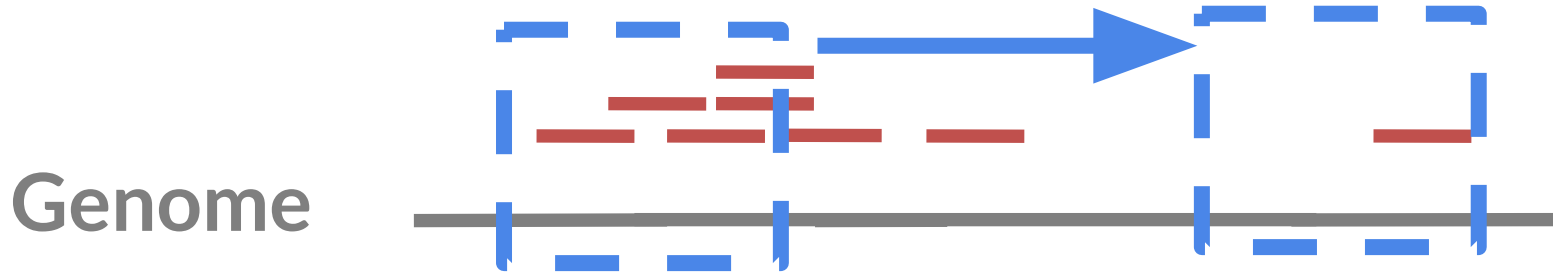
Genome



Step 2: Peak detection

Software:

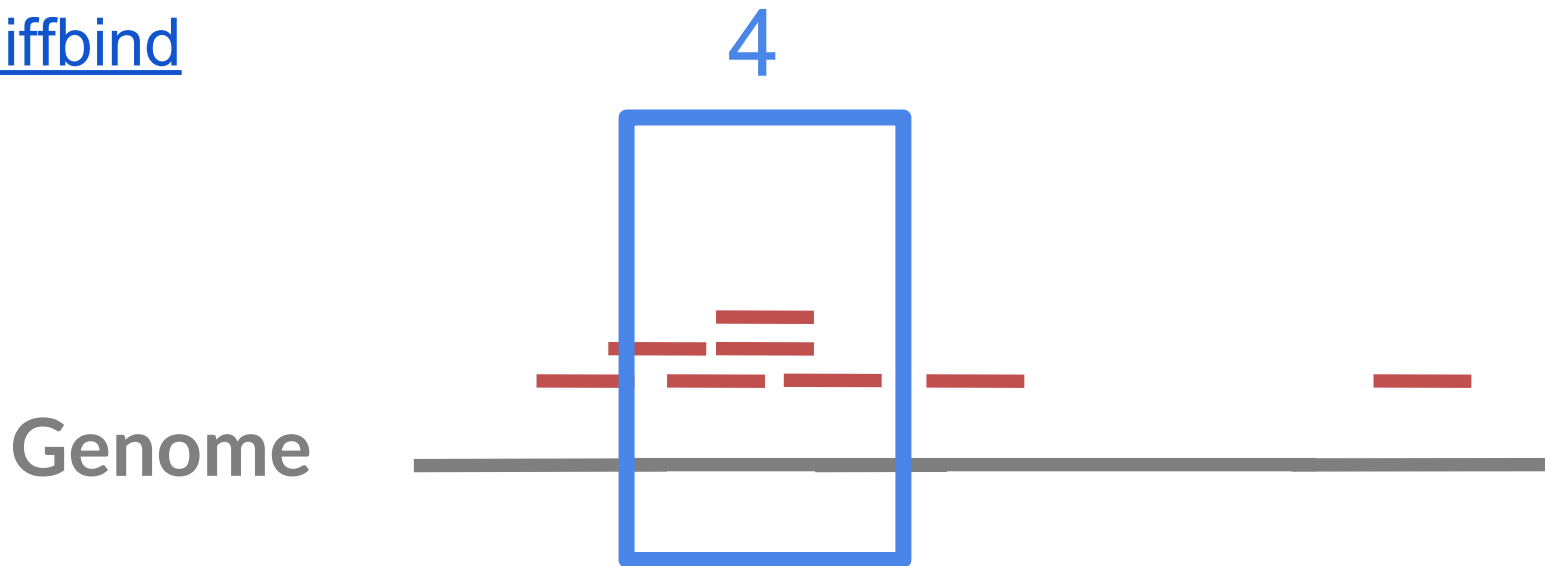
- [CisGenome](#)
- [MACS](#)
- [PICS](#)



Step 3: Counting

Software:

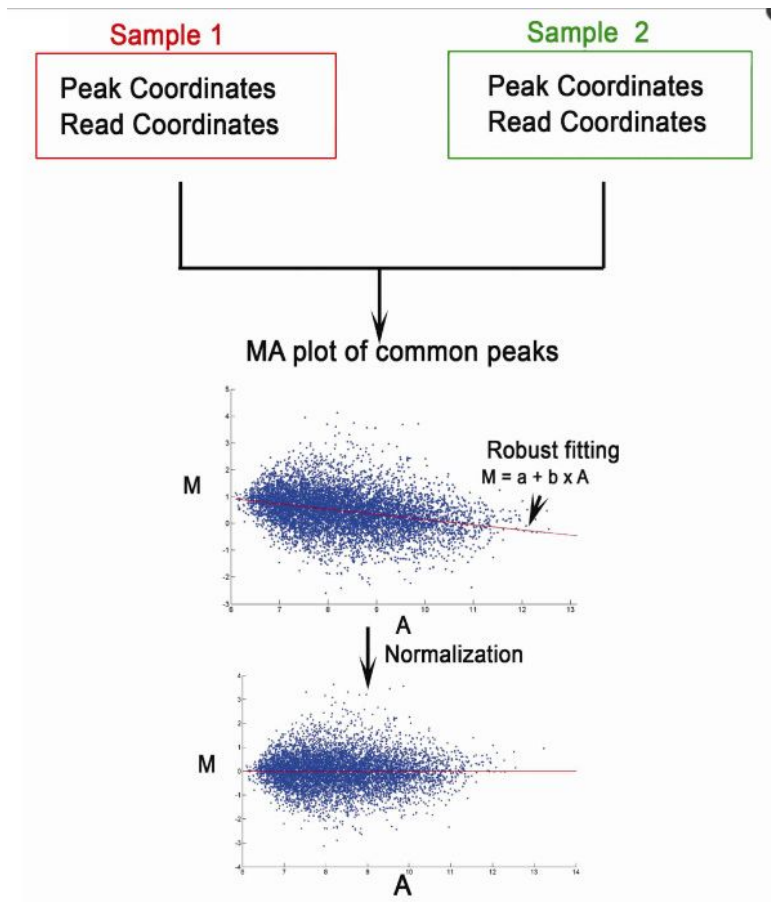
- [CisGenome](#)
- [MACS](#)
- [diffbind](#)



Step 4: Normalization

Software:

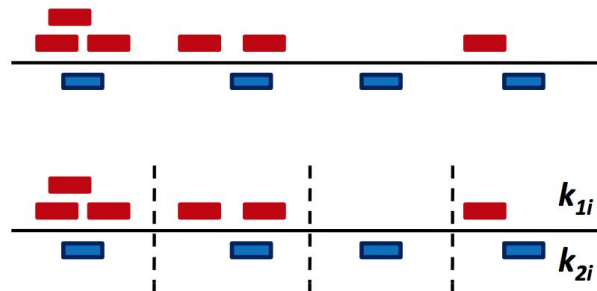
- [diffbind](#)
- [MAnorm](#)



Step 5: Statistical tests

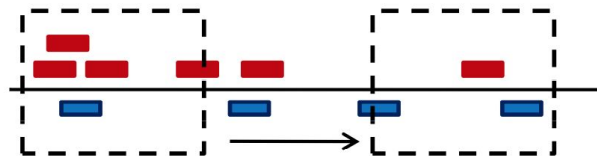
Software:

- [CisGenome](#)
- [MACS](#)
- [diffbind](#)



$$n_i = k_{1i} + k_{2i}$$

$$k_{1i} / n_i \sim \text{Binom}(n_i, p_0)$$



Step 6: Sequence motifs & Annotation

Software:

- [CisGenome](#)
- [meme-suite](#)
- [BioC Annotation Workflow](#)

DNA motif:

GTATGTACTTACTA**TF**TGGGTGGTCAACAAATCTATGTATGA
 TAACATGTGACTCCTATAACCTCTT**TF**TGGGTGGTACATGAA
 C**TF**TGGGAGGTCCCTCGGTTCCAGAGTCACAGAGCAGATAATCA
 TTAGAGGCACAATTGCT**TF**TGGGTGGTGCCACAAAAACAAG
 AACAGCCTTGATTAGCTGCTGGGGGGG**TF**TGAGTGGTCCCAC
 ATCAGAA**TF**TGGGTGGTCCATATATCCCAAAGAAGAGGGTAG

Transcription Factor Binding Sites (TFBS)

123456789

TGGGTGGTC

TGGGTGGTA

TGGGAGGTC

TGGGTGGTG

TGAGTGGTC

TGGGTGGTC

