ChIP-chip/seq/PET

胡勇

专业：生物化学与分子生物学

11-22-2011

Background

- After cloning of lots of genes that encoding TFs, what will we do next?

- What's the function of the non-protein coding sequences in the genome?

  23,000 protein-coding genes, only about 1.5% of the genome codes for proteins

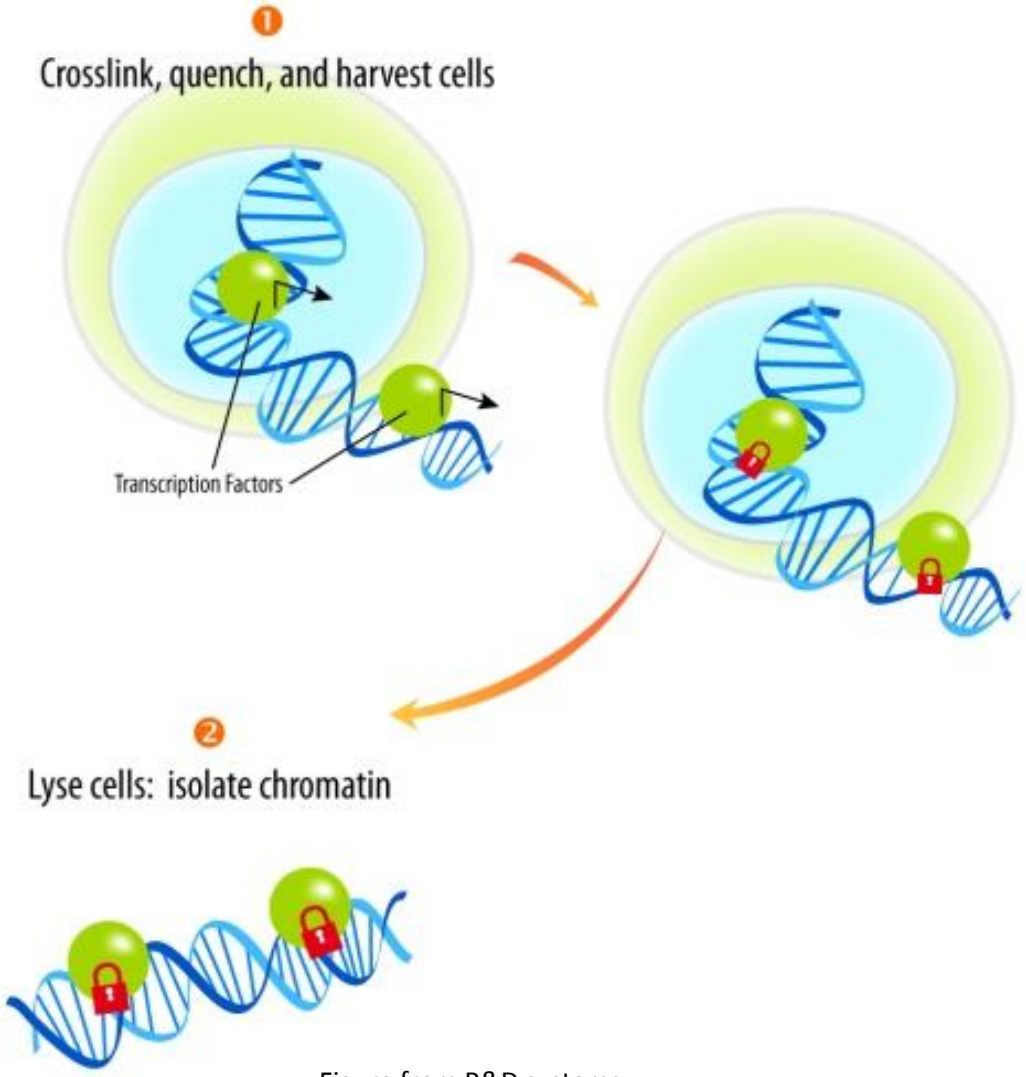
- Epigenetic marks is essential for transcription regulation.

ChIP-chip/seq/PET description

- all based on ChIP

- techniques for genome-wide profiling of DNA-binding proteins, histone modifications or nucleosomes

ChIP: Chromatin immunoprecipitation



a common method for detecting interactions between a protein and a DNA sequence in vivo

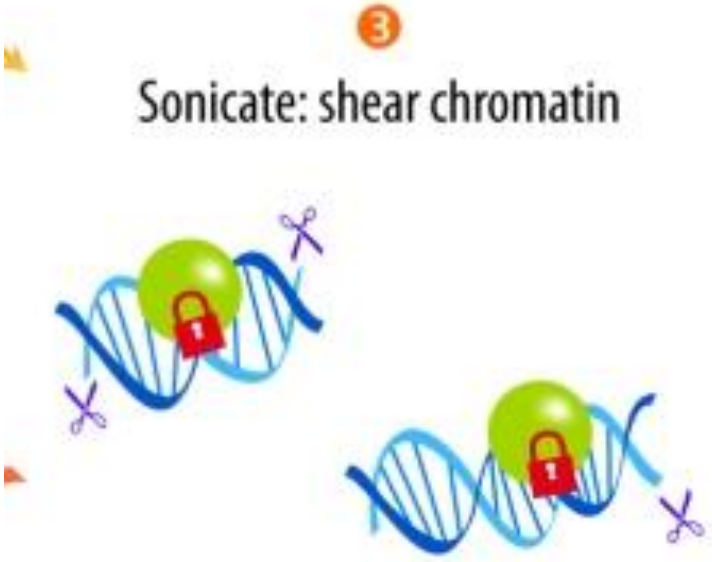
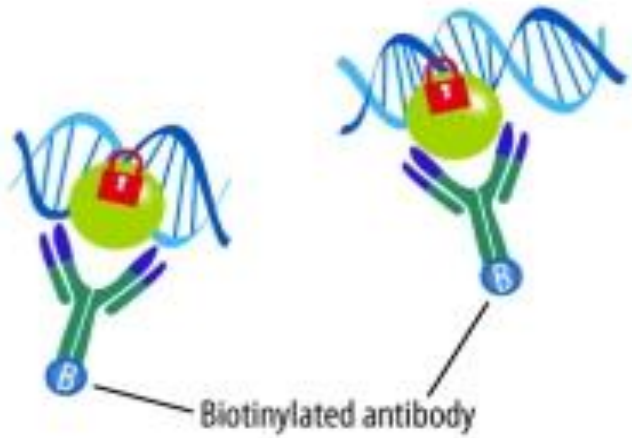


Figure from R&D systems

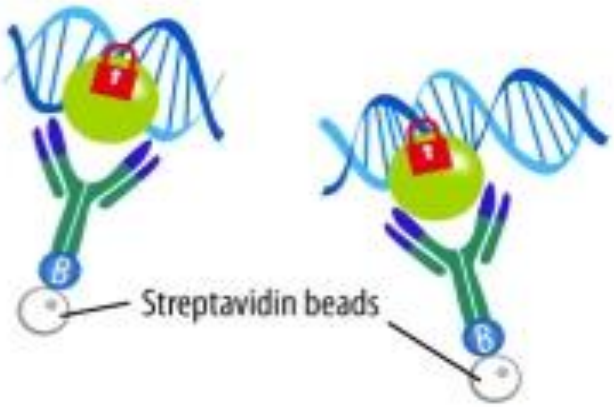
4

Immunoprecipitate target protein/DNA complexes



5

Capture immunocomplexes with streptavidin-conjugated magnetic or agarose beads



6

Purify DNA using chelating resin

Figures from R&D systems

ChIP
chromatin immunoprecipitation

+

DNA microarray

ChIP-chip

next generation sequencing

=

ChIP-seq

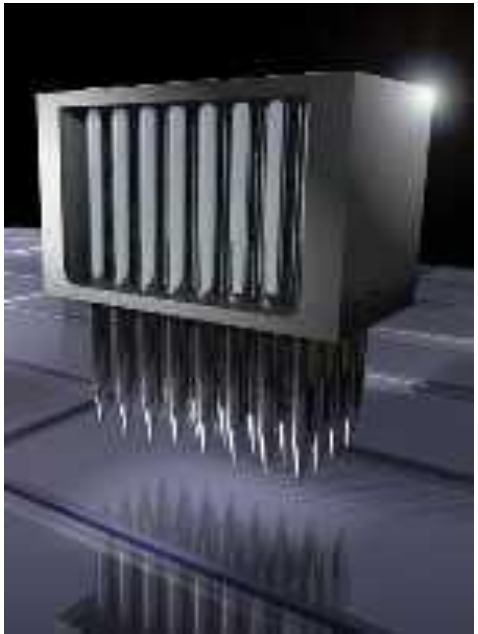
Paired-end tags sequencing

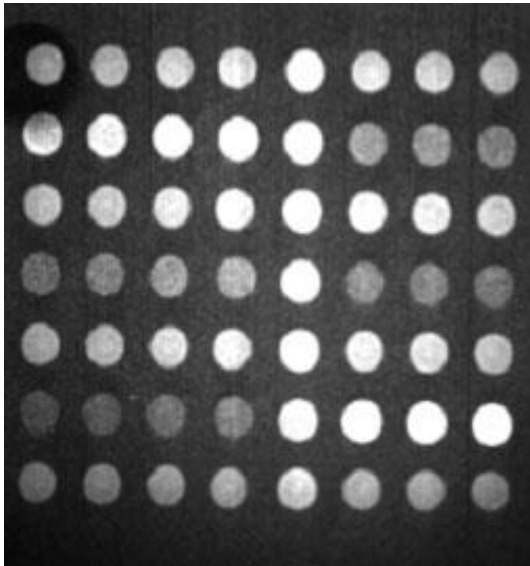
ChIP-PET

ChIP-chip

a. chip

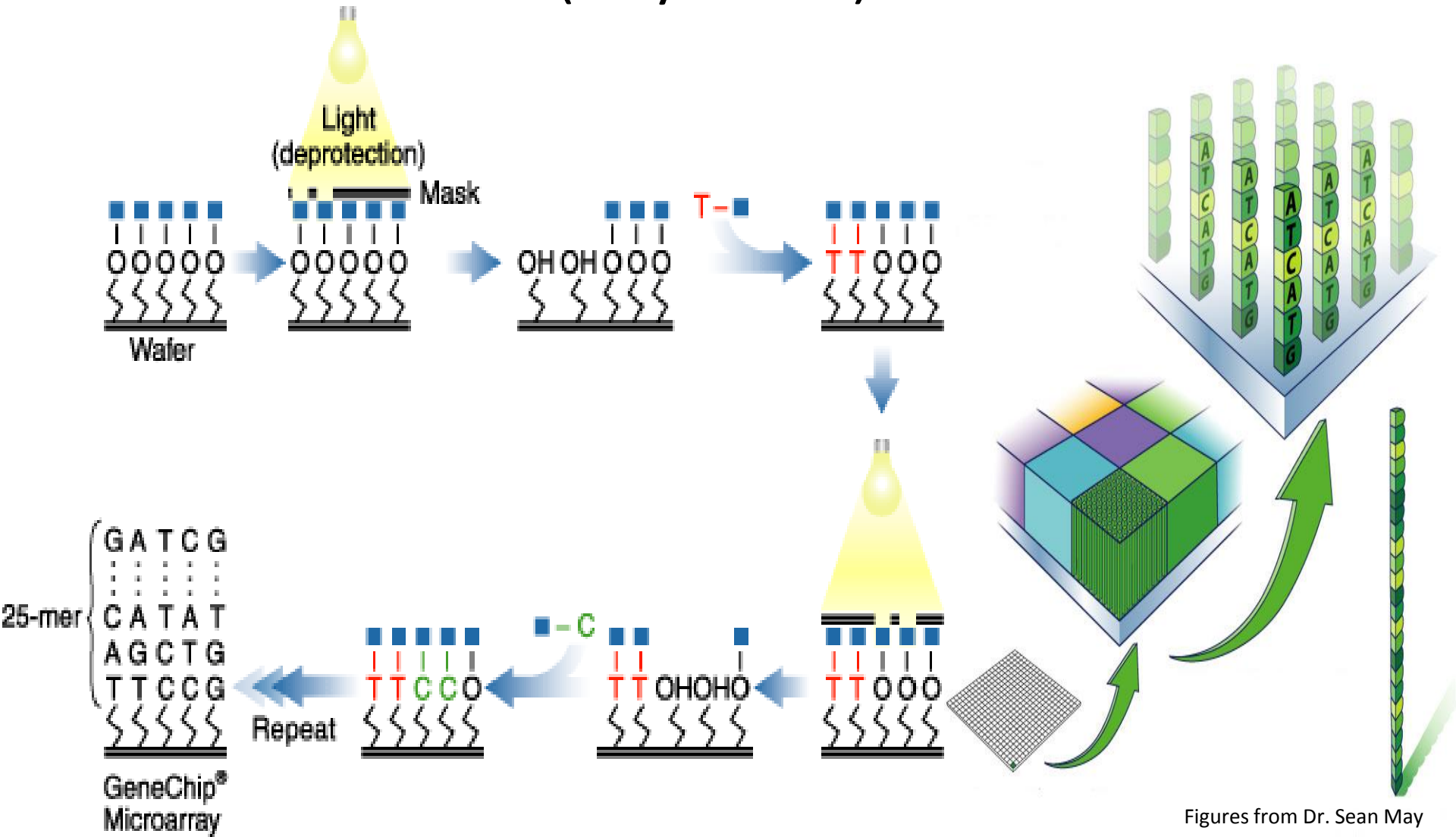
- cDNA probes (> 200 nt), usually produced by PCR, attached to either nylon or glass supports





Figures from Dr. Sean May

- Oligonucleotides (25-30 nt) synthesized in situ on silica wafers (Affymetrix)



Figures from Dr. Sean May

b. ChIP-chip outline

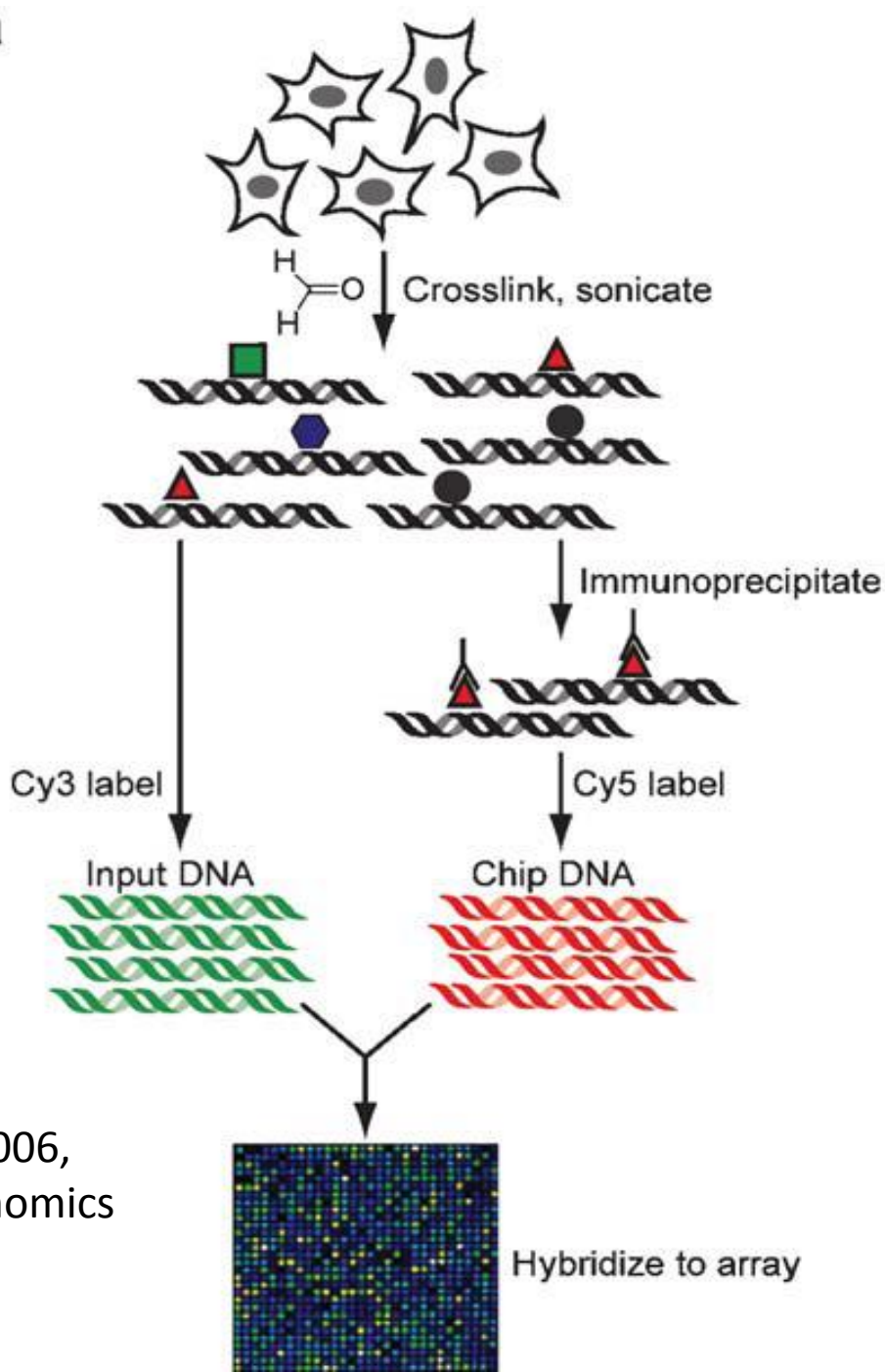
ChIP

PCR

Fluorescently labeling

Hybridization to chip

Scan



Kim and Ren 2006,
Annu. Rev. Genomics
Hum. Genet.

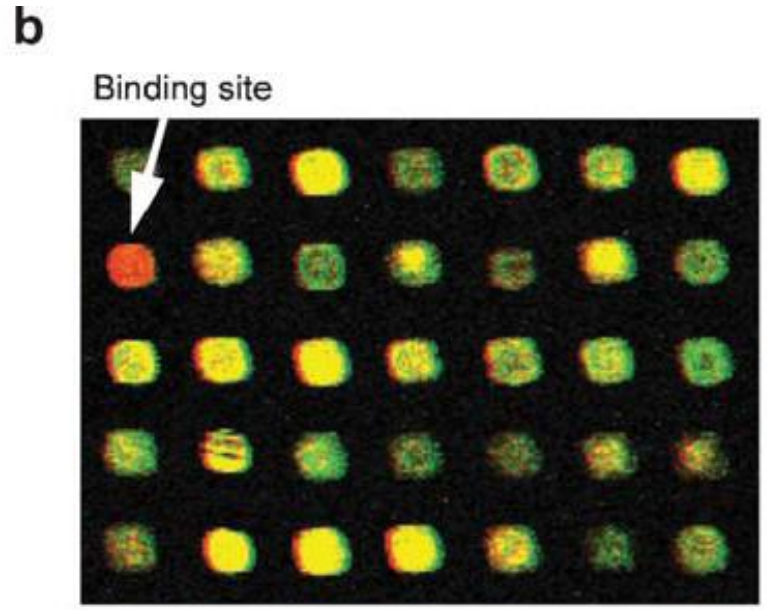
briefly

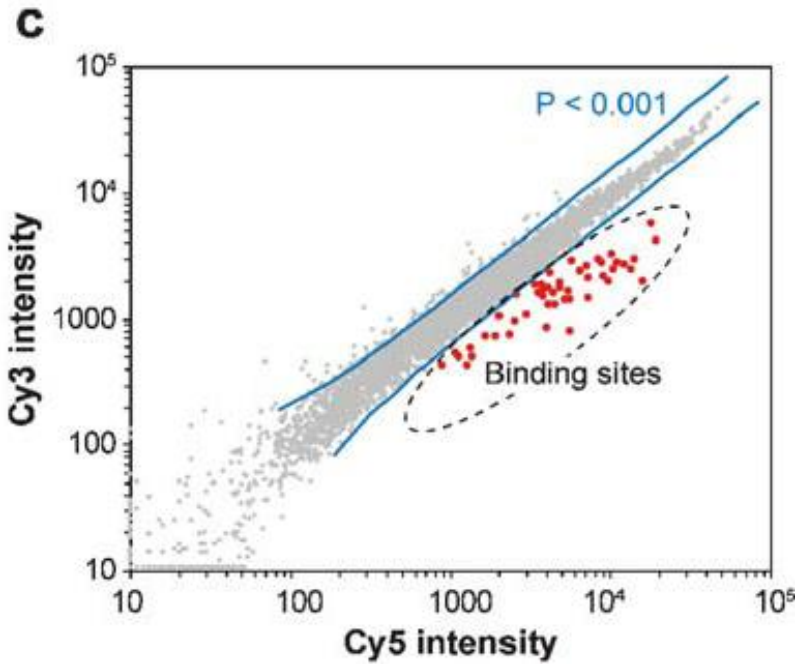
- Red: cy5>cy3

  binding sites
- Yellow: cy5=cy3

  non-specific DNA
- Green: cy5<cy3

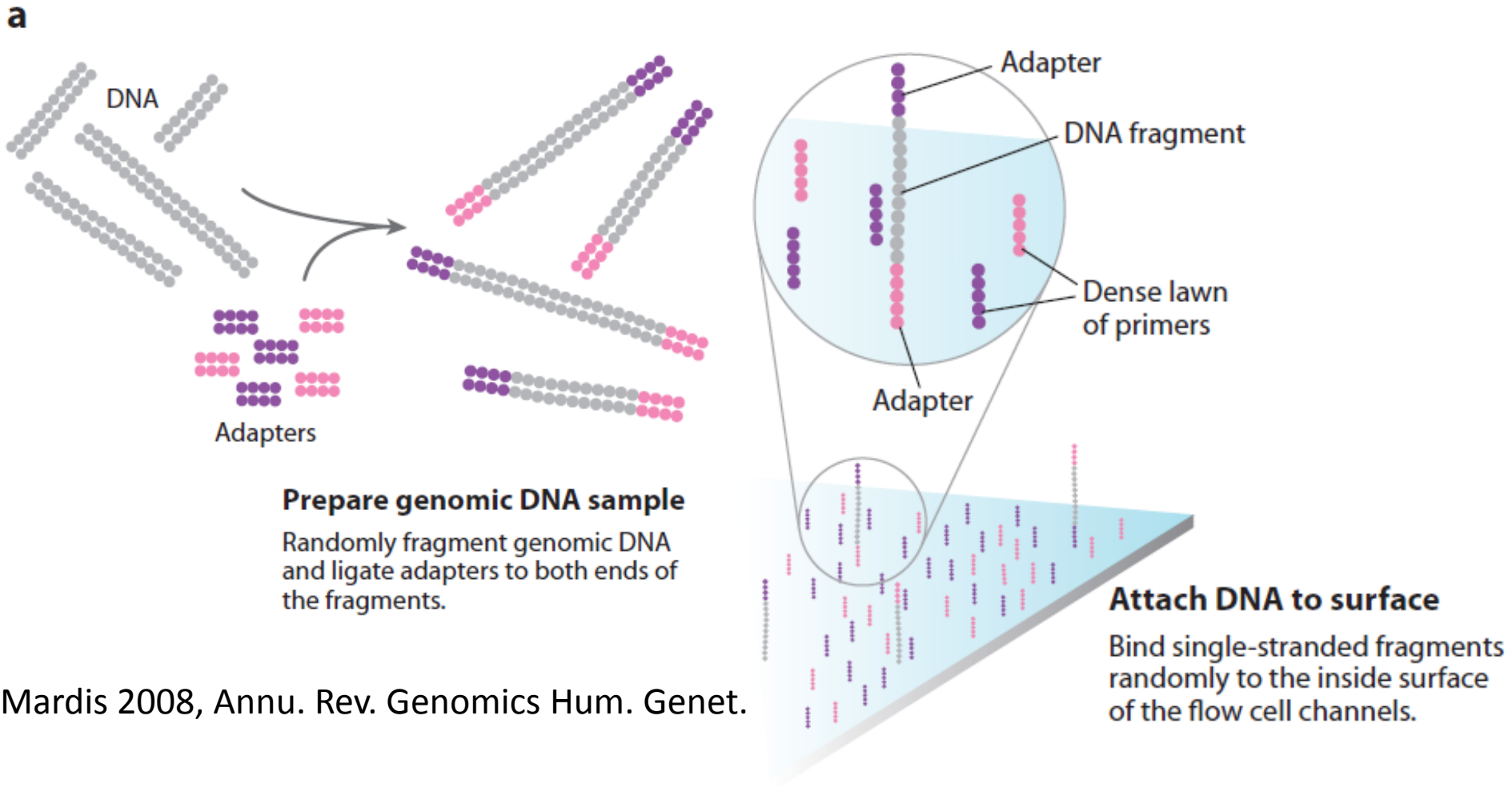
  input DNA



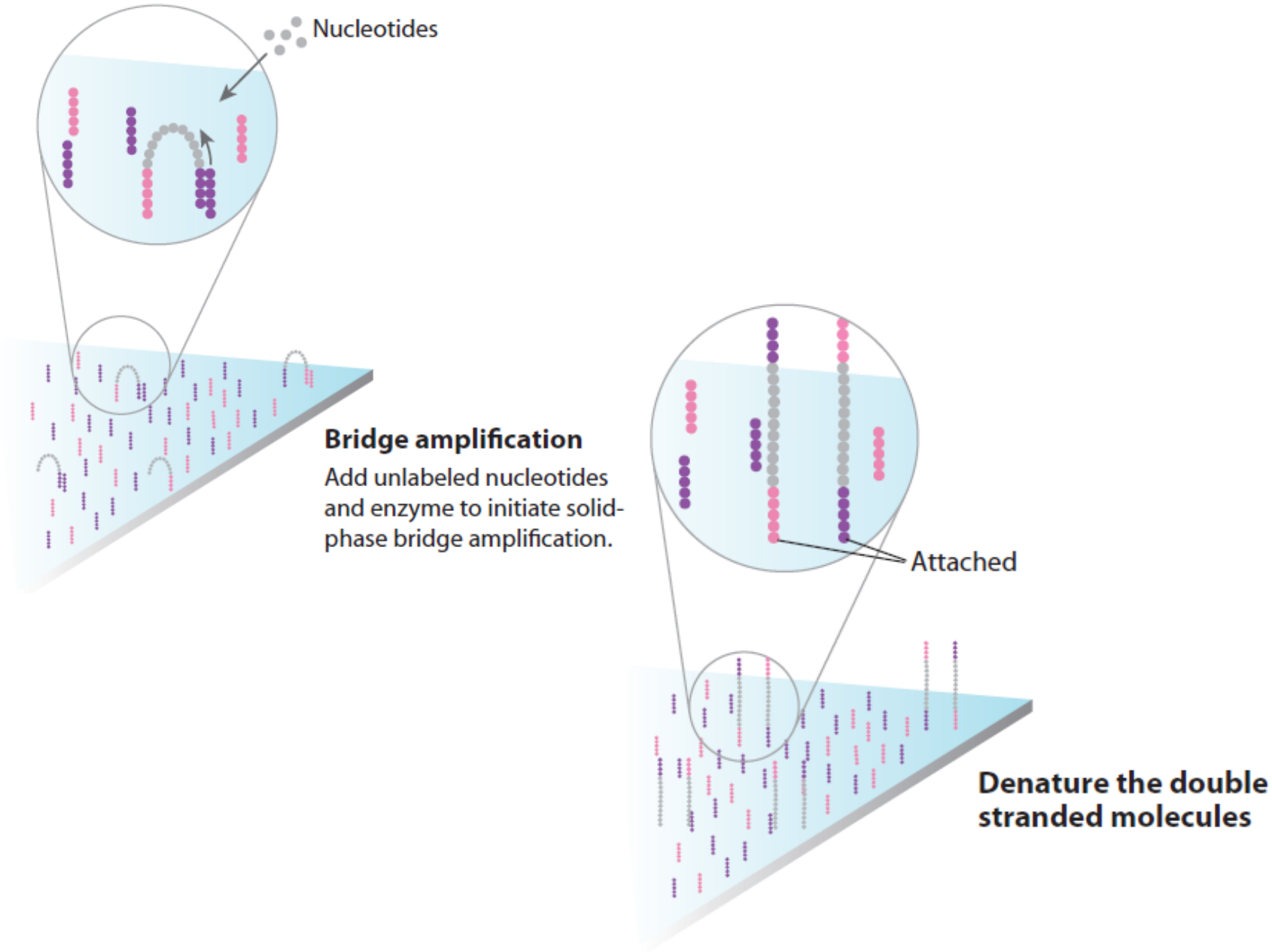


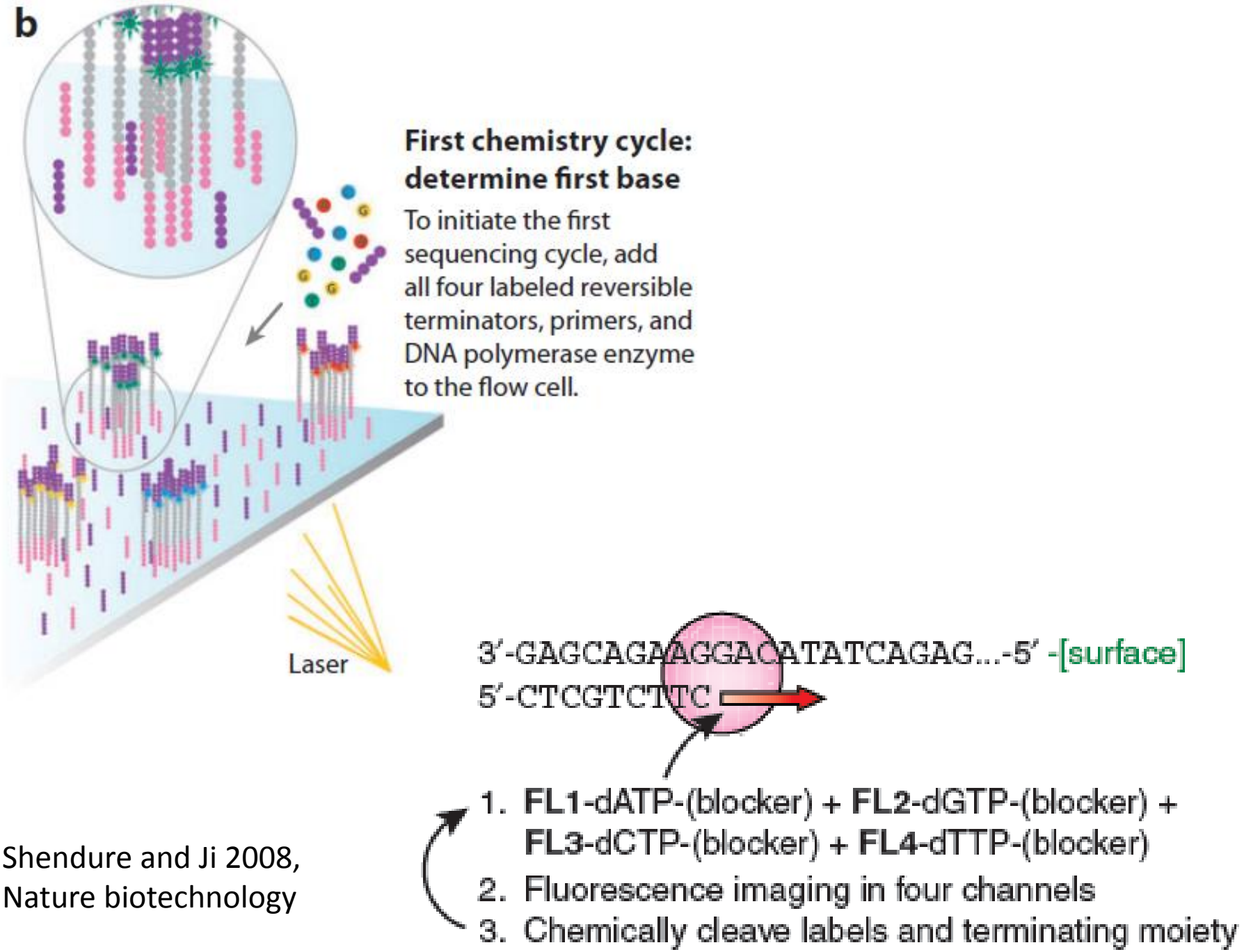
ChIP-seq

a. brief introduction of the next generation sequencing (solexa as example)



Mardis 2008, Annu. Rev. Genomics Hum. Genet.


Nucleotides
Bridge amplification
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.
Attached
Denature the double stranded molecules



Shendure and Ji 2008,
Nature biotechnology

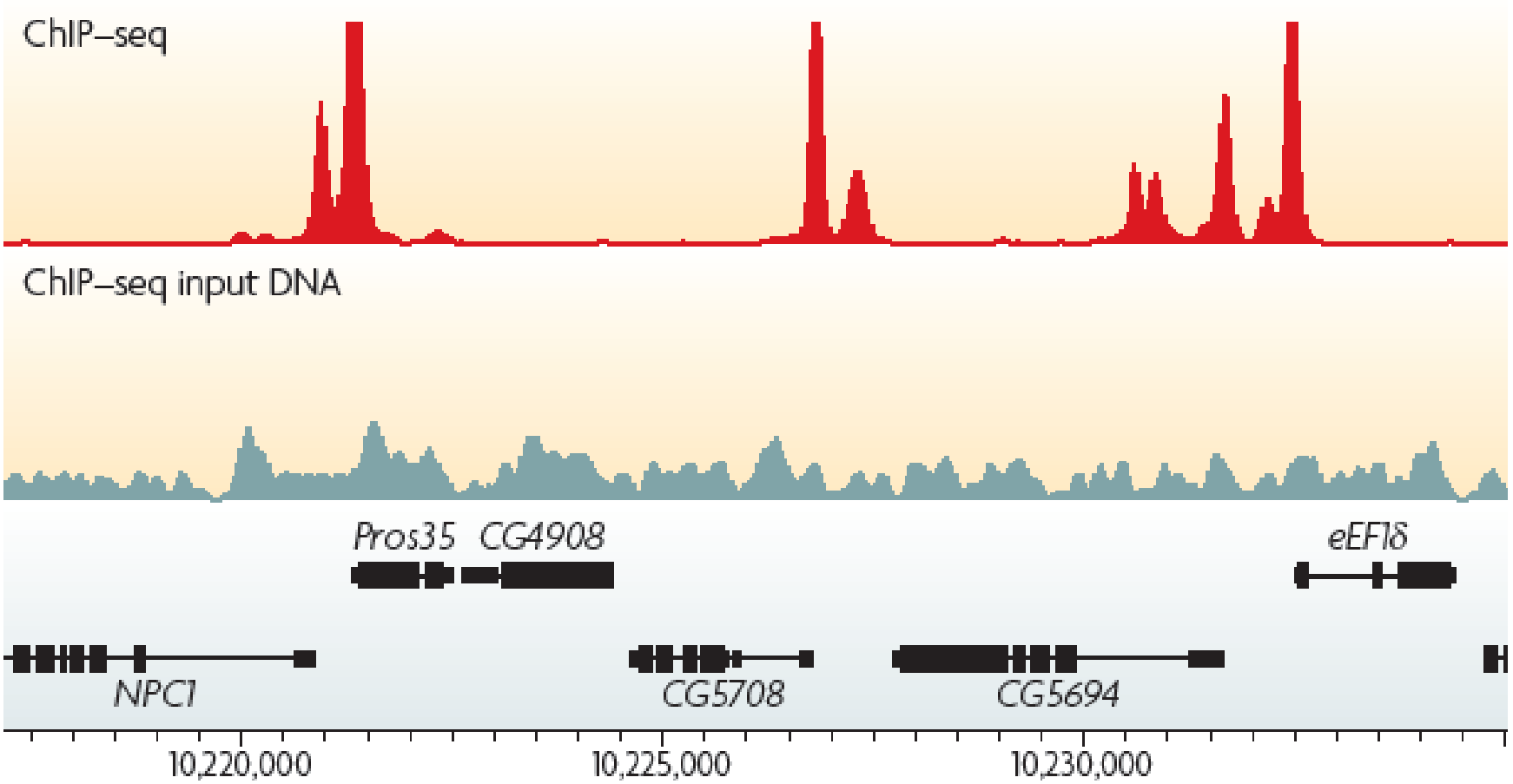
Other platforms and comparison

	Feature generation	Cost per megabase	Cost per instrument	Paired ends?	1° error modality	Read-length	References
454	Emulsion PCR	~$60	$500,000	Yes	Indel	250 bp	14,20
Solexa	Bridge PCR	~$2	$430,000	Yes	Subst.	36 bp	17,22
SOLiD	Emulsion PCR	~$2	$591,000	Yes	Subst.	35 bp	13,26
Polonator	Emulsion PCR	~$1	$155,000	Yes	Subst.	13 bp	13,20
HeliScope	Single molecule	~$1	$1,350,000	Yes	Del	30 bp	18,30

More and more cheap!!!

b. ChIP-seq result example

The sequencing result is mapped to the reference genome



ChIP-PET

a. PET

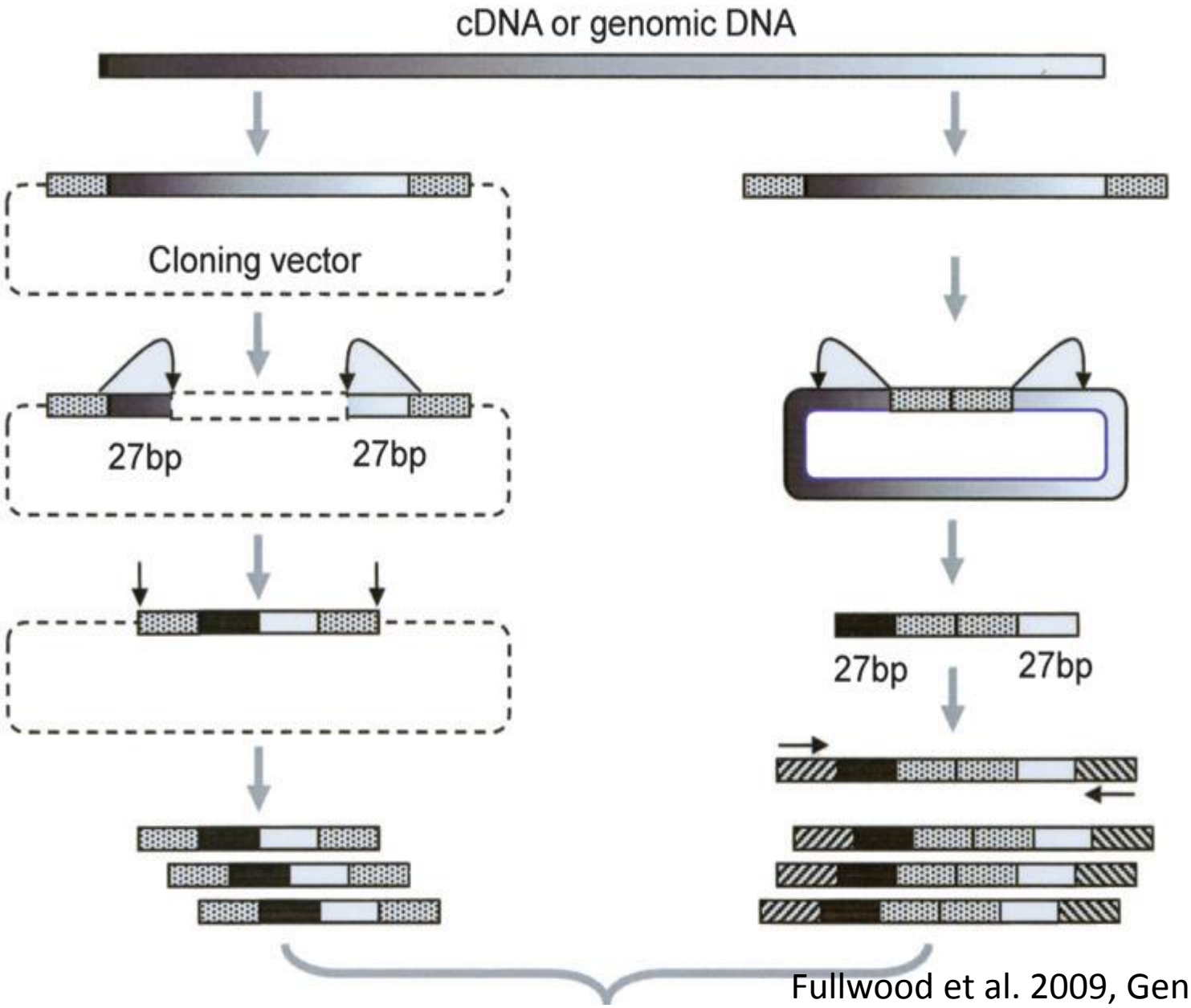
paired-end tag, two tags (20–30 bp) extracted from the two ends of the target DNA fragments as signature information

Type IIS restriction enzyme

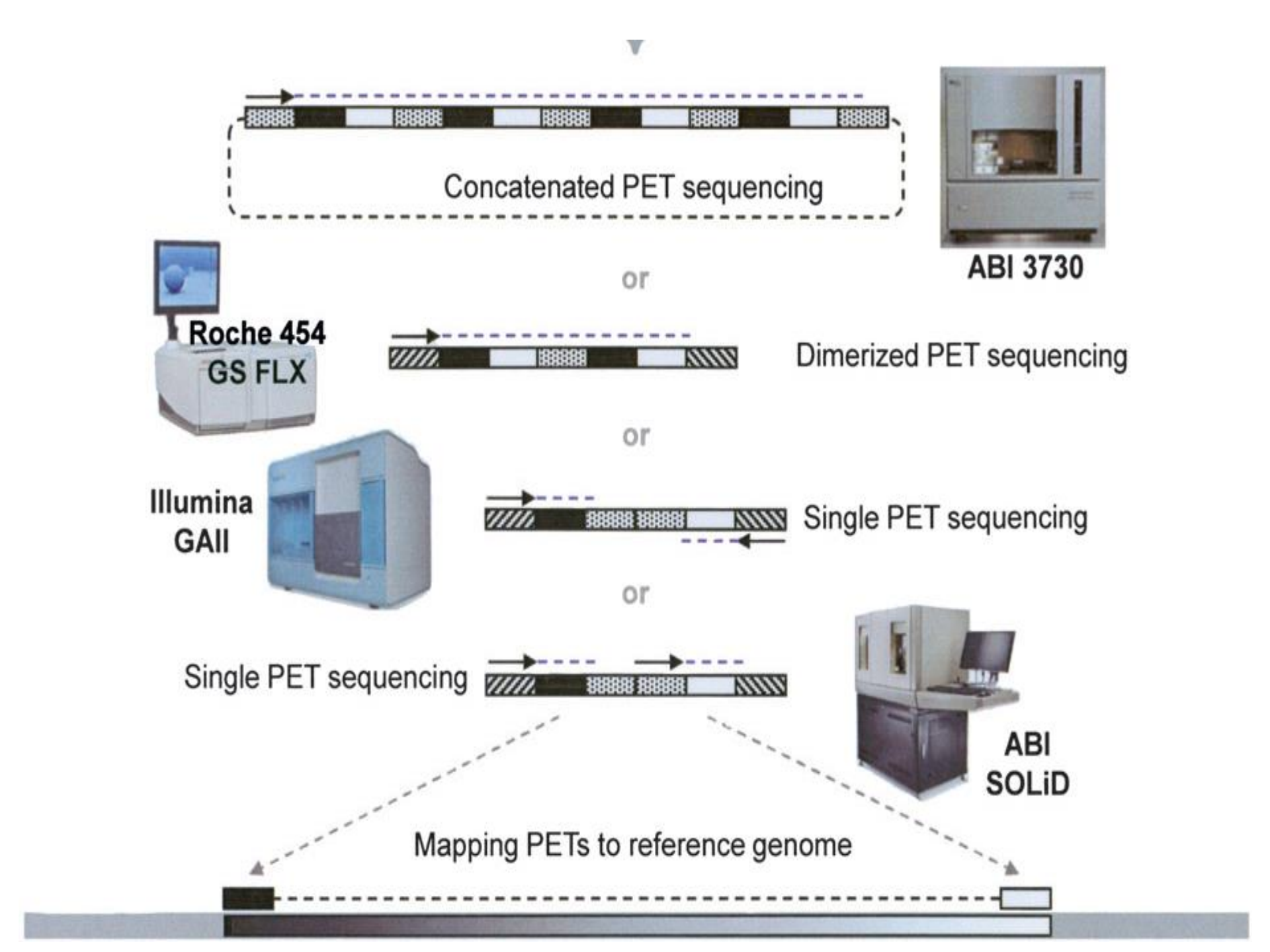
e.g. MmeI cuts DNA 18/20 bp down stream of its recognition site

EcoP15I 25/27bp

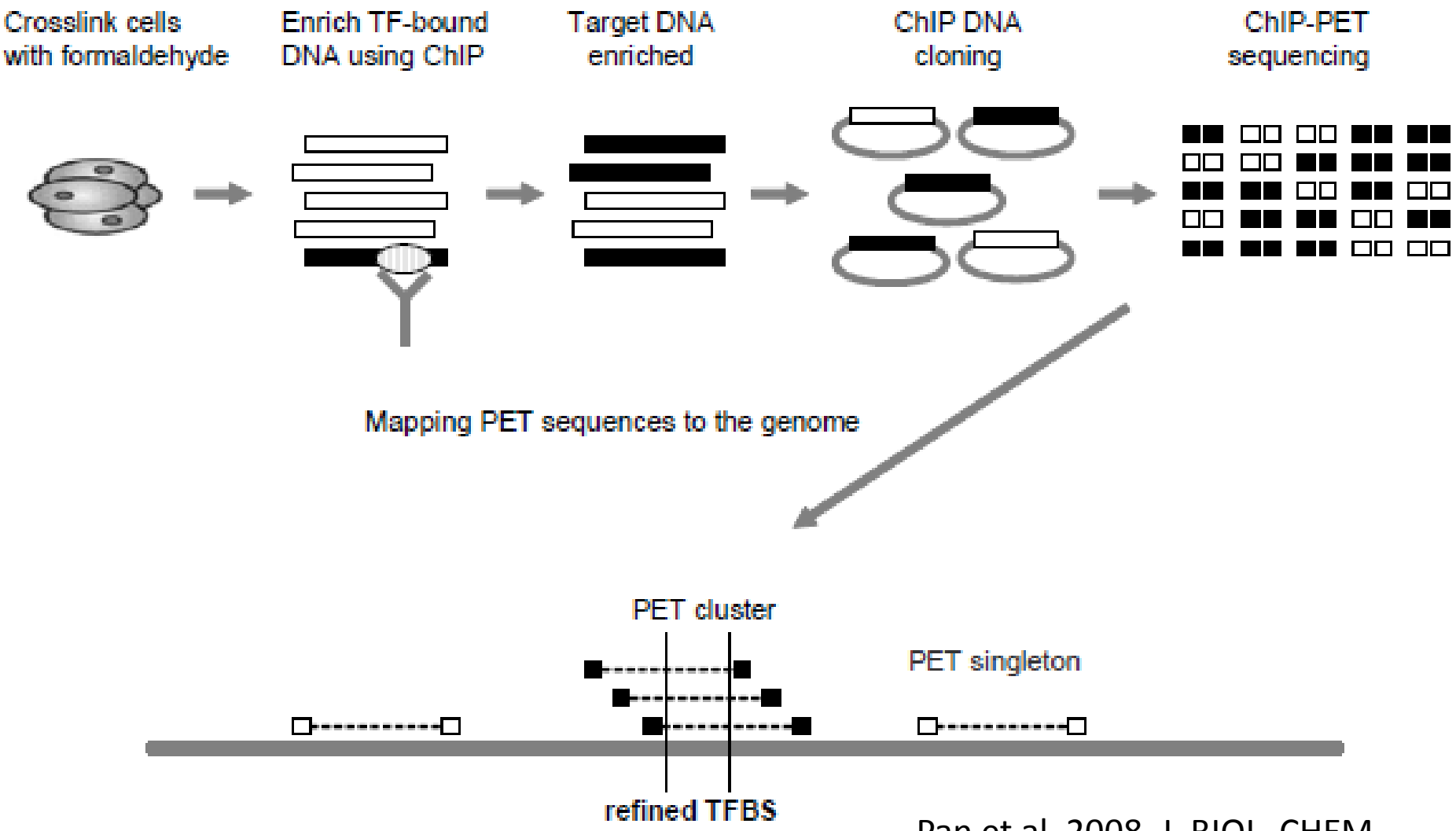
PET methodology



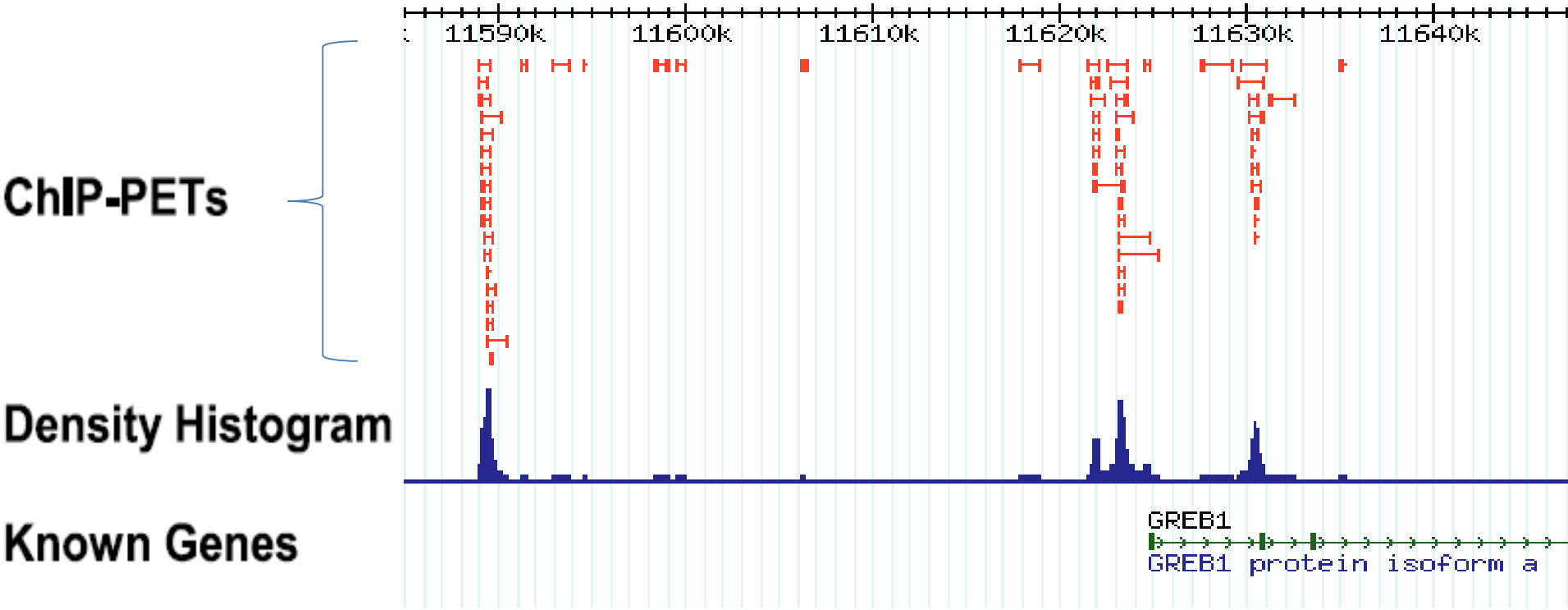
Fullwood et al. 2009, Genome Res.


Concatenated PET sequencing
ABI 3730
or
Roche 454
GS FLX
Dimerized PET sequencing
or
Illumina
GAII
Single PET sequencing
or
Single PET sequencing
ABI
SOLiD
Mapping PETs to reference genome

b.An overview of ChIP-PET



Pan et al. 2008, J. BIOL. CHEM.



so, sometimes the promoter region is longer than 3-4kb

the transcription regulation is far more complicated than what we've known.

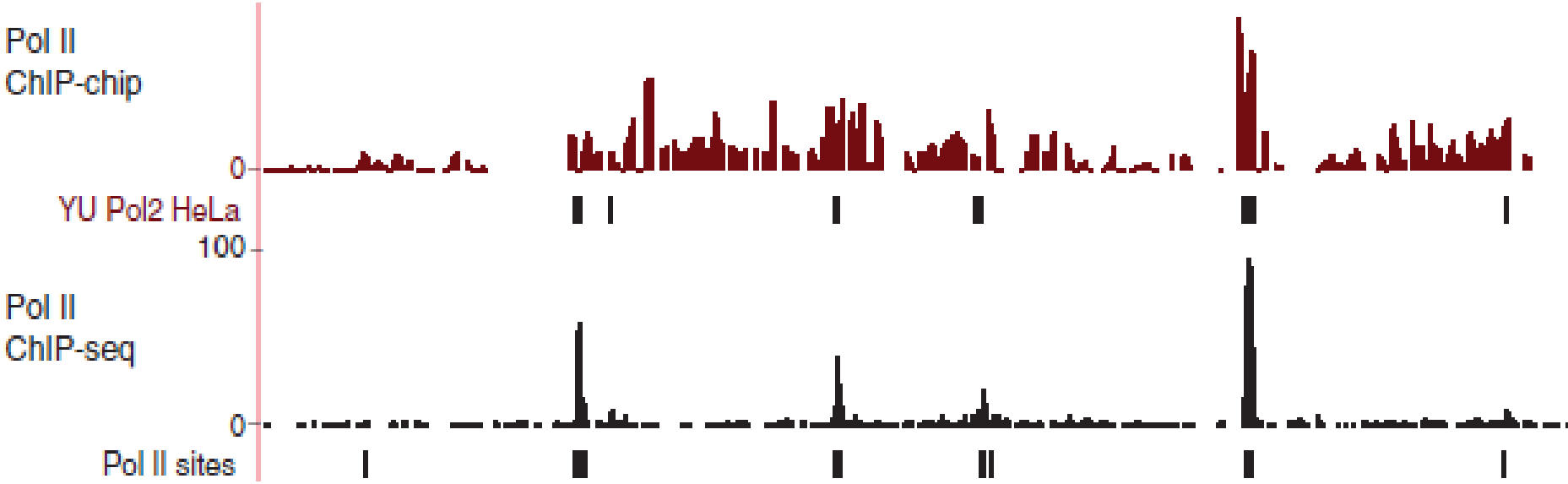
comparison of ChIP–chip and ChIP–seq

Table 1 | **Comparison of ChIP–chip and ChIP–seq**

	ChIP–chip	ChIP–seq
Maximum resolution	Array-specific, generally 30–100 bp	Single nucleotide
Coverage	Limited by sequences on the array; repetitive regions are usually masked out	Limited only by alignability of reads to the genome; increases with read length; many repetitive regions can be covered
Cost	US$400–800 per array (1–6 million probes); multiple arrays may be needed for large genomes	Currently US$1,000–2,000 per lane (using the Illumina Genome Analyzer); 6–15 million reads before alignment
Source of platform noise	Cross-hybridization between probes and nonspecific targets	Some GC bias can be present
Experimental design	Single- or double-channel, depending on the platform	Single channel
Cost-effective cases	Profiling of selected regions; when a large fraction of the genome is enriched for the modification or protein of interest (broad binding)	Large genomes; when a small fraction of the genome is enriched for the modification or protein of interest (sharp binding)
Required amount of ChIP DNA	High (a few micrograms)	Low (10–50 ng)
Dynamic range	Lower detection limit; saturation at high signal	Not limited
Amplification	More required	Less required; single-molecule sequencing without amplification is available
Multiplexing	Not possible	Possible

ChIP-seq surpasses ChIP-chip for identifying sites of transcription factor binding

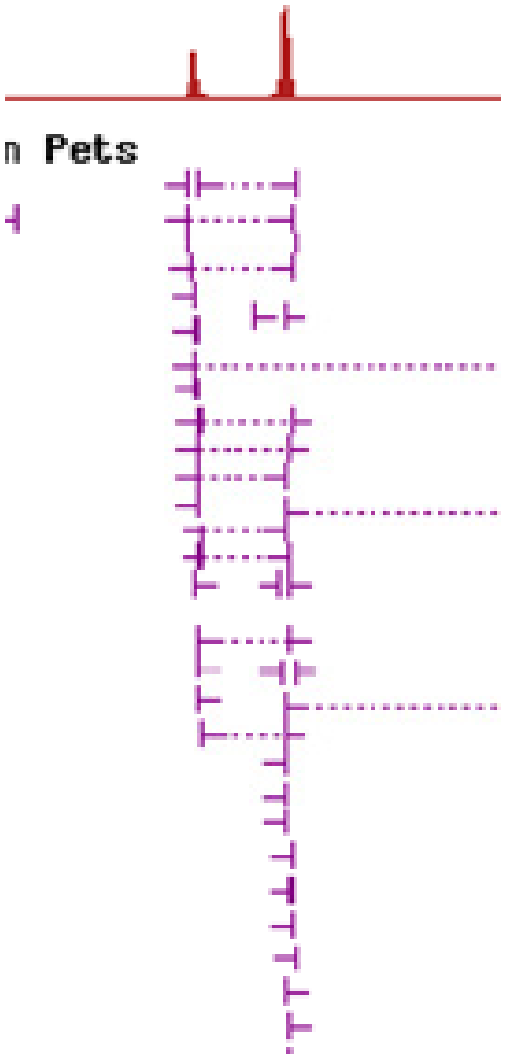
ChIP-seq data give finer resolution and a greater signal-to-noise ratio



Rozowsky et al. 2009, NAT. BIOTECHNOL.

comparison of ChIP–seq and ChIP–PET

The benefit of ChIP-PET over ChIP-seq is that it provides two connected DNA end tags for unambiguous identification of TFBS locations.



The application of ChIP based methouds

Transcription regulation network

Epigenome mapping

Histone modification maps

Nucleosome maps

Chromatin interaction

……

Thank you!