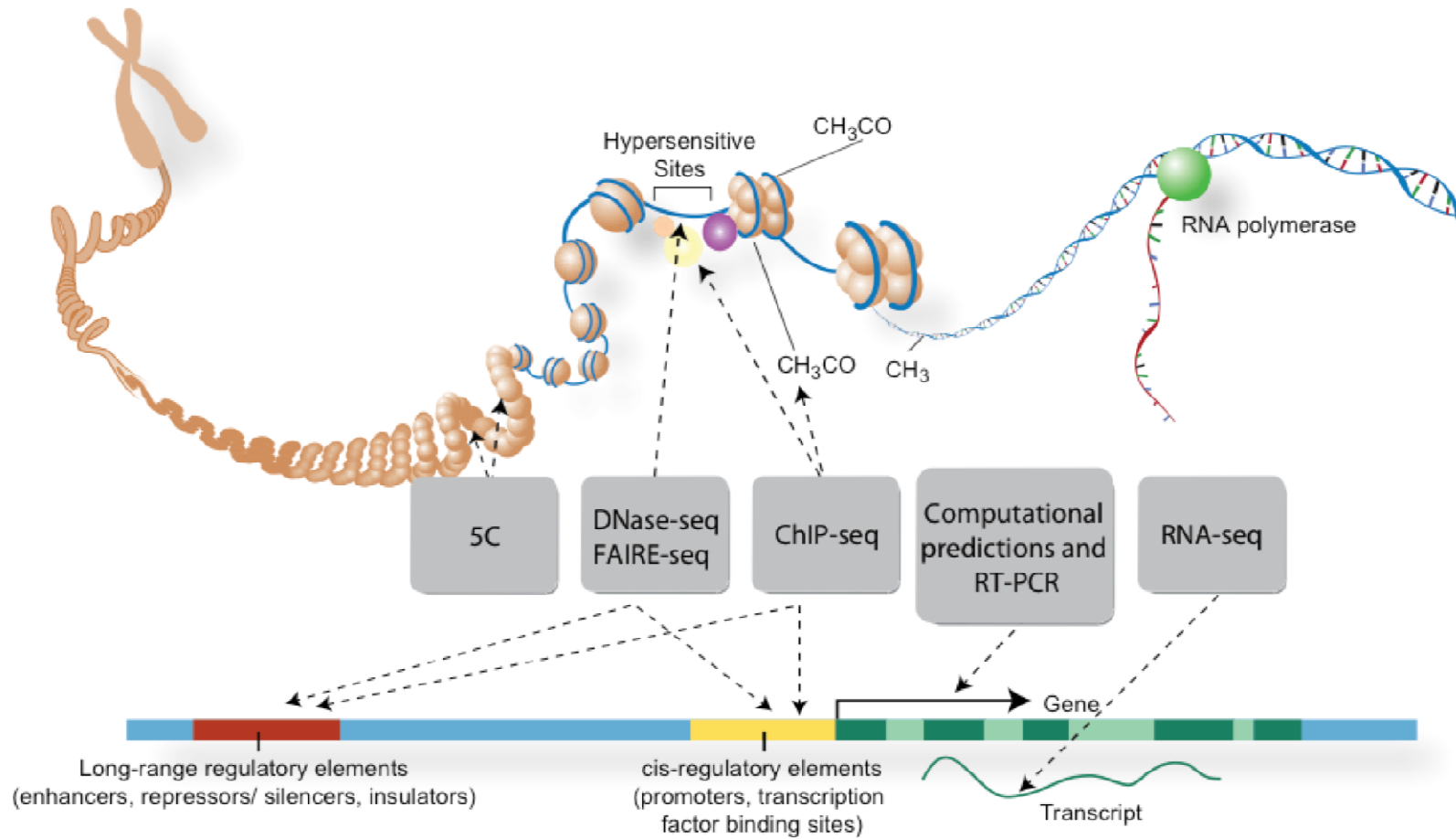


Introduction to ChIP-seq

Yingying Wei
April 16, 2013

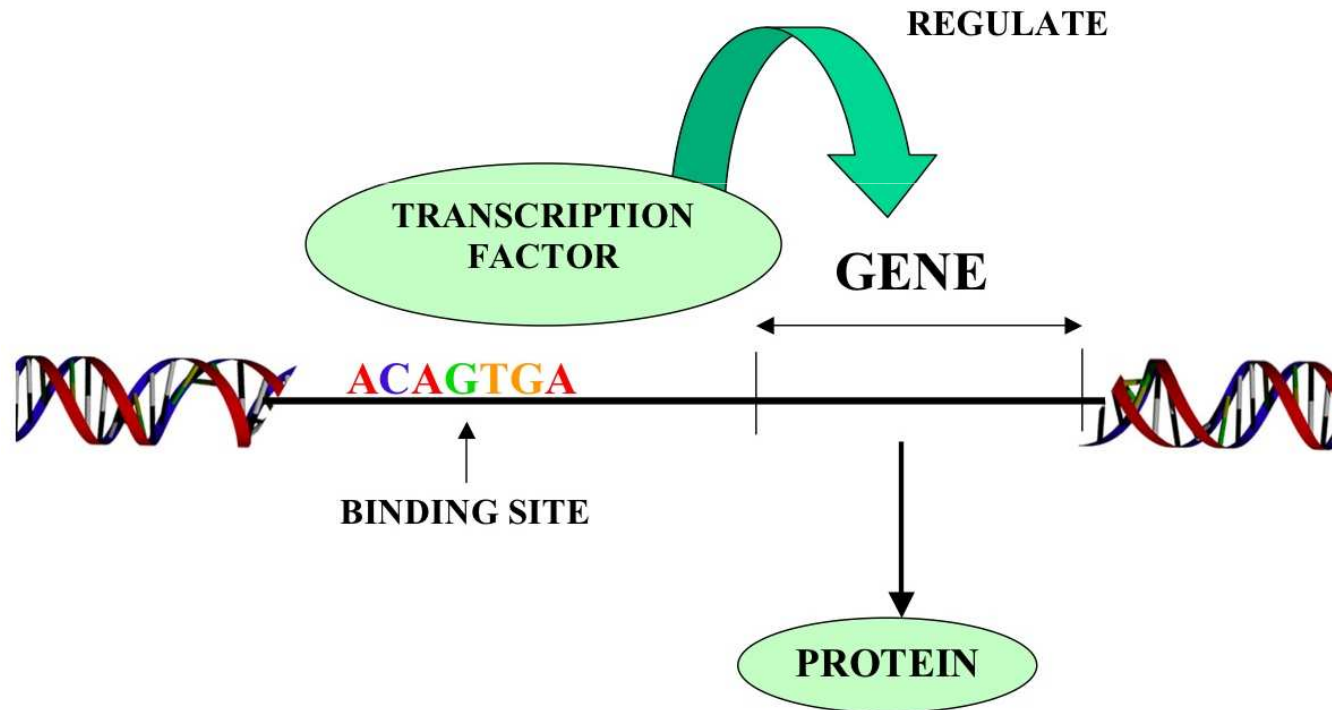
Regulatory element



<http://genome.ucsc.edu/ENCODE/aboutScaleup.html>

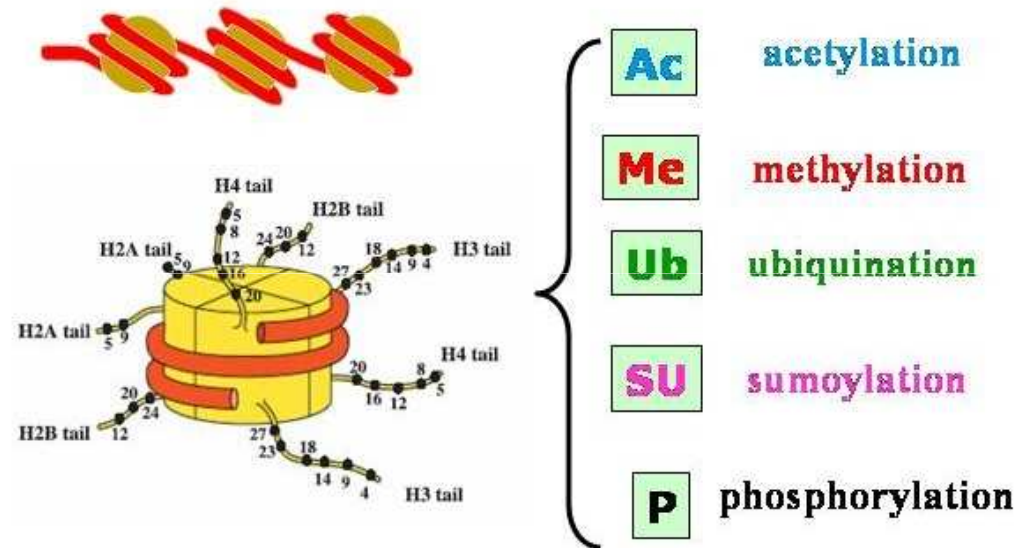
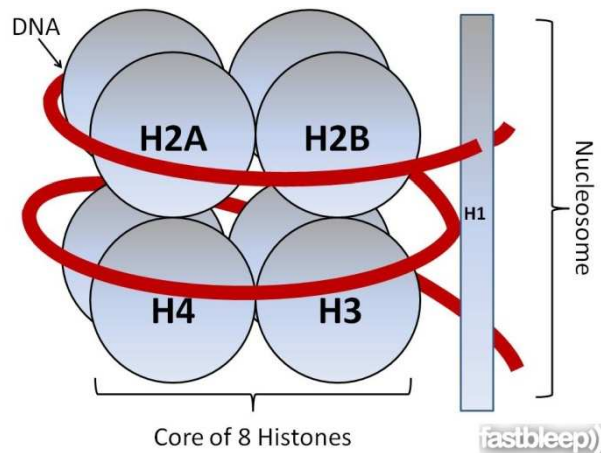
Transcription Factor

Legend: A transcription factor molecule binds to the DNA at its binding site, and thereby regulates the production of a protein from a gene.



http://www.bio.miami.edu/dana/pix/transcription_factor.jpg

Histone modification

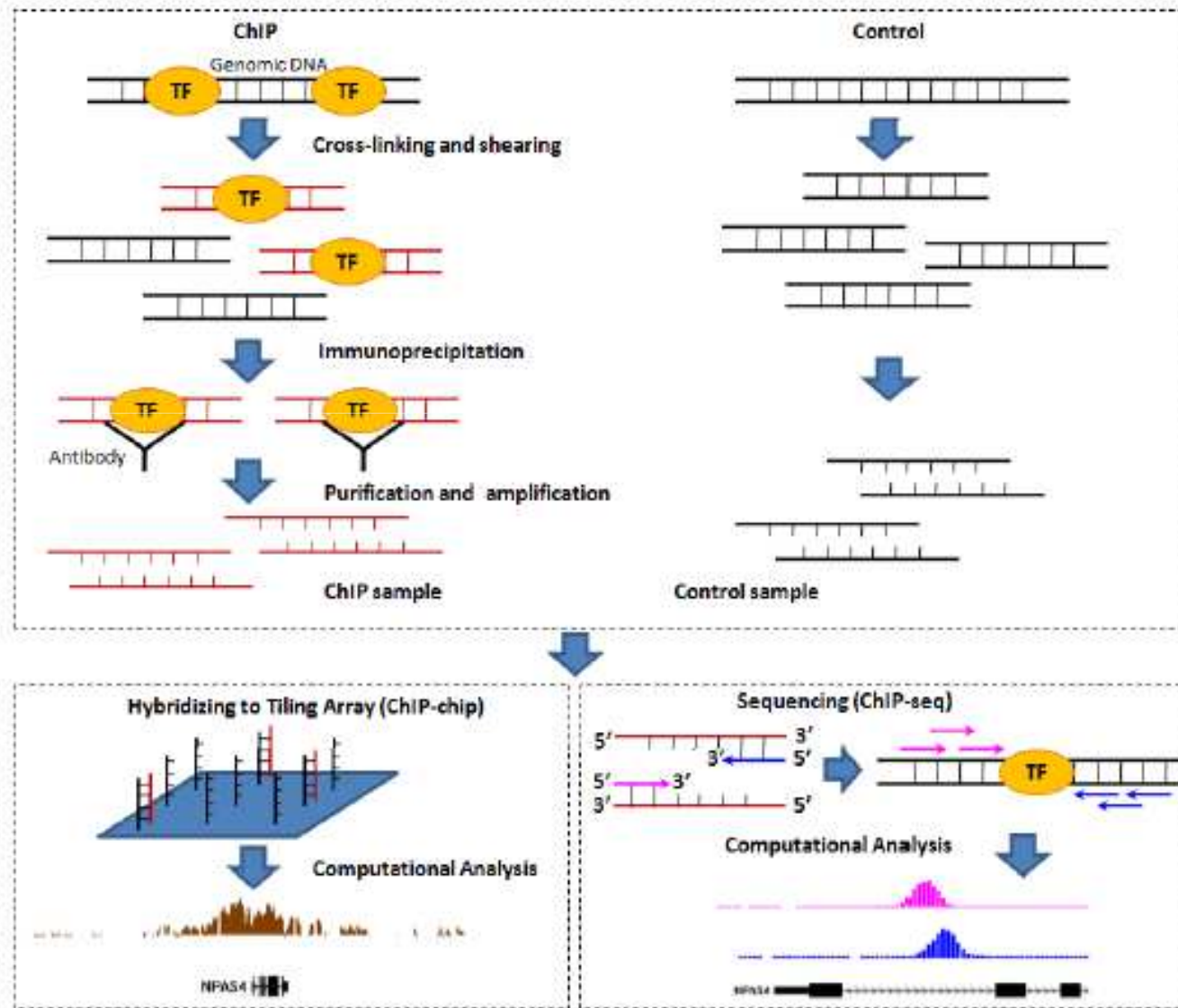


The figure illustrates nucleosome models and major posttranslational modifications which play essential roles in gene expression regulation and disease processes

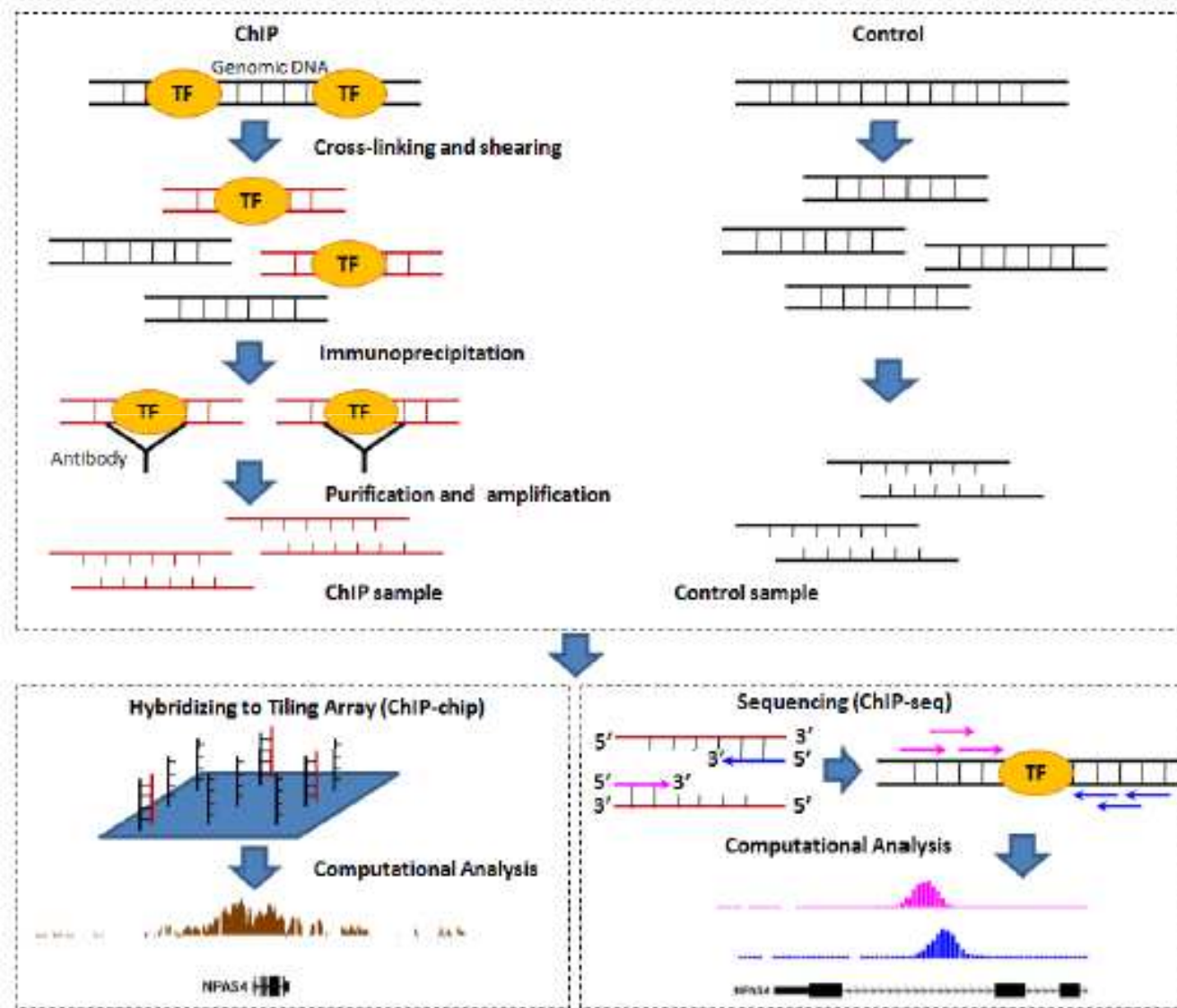
http://www.fastbleep.com/assets/notes/image/10257_1.jpg

<http://www.integratedhealthcare.eu/contentimages/nucleosome1.jpg>

ChIP-seq experiment



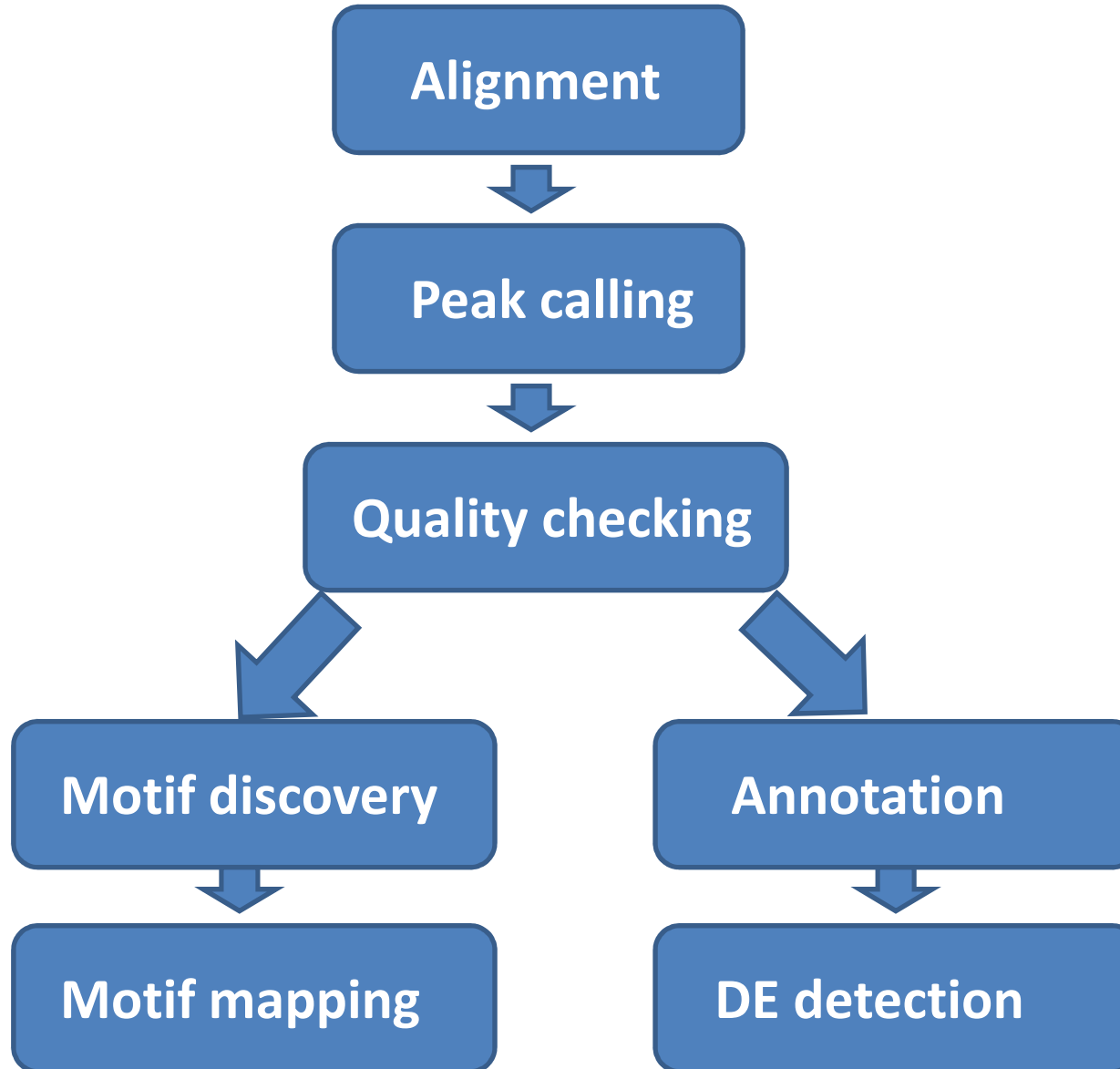
Chromatin ImmunoPrecipitation sequencing



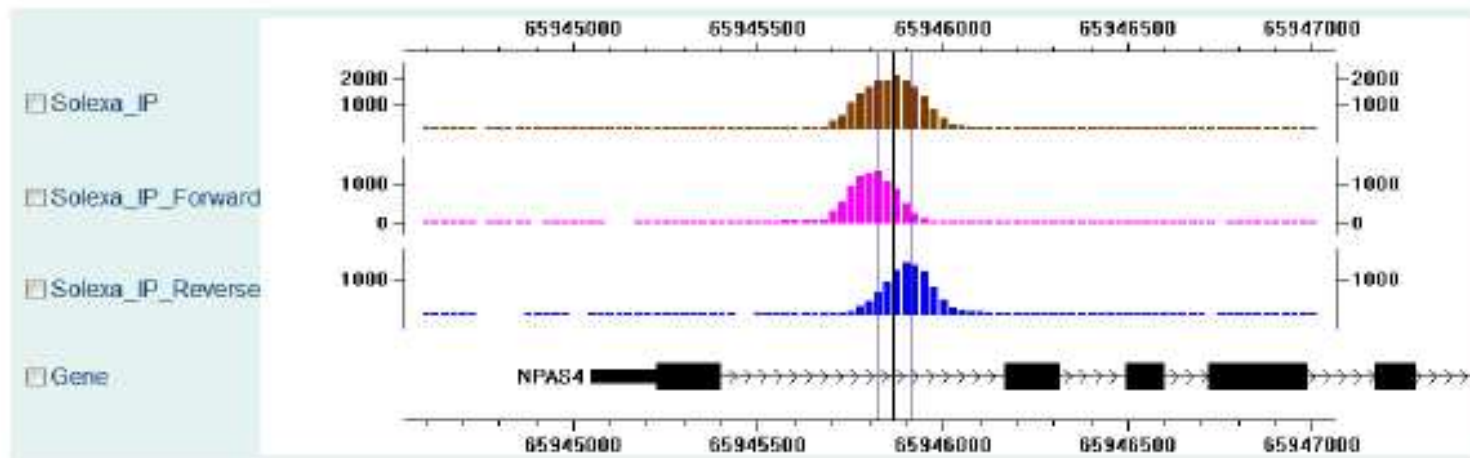
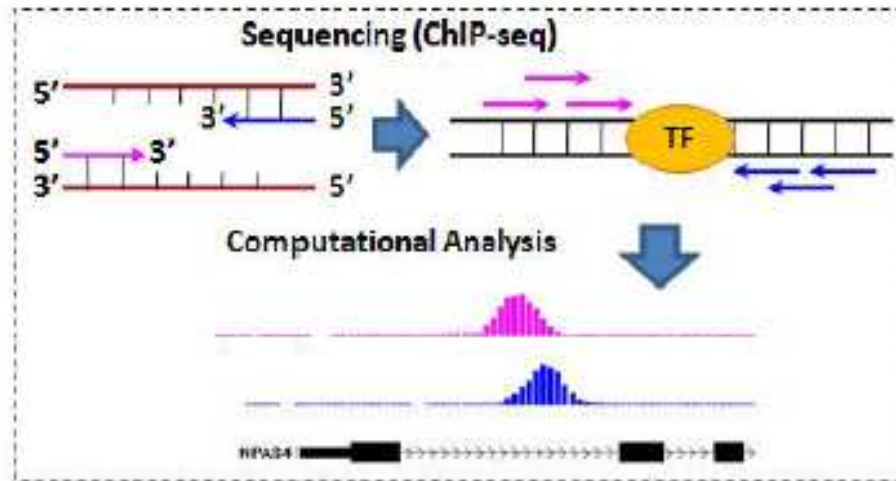
Different types of control

- Input DNA
- noIP control
 - Crosslink reversed, but immunoprecipitation bypassed
- Non-specific antibody
- Different tissue

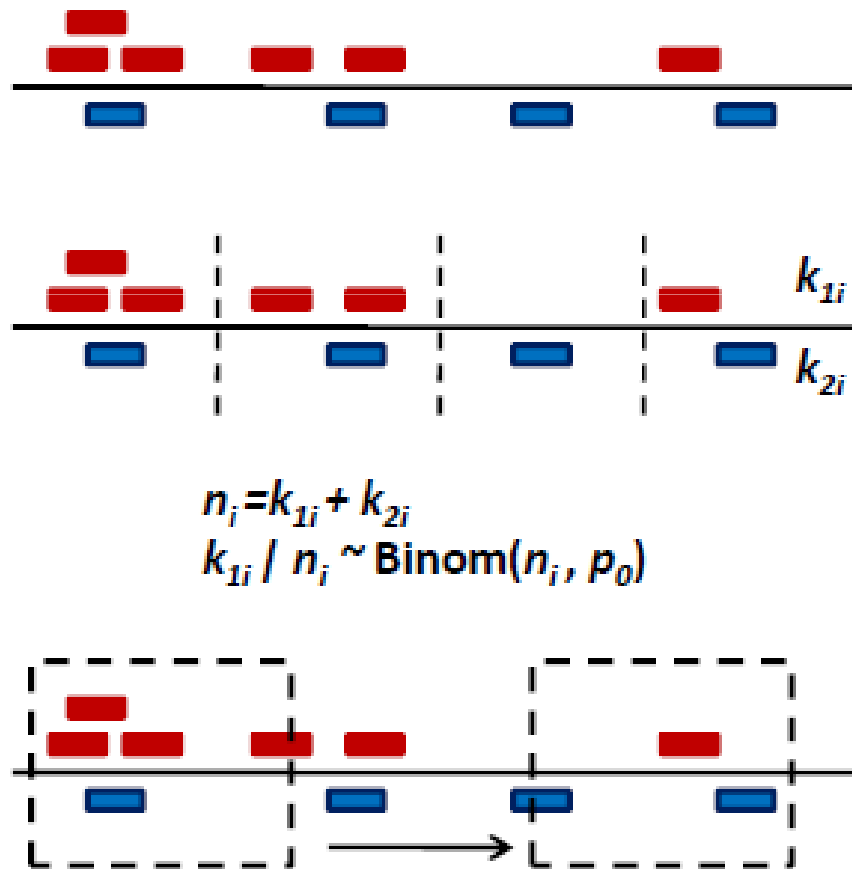
Analysis pipeline



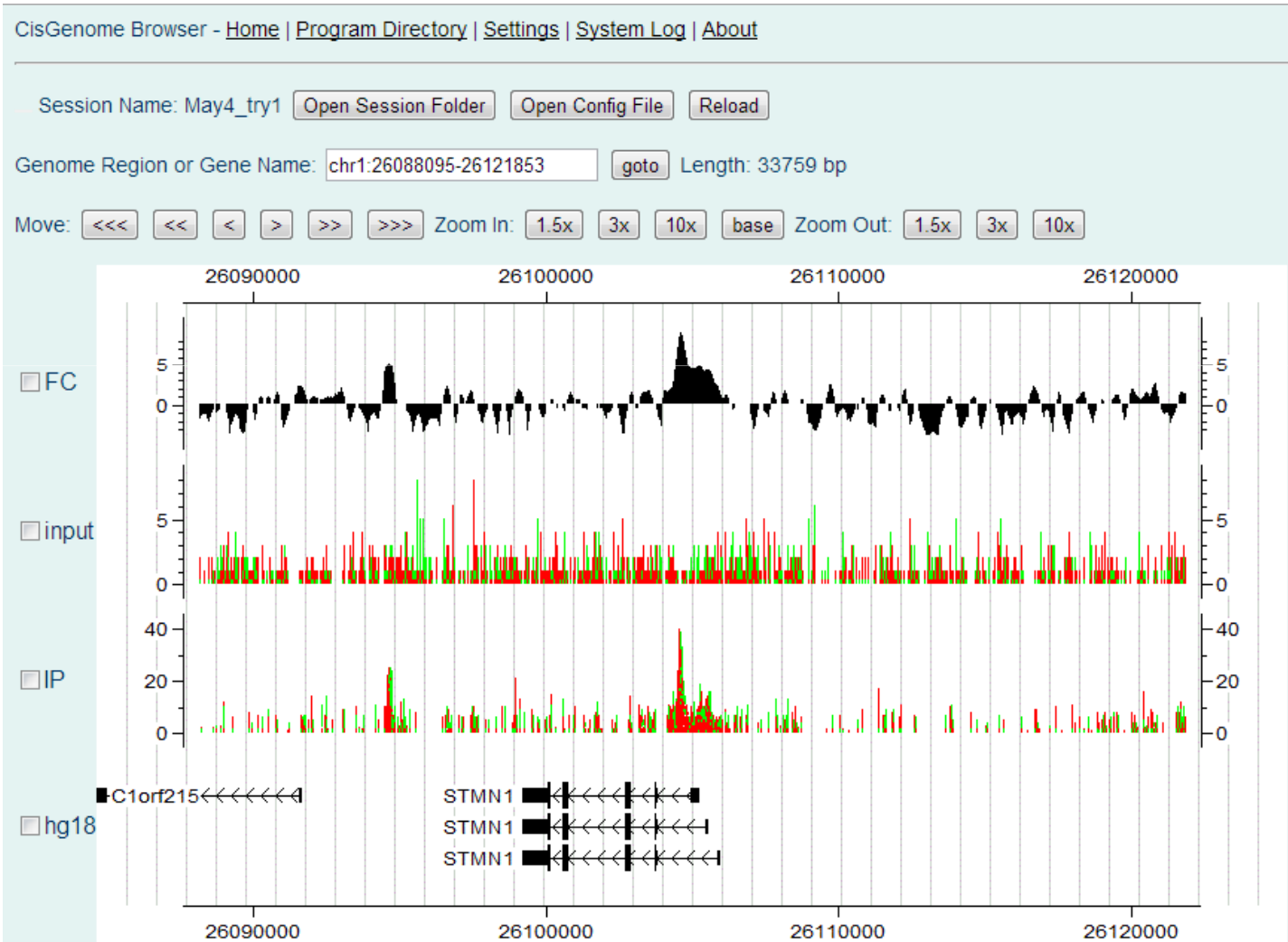
Peak refinement



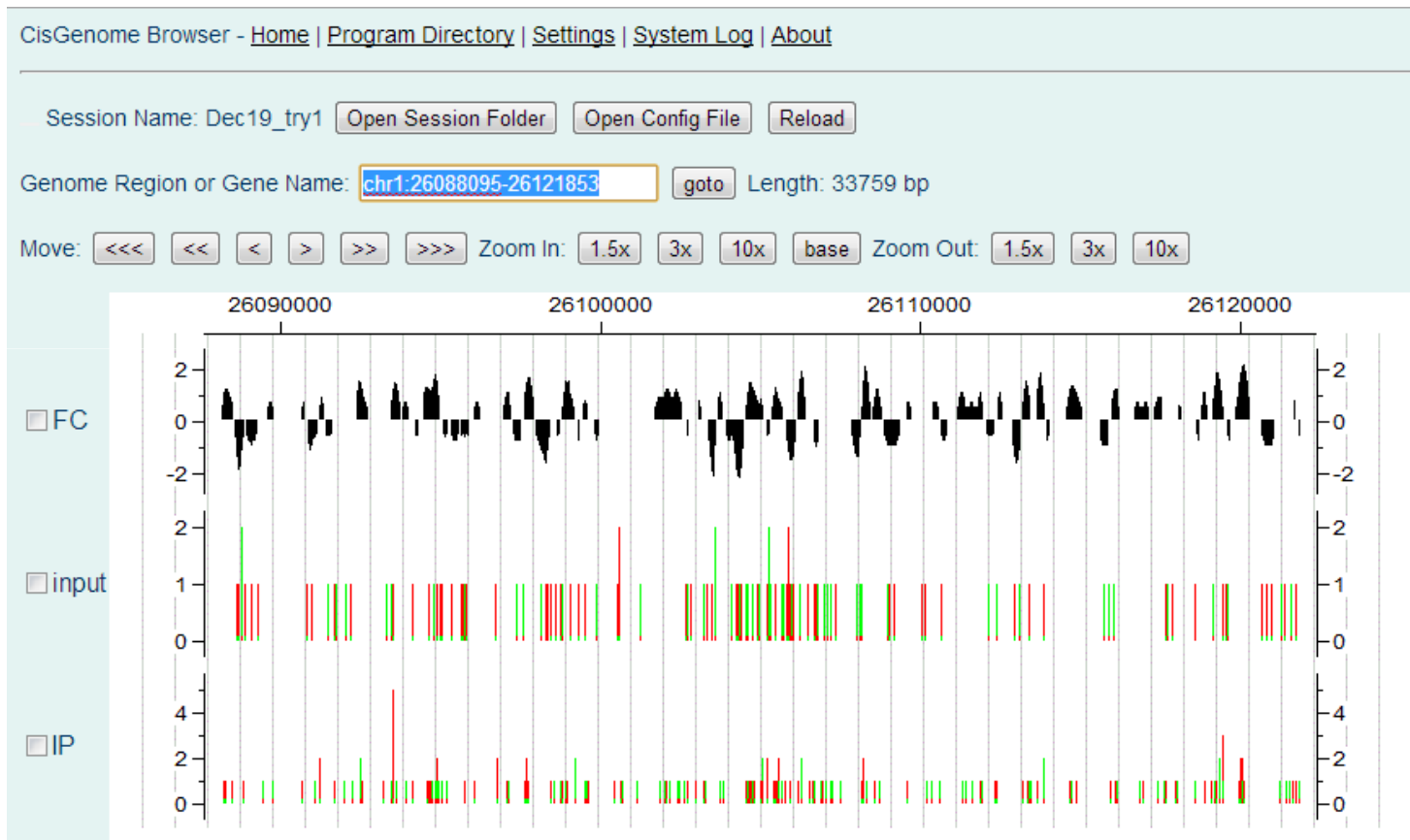
CisGenome two sample peak calling



Good data



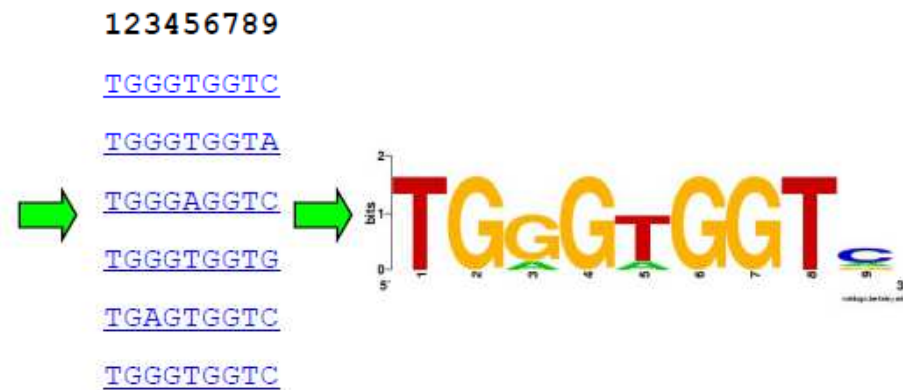
Bad data



Sequence motif- a pattern of nucleotide sequences

GTATGTA**CTTACTAT**TGGGTGGTCAACAAATCTATGTATGA
TAACATGTGACTCCTATAACCTCTTTGGGTGGTACATGAA
CTGGGAGGTCCCTCGGTTCAGAGTCACAGAGCAGATAATCA
TTAGAGGCACAATTGCTTGGGTGGTGCACAAAAAACAAG
AACAGCCTTGGATTAGCTGCTGGGGGGGTTGAGTGGTCCAC
ATCAGAATGGGTGGTCCATATATCCCAAAGAAGAGGGTAG

Transcription Factor Binding Sites (TFBS)



Motif discovery algorithm

	A	C	G	T
A	.3	.2	.2	.3
C	.2	.3	.3	.2
G	.2	.3	.3	.2
T	.3	.2	.2	.3

	1	2	3	4	5	6	7	8	9
A	0.00	0.00	0.17	0.00	0.17	0.00	0.00	0.00	0.17
C	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.66
G	0.00	1.00	0.83	1.00	0.00	1.00	1.00	0.00	0.17
T	1.00	0.00	0.00	0.00	0.83	0.00	0.00	1.00	0.00

Background: θ_0

Motif: Θ, w

q_0

q_1

$q = [q_0, q_1]$

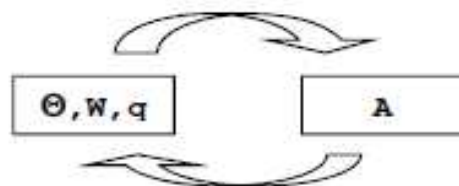


S: GTATGTA~~CTTACTAT~~TGGGTGGTCAACAAATCTATGTA~~TGACT~~TGGGAGGTCCTCGGTTTCAGAGTCACAGAGCA

A: 0000000000000001000000000000000000000000000000000001000

$$f(A, \Theta, W, q | S, \theta_0) \propto f(S, A | \Theta, W, q, \theta_0) \pi(\Theta, W, q)$$

Inference by iterative estimation/sampling



EM:

Lawrence and Reilly (1990)

Bailey and Elkan (1994), etc.

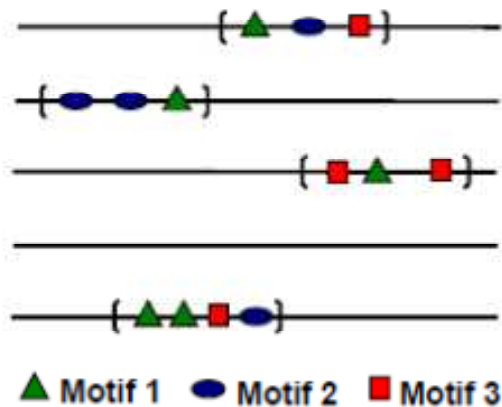
Gibbs Sampler:

Lawrence et al. (1993)

Liu (1994), Liu et al. (1995), etc.

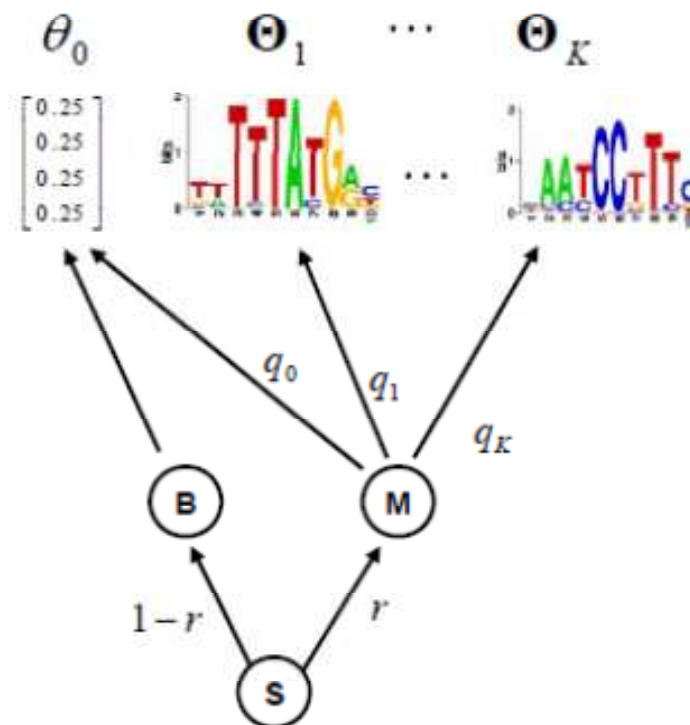
Motif module discovery

- Module structure: consider co-localization of motif sites.

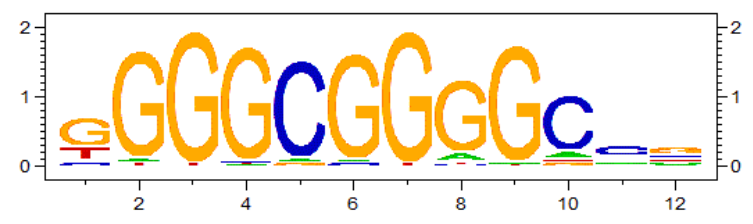
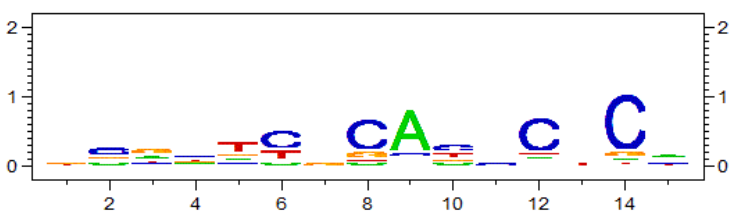
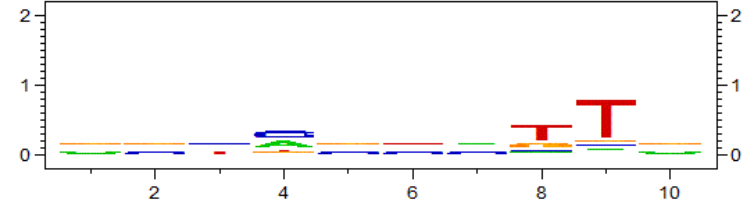
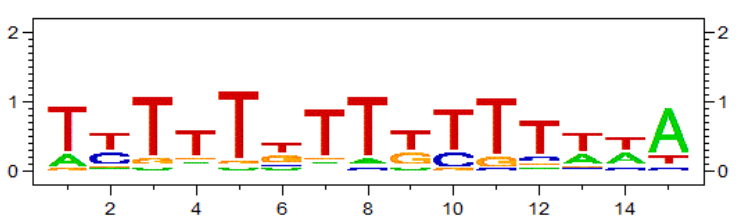
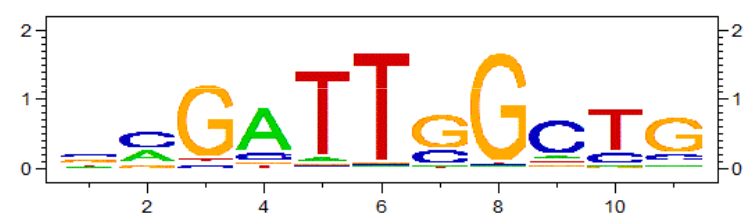
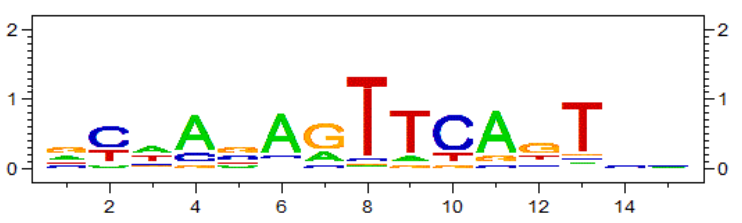
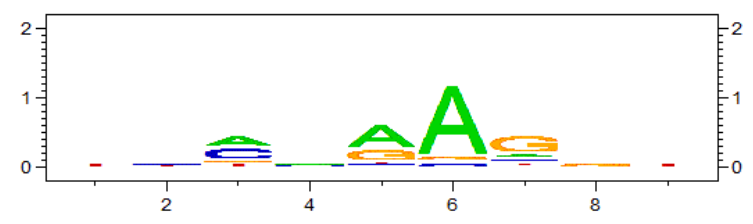
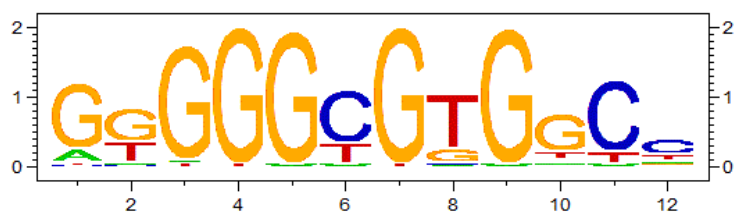
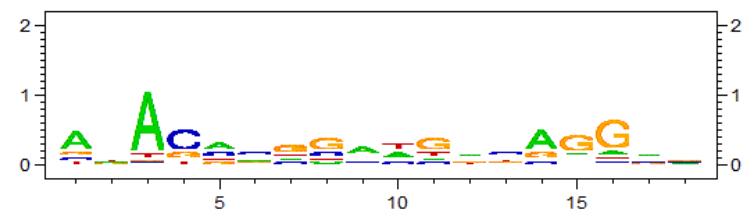
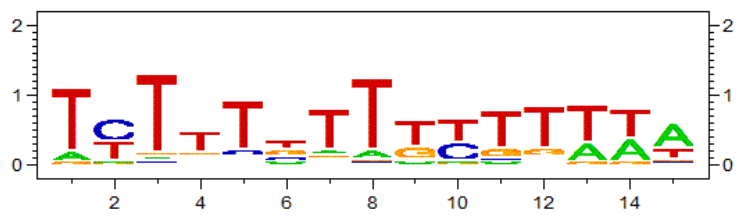


Hierarchical Mixture modeling →

K : # of motifs



Zhou Q and Wong W.H (2004) PNAS



Known motif mapping

Motif matrix mapping (CisGenome)

STEP 1: provide a motif and background model

STEP 2: specify a likelihood ratio cutoff (e.g. $LR \geq 500$)

STEP 3: scan the sequence

Background: θ_0

	A	C	G	T
A	.3	.2	.2	.3
C	.2	.3	.3	.2
G	.2	.3	.3	.2
T	.3	.2	.2	.3

Motif: θ

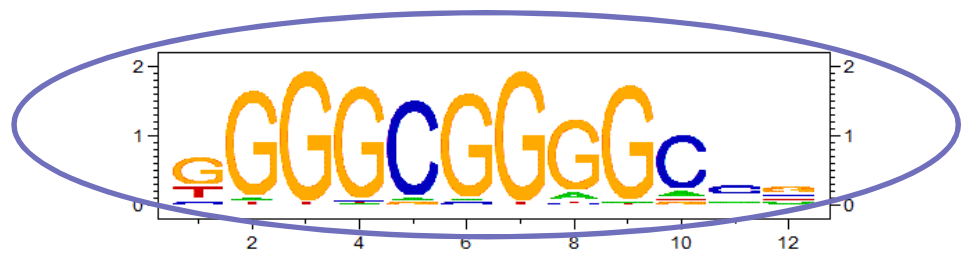
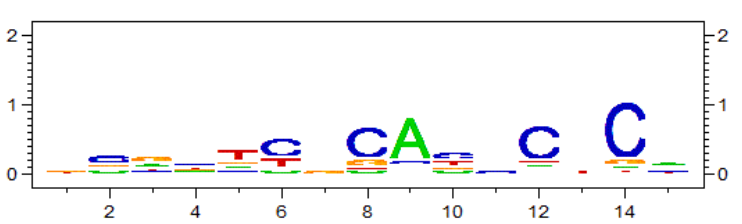
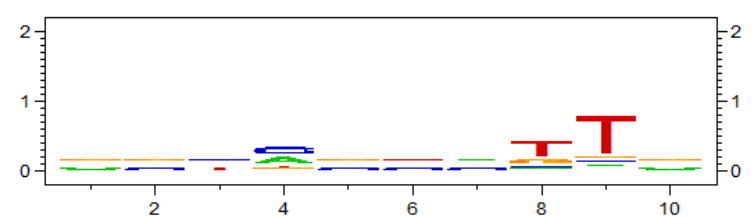
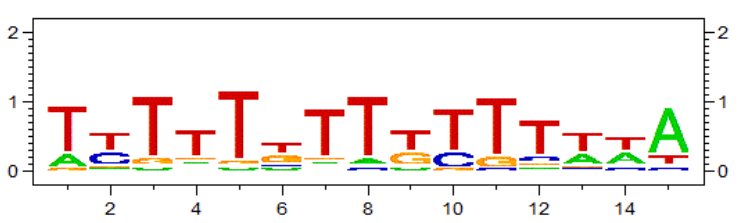
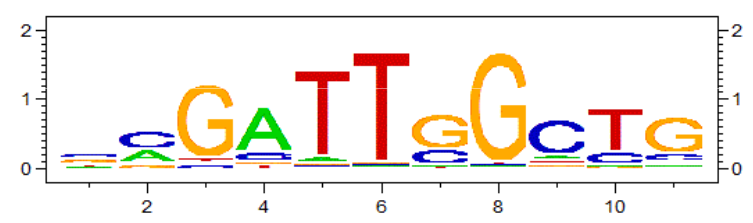
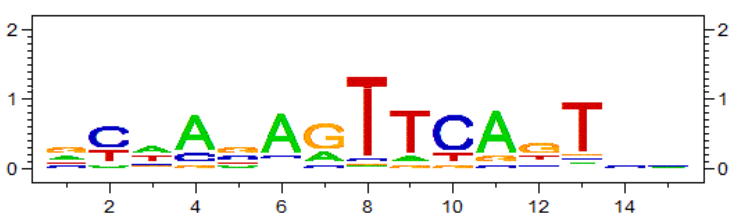
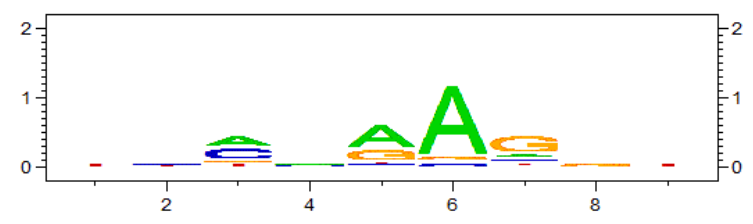
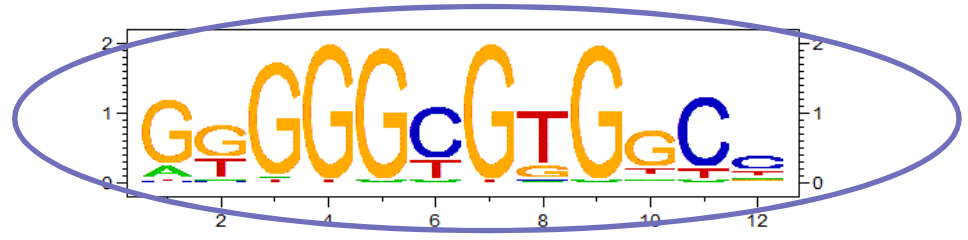
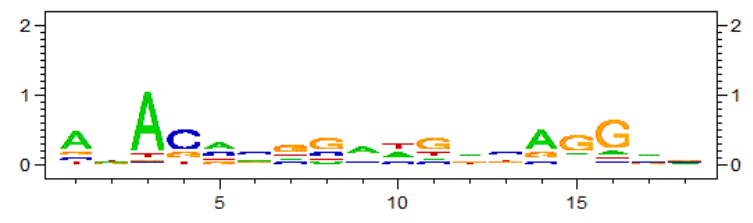
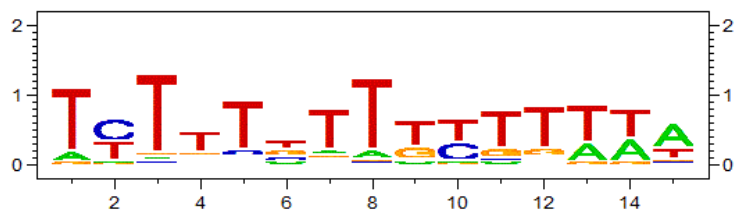


	1	2	3	4	5	6	7	8	9
A	0.00	0.00	0.17	0.00	0.17	0.00	0.00	0.00	0.17
C	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.66
G	0.00	1.00	0.83	1.00	0.00	1.00	1.00	0.00	0.17
T	1.00	0.00	0.00	0.00	0.83	0.00	0.00	1.00	0.00

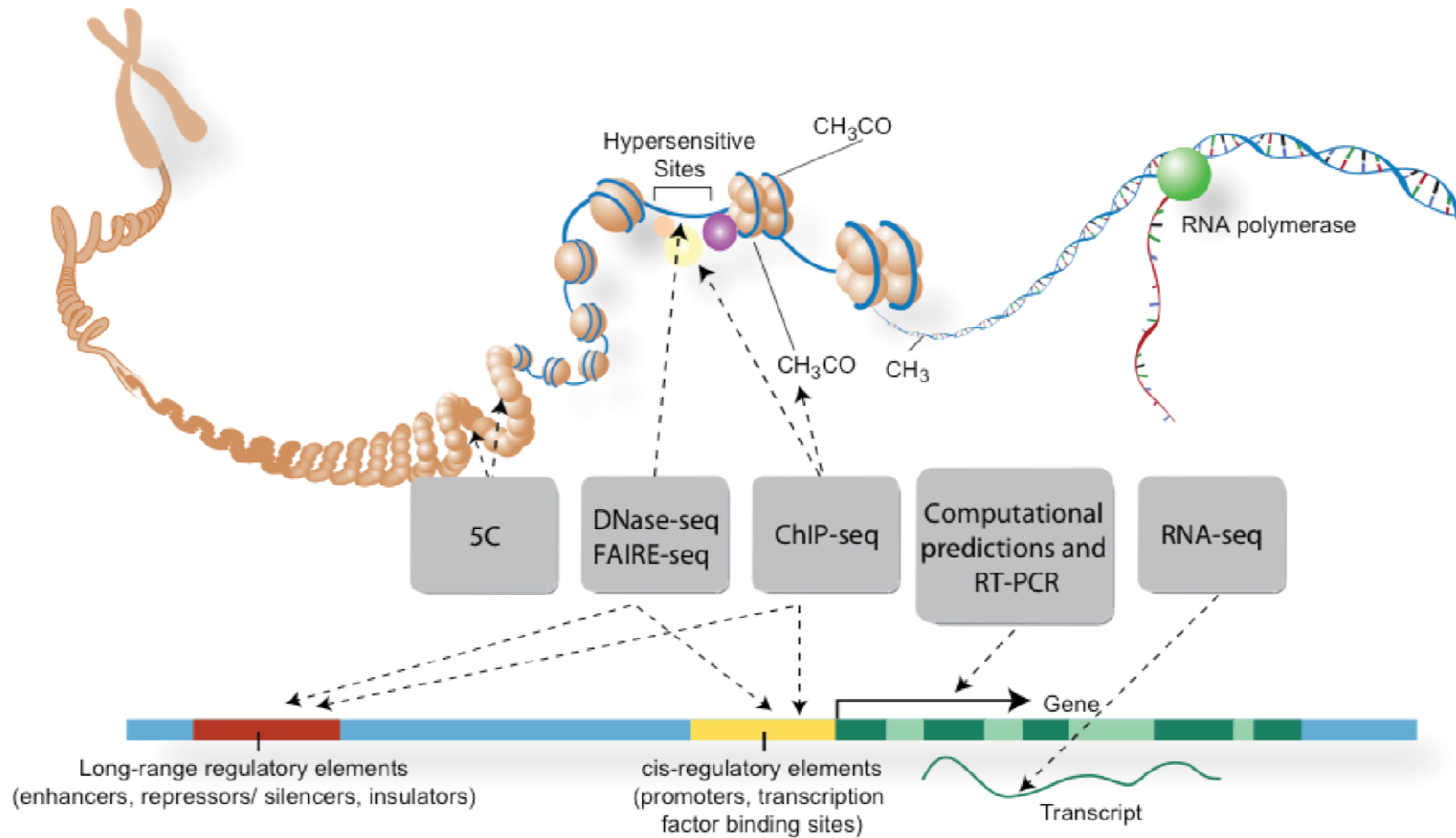
GTATGTA CTTACTA IGGGTGGTCAACAAATCTATGTATGACTGGGAGGTCCTCGGTTCAGAGTCACAGAGCA

LR > 500, yes

LR < 500, no



Other types of assays



<http://genome.ucsc.edu/ENCODE/aboutScaleup.html>

Acknowledgement

- **Thanks Dr. Ji and Dr. Leek for some slides!**
- **Thanks for your attention!**