ChIP-PET

主讲人：程赛风
导　师：周道绣
专　业：生物化学与分子生物学
11/5/2010

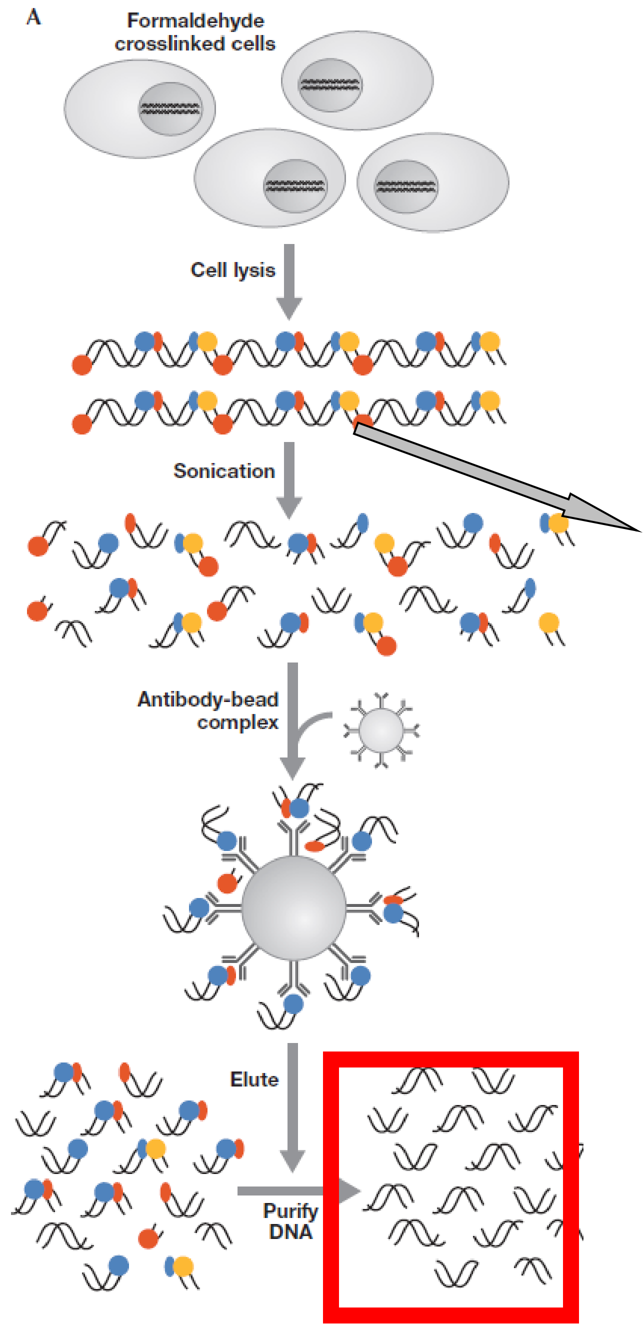
Background

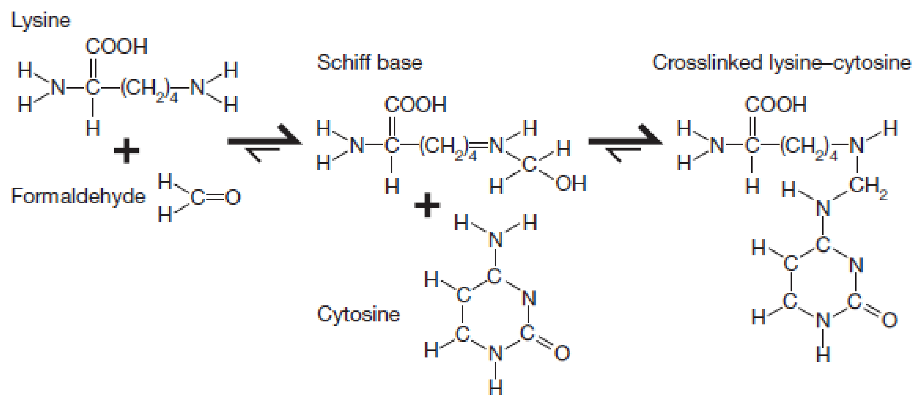
ChIP-PET:

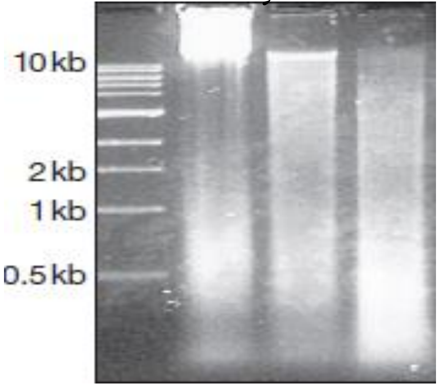
Chromatin Immunoprecipitation Paired-end diTag

- A ChIP-based high throughput sequencing method
- A robust approach for unbiased and precise global localization of transcription-factor binding site (TFBS) and histone modifications

Background





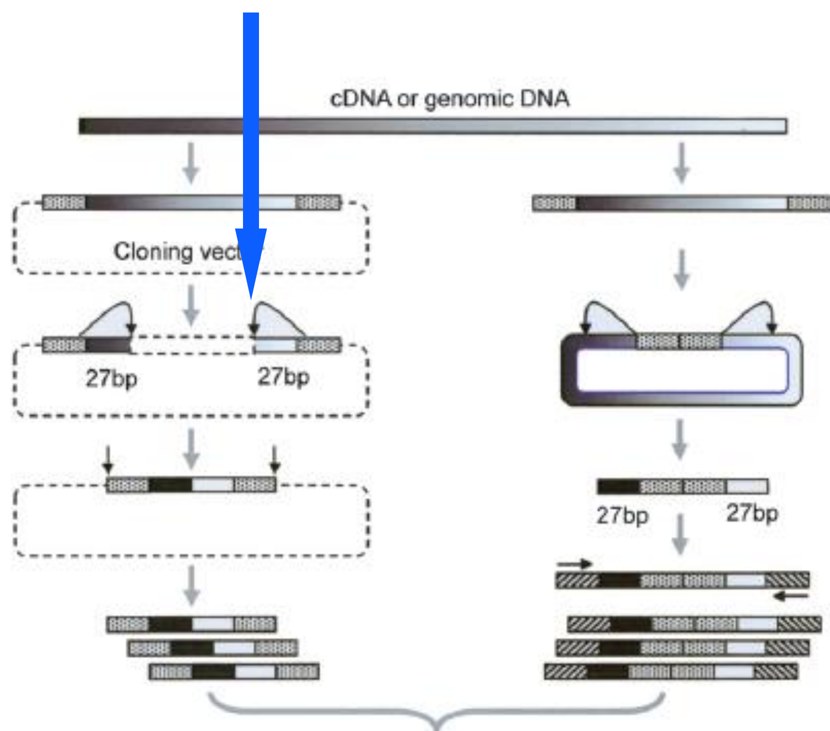


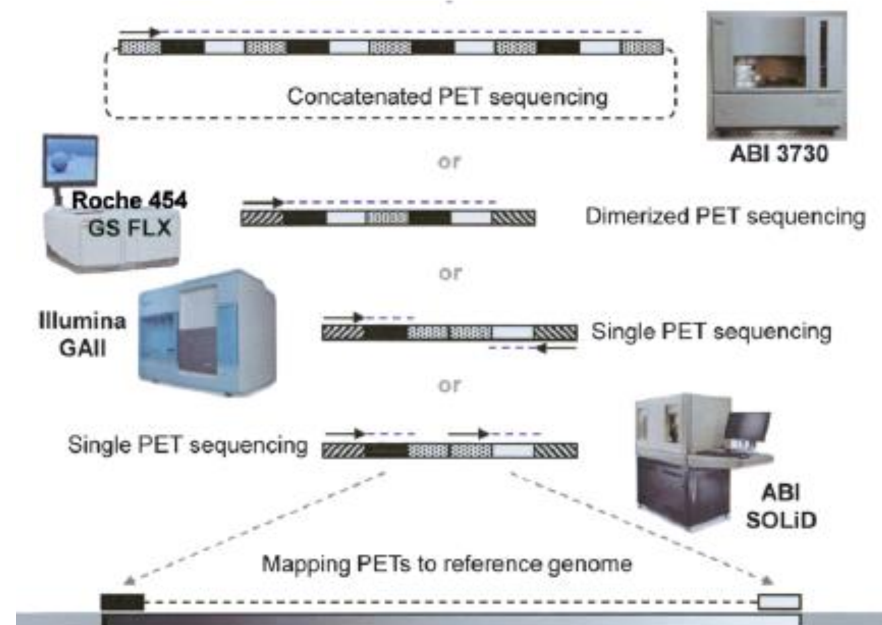
An overview of ChIP

EMBO *reports* VOL 9 | NO 4 | 2008

Schematic view of PET methodology

Type III restriction endonuclease that cut at the site N bps away from its recognition site





Genome Research 2006

Roche/454 Genome Sequencer FLX

Real Time Sequencing by Synthesis

Chemiluminescence detection in pico titer plates

Amplification: emulsion PCR

Pyrosequencing

up to 400,000 reads / run

on average 300-500 bases / read

up 500 Mb / run

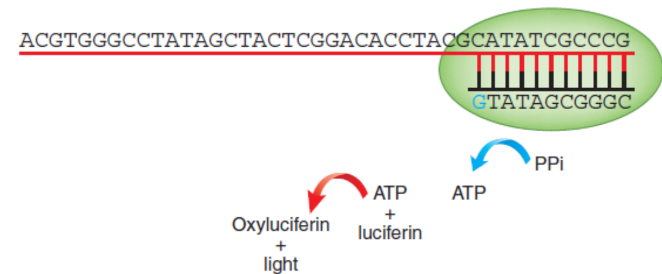


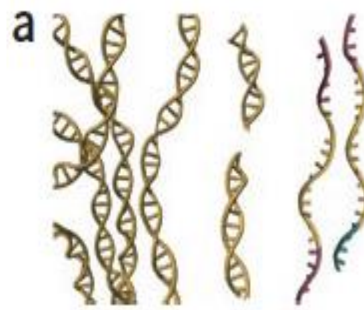
Diagram of the pysoequencing process

Pyrosequencing---454 Sequencing

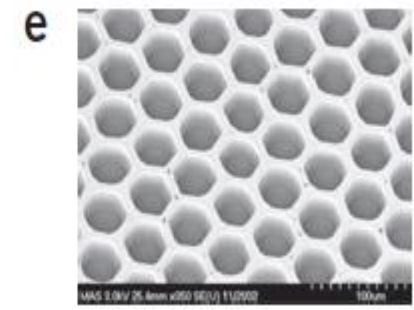
Figure 2 Overview of the 454 sequencing technology. (**a**) Genomic DNA is isolated, fragmented, ligated to adapters and separated into single strands. (**b**) Fragments are bound to beads under conditions that favor one fragment per bead, the beads are isolated and compartmentalized in the droplets of a PCR-reaction-mixture-in-oil emulsion and PCR amplification occurs within each droplet, resulting in beads each carrying ten million copies of a unique DNA template. (**c**) The emulsion is broken, the DNA strands are denatured, and beads carrying single-stranded DNA templates are enriched (not shown) and deposited into wells of a fiber-optic slide. (**d**) Smaller beads carrying immobilized enzymes required for a solid phase pyrophosphate sequencing reaction are deposited into each well. (**e**) Scanning electron micrograph of a portion of a fiber-optic slide, showing fiber-optic cladding and wells before bead deposition. (**f**) The 454 sequencing instrument consists of the following major subsystems: a fluidic assembly (object i), a flow cell that includes the well-containing fiber-optic slide (object ii), a CCD camera-based imaging assembly with its own fiber-optic bundle used to image the fiber-optic slide (part of object iii), and a computer that provides the necessary user interface and instrument control (part of object iii).

The development and impact of 454 sequencing Jonathan .M ,John .H. L et al, NATURE BIOTECHNOLOGY 2006

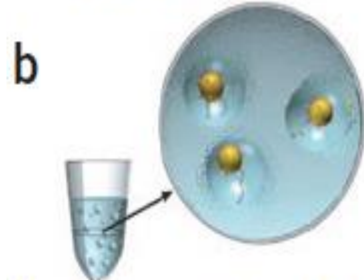
a



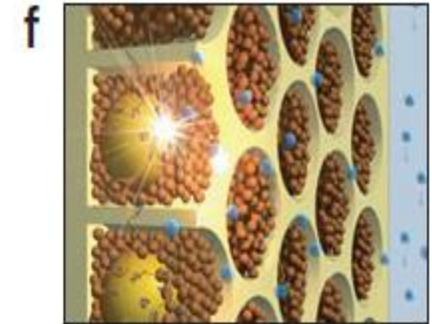
e



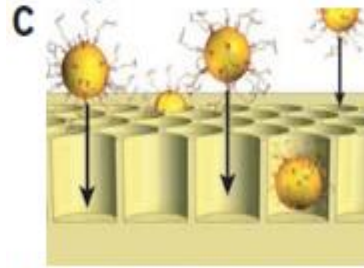
b



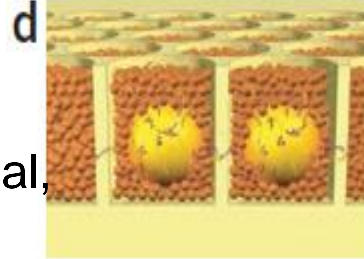
f

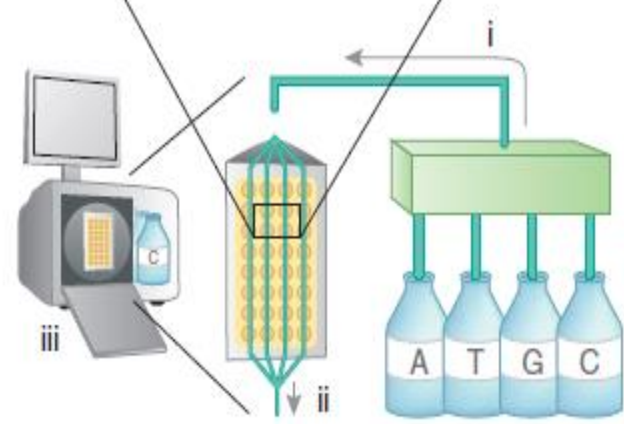


c

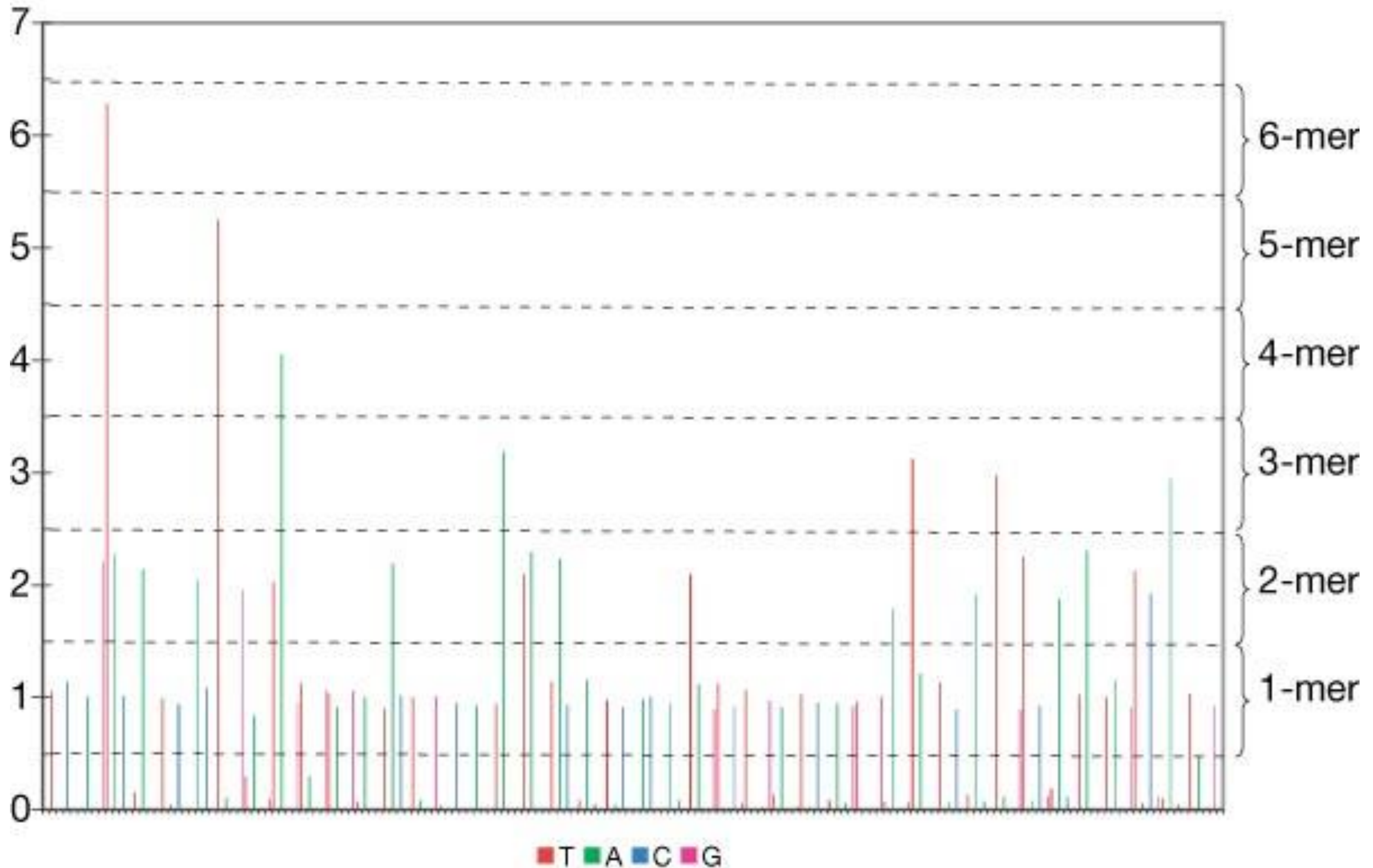


d





A section of Pyrosequencing reads

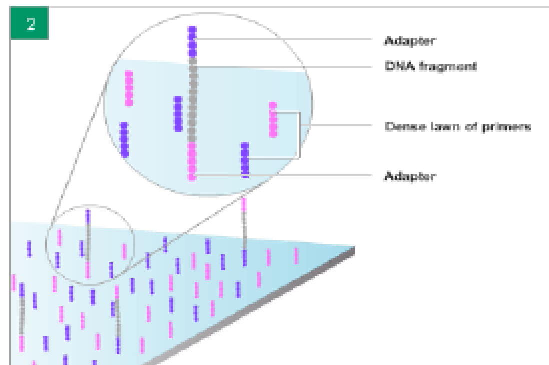


Illumina / Solexa: Genome Analyzer IIx

- Real Time Sequencing by Synthesis
- Clonal Single Molecule Array
- Amplification: bridging PCR
- 60 million reads / run
- up to 50 bases / read
- 2 Gb / run
- 8 channels, app. 5 mio reads / channel
- Fluorescent labels
- Reversible 3‘OH blocking

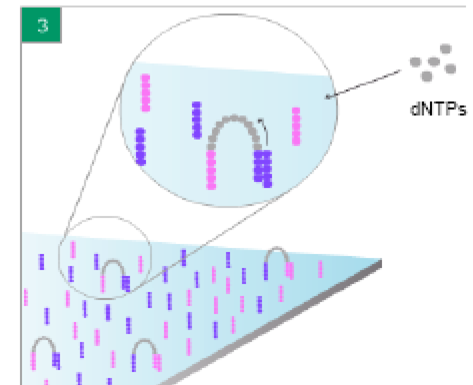
Genome Analyzer IIx

Attach primers and DNA molecule to the surface of flow cell channels



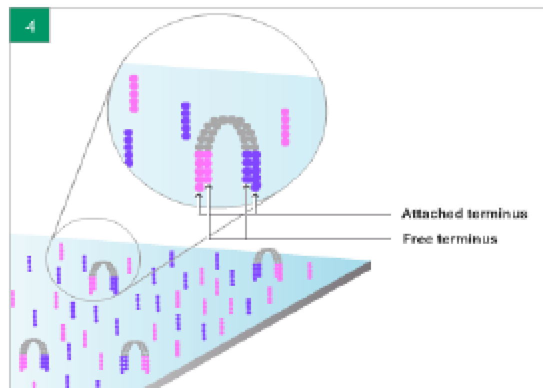
www.illumina.com

Cloning: "bridge" PCR amplification



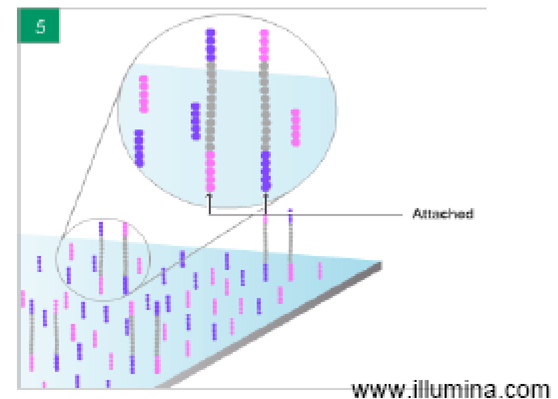
www.illumina.com

1st cycle of PCR amplification



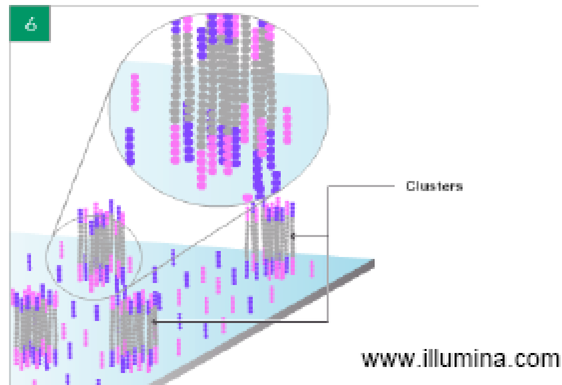
www.illumina.com

Denature the product



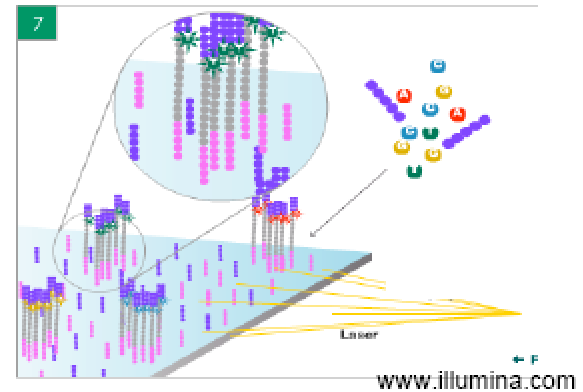
www.illumina.com

End result: PCR colonies (“Polonies”)



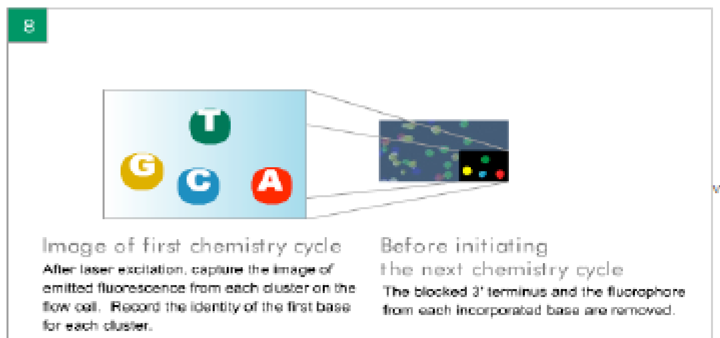
www.illumina.com

First sequencing cycle: determine 1st base



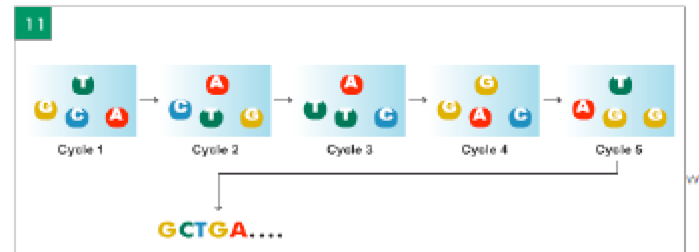
www.illumina.com

Image the base added at each “polony”



www.illumina.com

After additional sequencing cycles



www.illumina.com

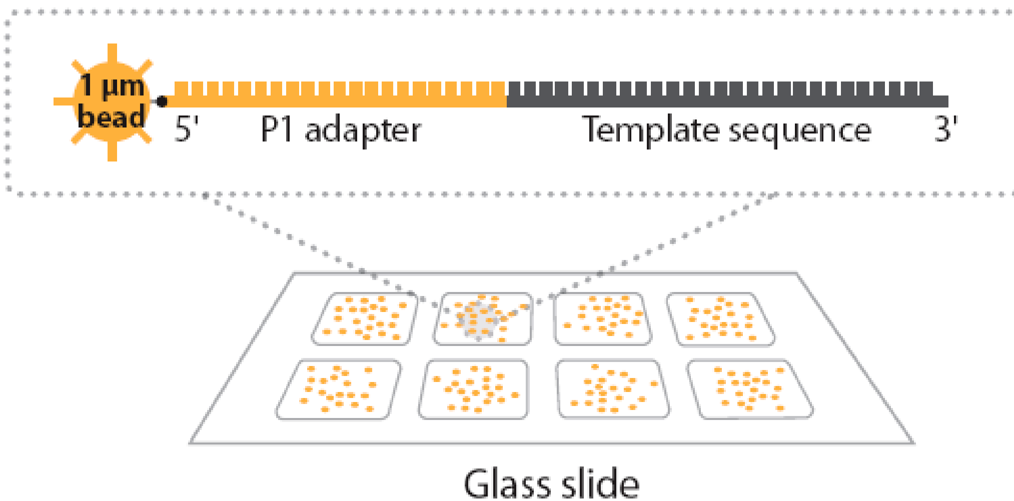
ABI SOLiD

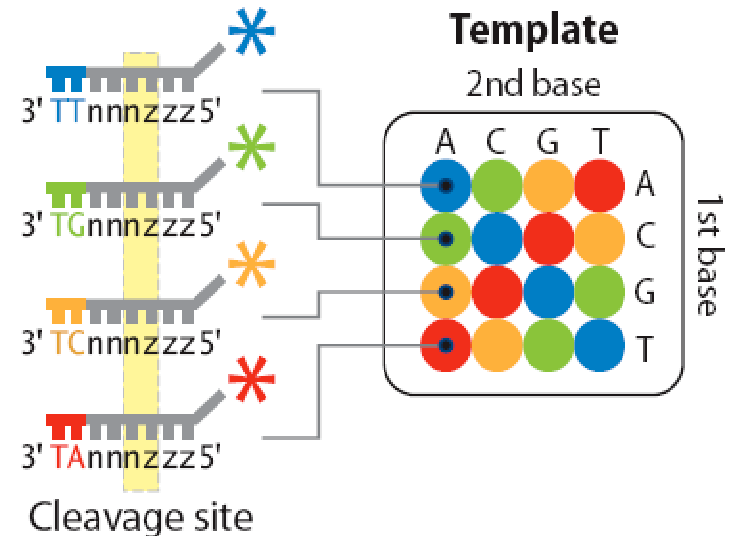
- Obtain more than 2 Gb of data per run
- High quality 35 bp reads generated with high confidence di-base encoding
- Single molecule clonal amplification of templates avoids cloning bias
- Ability to run up to 8 samples on separate channels
- Paired end library approaches
- Robust chemistry for accurate base calling

Applied Biosystems SOLiD

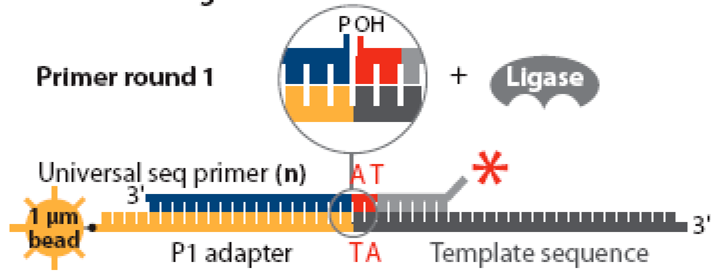
SOLiD:

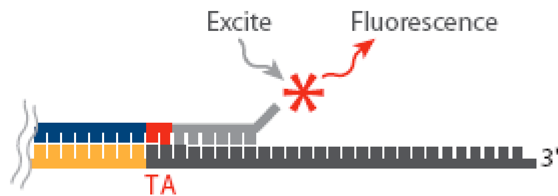
- Substrate attachment; dibase probes
- Make sequencing library by shearing and adapter ligation
- Attach DNA fragments to beads and amplify polonies in emulsion
- Attach beads to slide

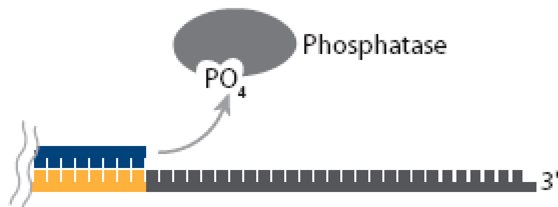


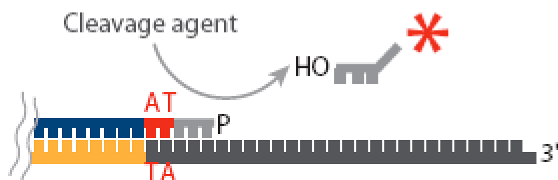


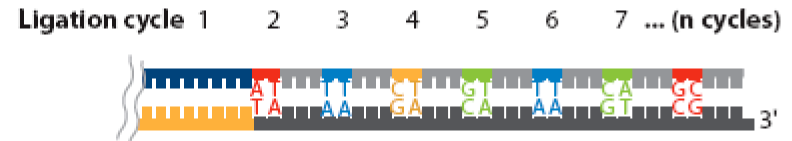
SOLiD: Sequencing ligation cycles

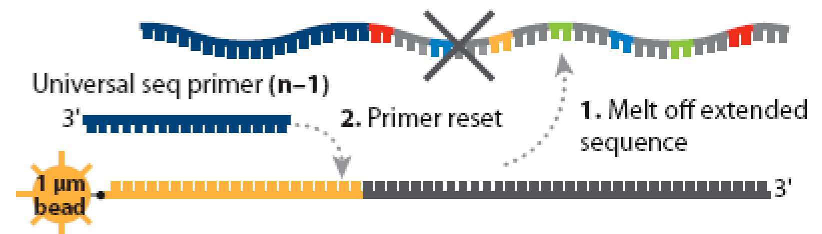


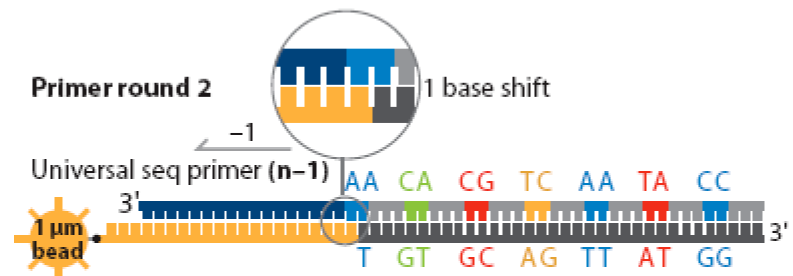




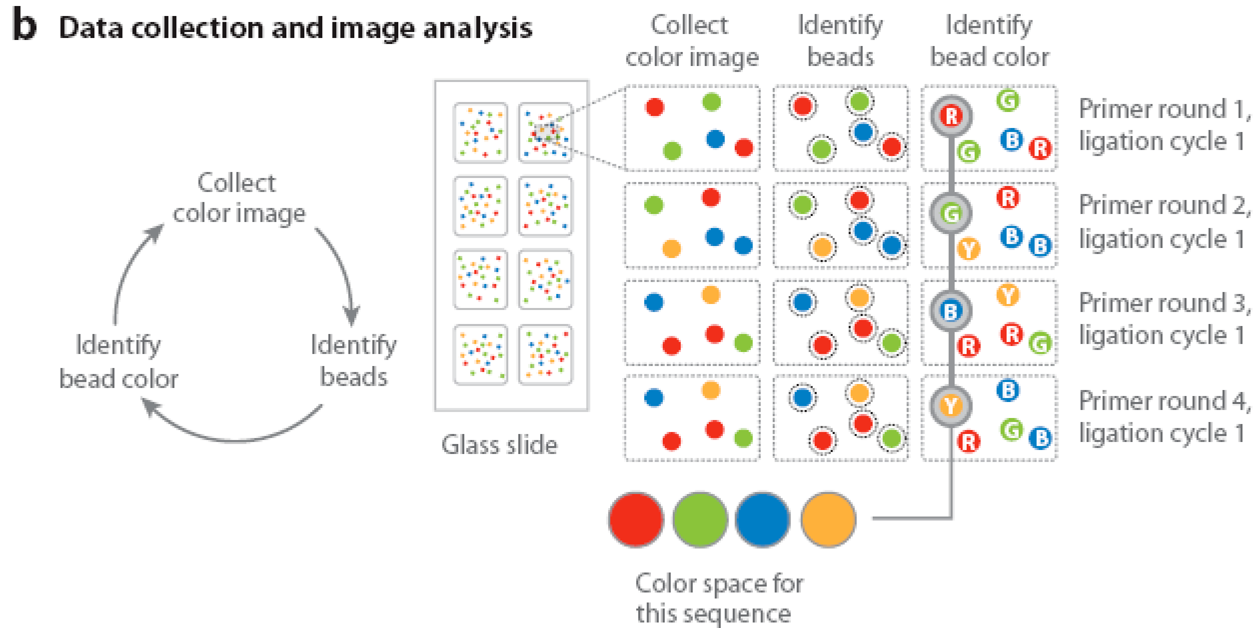


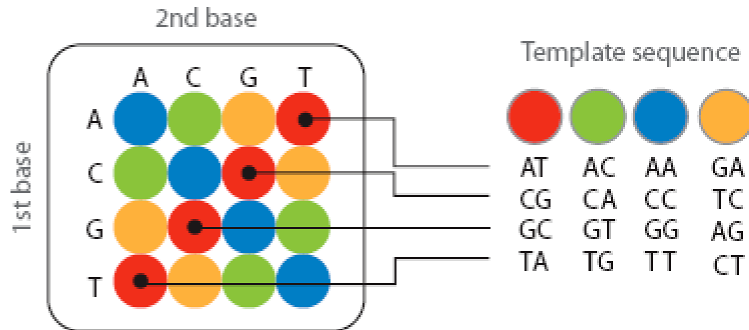






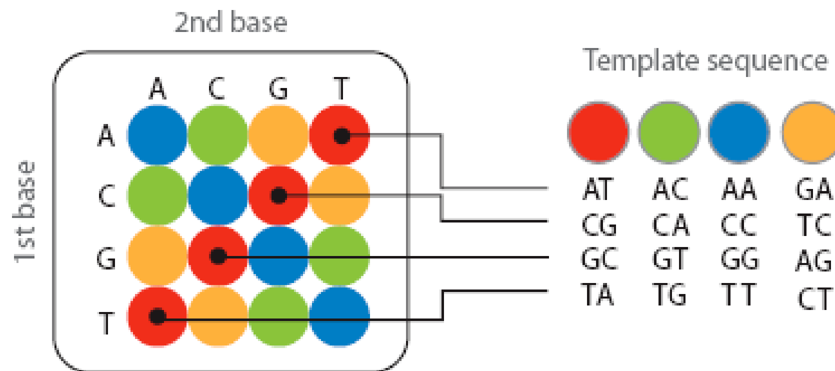
SOLiD: Data Collection and Image Analysis





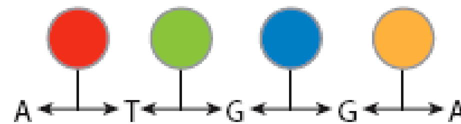
Data decoding

Possible dinucleotides encoded by each color

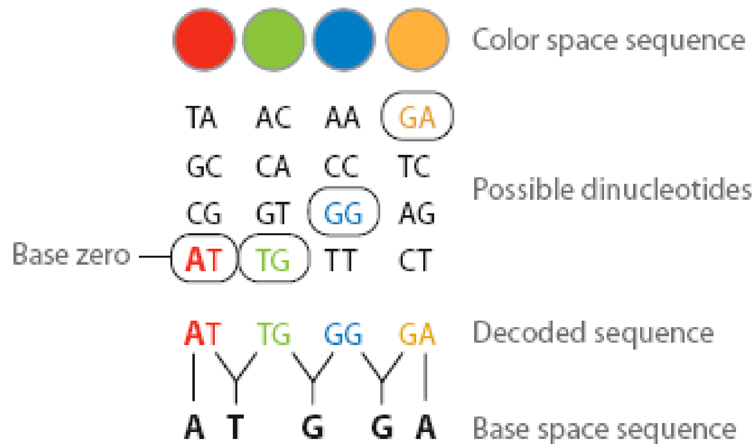


Double interrogation

With 2 base encoding each base is defined twice



Decoding

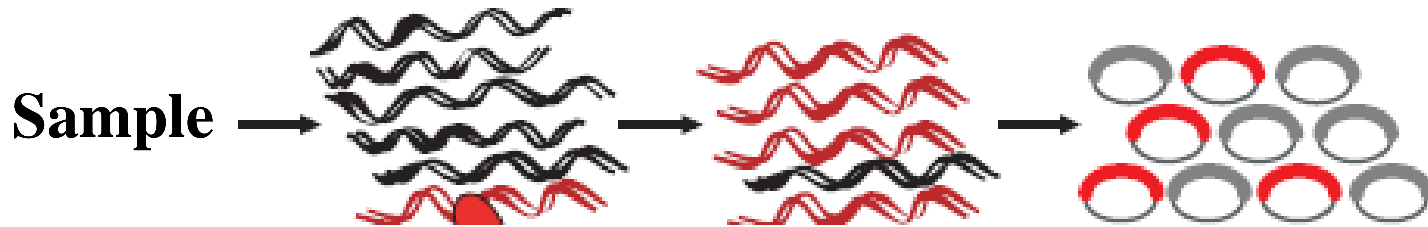


ChIP-PET

ChIP combines PET

ChIP-PET

Sample

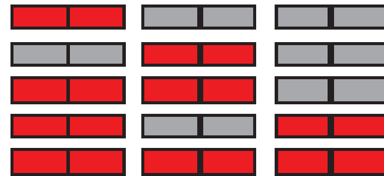


NATURE GENETICS VOLUME 38 | NUMBER 4 | APRIL 20

immunoprecipitation

DNA

Cloning

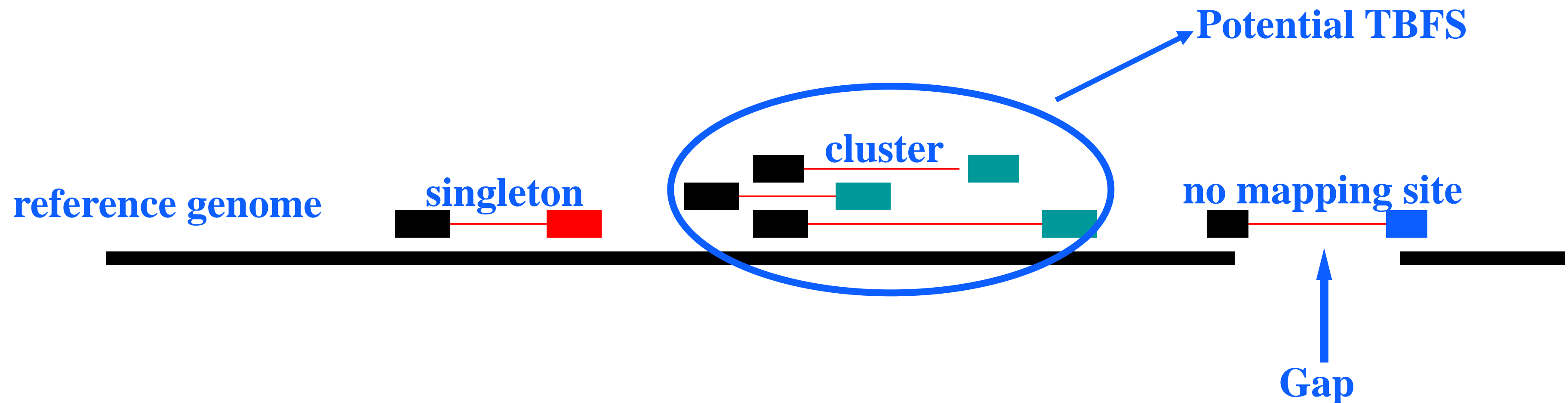


PET library construction
& sequencing to capture
paired end tags

mapping to a reference genome

The raw PETs after sequencing

classified into 3 classes:



some definitions:

cluster: the overlapping PET DNA sequence fragments mapped to the genome

cluster overlap: the most common PET DNA fragment in overlapping PETs
in the given cluster

peak: the number of overlapping PETs in the given cluster

Background noisy

- most polyclonal antibodies are far from specific
- monoclonal antibodies that bind to one epitope can not work well in every biological repeat
- sonication
- amplification biases in the PET library preparation (artifact)

Statistical analysis

1. Saturation Kinetics analysis of ChIP-PET Library

   to determine that the library contains most all the origin DNA fragment pulled down by ChIP with Hill Function (Kuznetsov et al., 2002)

2. Mapping Simulation of Overlapping PET Clusters

   A Monte Carlo simulation was performed to assess the background level of overlapping PET sequences when mapped to the genome

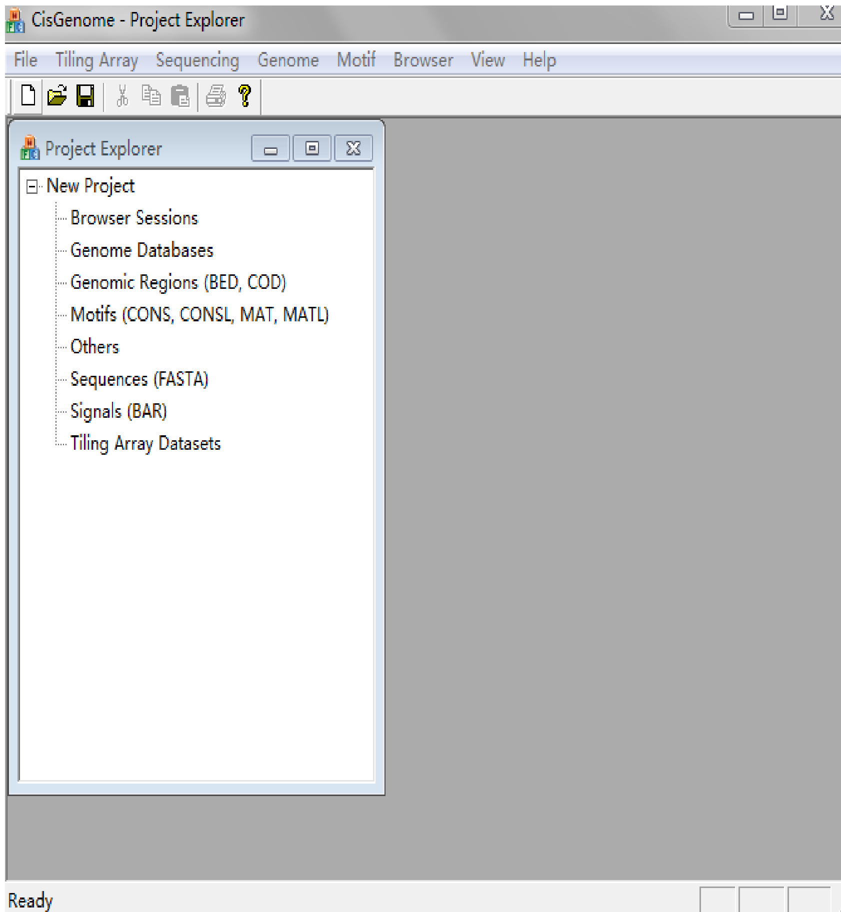
3. Goodness-of-Fit Analysis for Assessing the Reliability of PET Clusters

   to assess the reliability of using PET clusters as a readout for the entity

4. ChIP-PCR or Microarrays to further confer the PET we get

Software

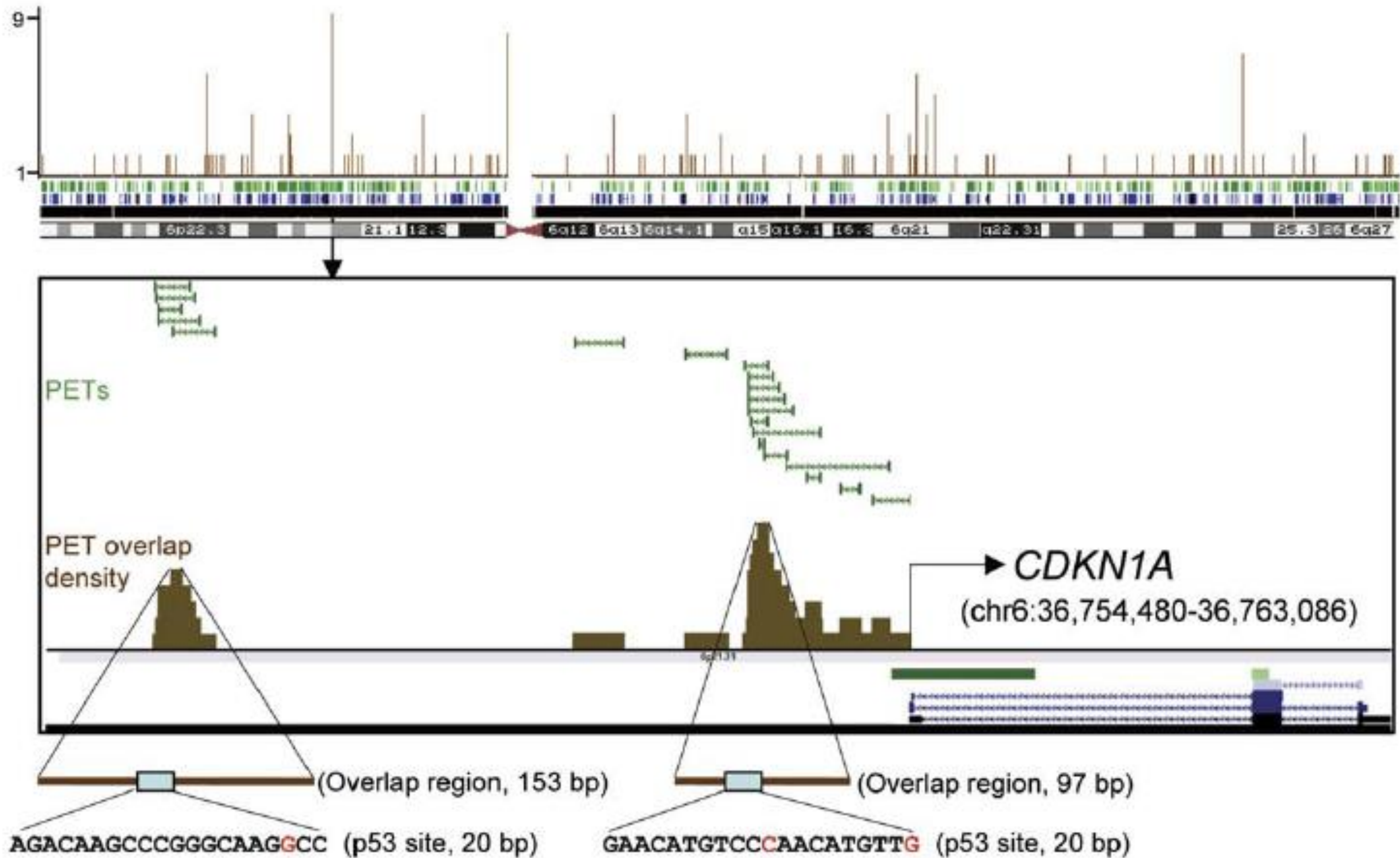
- CisGenome, FindPeaks and so on



CisGenome

FindPeaks

A brief look of the mapping result



Cell *124*, 20219.Juanury 13, 2006

De novo motif/motifs discovery

Position weight matrix

$$S = \frac{\log(n) \sum_{i=1}^{W} \sum_{j=A}^{T} p_{ij} \log(p_{ij}/q_j)}{W}$$

Here, n is the number of aligned sites that are used to construct the matrix, i.e. $n \equiv n_i = n_{iA} + n_{iC} + n_{iG} + n_{iT}$, where n_{ij} (j = A, C, G, T) is the number of occurrences of nucleotide j at the i-th position of the motif. A pseudocount 0.5 was added to each n_{ij} to avoid zero. $p_{ij}=n_{ij}/n_i$. q_j is the occurrence frequency of nucleotide j in the background sequences (derived from all input sequences). W is the length of the motif. This score is essentially the motif score used by MDSCAN under a zeroth-order Markov background model

Nucleic Acid Research 2006

Comparison with ChIP-chip

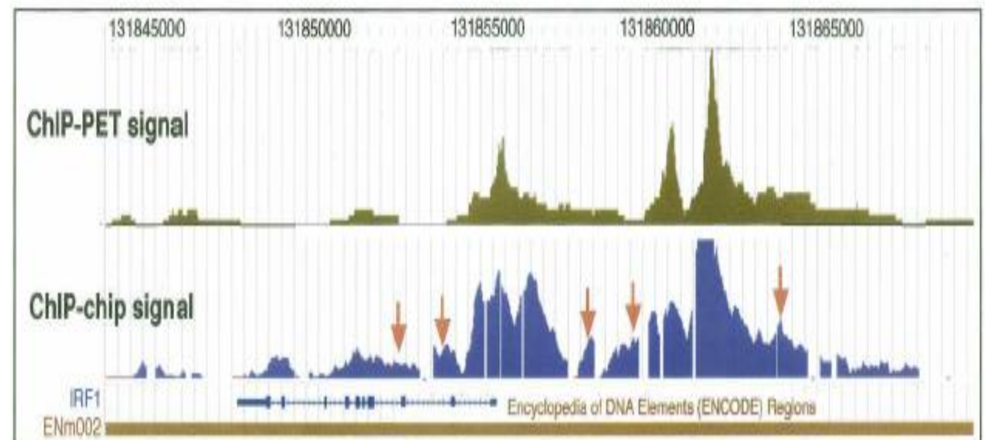
1.high resolution than ChIP-chip

2.do not require a reference sample, while the DNA associated with a TF is compared to the genomic DNA or a negative sample by ChIP-chip generally

3.a lower concentration of DNA is needed for sequencing than ChIP-chip

4.better at distinguish repetitive fragment than ChIP-chip

5. very small targets flanked by large adjacent repeats are likely to be missed by ChIP-PET but might be detected by ChIP-chip

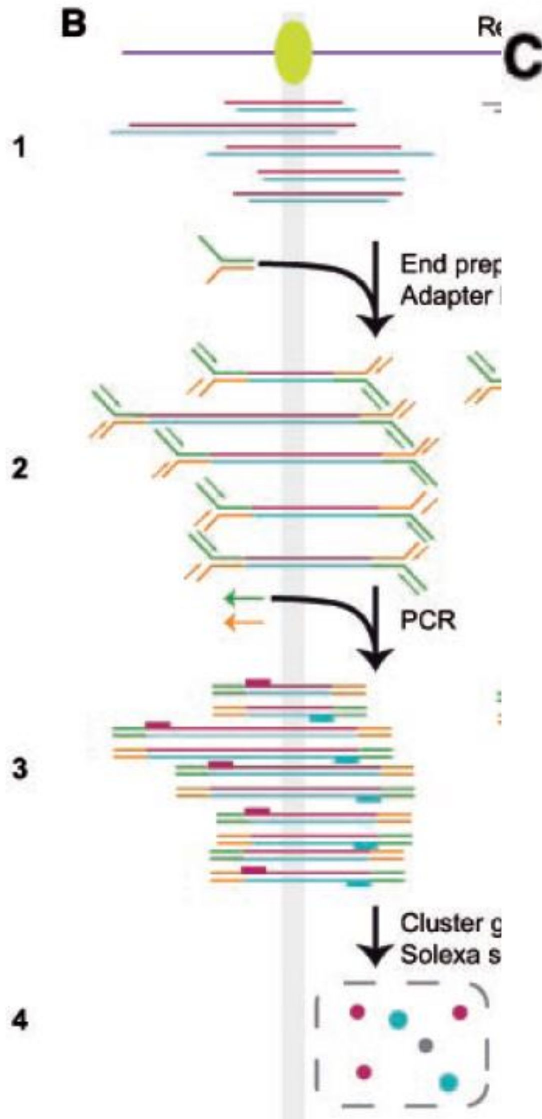


Comparison with ChIP-seq

1. marks both the start and the end of each DNA fragment
2. duplicated PET fragments arising from ChIP-PET library preparation can be easily removed

A brief look of ChIP-seq

ChIP



Application

- histone modification
- TFBS
- nucleosome map
- ……

to be discovered

THANKS