Wang et al. Genome Biology 2014, 15:R11
http://genomebiology.com/2014/15/1/R11

Genome Biology

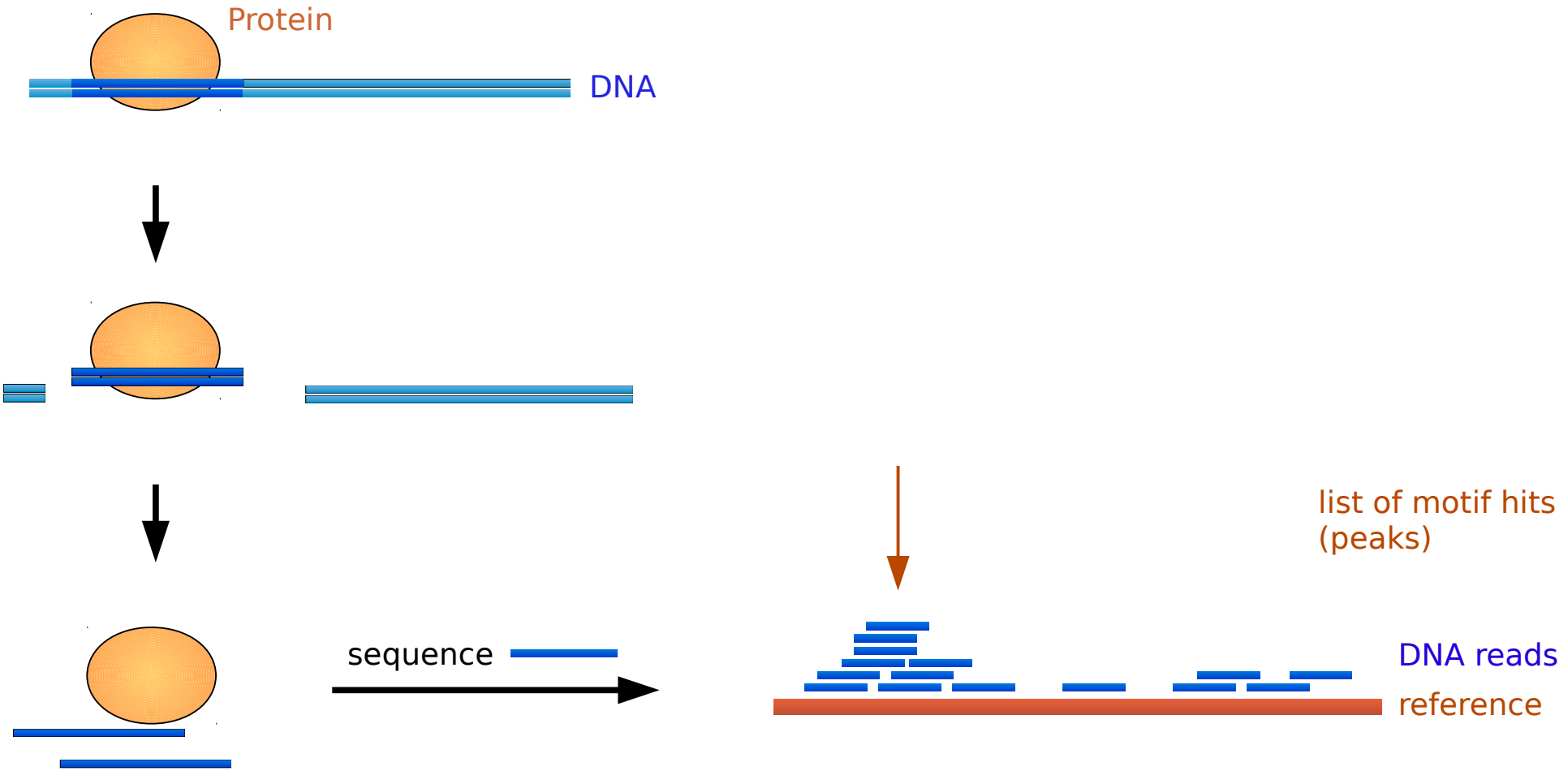
SOFTWARE Open Access

dCLIP: a computational approach for comparative CLIP-seq analyses

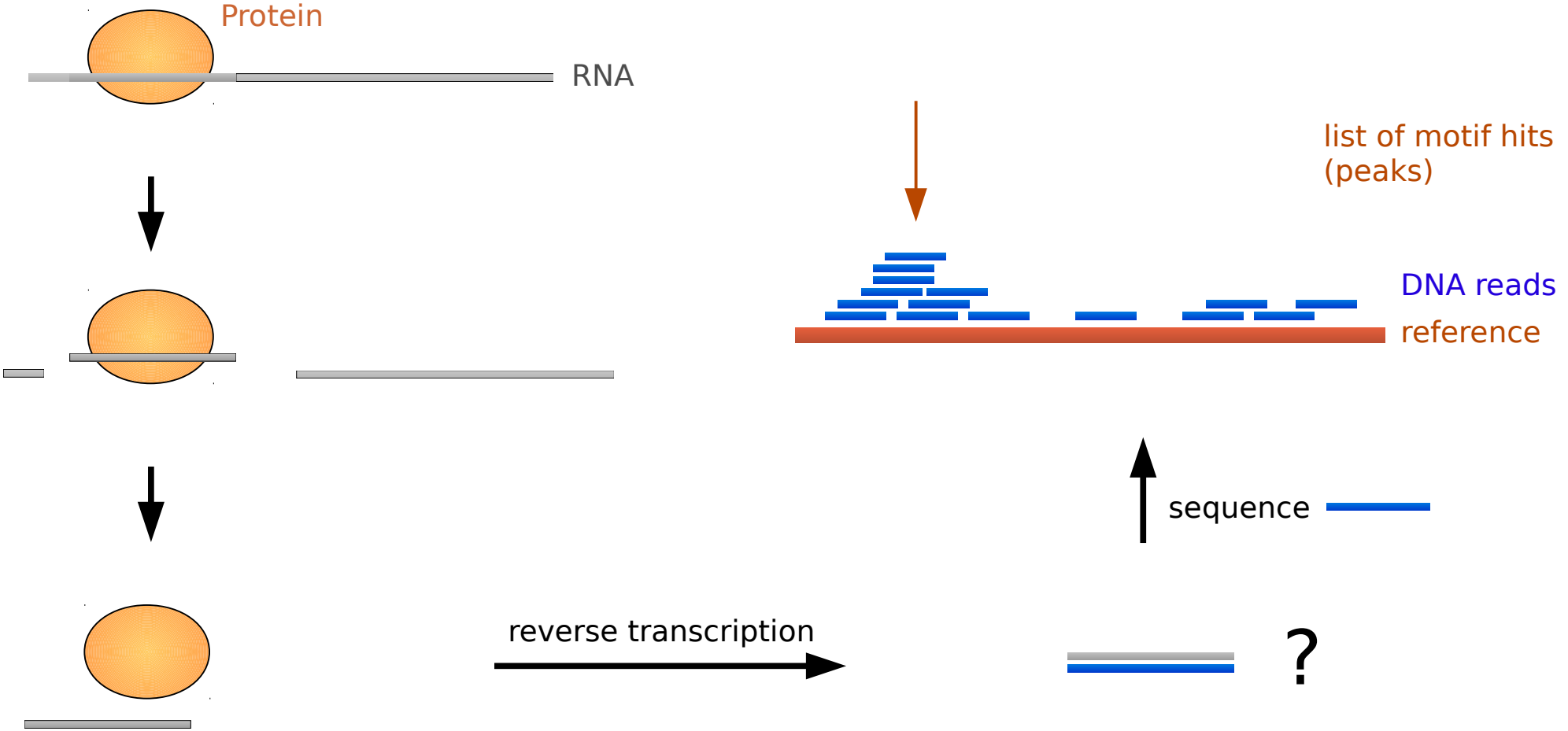
Tao Wang[1], Yang Xie[1,2] and Guanghua Xiao[1*]

presented to you by Jakob Schulze

ChIP-seq


Protein
DNA
sequence
list of motif hits
(peaks)
DNA reads
reference

CLIP-seq


Protein
RNA
list of motif hits
(peaks)
DNA reads
reference
sequence
reverse transcription
?

CLIP-Seq overview

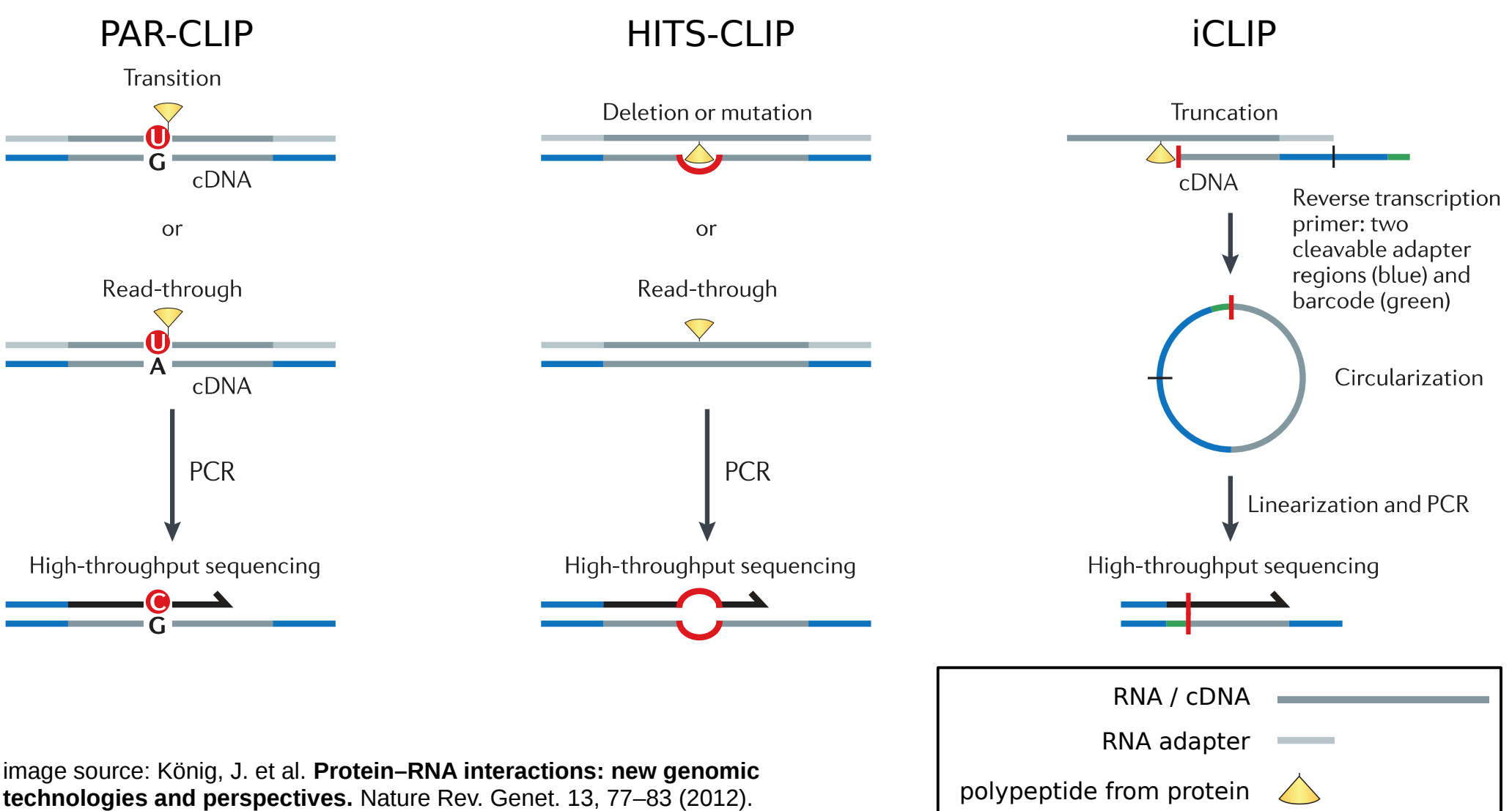


image source: König, J. et al. **Protein–RNA interactions: new genomic technologies and perspectives.** Nature Rev. Genet. 13, 77–83 (2012).

Specific mutations

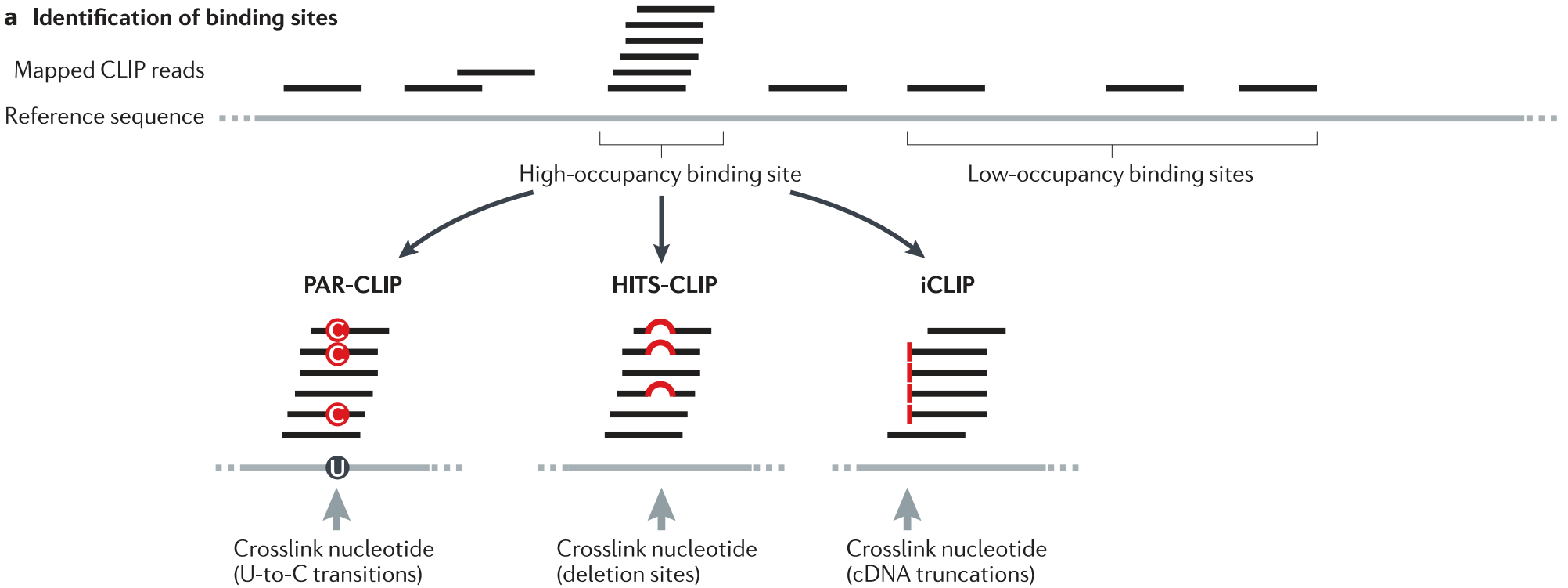


image source: König, J. et al. **Protein–RNA interactions: new genomic technologies and perspectives.** Nature Rev. Genet. 13, 77–83 (2012).

Tools so far ...

Freie Universität Berlin

single CLIP-Seq analyser:

- PARalyzer
- CLIPZ
- wavClusteR
- miRTarClip

comparative ChIP-Seq analyser:

- ChIPDiff
- ChIPnorm
- MAnorm
- dPCA

Tools so far ...

Freie Universität Berlin

single CLIP-Seq analyser:

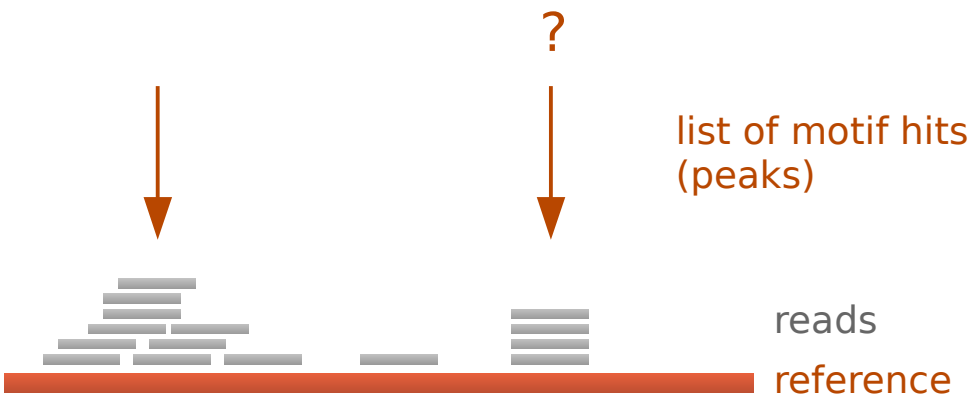
- PARalyzer
- CLIPZ
- wavClusteR
- miRTarClip

comparative ChIP-Seq analyser:

- ChIPDiff
- ChIPnorm
- MAnorm
- dPCA

comparative CLIP-Seq analyser:

- Piranha ???



hit or no hit ???

Data normalization - motivation

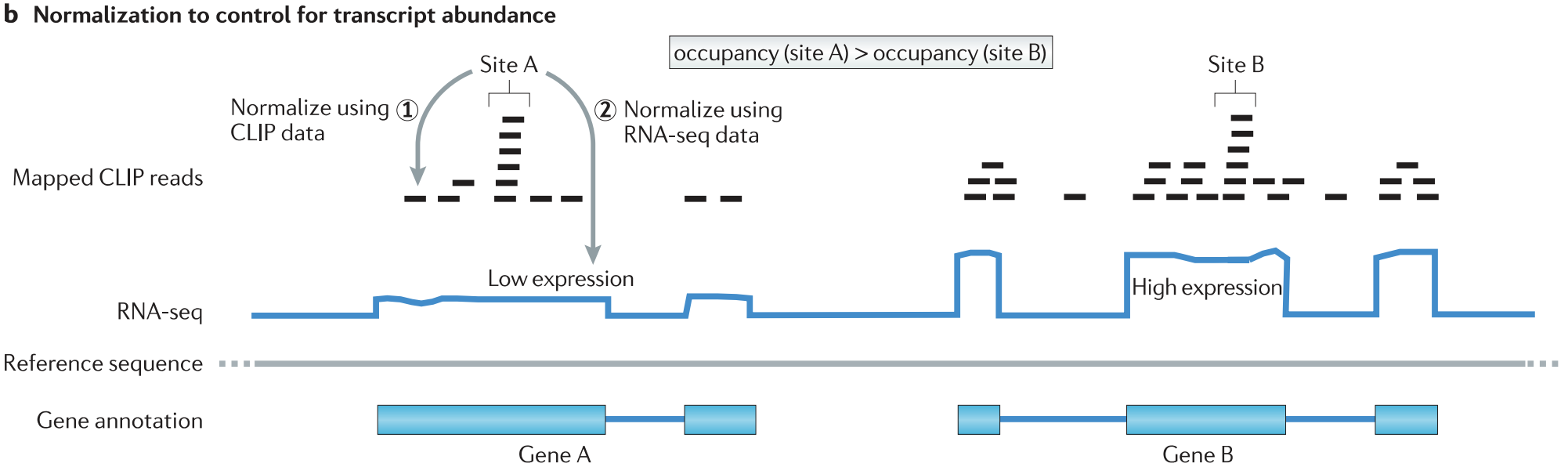


image source: König, J. et al. **Protein–RNA interactions: new genomic technologies and perspectives.** Nature Rev. Genet. 13, 77–83 (2012).

Data normalization - procedure

reads condition X

reference

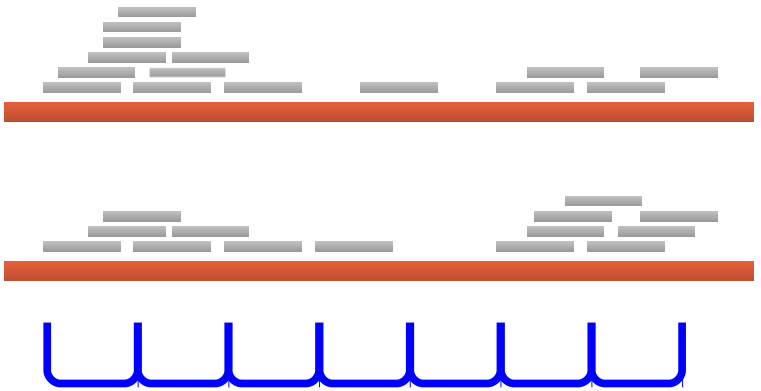
reads condition Y

bins

6	3	1	1	0	2	2	counts X
3	2	1	1	0	4	3	counts Y
3.2	2.6	0.2	-0.3	0.3	-2.4	-2.0	M value

$$M = \log X - \log Y$$

Data normalization - procedure



reads condition X

reference

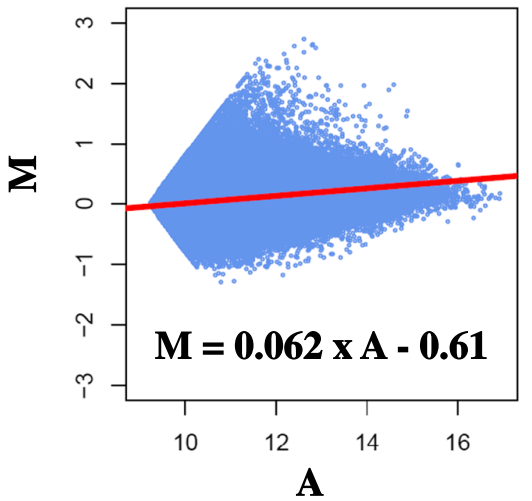
reads condition Y

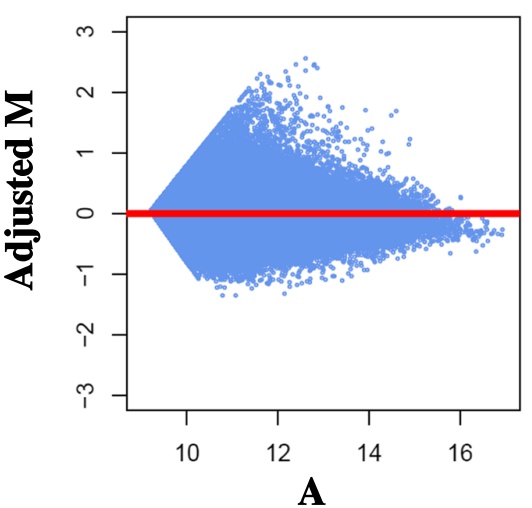
bins

6	3	1	1	0	2	2	counts X
3	2	1	1	0	4	3	counts Y
3.2	2.6	0.2	-0.3	0.3	-2.4	-2.0	M value

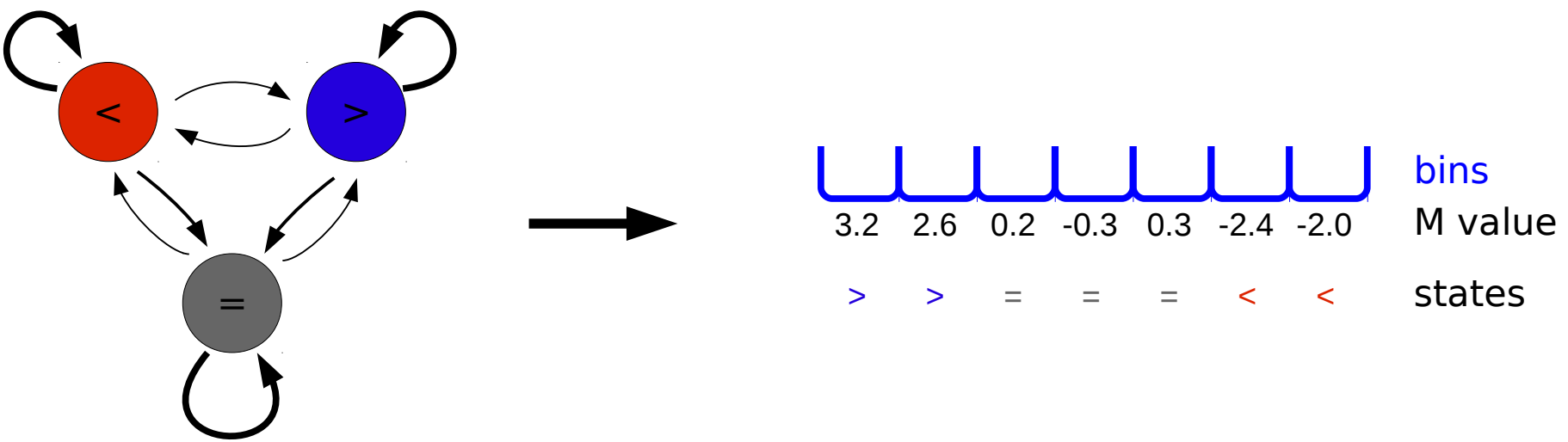
$$M = \log X - \log Y$$

$$A = \log X + \log Y$$

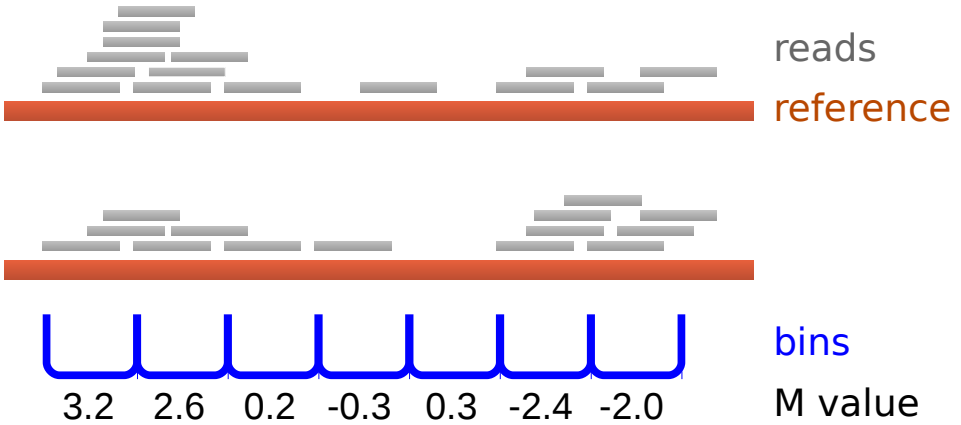




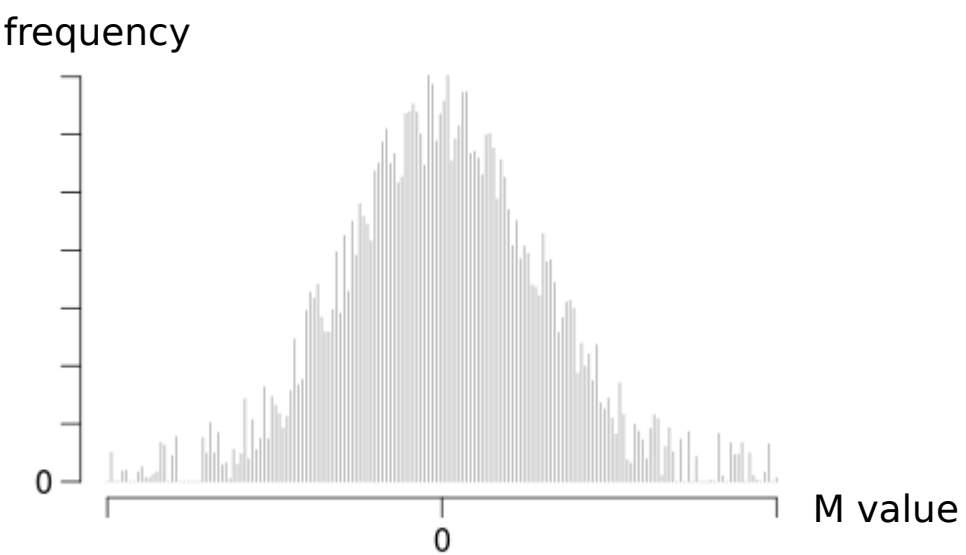
- states emit M values
- hidden states model differential binding motifs



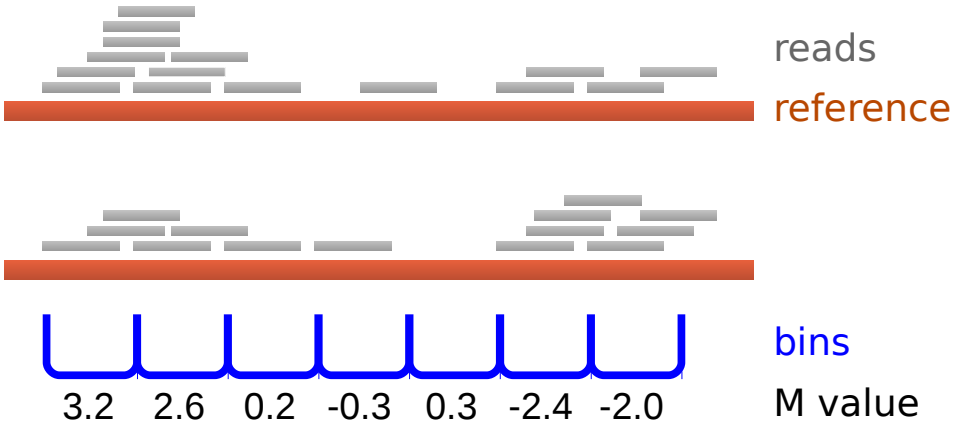
Hidden Markov model - emissions



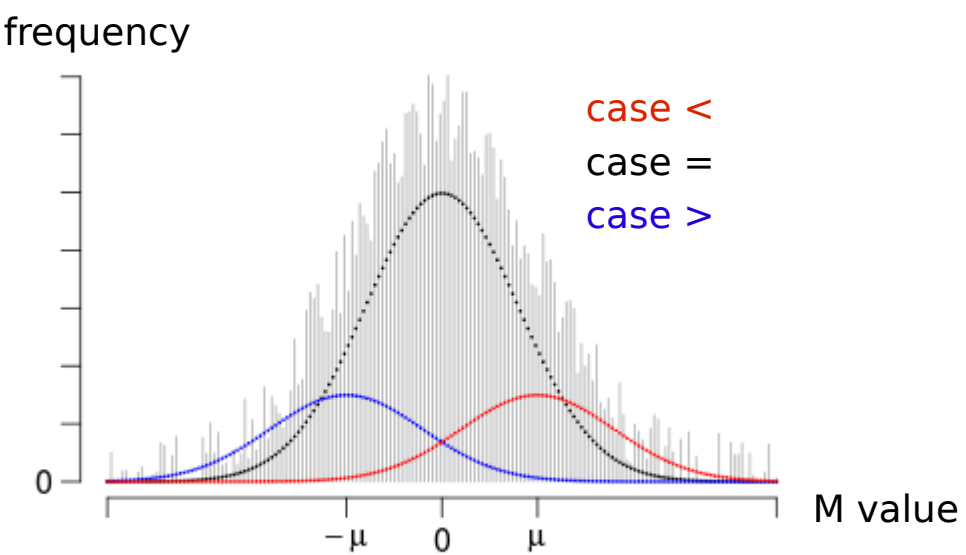
- take M values from all bins



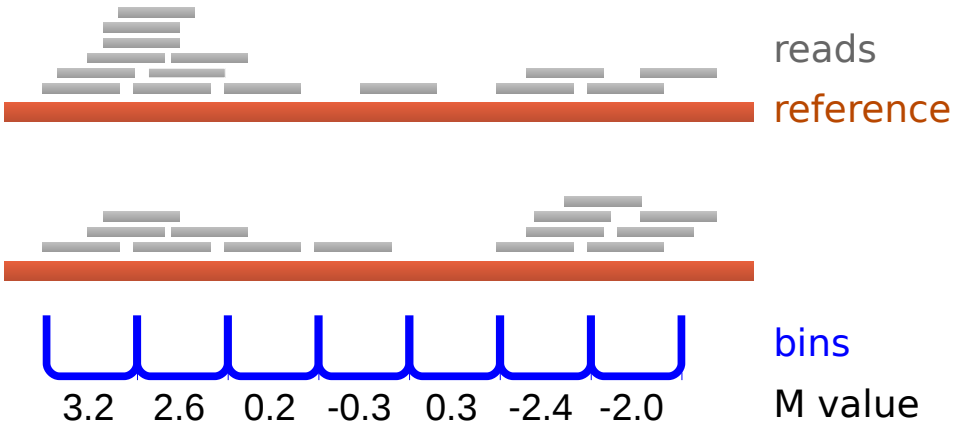
Hidden Markov model - emissions



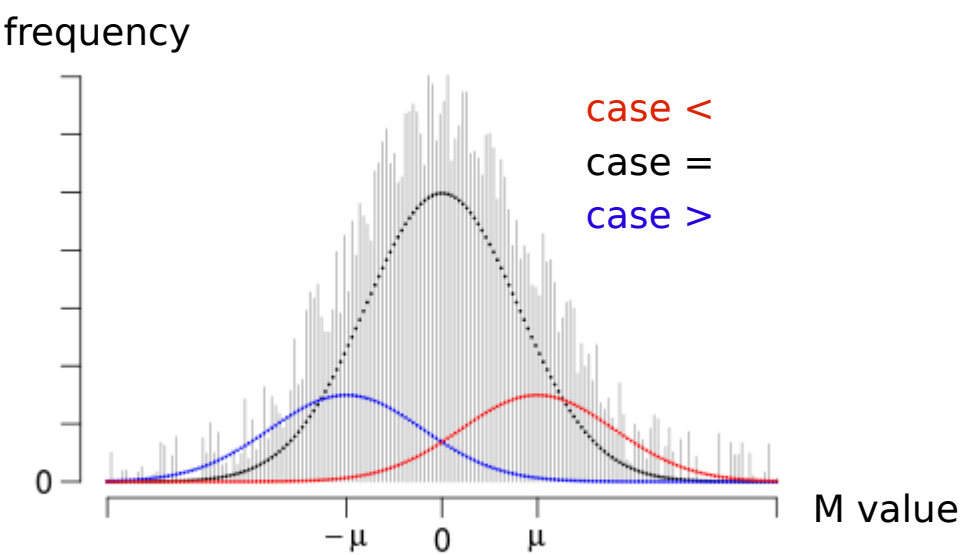
- take M values from all bins
- fit three component mixture model

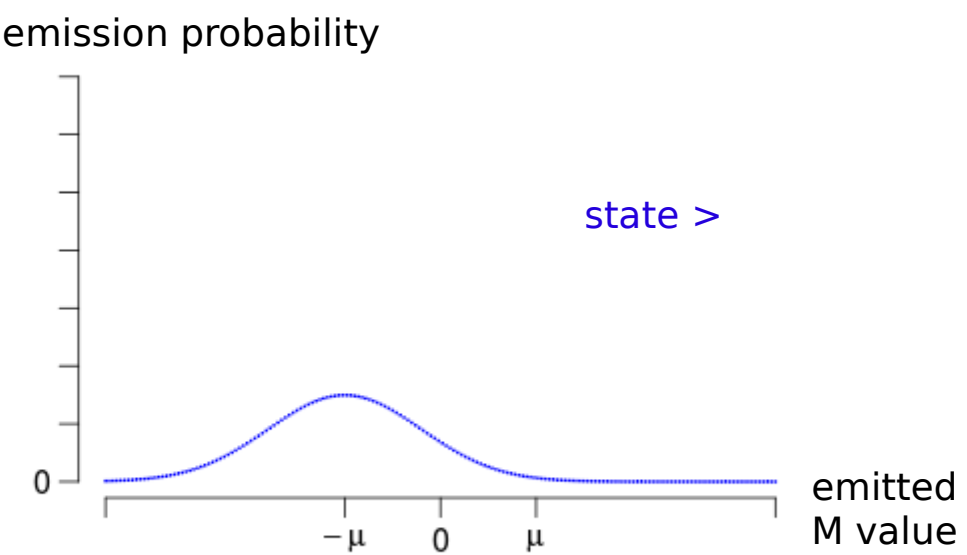


Hidden Markov model - emissions

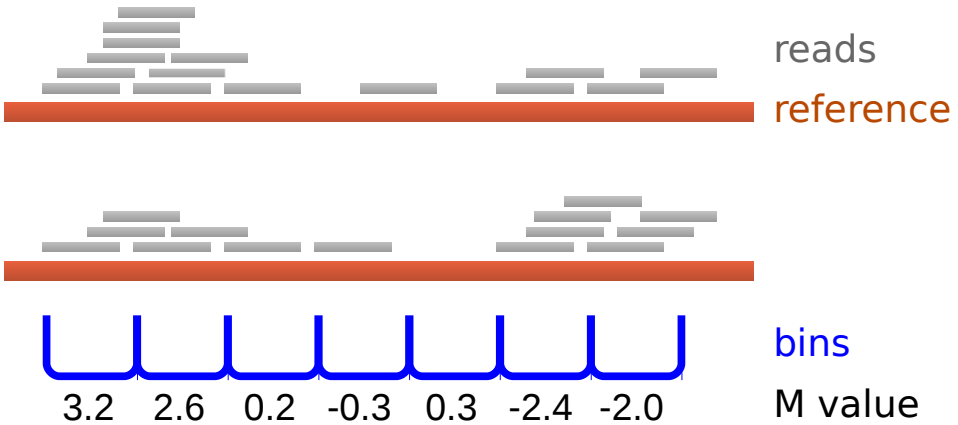


- take M values from all bins
- fit three component mixture model
- use three component mixture model as emission function

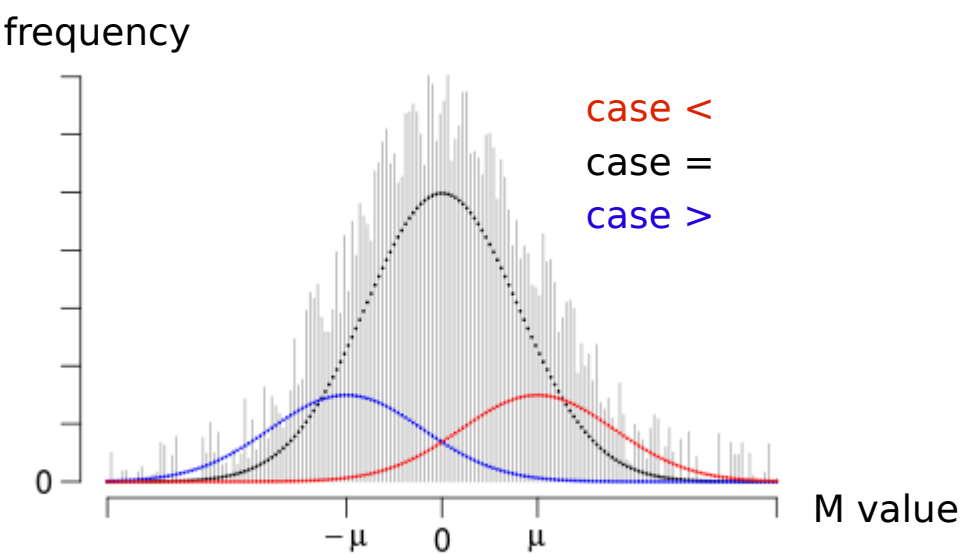


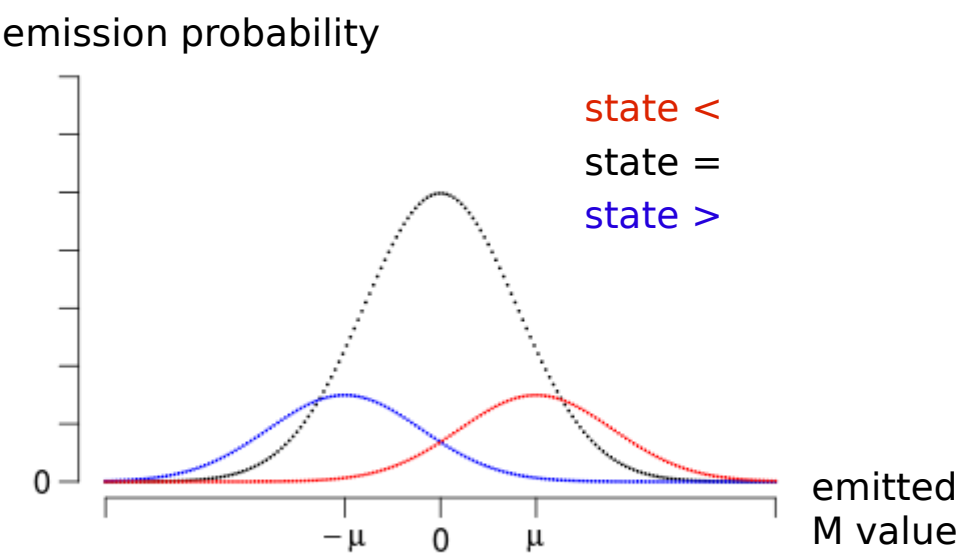


Hidden Markov model - emissions



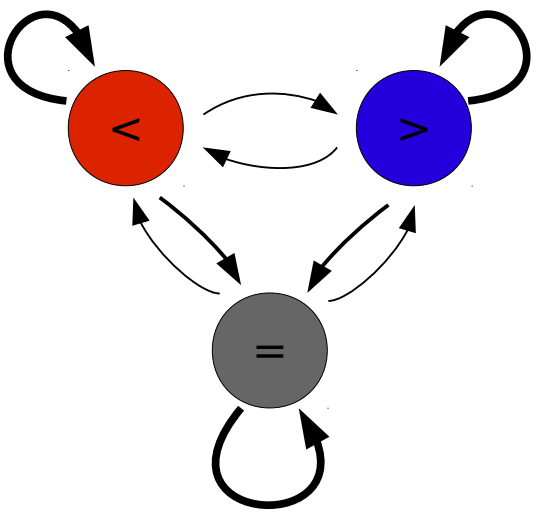
- take M values from all bins
- fit three component mixture model
- use three component as emission function

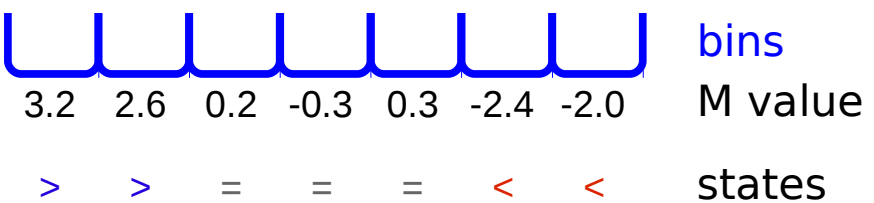




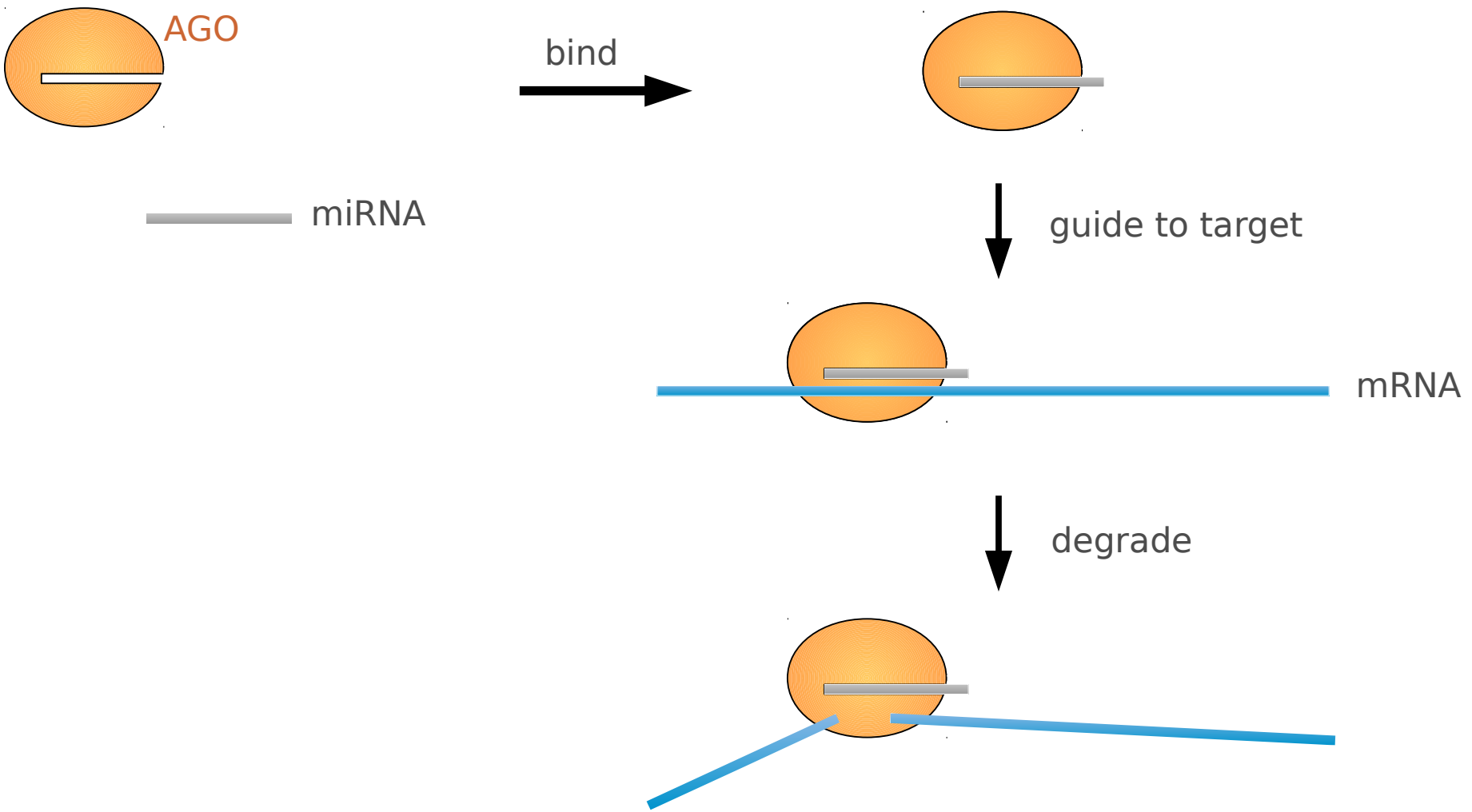
Hidden Markov model - summary

- emission probabilities → three component mixture model
- transition probabilities → Baum-Welch algorithm
- most likely hidden state sequence → Viterbi algorithm



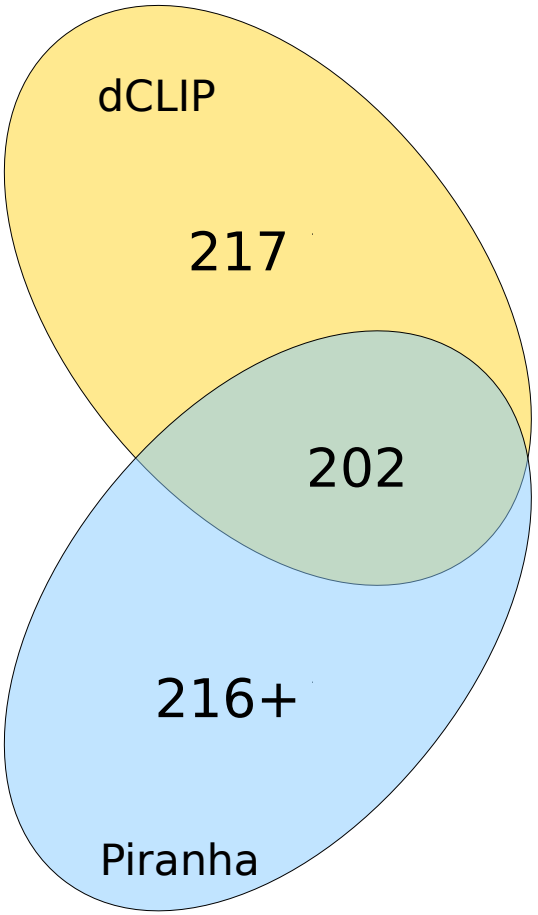


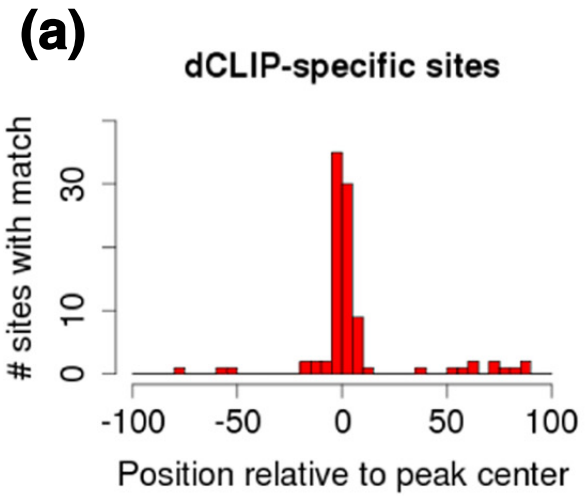
Review – RNA interference

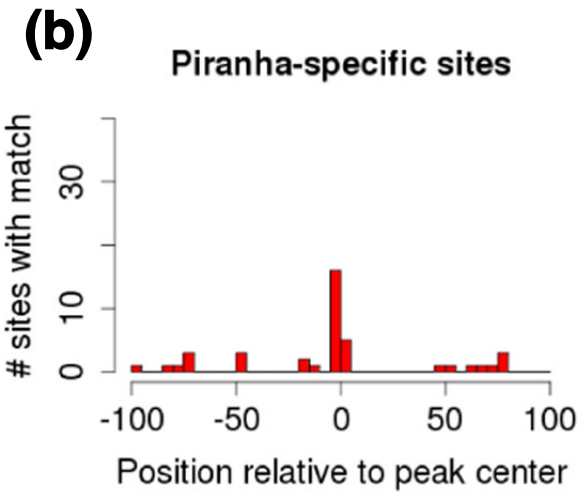


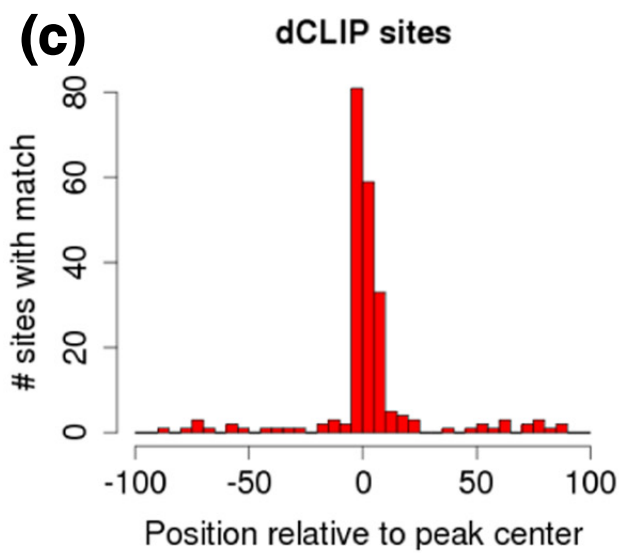
dCLIP vs. Piranha – miR-124/AGO HITS-CLIP dataset

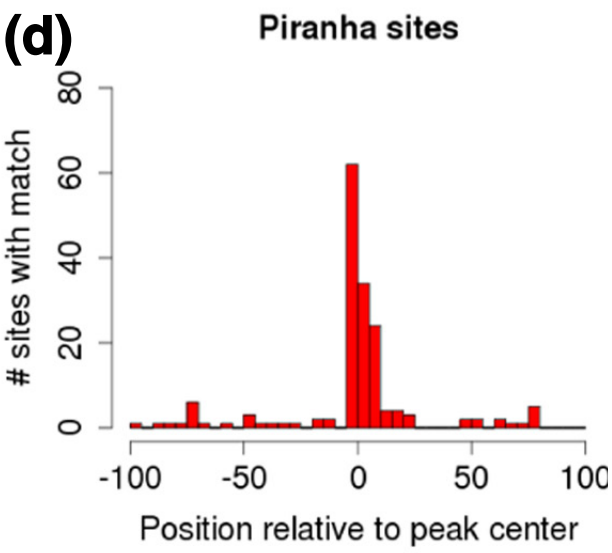
identified sites:





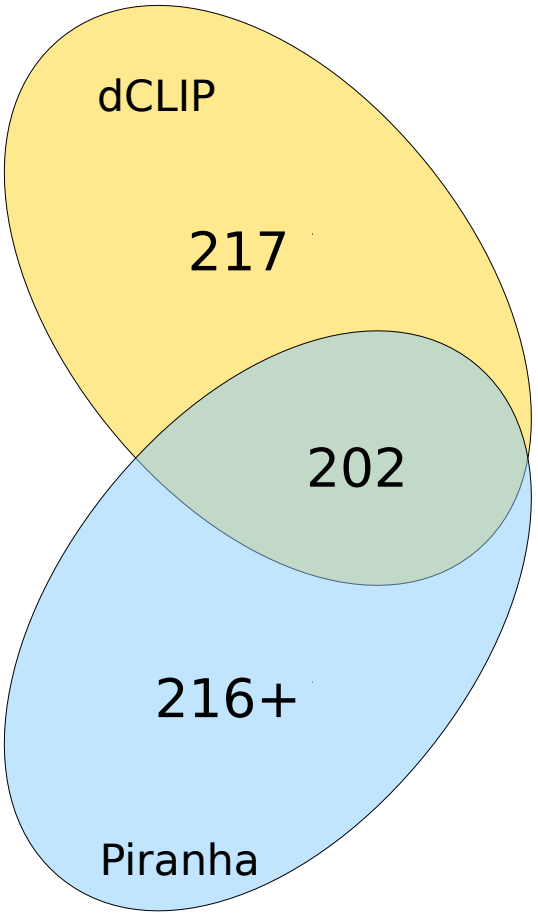


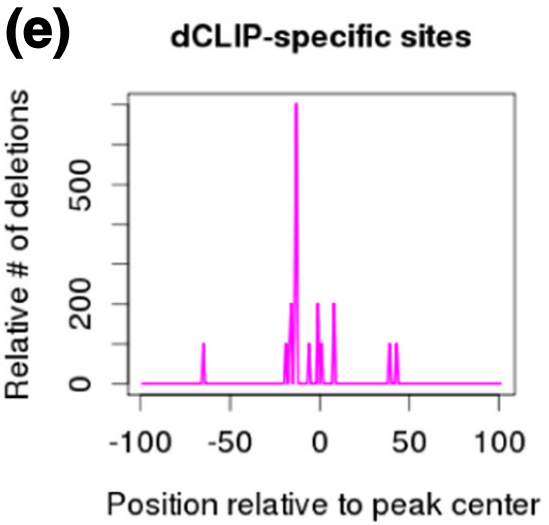




dCLIP vs. Piranha – miR-124/AGO HITS-CLIP dataset

identified sites:





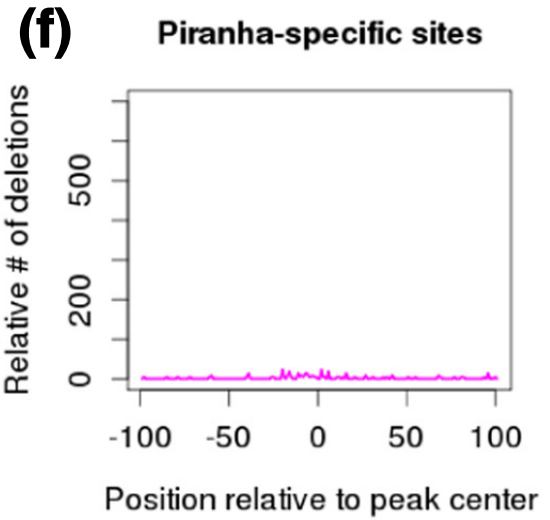
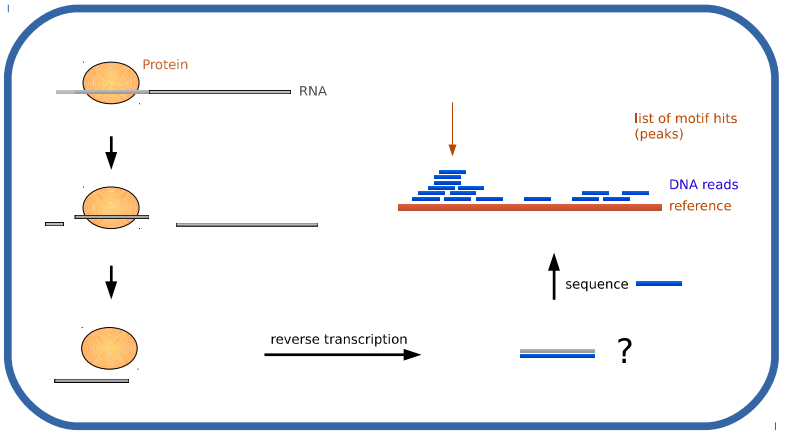
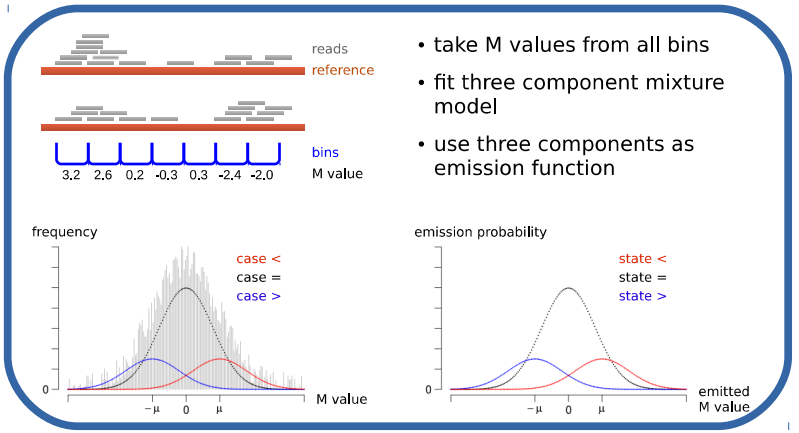


image source of all 6 plots:
Wang et al. **dCLIP: a computational approach for comparative CLIP-seq analyses.** Genome Biology 2014, 15:R11

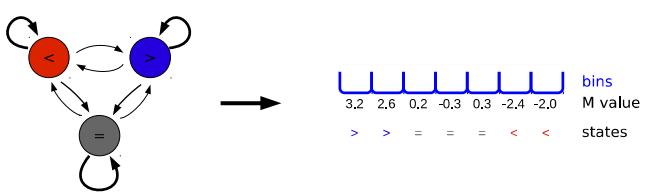
In a nutshell ...

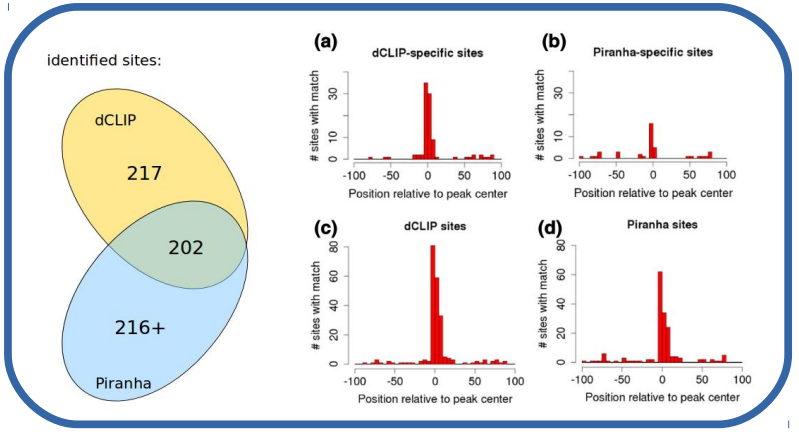




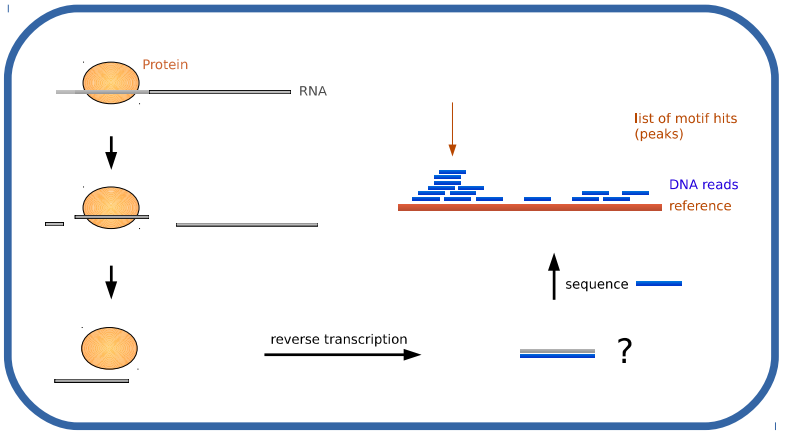
- take M values from all bins
- fit three component mixture model
- use three components as emission function

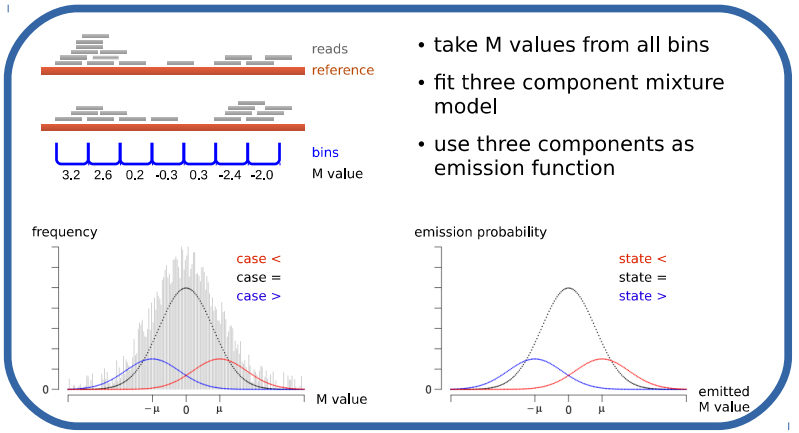
- emission probabilities → three component mixture model
- transition probabilities → Baum-Welch algorithm
- most likely hidden state sequence → Viterbi algorithm



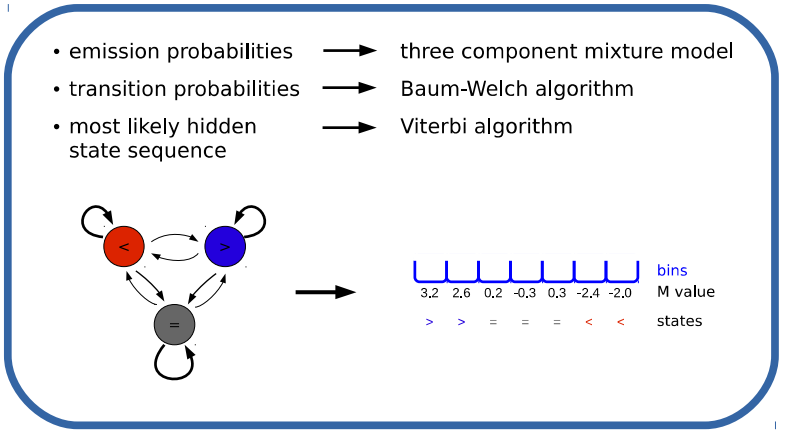


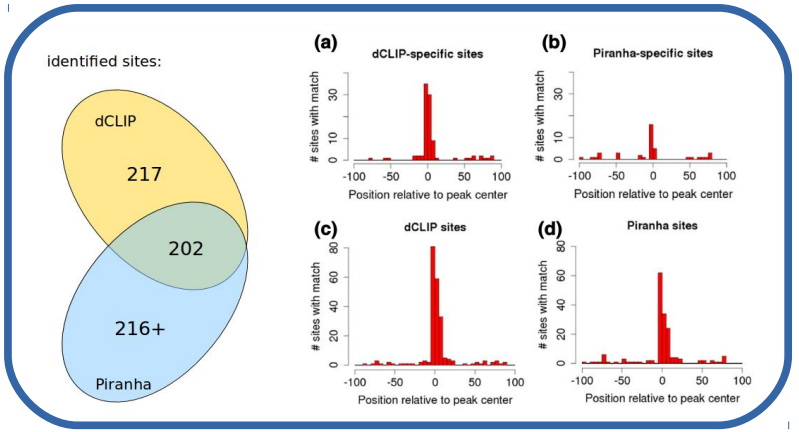
In a nutshell ...





- take M values from all bins
- fit three component mixture model
- use three components as emission function





Thank you for listening!