

Next Generation Sequencing ?

Not anymore.

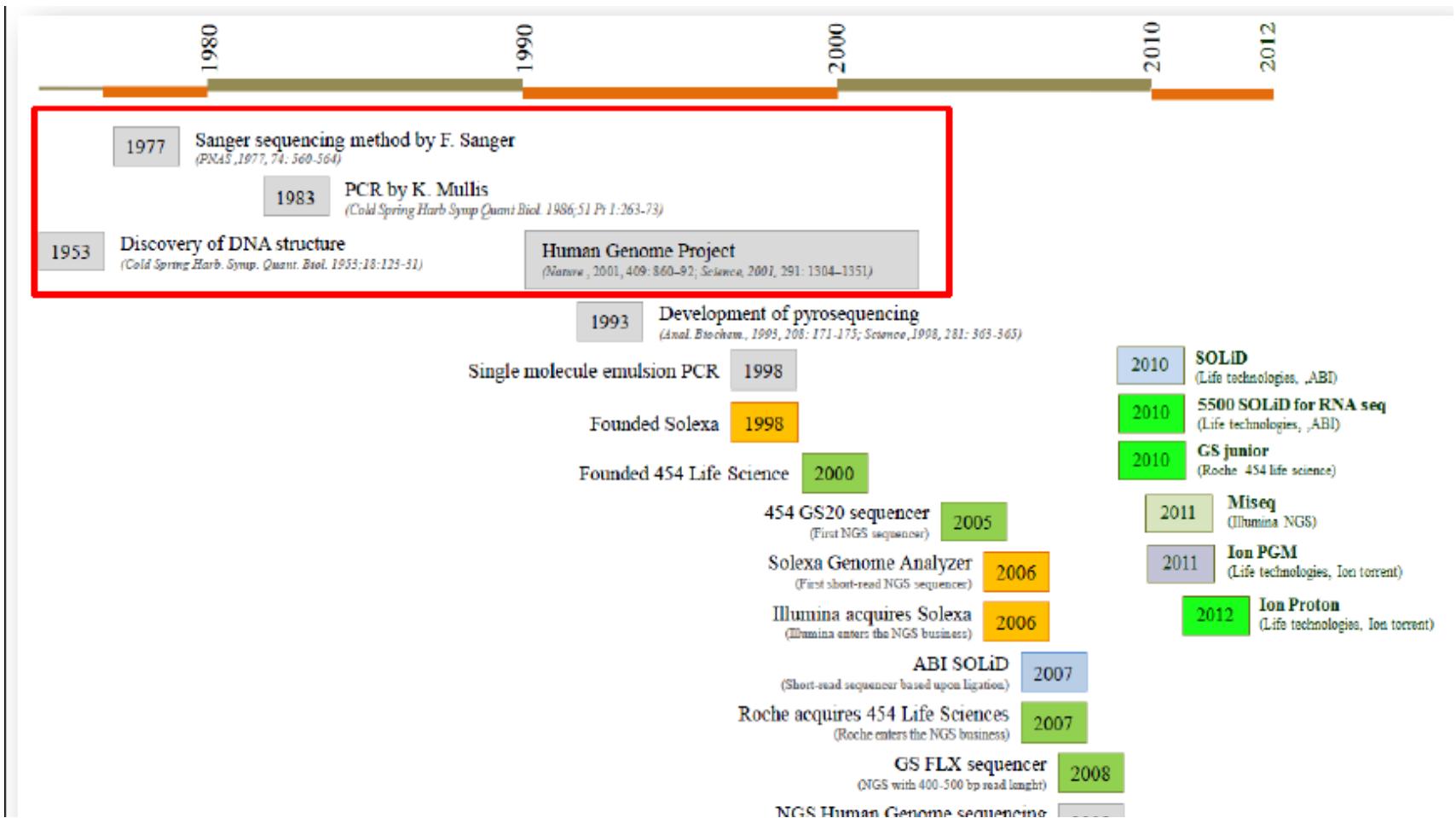


이 우 제

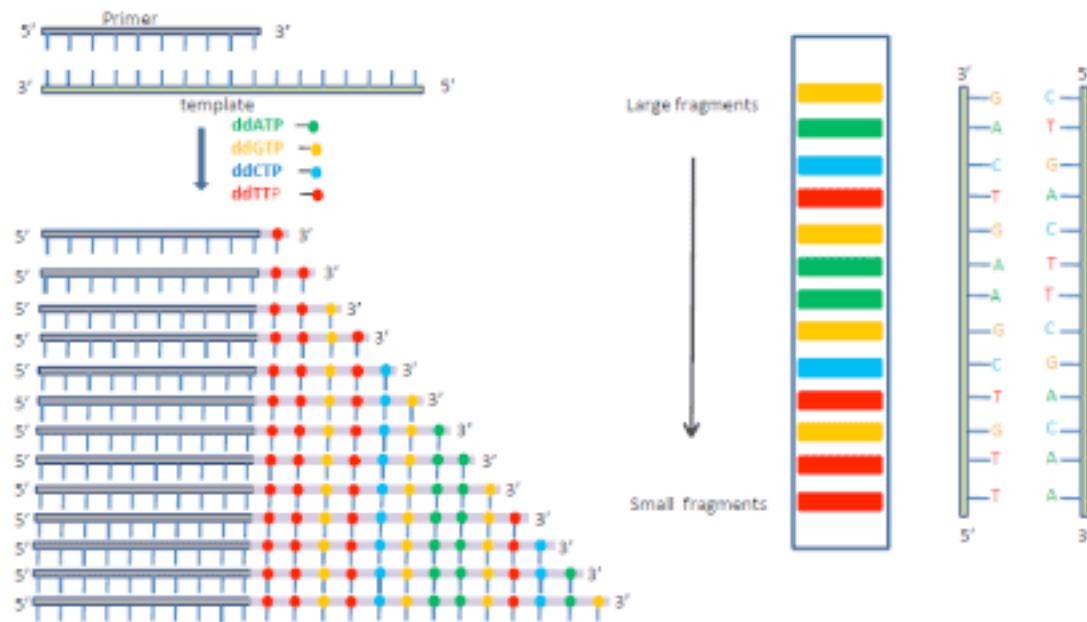
DNA sequencing



History of Sequencing Technology



Dye-terminator Sanger Sequencing





Fruit Fly

44%



Yeast

26%



Plant

18%



Mouse

92%



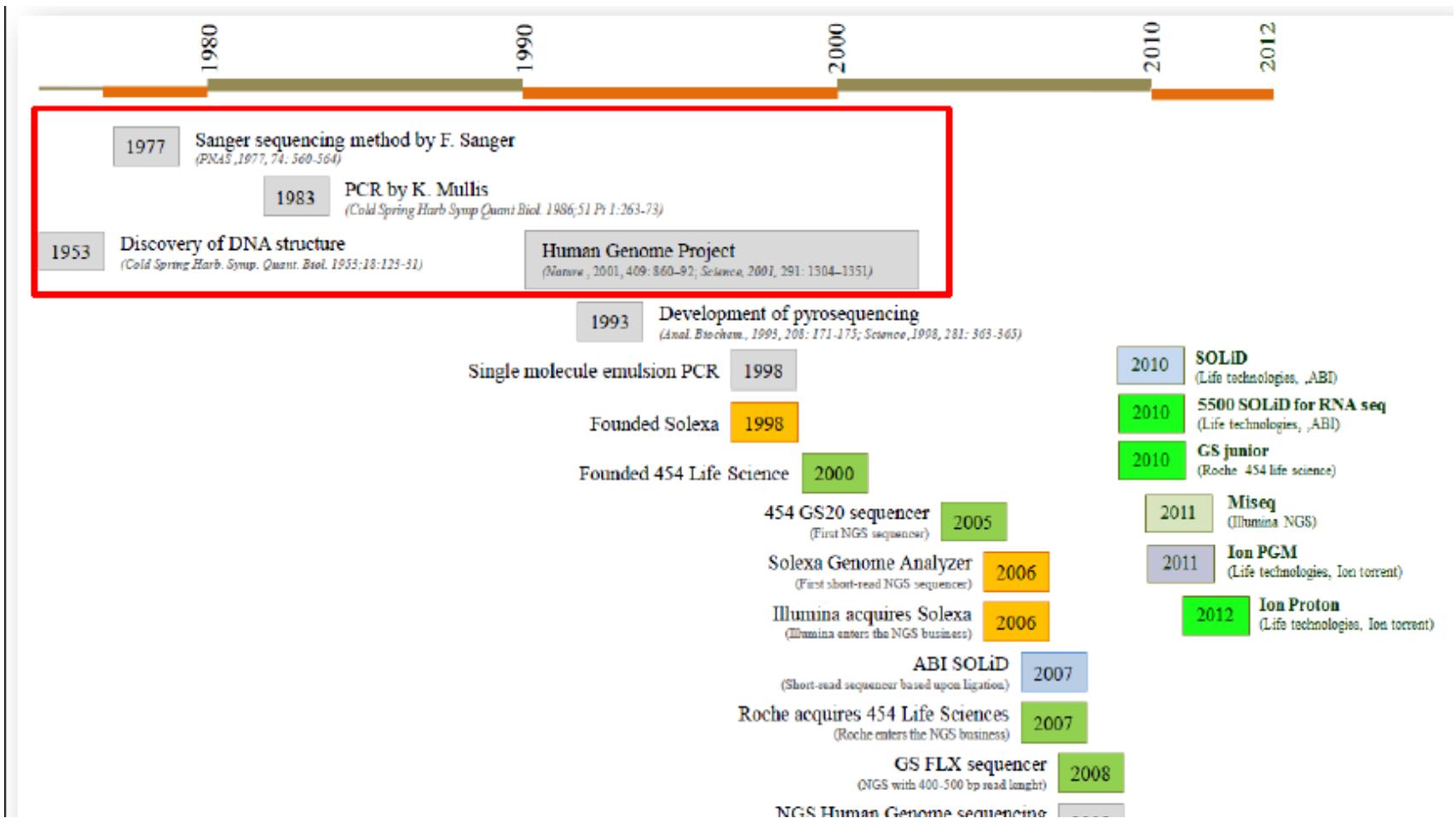
Chimp

98%

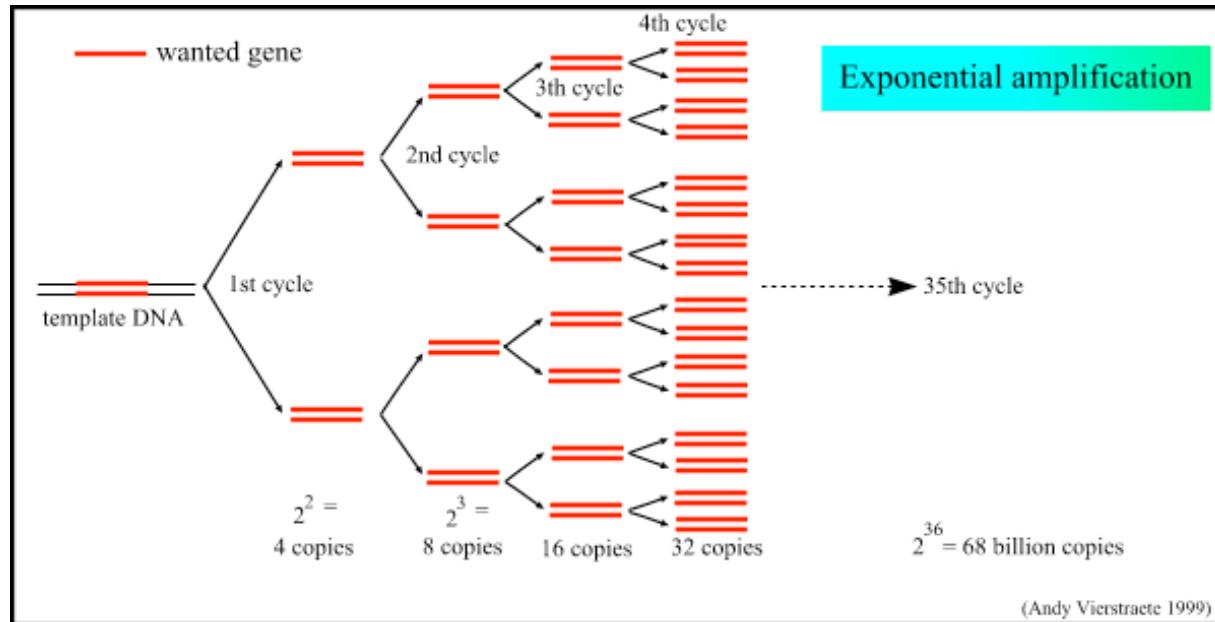
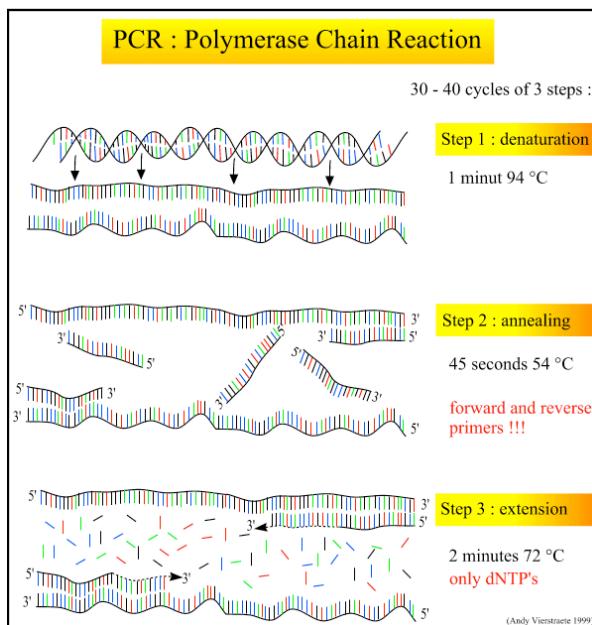


**What percent
of your genes
do you share?**

History of Sequencing Technology

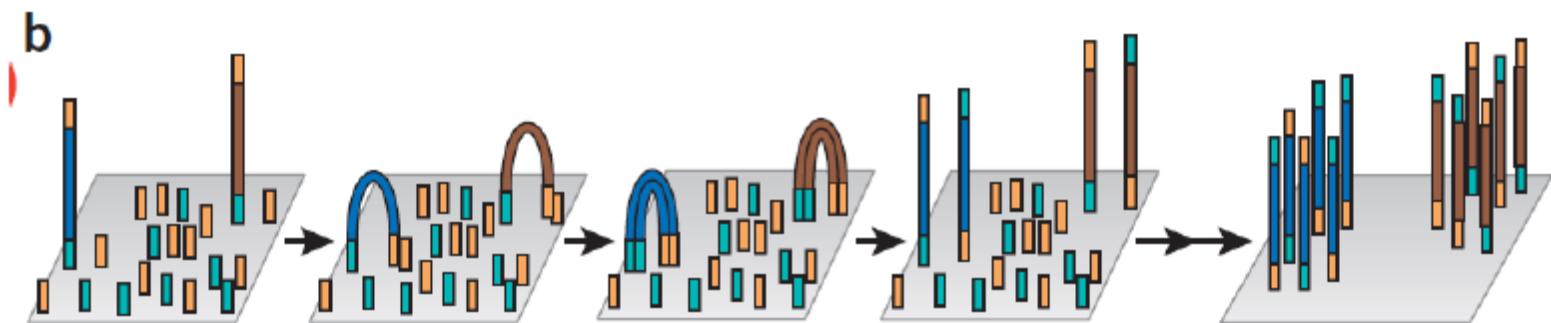


Polymerase Chain Reaction: PCR



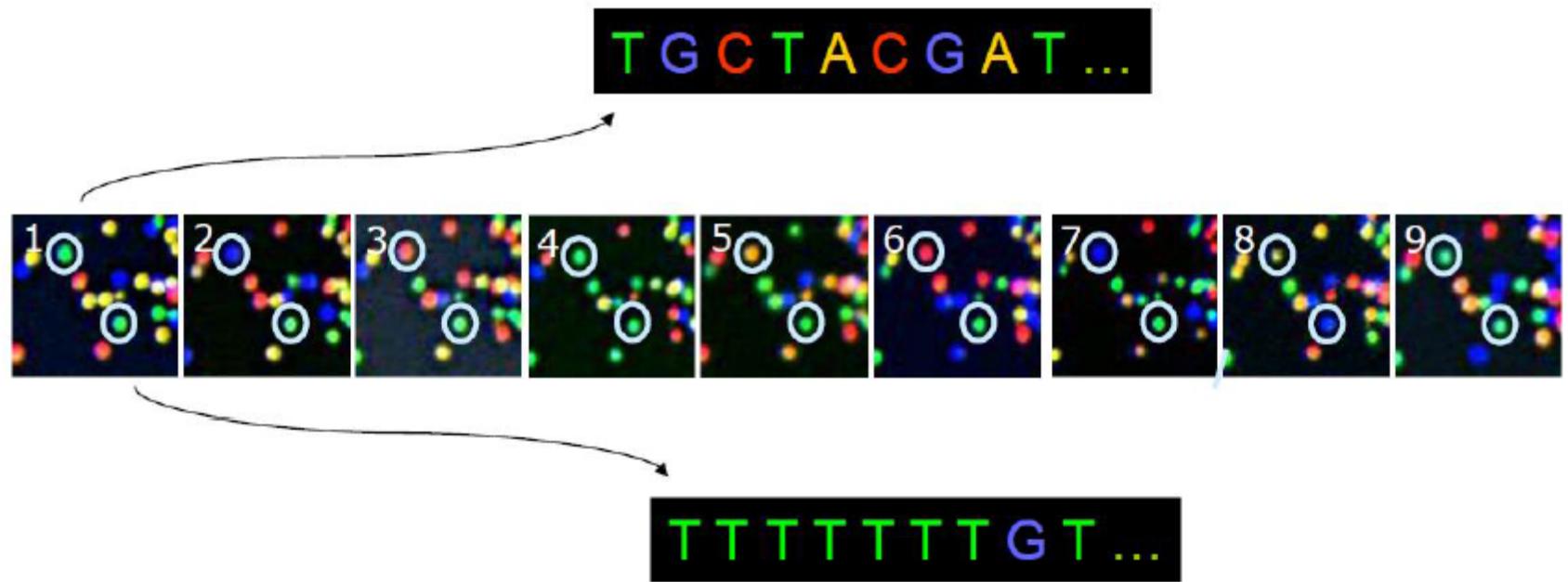
Illumina/Solexa

- Bridge amplification



- DNA fragments are flanked with adaptors.
- A flat surface coated with two types of primers, corresponding to the adaptors.
- Amplification proceeds in cycles, with one end of each

Base Calling from image(Illumina)



NGS platforms

Illumina
(~10days)



HiSeq



Genome Analyzer IIx



MiSeq

600 Gb : 3 billion
total
187 million reads/
lane
 2×1

95 Gb: 320 million total : 7 Gb: 15 million
40 million / lane
 $2 \times 150\text{bp}$

total :
 $2 \times 250\text{bp}$

SOLiD
(~8days)



SOLiD 5500xl



SOLiD 5500



SOLiD 4

Ion Torrent
(1day)

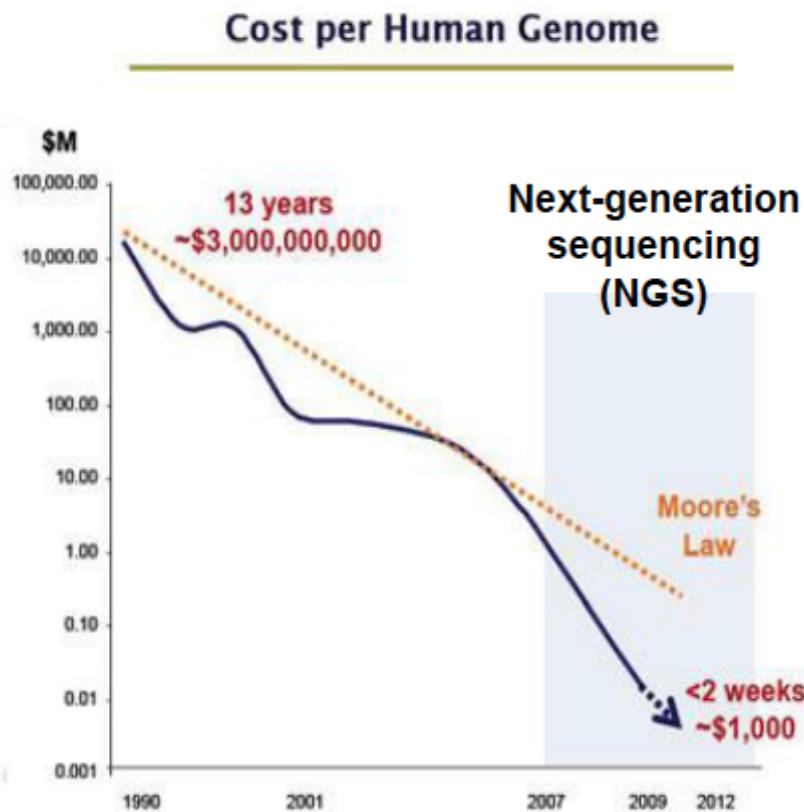


Ion Proton
10-
100Gb



Ion PGM
1Gb-
20Mb

Next-Generation Sequencing



1000 Genomes Project (2008~)
: human genetic variation



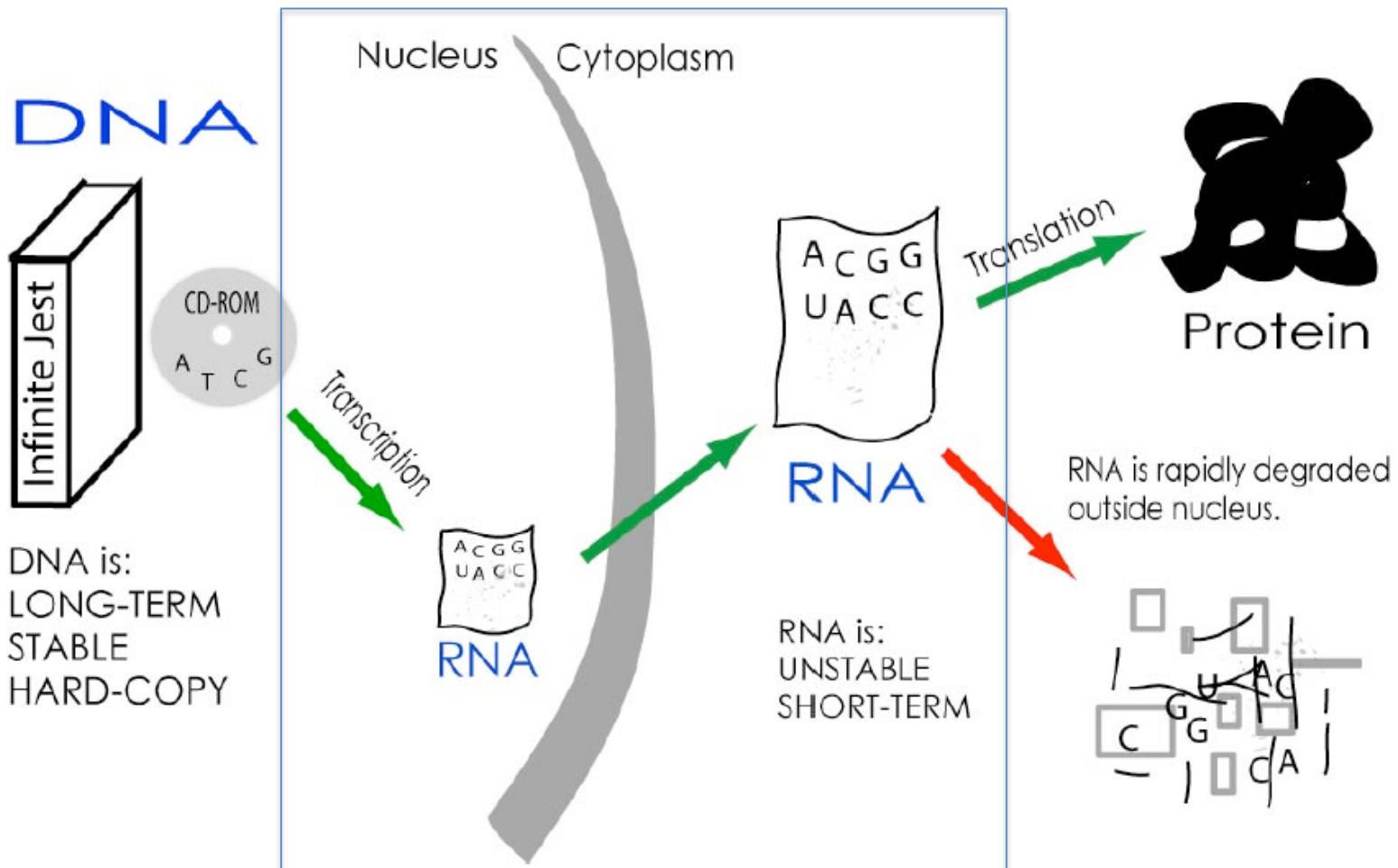
Did You Know There Are (At Least) 20 Applications of Next-generation DNA Sequencing?

Here is the 20 different applications of Next-Gen Sequencing with references from the paper.

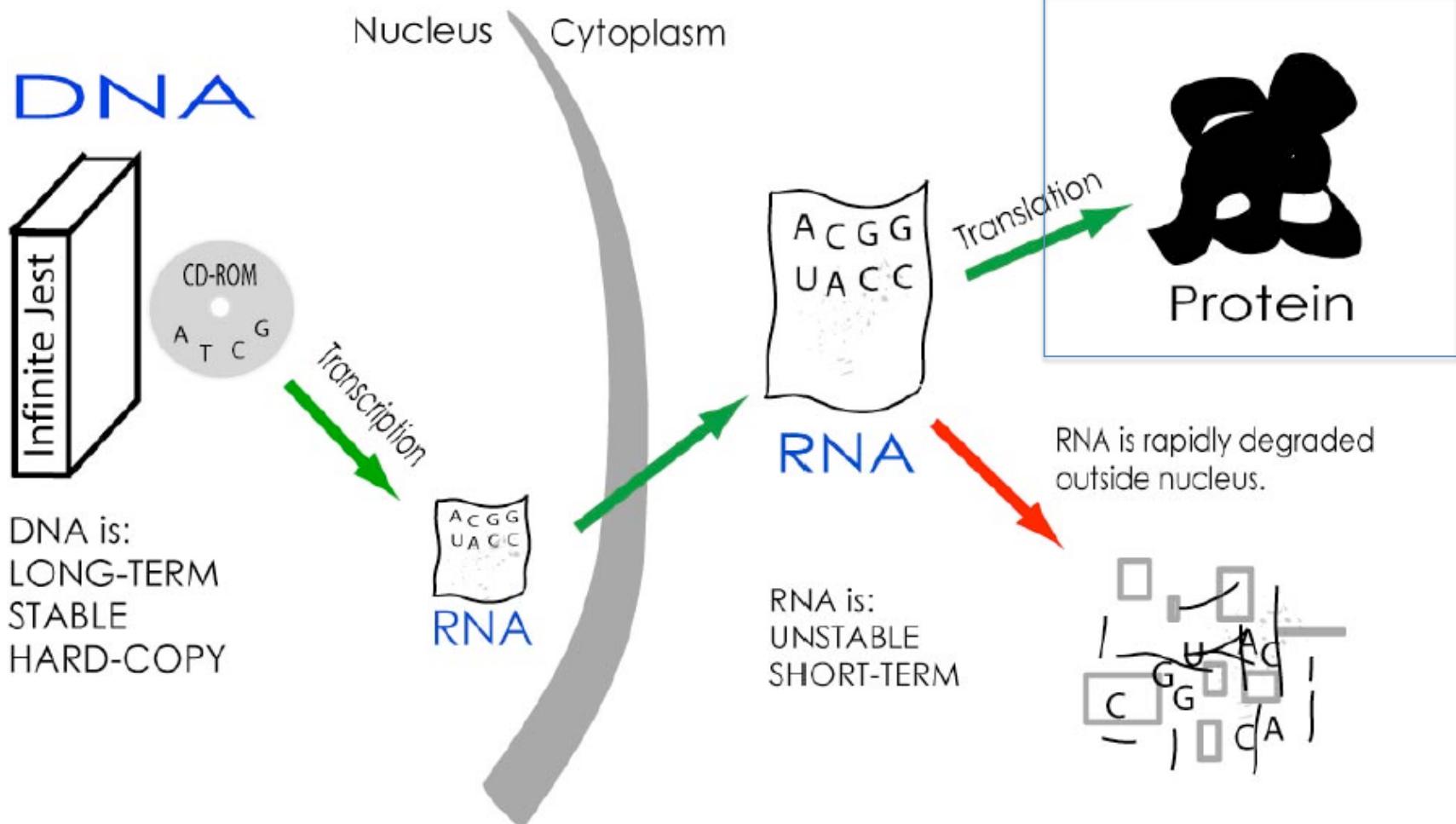
DNA-Seq, Targeted DNA-Seq, Methyl-Seq, Targeted methyl-Seq, DNase-Seq, Sono-Seq and FAIR-E-Seq, MAINE-Seq, ChIP-Seq, RIP-Seq, CLIP-Seq, HITS-CLIP, RNA-Seq, FRT-Seq, NET-Seq (Native Elongating Transcript Sequencing), Hi-C, ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag sequencing), Ribo-Seq, TRAP, PARS (Parallel Analysis of RNA Structure), Synthetic saturation mutagenesis, Immuno-Seq, Deep protein mutagenesis, PhIT-Seq

<http://nextgenseek.com/2012/11/did-you-know-there-are-at-least-20-applications-of-next-generation-dna-sequencing/>

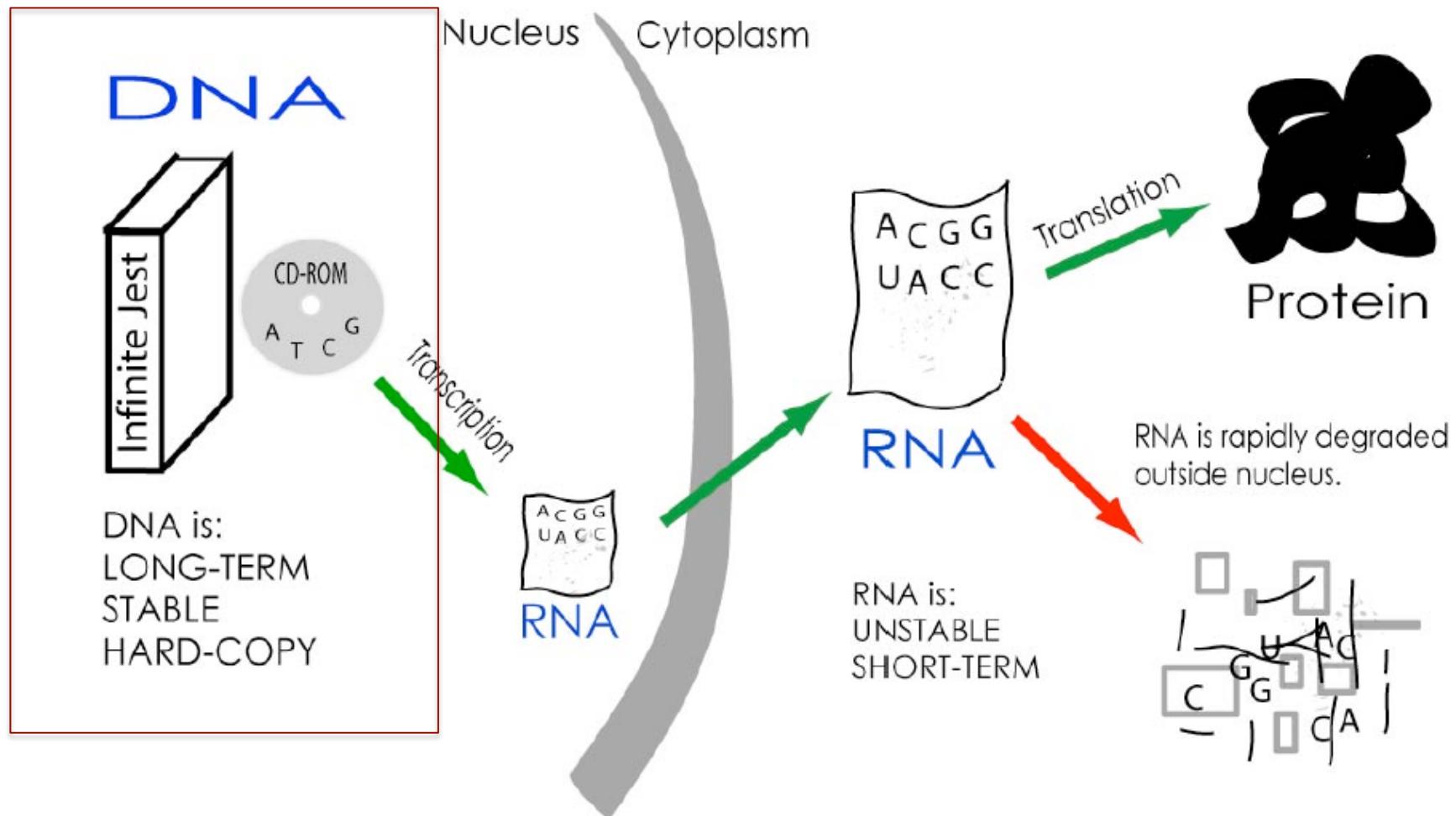
RNA-Seq



Proteomics



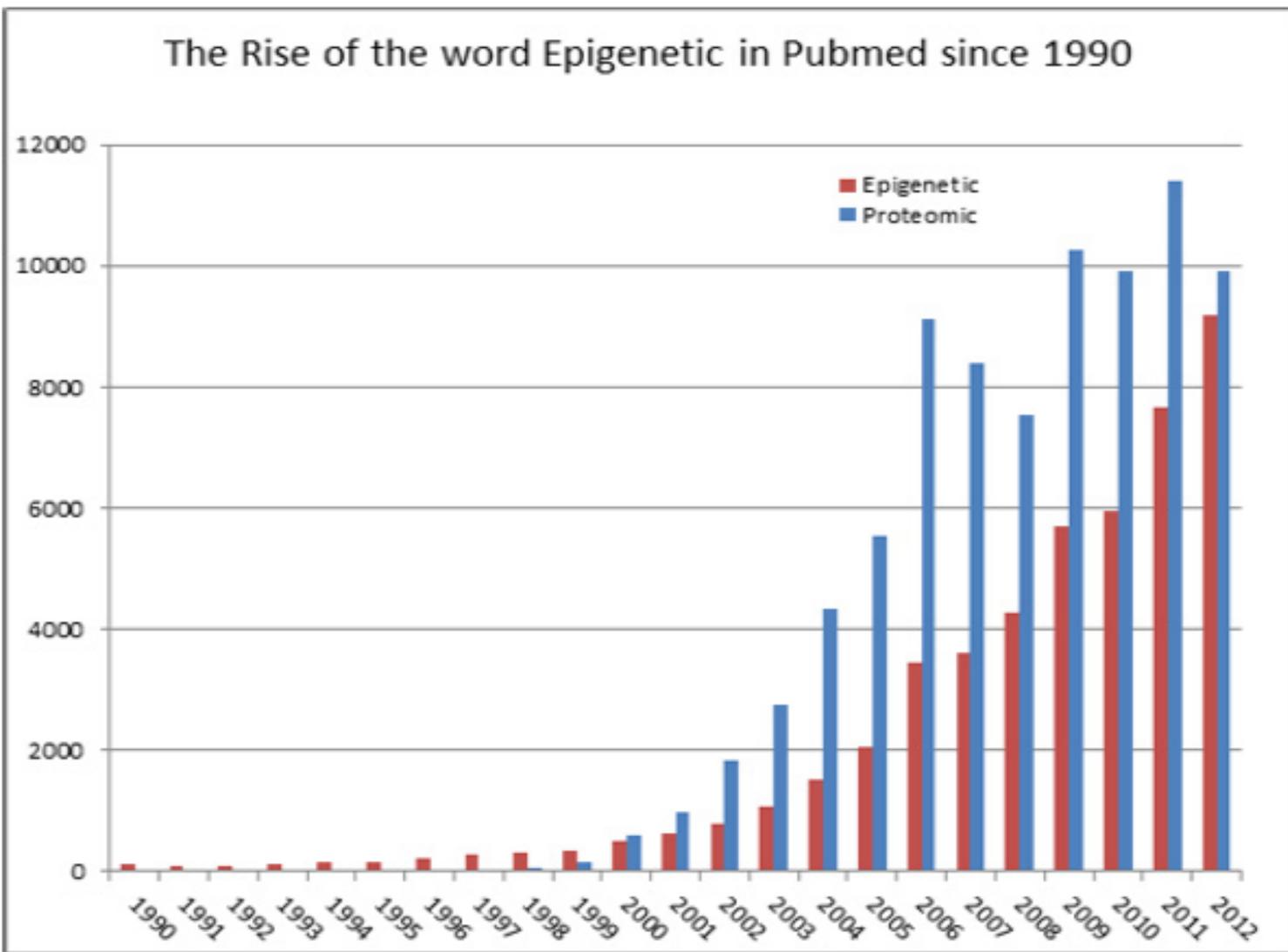
ChIP-Seq



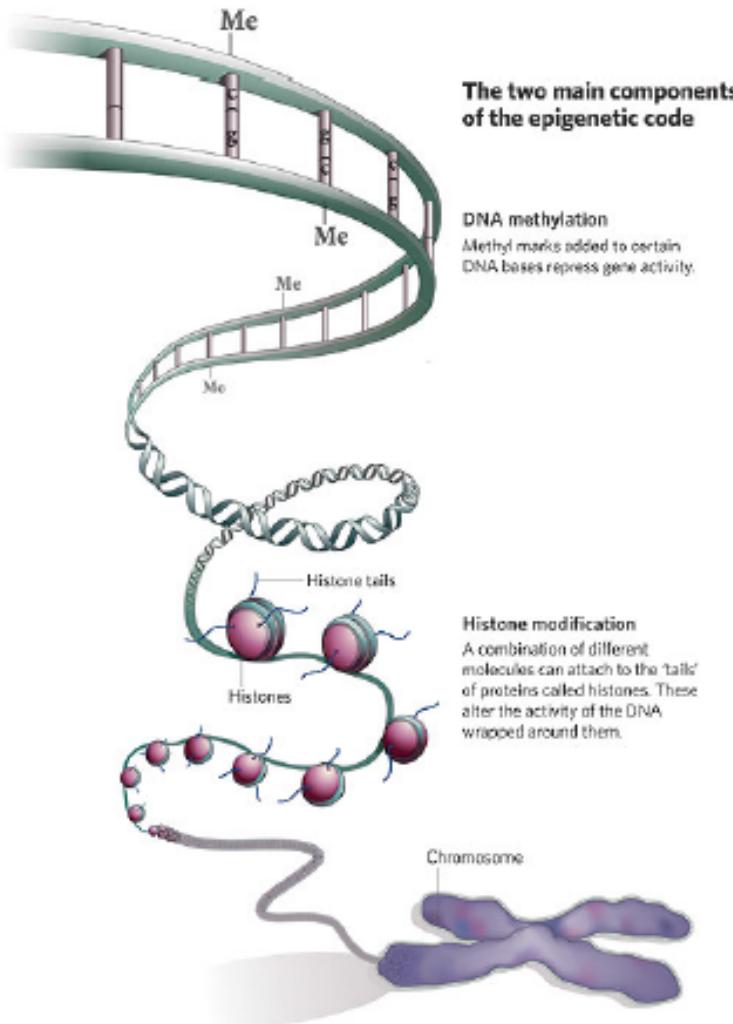
What is epigenetics?

- Greek, epi = above, upon - genetics
- The study of heritable changes in gene function that occur without a change in the DNA sequence.
- the study of changes in gene silencing that occur without changes in the genes themselves. Many genes in the body are permanently turned off as part of normal development. But sometimes that process goes awry, turning off genes that should otherwise remain active. ...
- **Epigenome** represents all epigenetic phenomenon across the Genome

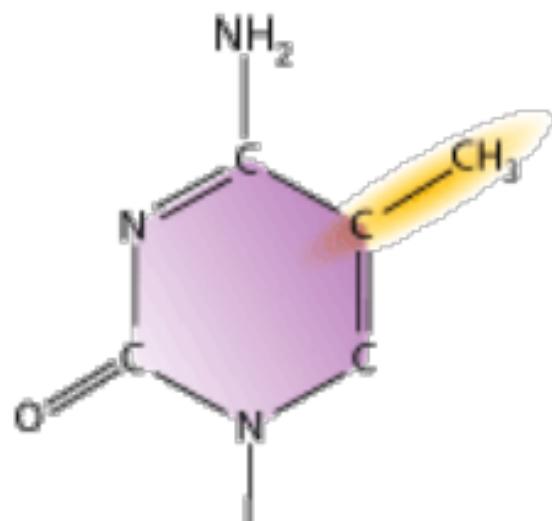
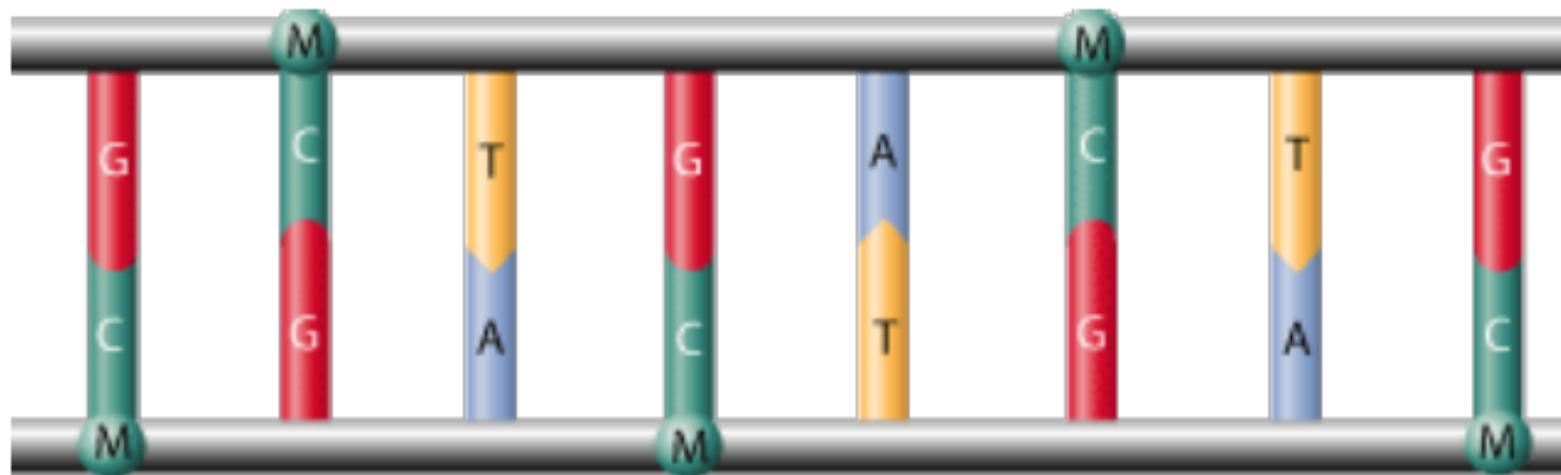
Epigenetics publications



Epigenetics

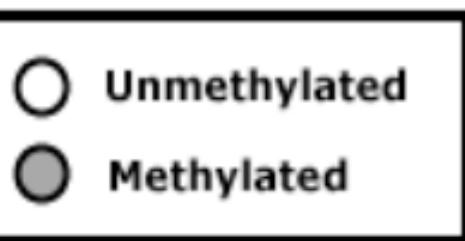


DNA methylation

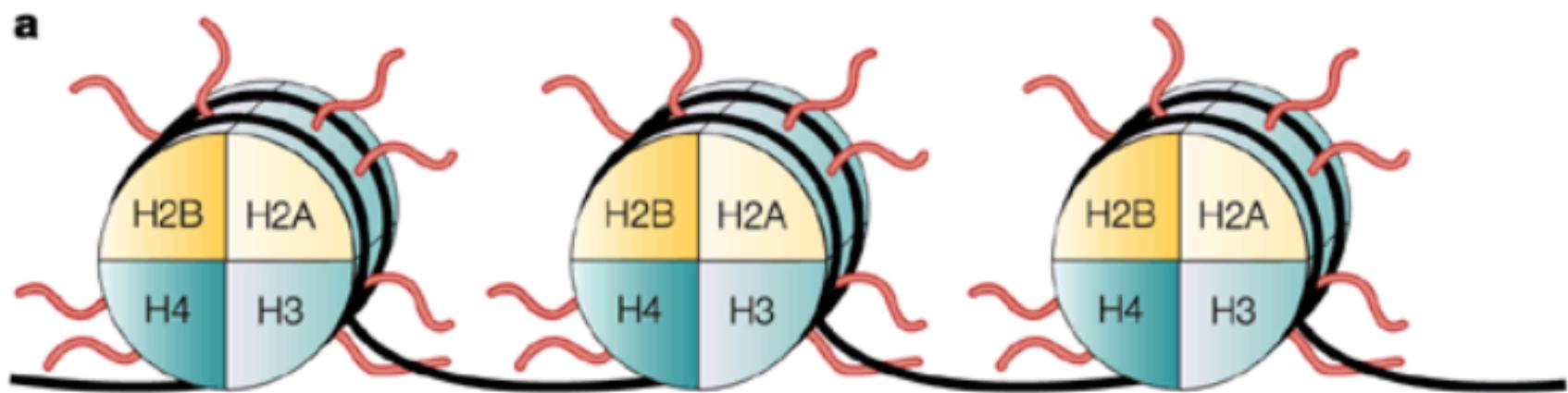


DNA methylation is the addition of a methyl group (M) to the DNA base cytosine (C).

Methylation of CpG islands



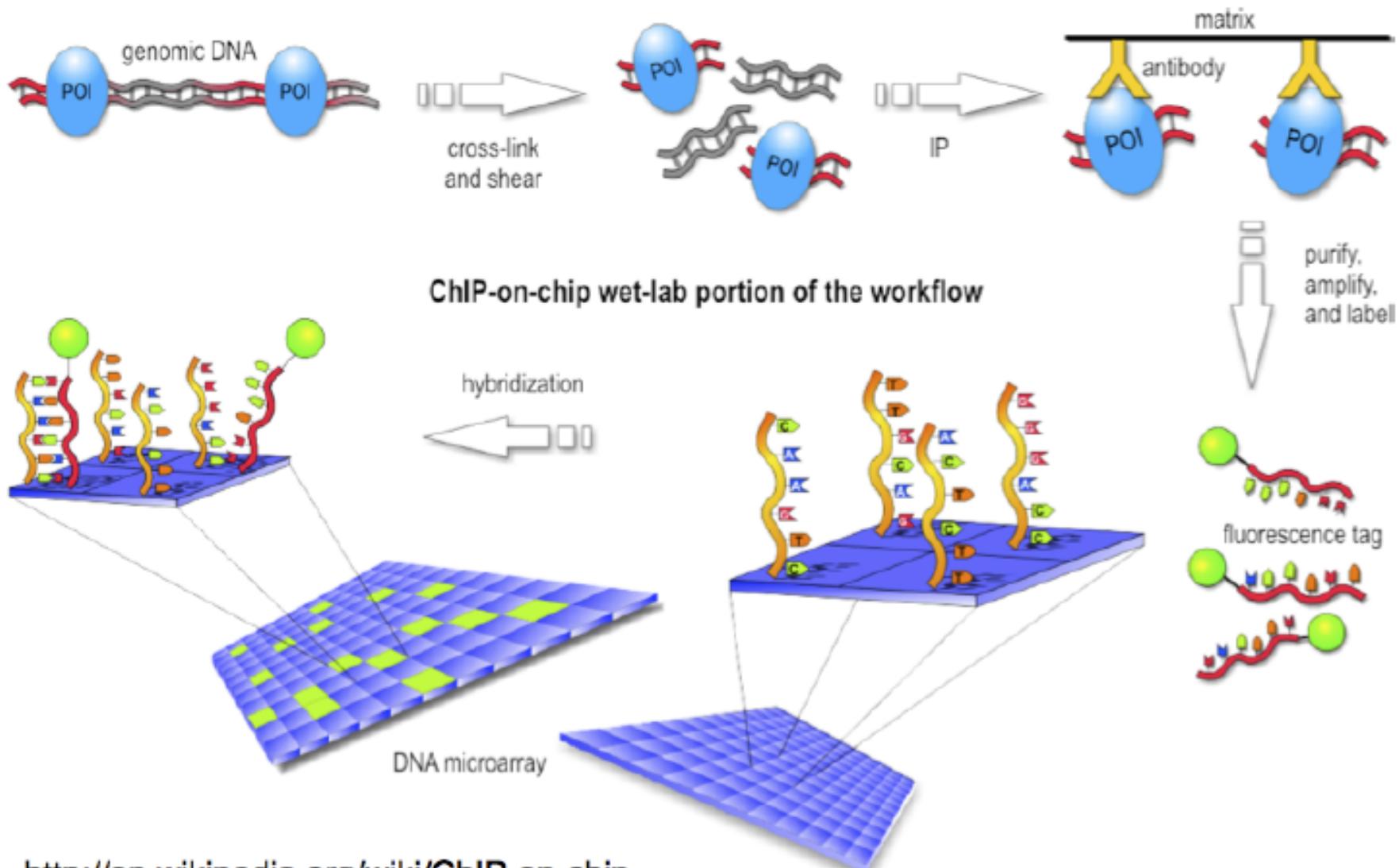
Histone modifications



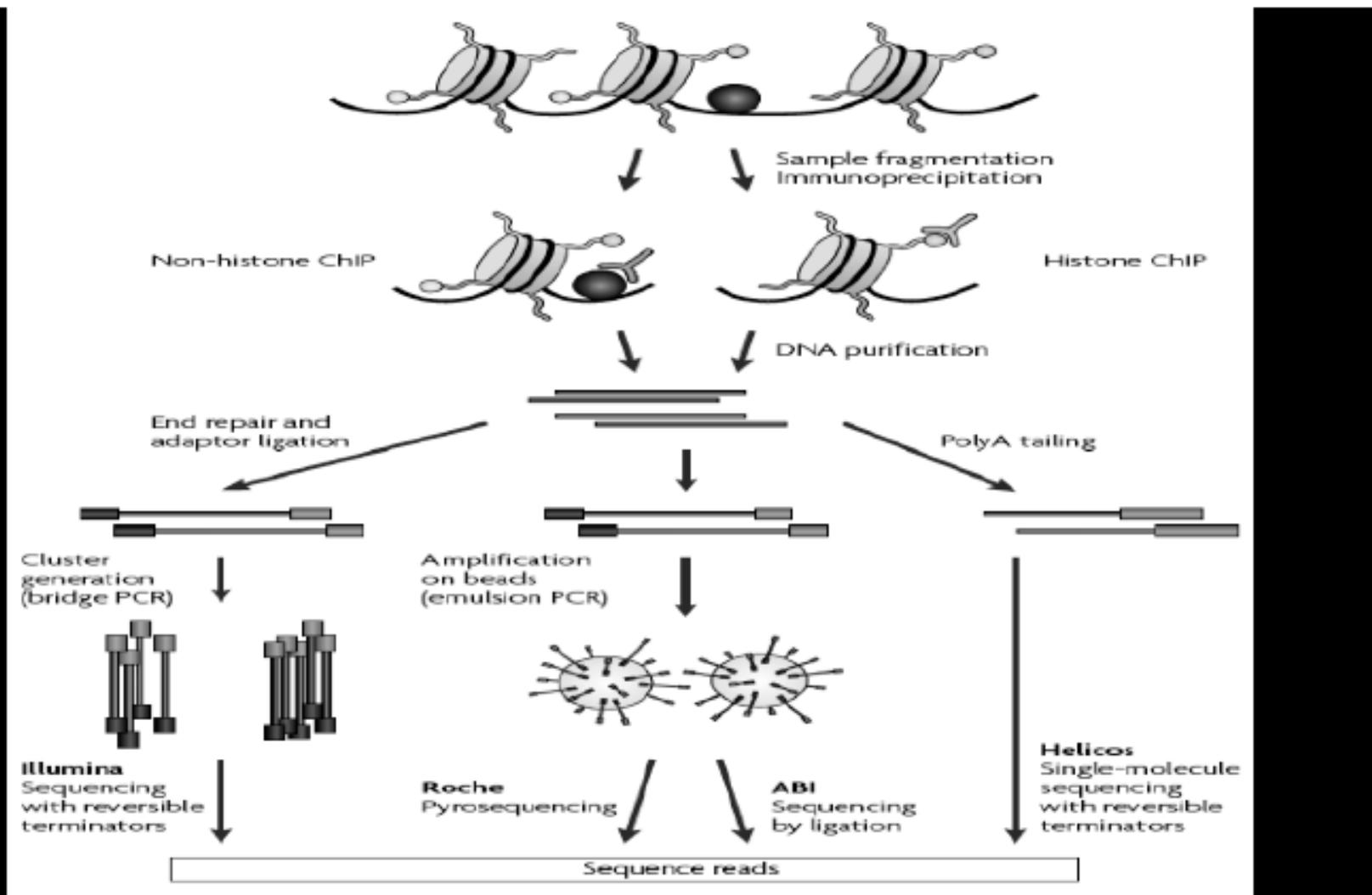
Interplay between histone marks

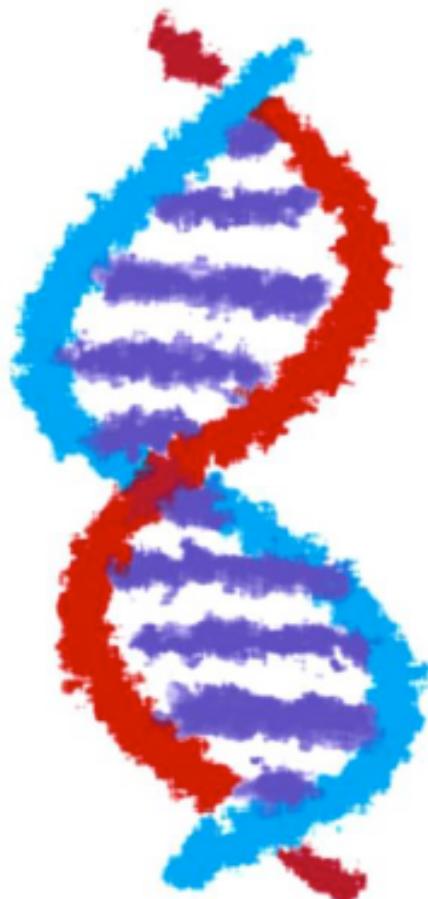


chip-chip

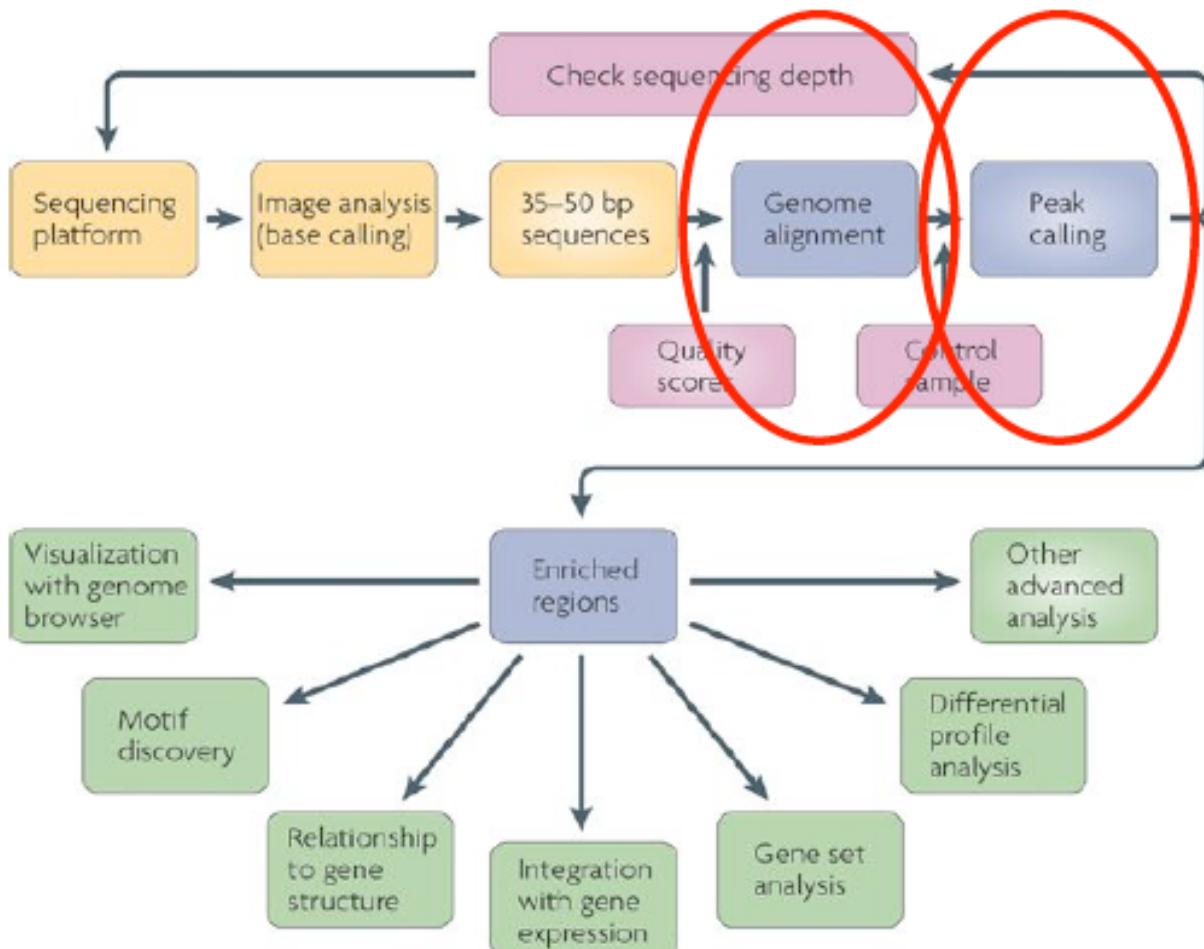


chip-seq

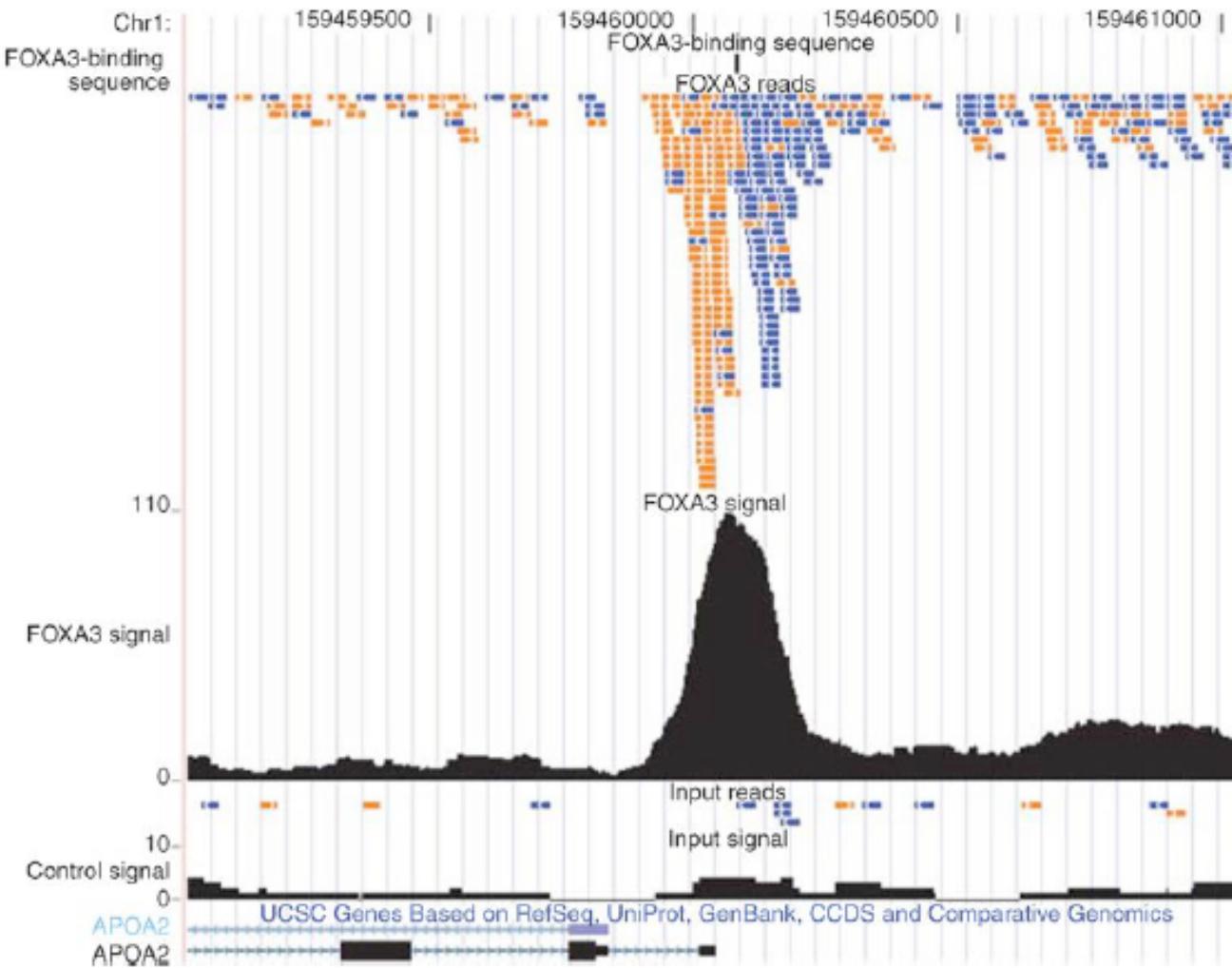




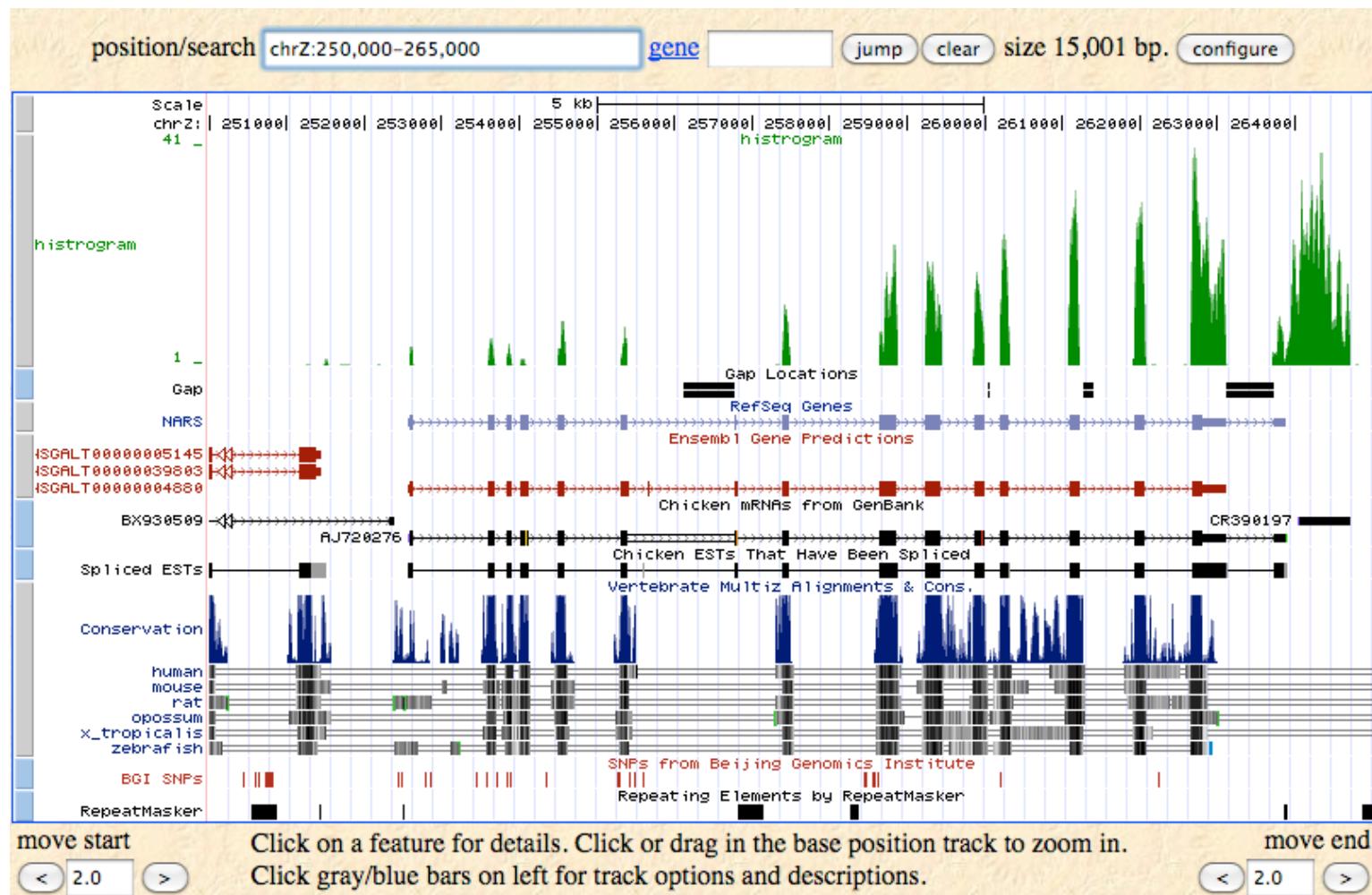
The ChIP-Seq pipeline



Peak finding



UCSC Genome Browser



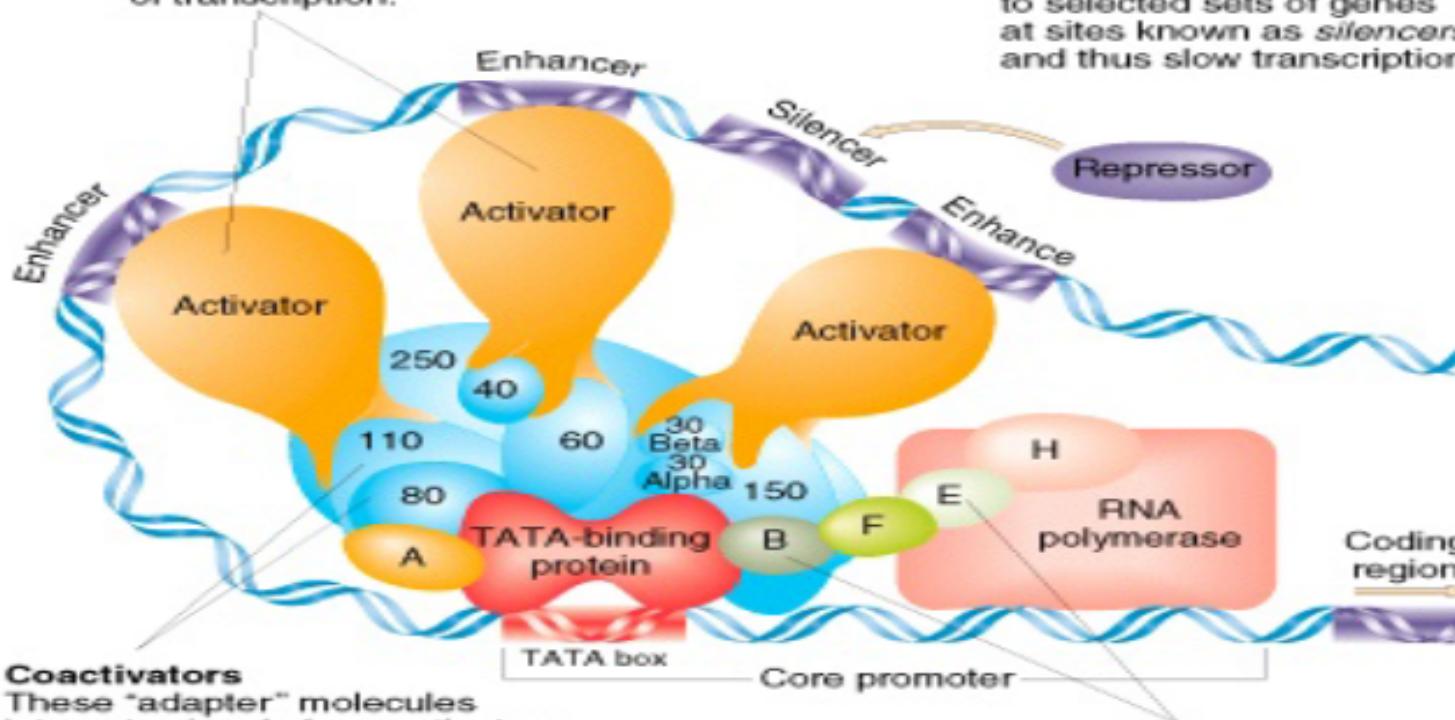
Promoters, enhancers, silencers etc.

Activators

These proteins bind to genes at sites known as *enhancers* and speed the rate of transcription.

Repressors

These proteins bind to selected sets of genes at sites known as *silencers* and thus slow transcription.



Coactivators

These "adapter" molecules integrate signals from activators and perhaps repressors.

Basal transcription factors

In response to injunctions from activators, these factors position RNA polymerase at the start of transcription and initiate the transcription process.

Example of ChIP-Seq data

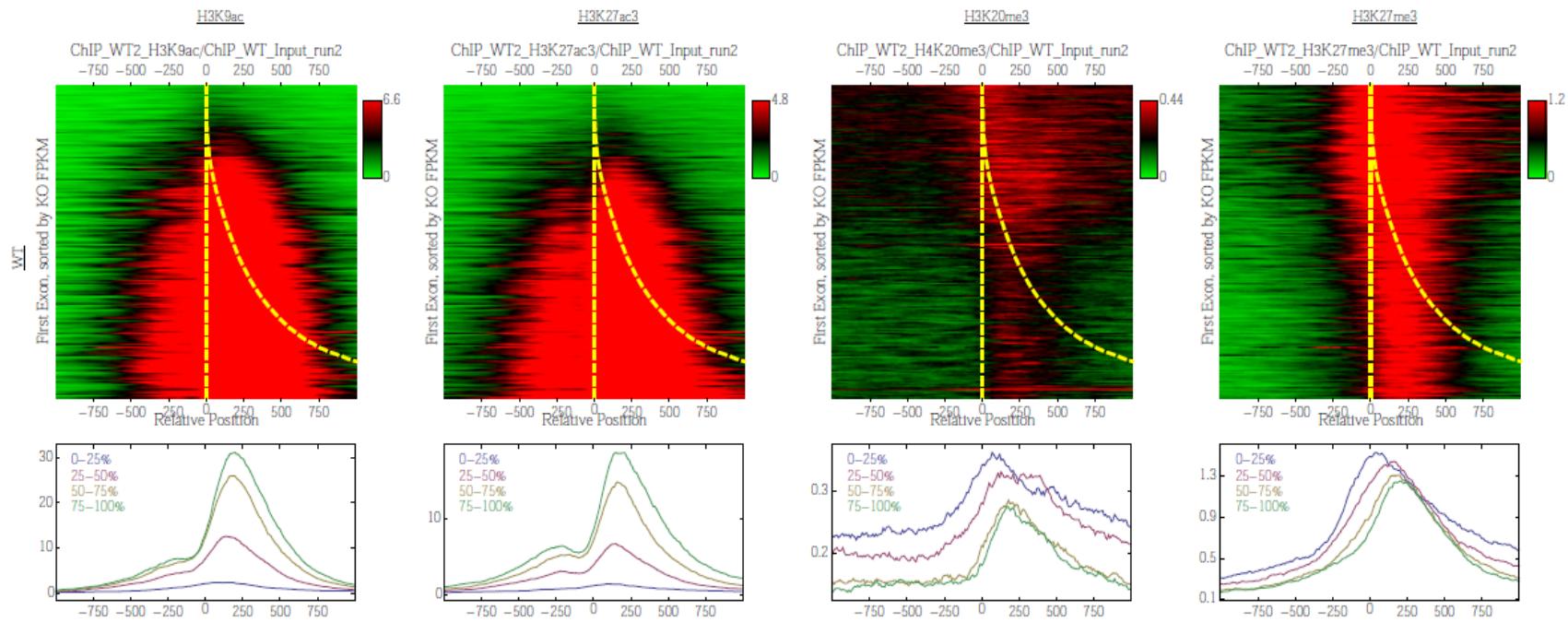
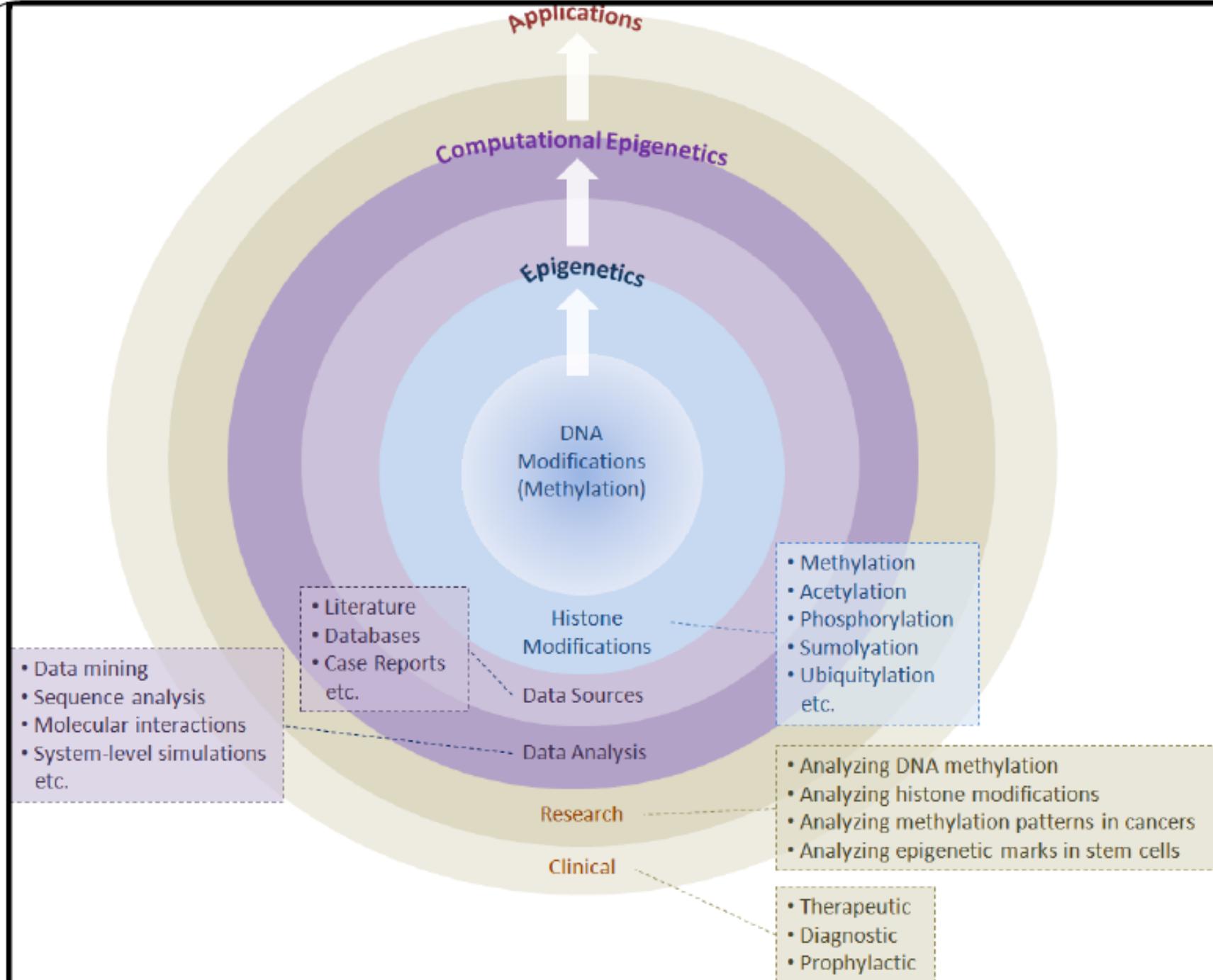


Table 1 | Publicly available ChIP-seq software packages discussed in this review

								Artifact filtering: strand-based/ duplicate ^a	Refs.
	Profile	Peak criteria ^a	Tag shift	Control data ^b	Rank by	FDR ^c	User input parameters ^d		
CisGenome v1.1	Strand-specific window scan	1: Number of reads in window 2: Number of ChIP reads minus control reads in window	Average for highest ranking peak pairs	Conditional binomial used to estimate FDR	Number of reads under peak	1: Negative binomial 2: conditional binomial	Target FDR, optional window width, window interval	Yes / Yes	10
ERANGE v3.1	Tag aggregation	1: Height cutoff High quality peak estimate, per-region estimate, or input	High quality peak estimate, per-region estimate, or input	Used to calculate fold enrichment and optionally P values	P value	1: None 2: # control # ChIP	Optional peak height, ratio to background	Yes / No	4, 18
FindPeaks v3.1.9.2	Aggregation of overlapped tags	Height threshold	Input or estimated	NA	Number of reads under peak	1: Monte Carlo simulation 2: NA	Minimum peak height, subpeak valley depth	Yes / Yes	19
F-Seq v1.82	Kernel density estimation (KDE)	s.s.d. above KDE for 1: random background, 2: control	Input or estimated	KDE for local background	Peak height	1: None 2: None	Threshold s.d. value, KDE bandwidth	No / No	14
GLITR	Aggregation of overlapped tags	Classification by height and relative enrichment	User input tag extension	Multiply sampled to estimate background class values	Peak height and fold enrichment	2: # control # ChIP	Target FDR, number nearest neighbors for clustering	No / No	17
MACS v1.3.5	Tags shifted then window scan	Local region Poisson P value	Estimate from high quality peak pairs	Used for Poisson fit when available	P value	1: None 2: # control # ChIP	P-value threshold, tag length, mfold for shift estimate	No / Yes	13
PeakSeq	Extended tag aggregation	Local region binomial P value	Input tag extension length	Used for significance of sample enrichment with binomial distribution	q value	1: Poisson background assumption 2: From binomial for sample plus control	Target FDR	No / No	5
QuEST v2.3	Kernel density estimation	2: Height threshold, background ratio	Mode of local shifts that maximize strand cross-correlation	KDE for enrichment and empirical FDR estimation	q value	1: NA 2: # control # ChIP as a function of profile threshold	KDE bandwidth, peak height, subpeak valley depth, ratio to background	Yes / Yes	9
SICER v1.02	Window scan with gaps allowed	P value from random background model, enrichment relative to control	Input	Linearly rescaled for candidate peak rejection and p values	q value	1: None 2: From Poisson P values	Window length, gap size, FDR (with control) or E-value (no control)	No / Yes	15
SiSSRs v1.4	Window scan	$N_+ - N_-$ sign change, N_+ + N_- threshold in region ^f	Average nearest paired tag distance	Used to compute fold-enrichment distribution	P value	1: Poisson 2: control distribution	1: FDR 1, 2: N_+ + N_- threshold	Yes / Yes	11
spp v1.0	Strand specific window scan	Poisson P value (paired peaks only)	Maximal strand cross-correlation	Subtracted before peak calling	P value	1: Monte Carlo simulation 2: # control # ChIP	Ratio to background	Yes / No	12
USeq v4.2	Window scan	Binomial P value	Estimated or user specified	Subtracted before peak calling	q value	1, 2: binomial 2: # control # ChIP	Target FDR	No / Yes	20



감사합니다.