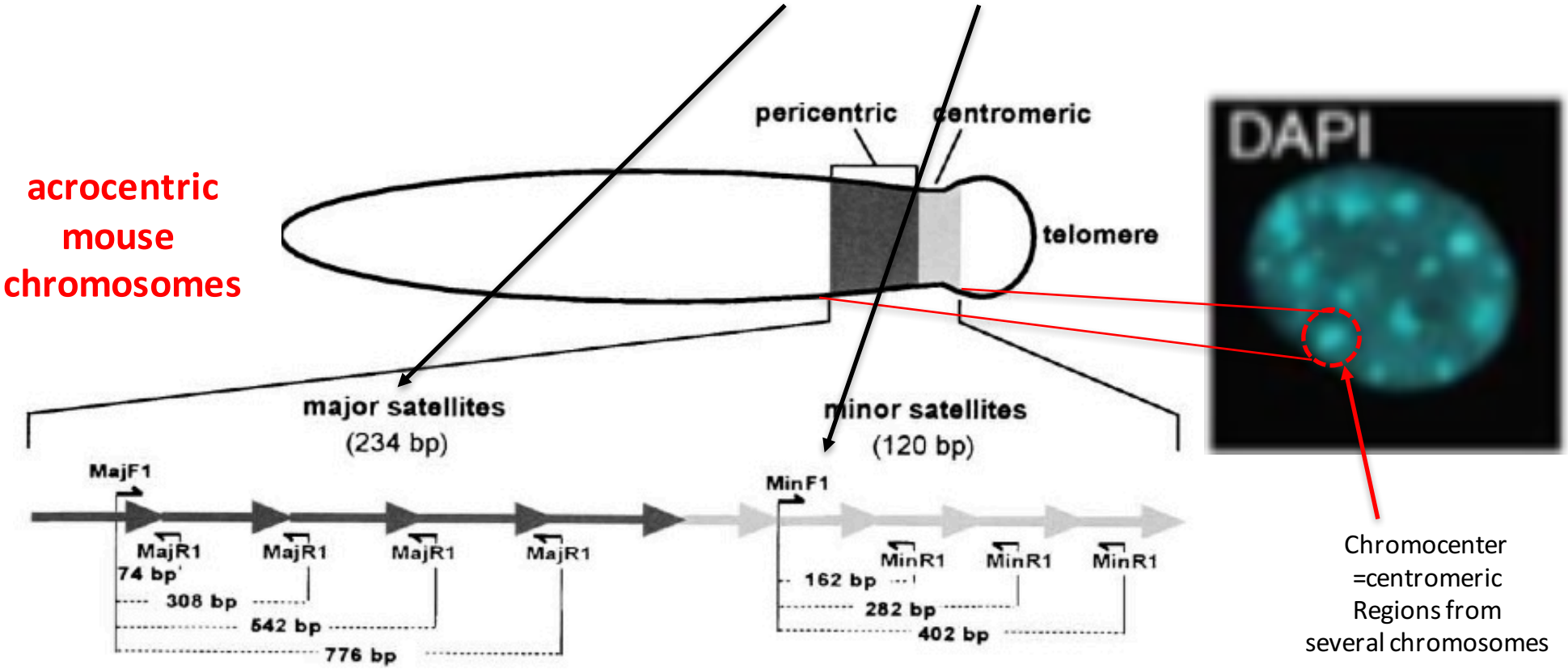
HOW TO STUDY EPIGENETIC MODIFICATIONS

-- STRATEGIES USING SUV39H1 AS A HALLMARK MODEL FOR EPIGENTIC REGUALTION --

What are the target sites for Suv39h1 and H3K9me???

CHROMATIN IMMUNOPRECIPITATION FOLLOWED BY QUANTITTIVE PCR

Design PCR oligos that amplify major and minor satellite repeats

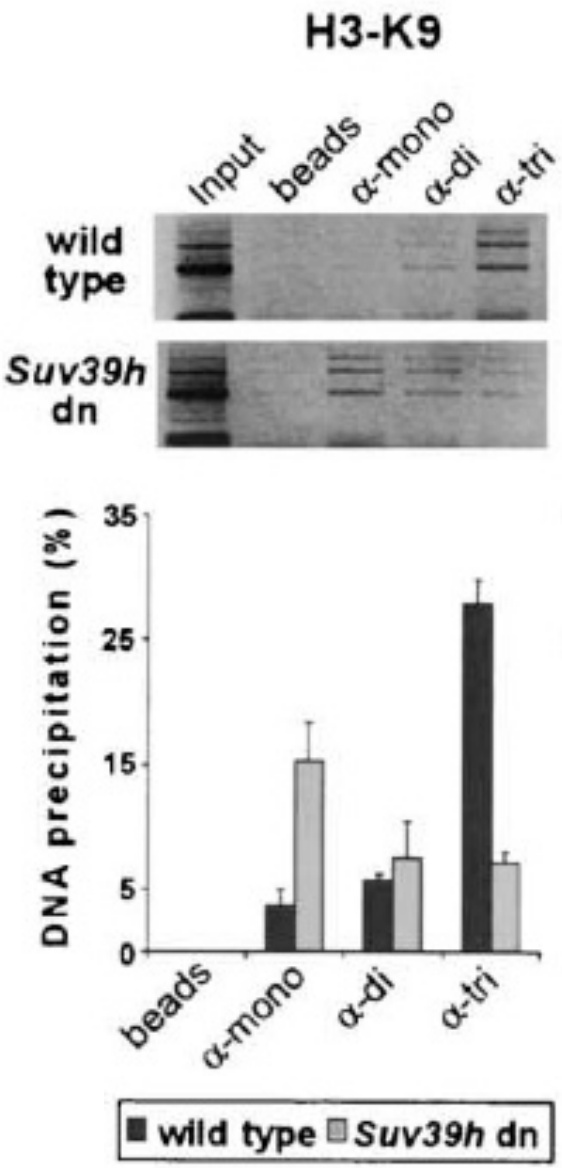


EXAMPLE: Pericentric heterochromatin in mouse cells

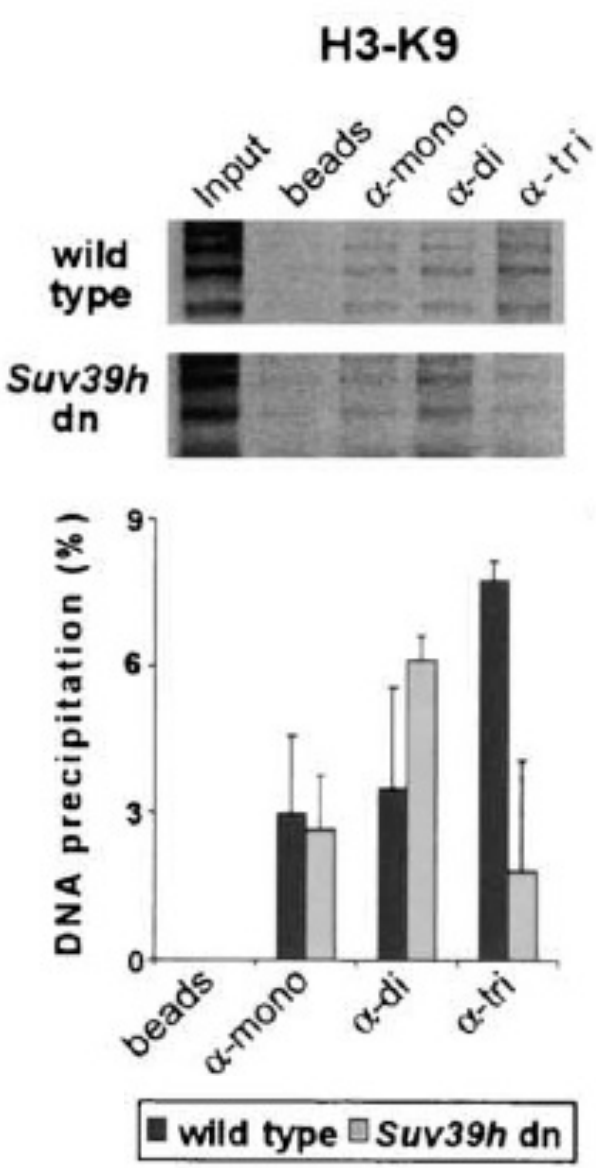
Min/Maj F1…: forward primer in unique region
Rev primer bind in every repeat

CHROMATIN IMMUNOPRECIPITATION (ChIP) COUPLED WITH PCR
→ H3K9me3 is enriched at pericentric (major+minor) repeats in mouse cells

Major satellite repeats



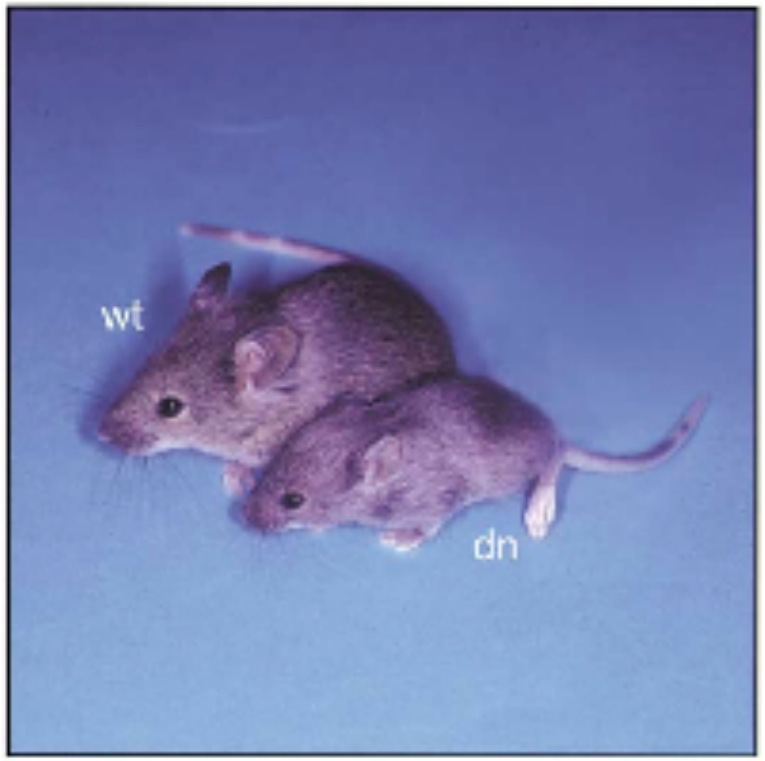
Minor satellite repeats



PCR amplification of major/minor satellite Repeats after ChiP using Antibodies that are specific for H3K9me1; H3K9me3; H3K9me3

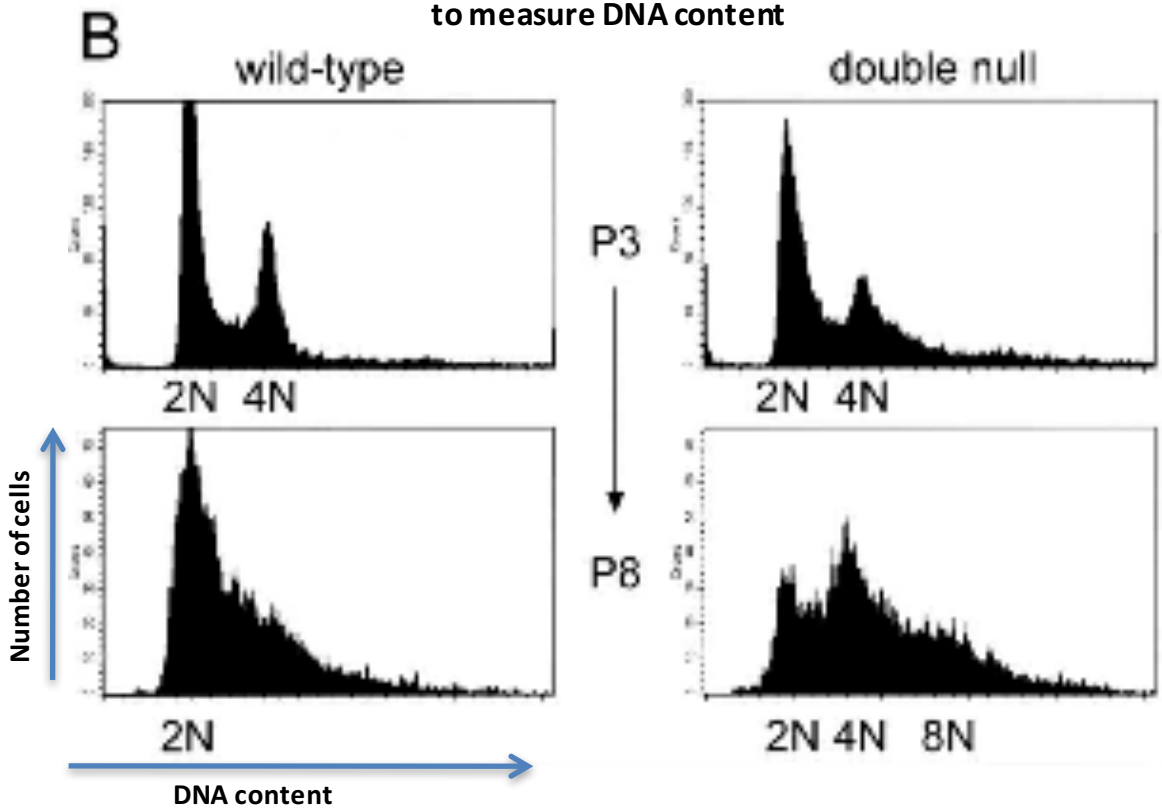
Suv39h1 is required for imposition of H3K9me3 at pericentric repeats

Lack of SUV39h HMTase activity results in genomic instability



A knock-out model system
for Suv39h1 and Suv39h2
- Loss of Suv39h1/2: smaller body size

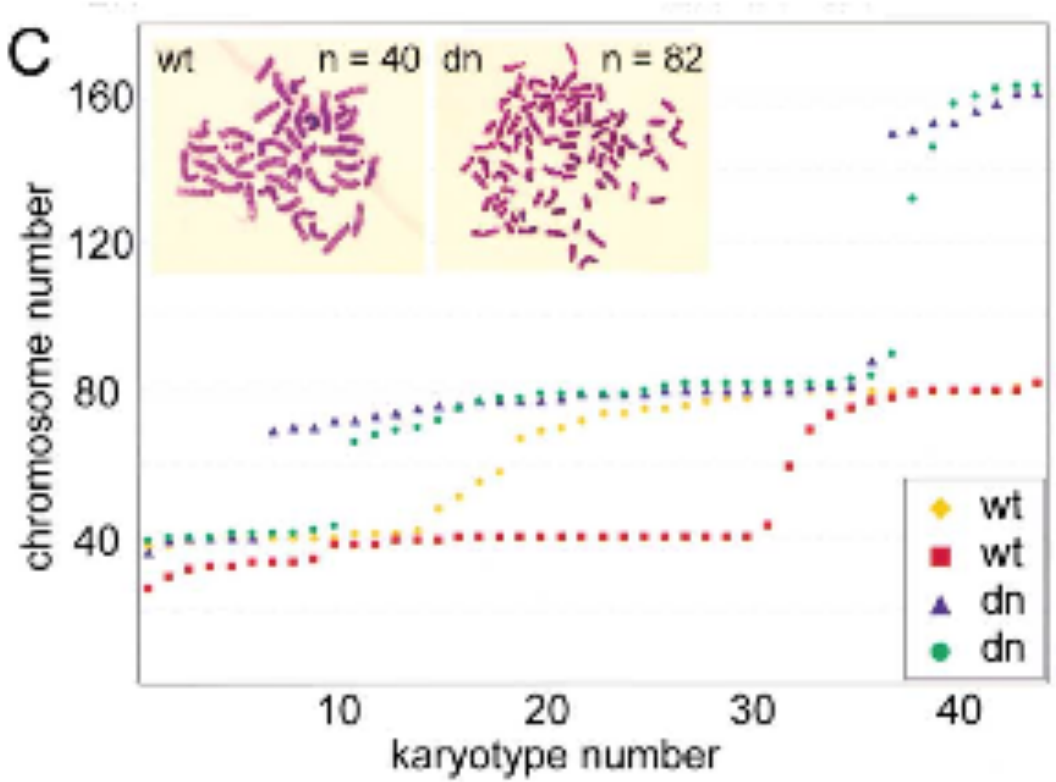
Fluorescence activated cell sorting (FACS)
to measure DNA content



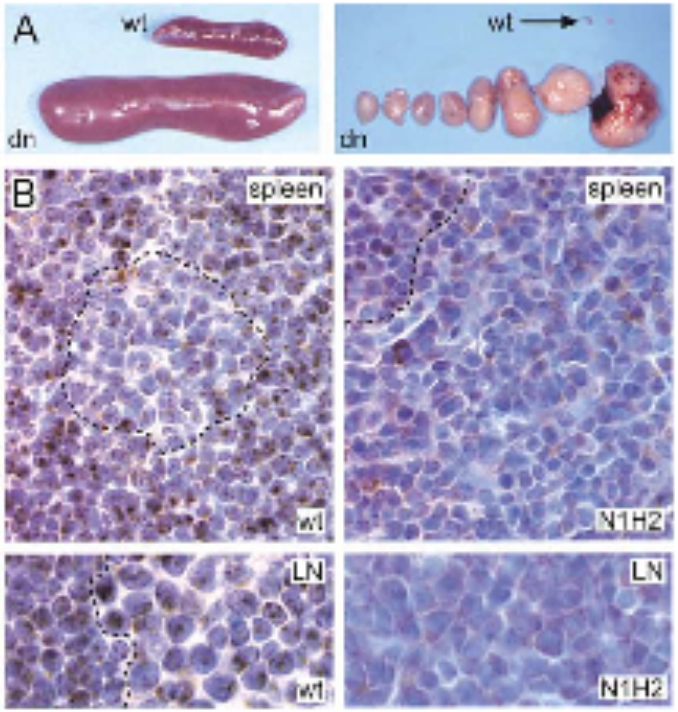
Fibroblasts from Suv39h1/2 null mice
are aneuploidy

Lack of SUV39h HMTase activity results in genomic instability

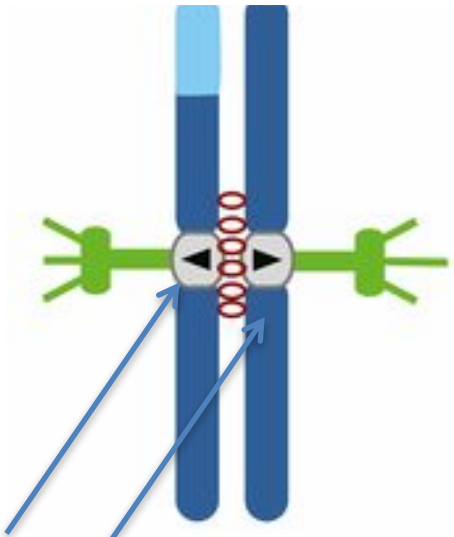
Loss of Suv39h1/2 results in increased chromosome numbers



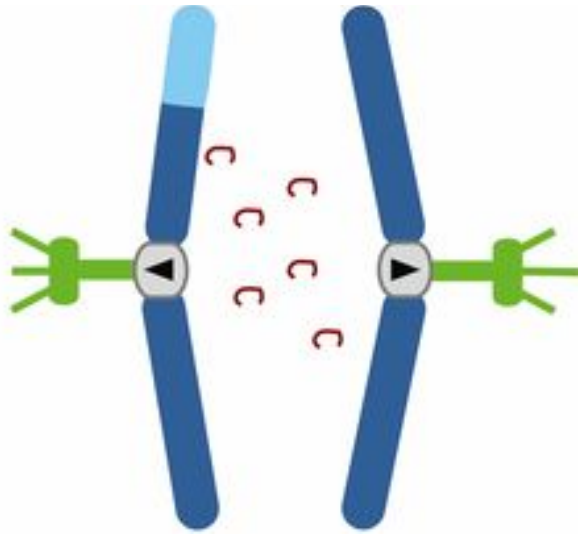
Genomic instability in Suv39h1/2 mice increases lymphomas



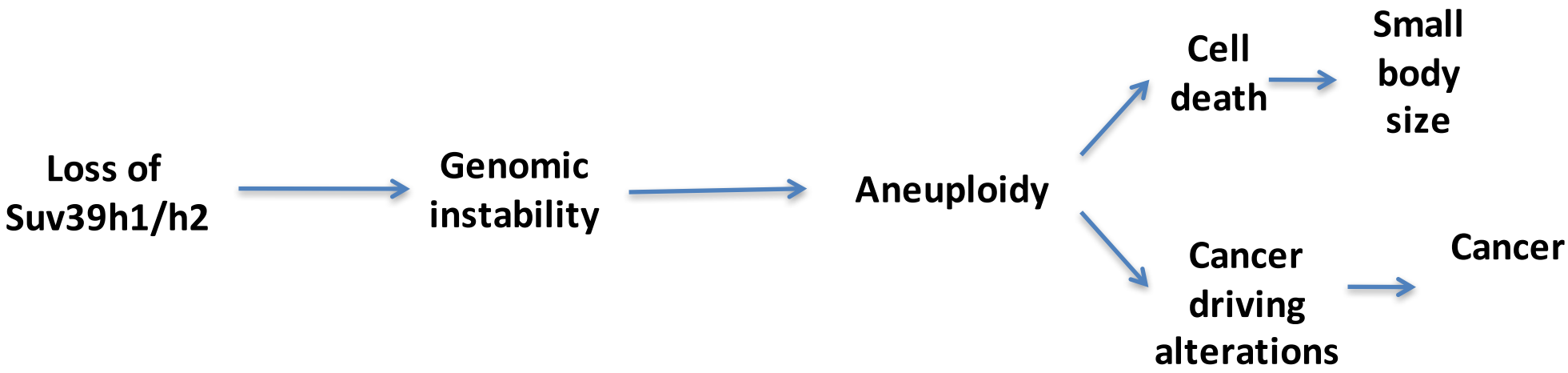
SUV39h HMTase activity is essential for fidelity in mitosis/meiosis



perientric heterochomatin



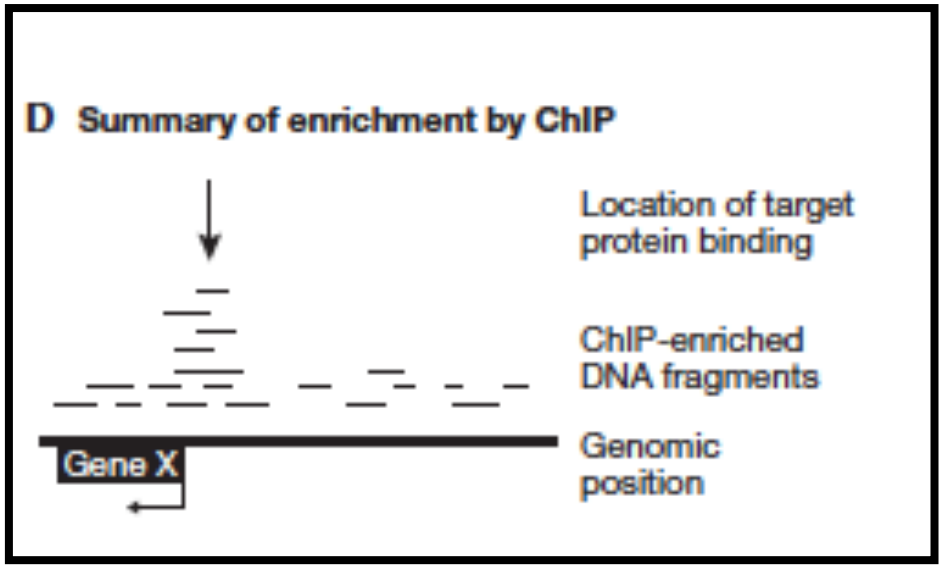
Correct heterochromatin structure supports chromosome segregation by the mitotic/meiotic spindle apparatus



CHROMATIN IMMUNOPRECIPITATION (ChIP) → DETAILLED ANALYSIS → Mapping modifications along the genome

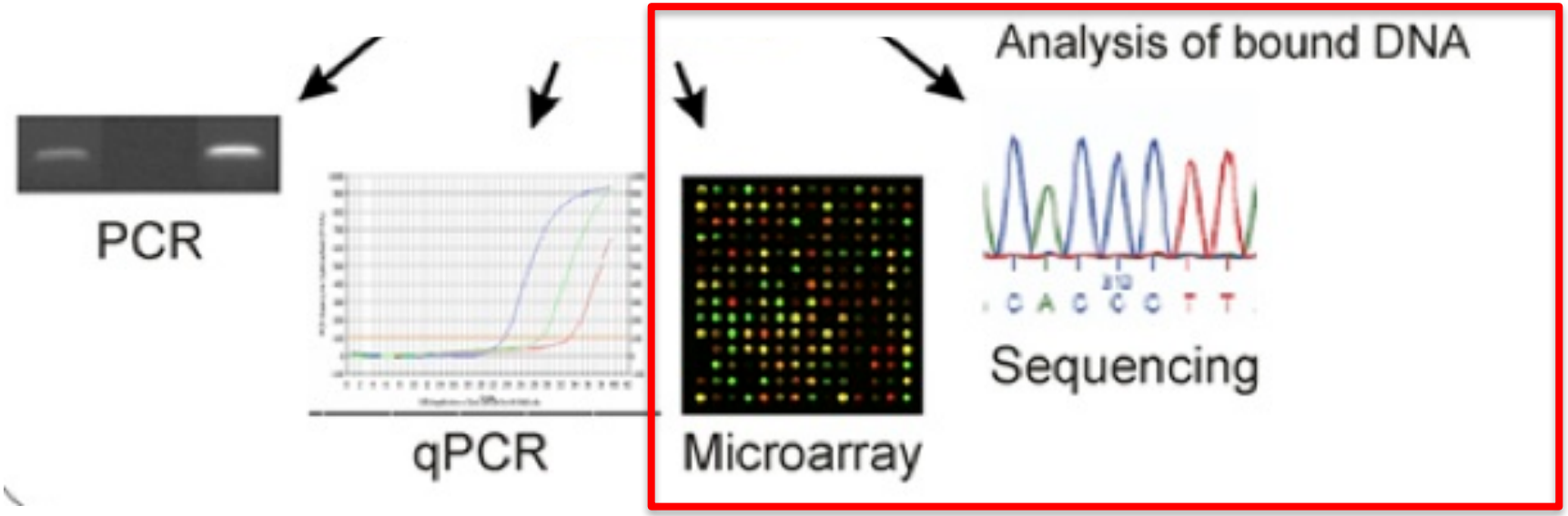
5. Analysis of ChIP DNA

Identification of DNA regions associated with the protein/modification of interest

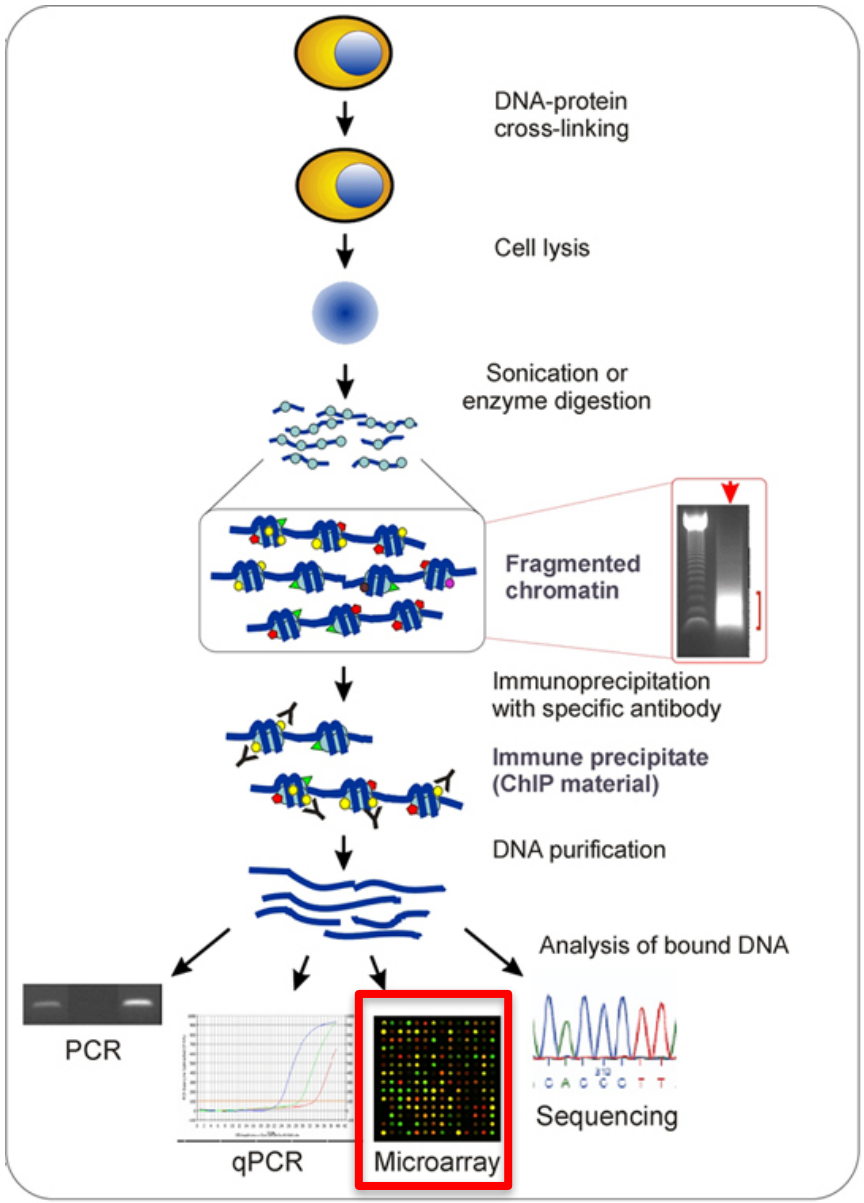


DNA from chromatin enriched in H3K9me
→ Correspond to precise sites in the genome

Methods of ChiP analysis



CHROMATIN IMMUNOPRECIPITATION (ChIP) → DETAILLED ANALYSIS → Mapping modifications along the genome



Cell model system:
i.e. Wild-type or Suv39 dn cells
that grow in cell culture dish

1. Crosslink chromatin
(treatment of cells with Paraformaldehyde

2. Sonicate crosslinked cells

3. Incubate chromatin fragments with antibodies raised against
H3K9me3

4. Recover antibody-chromatin complexes using a resin that binds to the constant region of antibodies

5. Elute chromatin at high salt concentration and revert crosslinks at high temperature

6. Digest protein with protease K and RNA with RNase

5. Elute chromatin at high salt concentration and revert crosslinks at high temperature

6. Digest protein with protease K and RNA with RNase

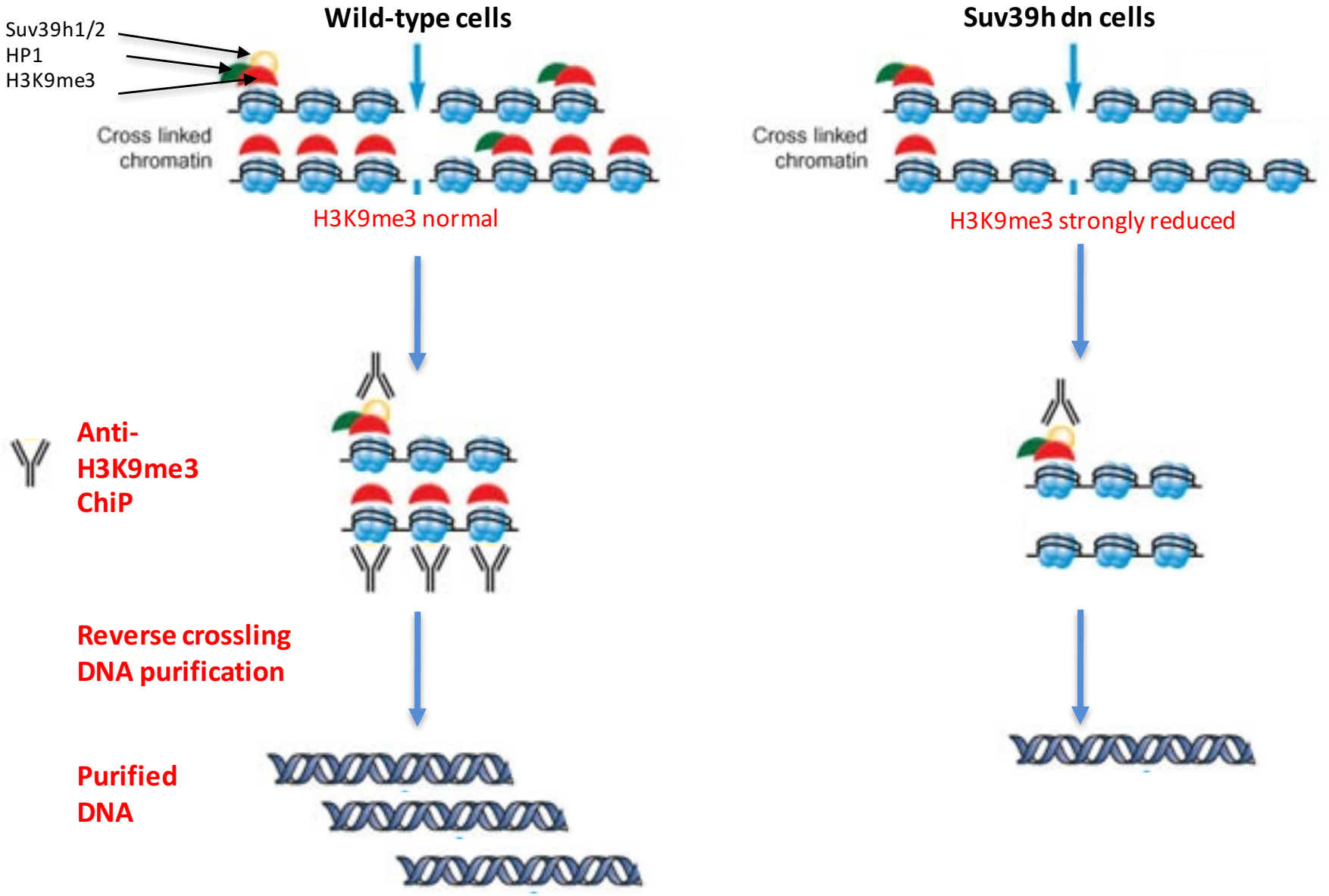
7. Purify DNA and precipitate DNA

8. Measure the amount of immunoprecipitated DNA In control versus Suv39h dn cells

METHODS:

Quantitative PCR: design PCR oligos to amplify defined sequences in immunoprecipitates. Per PCR reaction only one Locus can be examined by real-time PCR

ChIP on ChIP: Analysis of epigenetic information across a high number of genomic sites in the genome

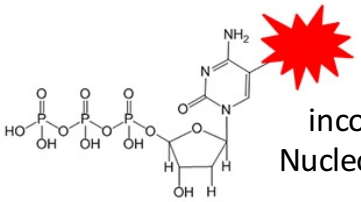


ChIP on ChIP: Analysis of epigenetic information across a high number of genomic sites in the genome

Purified DNA From ChIP

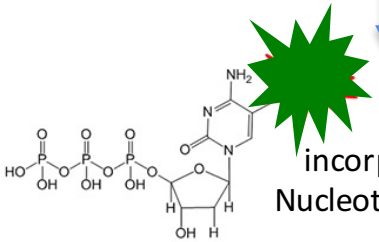
Wild-type cells

Suv39h dn cells



Fluorescently labeled probe

Denaturalization of DNA; priming of DNA synthesis; incorporation of fluorescently labeled Nucleotides into newly synthesized strand



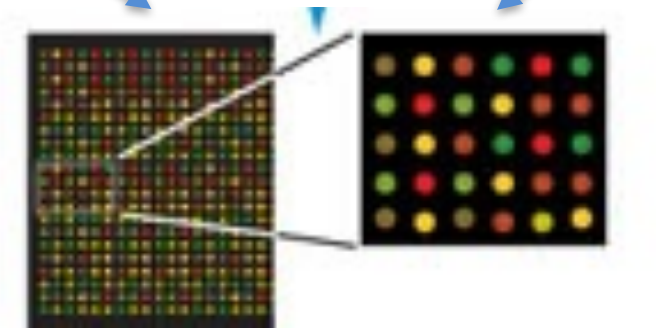
Denaturalization of DNA; priming of DNA synthesis; incorporation of fluorescently labeled Nucleotides into newly synthesized strand

Cy5- labeled (~650 excitation, 670 nm emission)

Cy3- labeled (550 nm excitation, ~570 nm emission)

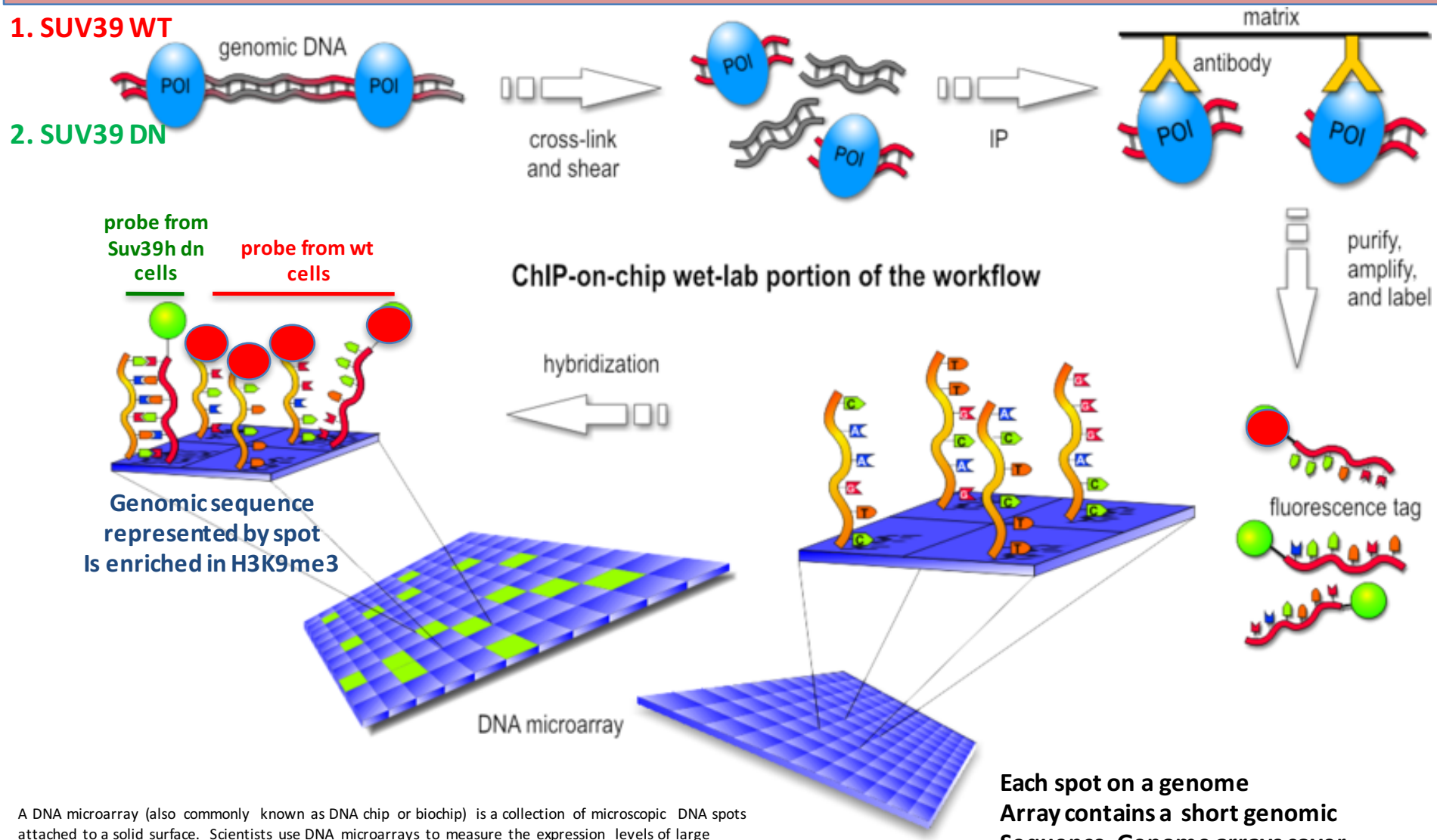
Mix probes

Hybridize to a genome array



Measure Green/Red ratio

ChIP on ChIP: Analysis of epigenetic information across a high number of genomic sites in the genome



A DNA microarray (also commonly known as DNA chip or biochip) is a collection of microscopic DNA spots attached to a solid surface. Scientists use DNA microarrays to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome. Each DNA spot contains picomoles (10−12 moles) of a specific DNA sequence, known as probes (or reporters or oligos). These can be a short section of a gene or other DNA element that are used to hybridize a cDNA or cRNA (also called anti-sense RNA) sample (called target) under high-stringency conditions. Probe-target hybridization is usually detected and quantified by detection of fluorophore-, silver-, or chemiluminescence-labeled targets to determine relative abundance of nucleic acid sequences in the target

Each spot on a genome Array contains a short genomic Sequence. Genome arrays cover a large part of the Genome
A ChIP holds ca 20000 spots

ChIP on ChIP: Analysis of epigenetic information across a high number of genomic sites in the genome

Purified DNA From ChIP

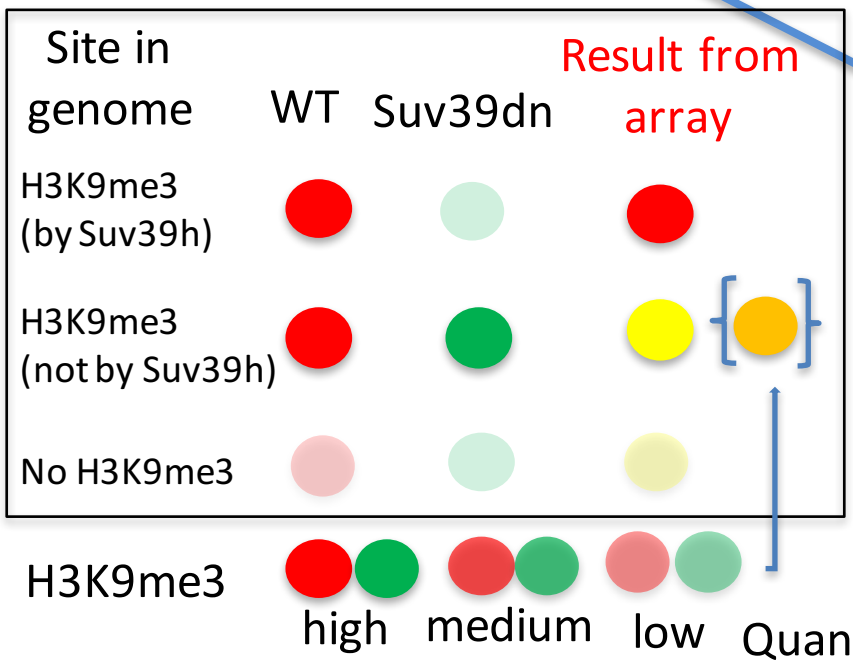
Wild-type cells

Suv39h dn cells

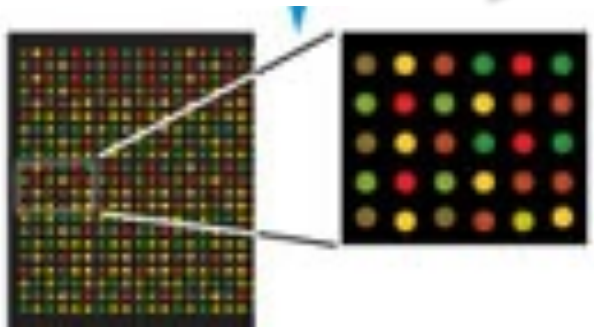
Fluorescently labeled probe

Cy5- labeled
(~650 excitation, 670 nm emission)

Cy3- labeled
(550 nm excitation, ~570 nm emission)



Mix probes

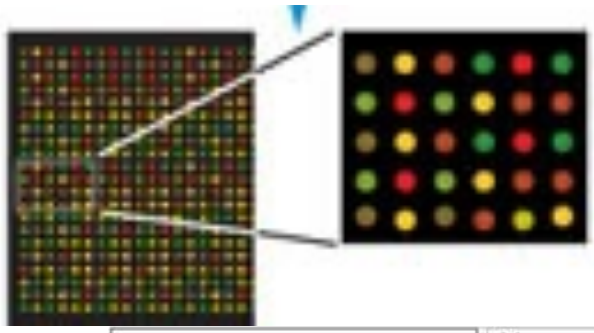


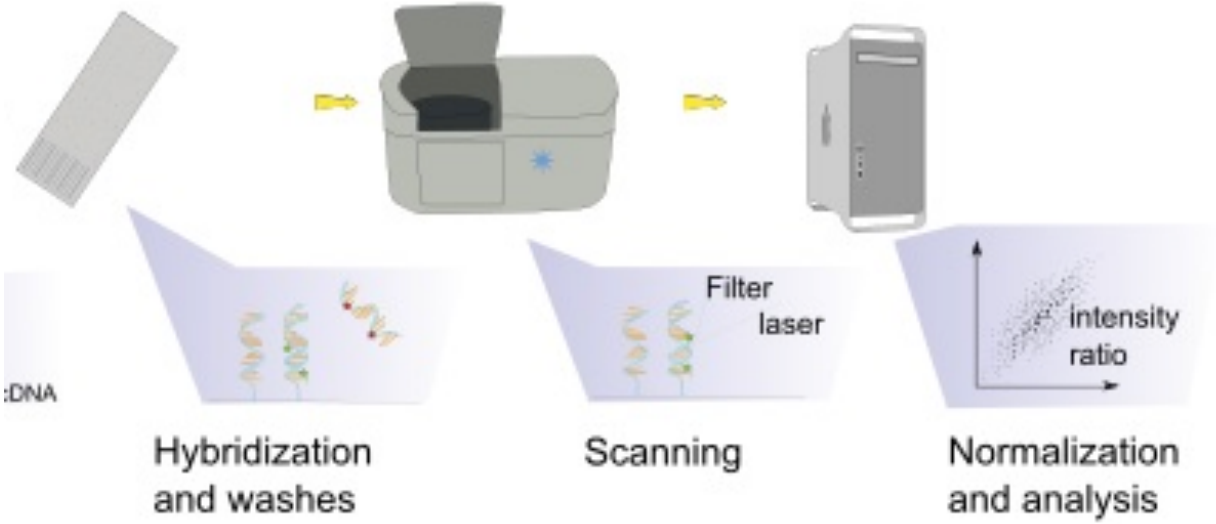
Measure Green/Red ratio
Hybridize to a genome array

Scan:

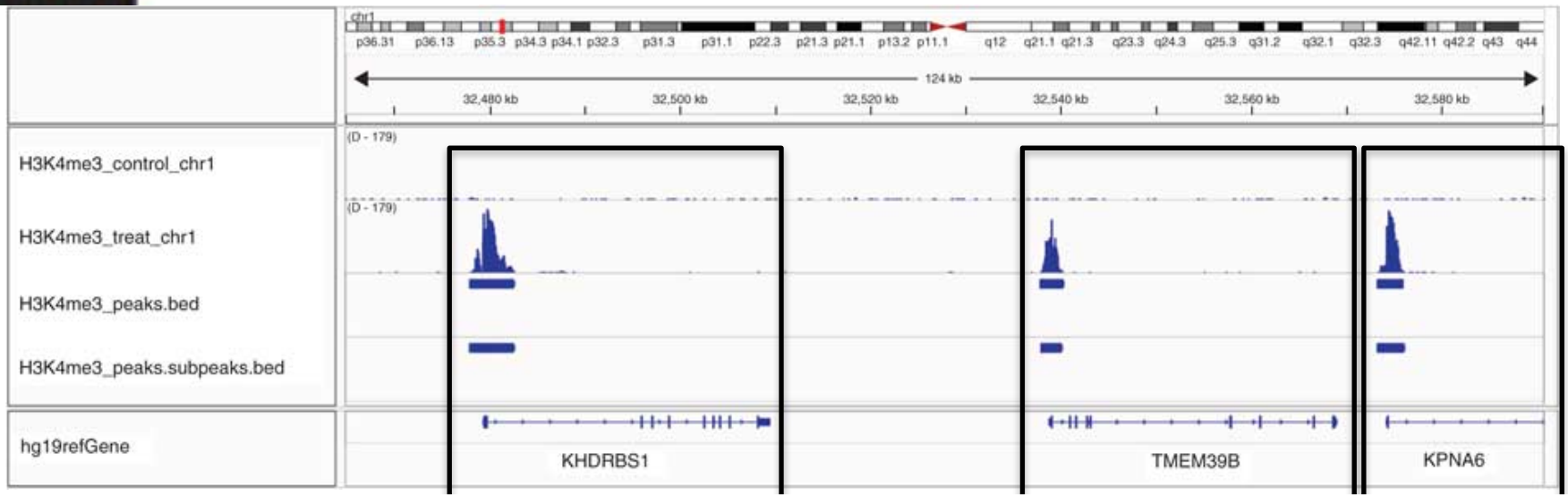
1. *Red channel (+normalize signal)*
2. *Green channel (+normalize signal)*

ChIP on ChIP: Analysis of epigenetic information across a high number of genomic sites in the genome

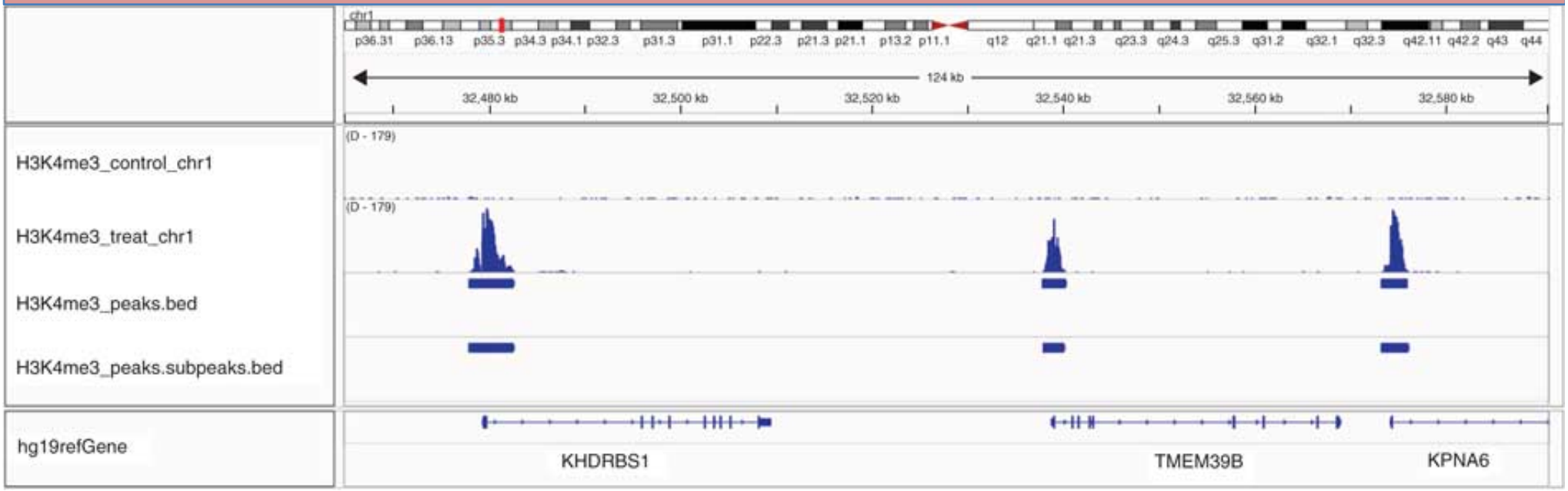




Arrays do not contain the entire human/mouse genome
Are enriched for classes of seqeunces
(promoter, enhancer, gene bodies etc…)



ChIP on ChIP: Analysis of epigenetic information across a high number of genomic sites in the genome

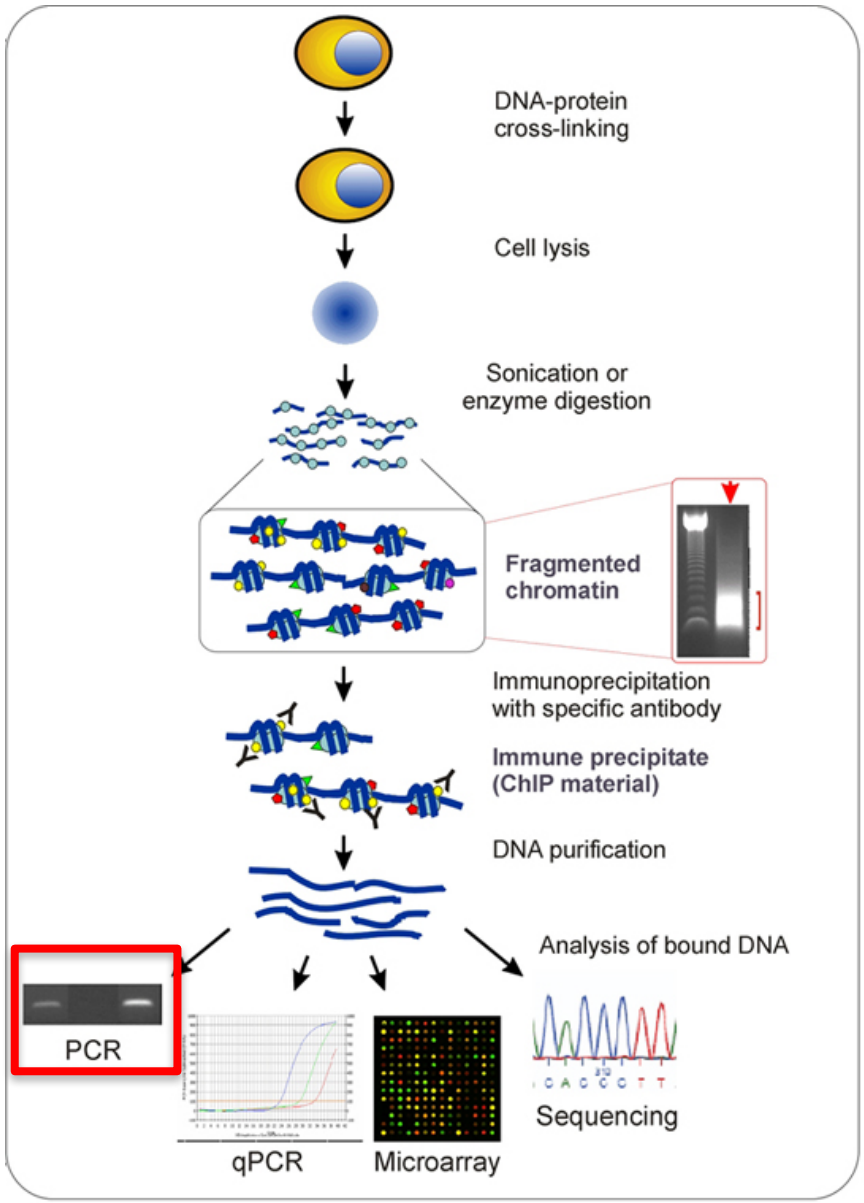


Advantage: low tech, cheap

Disadvantage: low resolution, no data on number of molecules – just proportions; laborious to reach a good genome coverage

Already outdated → state of the art: ChIP seq

ChIP seq: Analysis of epigenetic information on the single nucleotide level
→ GENERATION OF GENOME WIDE EPIGENTIC MAPS



Cell model system:
i.e. Wild-type or Suv39 dn cells
that grow in cell culture dish

1. Crosslink chromatin
(treatment of cells with Paraformaldehyde

2. Sonicate crosslinked cells

3. Incubate chromatin fragments with antibodies raised against
H3K9me3

4. Recover antibody-chromatin complexes using a resin that binds to the constant region of antibodies

5. Elute chromatin at high salt concentration and revert crosslinks at high temperature

6. Digest protein with protease K and RNA with RNase

5. Elute chromatin at high salt concentration and revert crosslinks at high temperature

6. Digest protein with protease K and RNA with RNase

7. Purify DNA and precipitate DNA

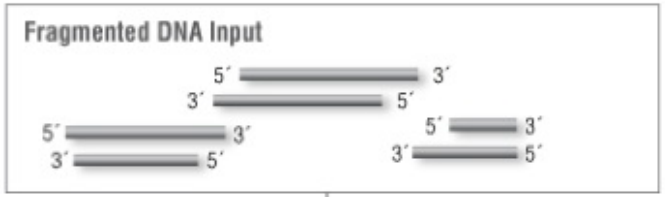
8. Measure the amount of immunoprecipitated DNA In control versus Suv39h dn cells

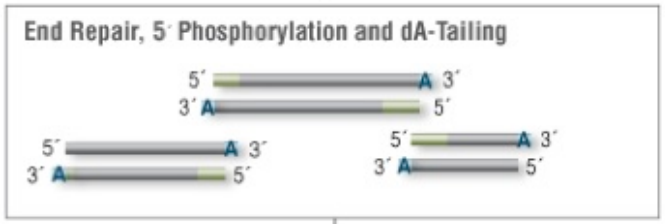
METHODS:

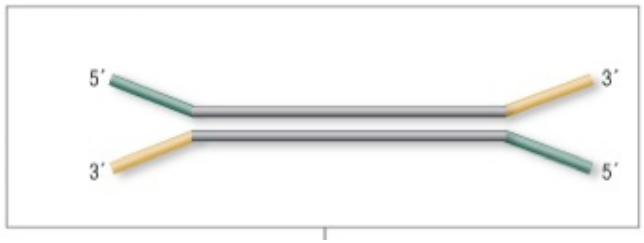
Massive parallel sequencing of immunoprecipitated DNA
Permits to obtain epigenetic information on the single nucleotide level

ChIP seq: Analysis of epigenetic information on the single nucleotide level
→ GENERASTION OF GENOME WIDE EPIGENTIC MAPS

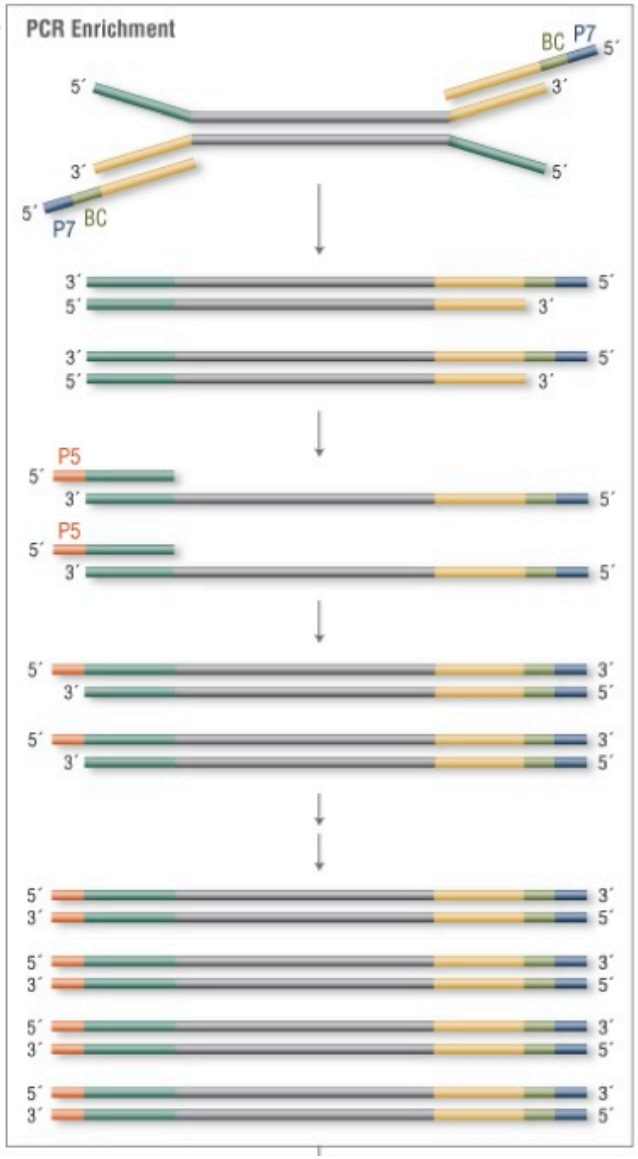
Use immunoprecipitated DNA to generate DNA libraries:







fusion of DNA with linker oligo



Linkers serve as uniform primer binding sites. This allows the amplification of the entire DNA library using only 2 types of oligonucleotides

Amplified library

READY FOR MASSIVE PARALLEL SEQEUNCING

ChIP seq: Analysis of epigenetic information on the single nucleotide level
→ GENERASTION OF GENOME WIDE EPIGENTIC MAPS

Illumina Massively Parallel Sequencing

https://www.illumina.com/company/video-hub/pfZp5Vgsbw0.html

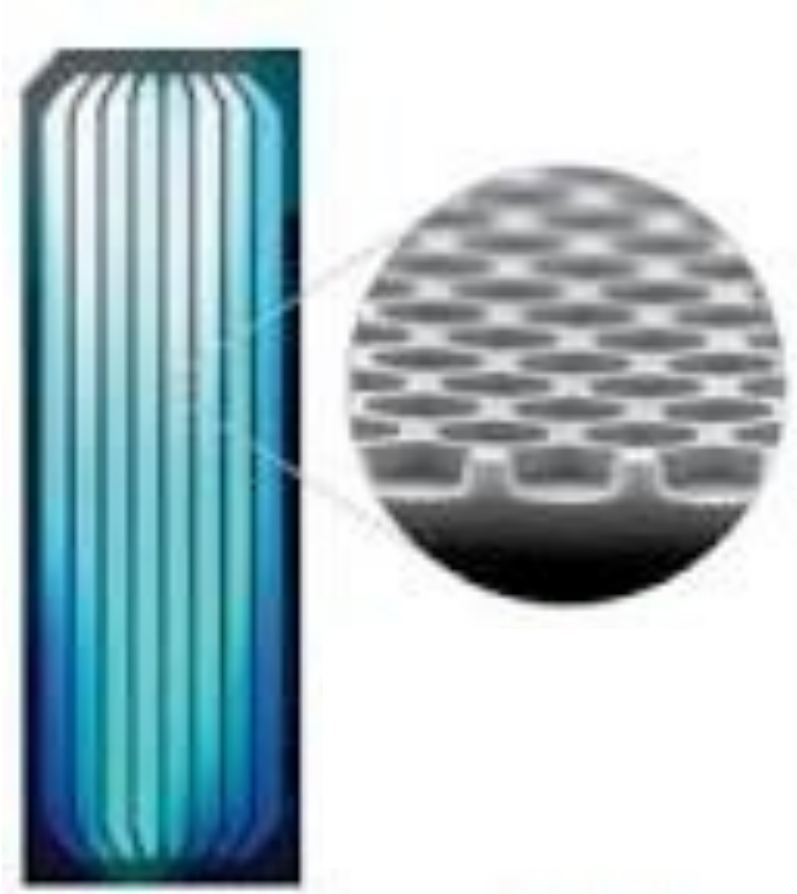
The heart of the Illumina Massive Parallel Sequencer is the "FLOW-CELL". A surface with millions of small wells that allow thousands of Sanger-sequencing reaction In parallel = "massive parallel sequencing". In each well a SINGLE MOLECULE of DNA Is amplified and sequenced

Illumina offers the most potent massive sequencing instruments – leader on the market

https://www.youtube.com/watch?v=pfZp5Vgsbw0

ChIP seq: Analysis of epigenetic information on the single nucleotide level
→ GENERASTION OF GENOME WIDE EPIGENTIC MAPS

CLUSTER AMPLIFICATION:



Flow cell contains surface with millions of wells

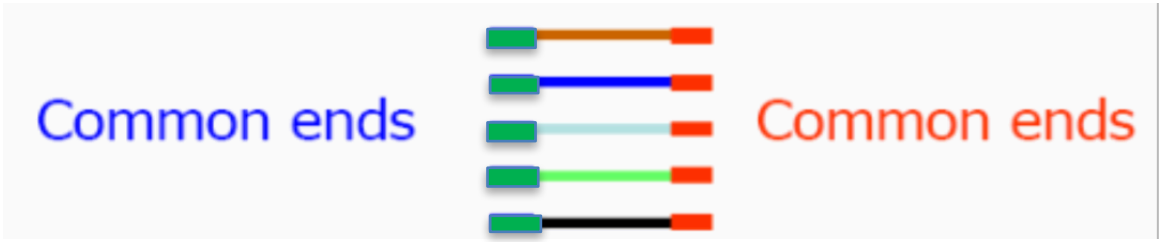
→Each well contains beads mounted with 2 species of oligonucleotides that hybridize with adaptor oligos of DNA library

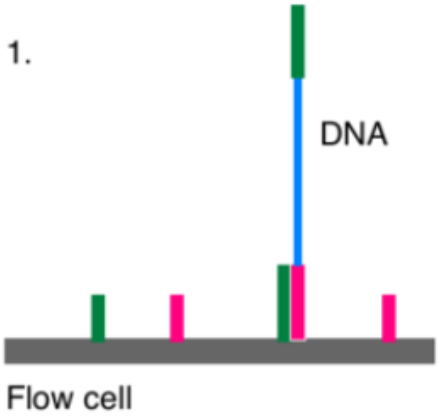
→DNA library will be loaded onto the flow cell in a determined concentration:

ONLY ONE MOLECULE OF DNA WILL BE PROCESSED FOR SEQUENCING IN A SINGLE WELL

CLUSTER AMPLIFICATION:

-making DNA library (~300bp fragments)
-ligation of adapters **A** and **B** to the fragments





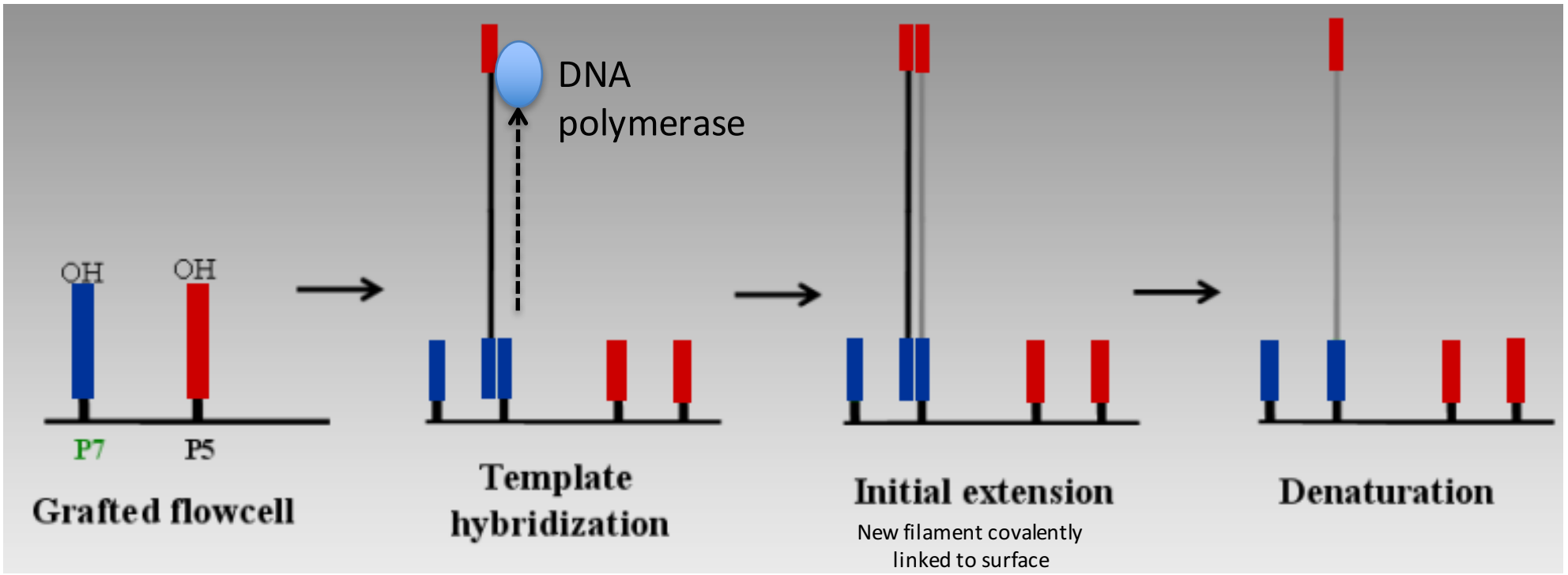
1 well, covered with
millions of 2 types of oligos

1 well in a flow-cell with
billions of wells

- ***complementary*** primers are ligated to the surface
- pairing with ChiP ed ssDNA at random position in the well of the flow cell

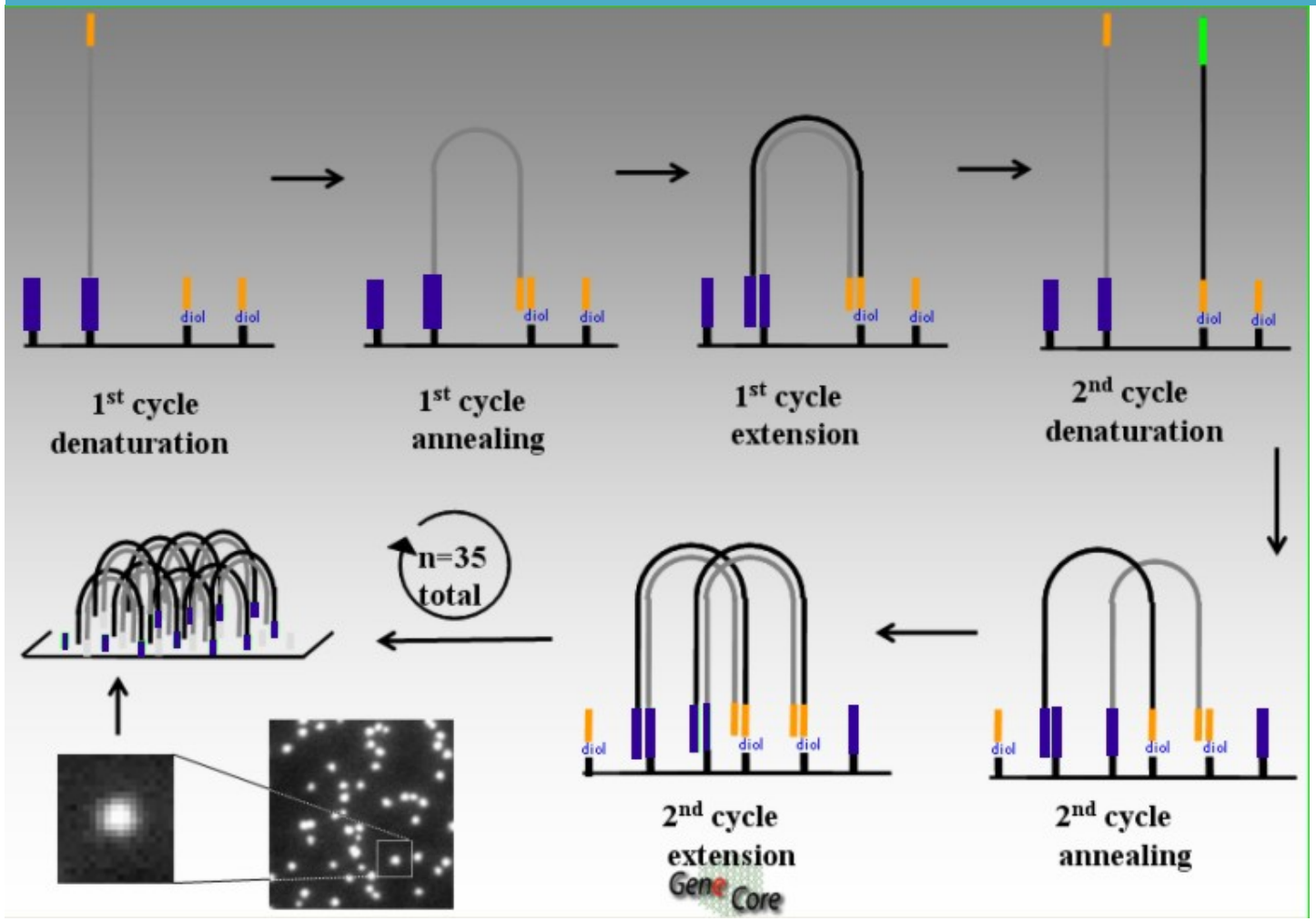
CLUSTER AMPLIFICATION:

Bridge amplification: takes place on surface of beads (each bead is mounted with 2 species of oligos; each oligo can hybridize to a DNA library fragment): initiation



On the surface: complementary oligos

CLUSTER AMPLIFICATION:



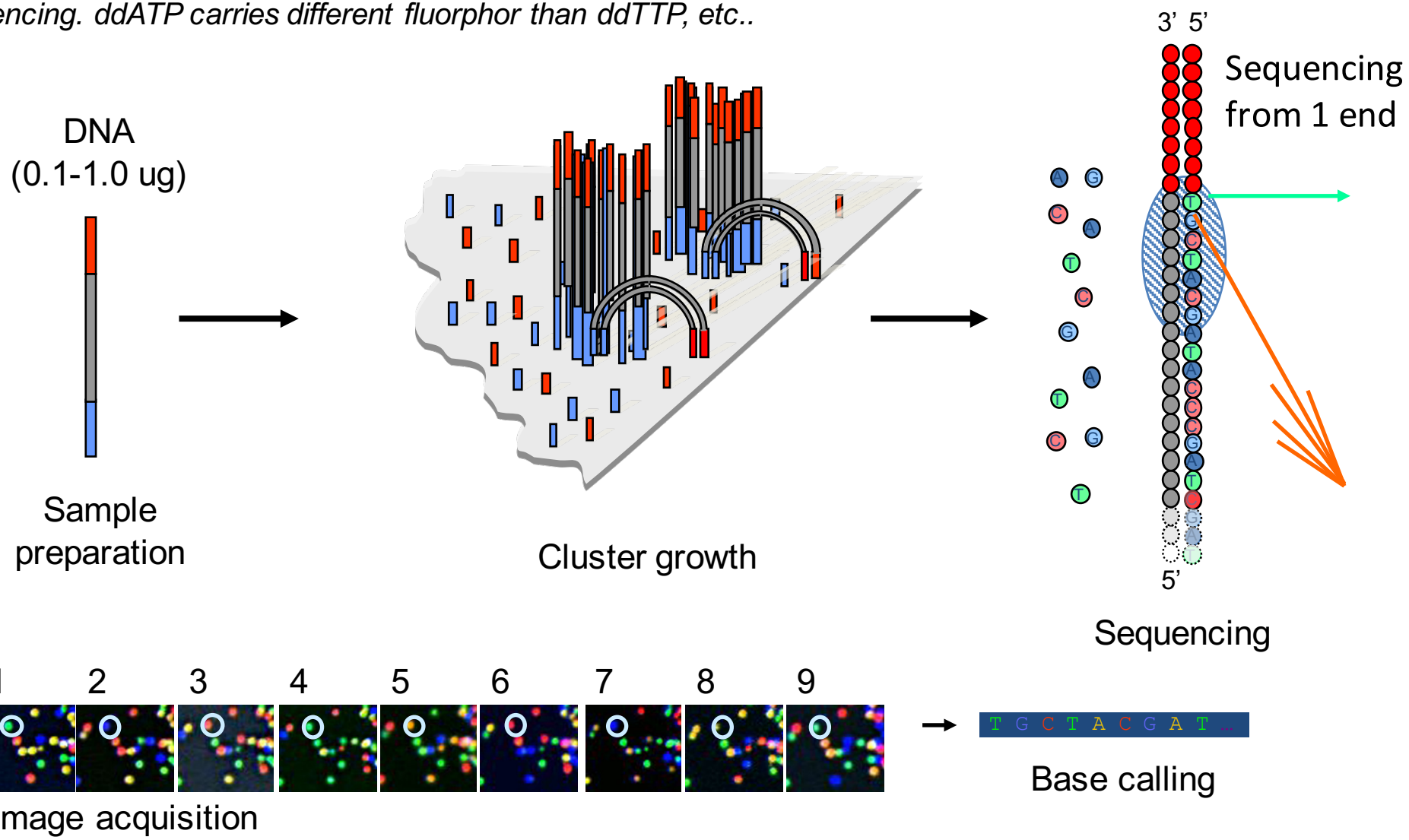
Reverse and Forward strand amplification by "bridge amplification"

Reverse strands are cleaved of → only forwards strand remains for sequencing

Illumina Sequencing Technology

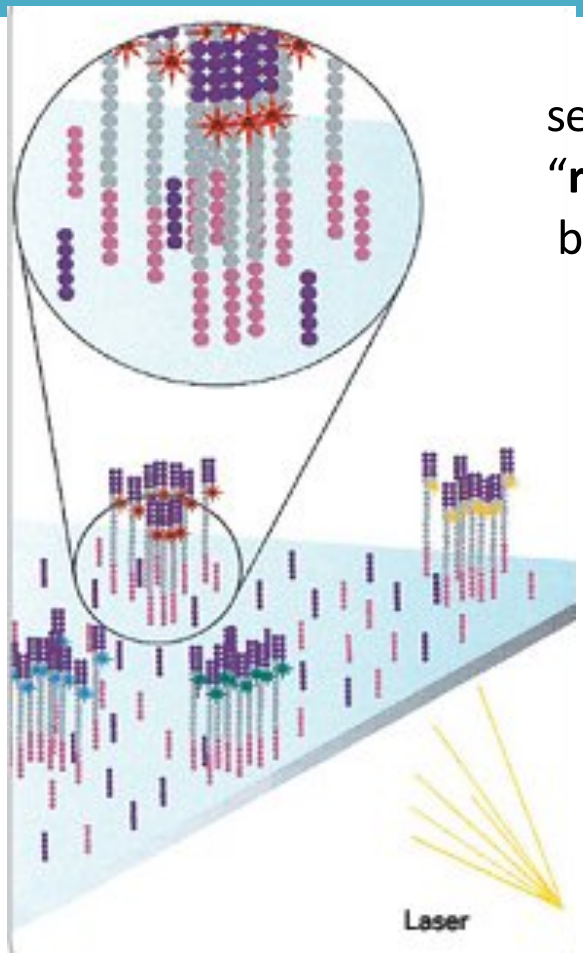
Robust Reversible Terminator Chemistry Foundation

In each round of sequencing a fluorescently labelled ddNTP will be used for sequencing. ddATP carries different fluorphor than ddTTP, etc..

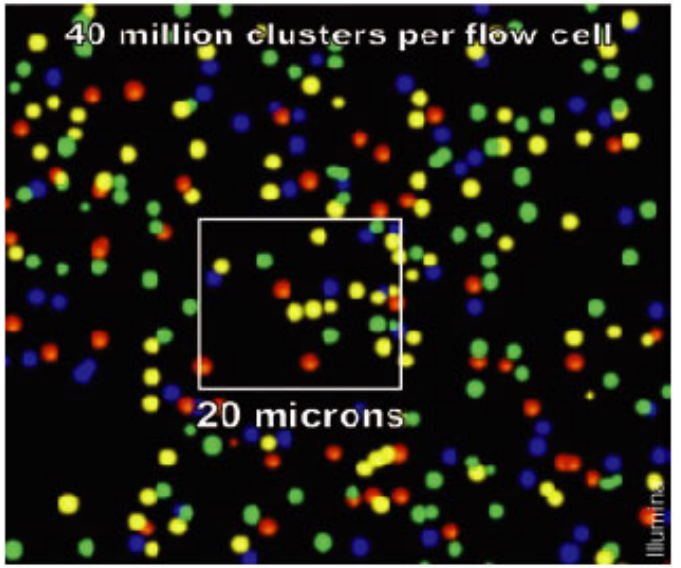


= "sequencing by synthesis"

Illumina: massive parallel sequencing:



sequencing by synthesis:
"**reverible terminator**" nucleotides
blocked + fluorescently labeled



1. **Synthesis = incorporation of fluorescent nucleotide: blocking synthesis**
2. **4. Scanning of fluorescent signal**
3. **dye cleavage + elimination**
4. **wash step**
1. **Synthesis = incorporation of fluorescent nucleotide: blocking synthesis**

READ LENGTH: ca: 150nt from each primer (2x150nt = 300nt)

illumina.com

Illumina: paired end sequencing increases information content

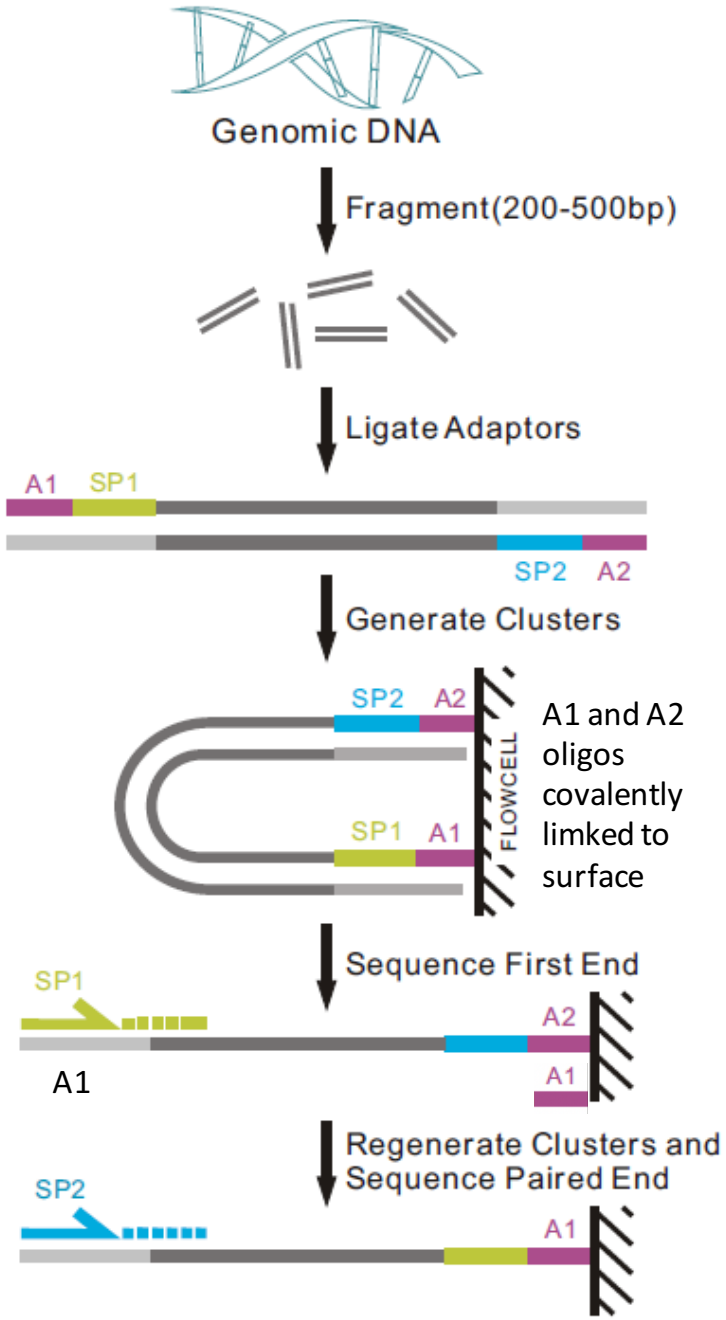


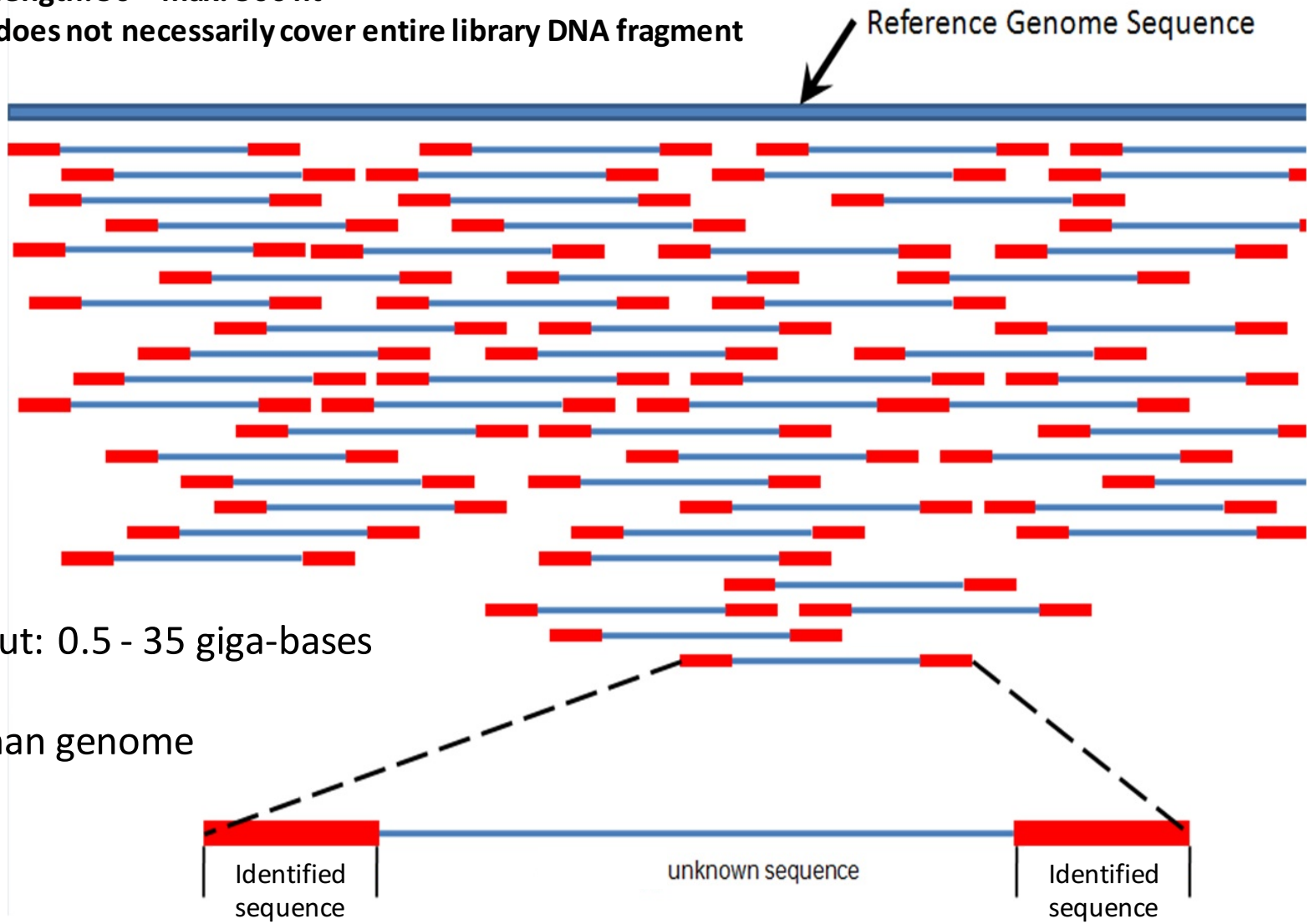
Figure 1-2-1 Pipeline of paired-end sequencing (www.illumina.com)

After 1° forward strand sequencing, A1 anneals to A1 in nanowell
→ reverse strand synthesis
→ forward (template strand= cleaved off
→ sequencing by synthesis of reverse strand

https://www.youtube.com/watch?v=9YxExTSwgPM

Data analysis: obtained sequence reads are aligned along genomic DNA sequence → high number of reads necessary to obtain full sequence coverage

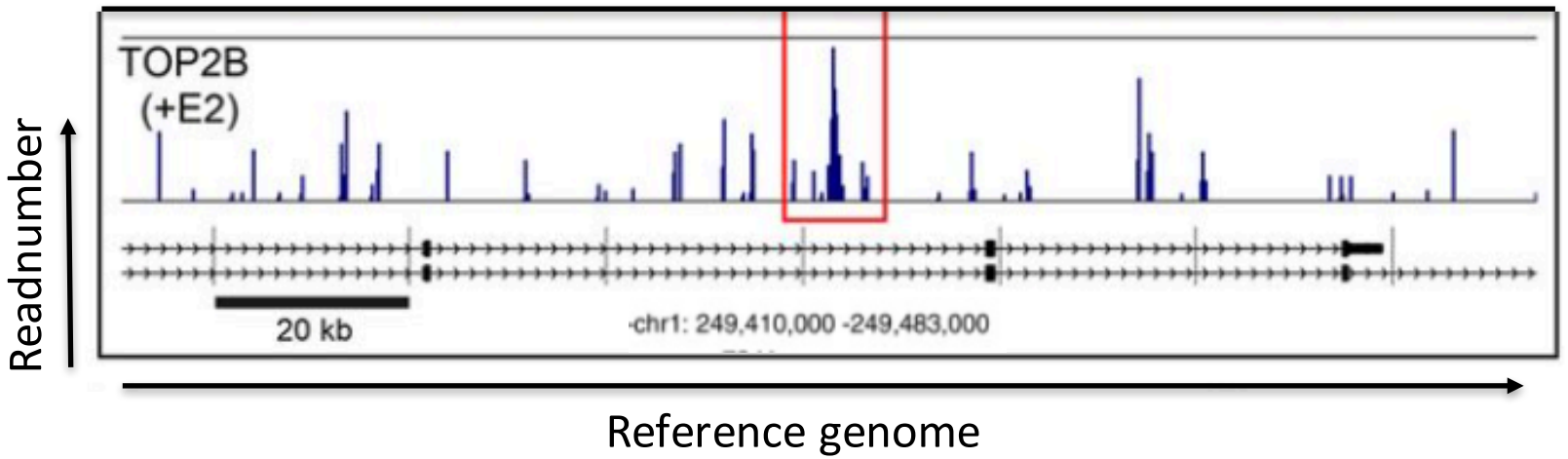
Read length: 50 – max. 300 nt
Read does not necessarily cover entire library DNA fragment



Max. output: 0.5 - 35 giga-bases
$=3.5*10^{10}$
= 10x human genome

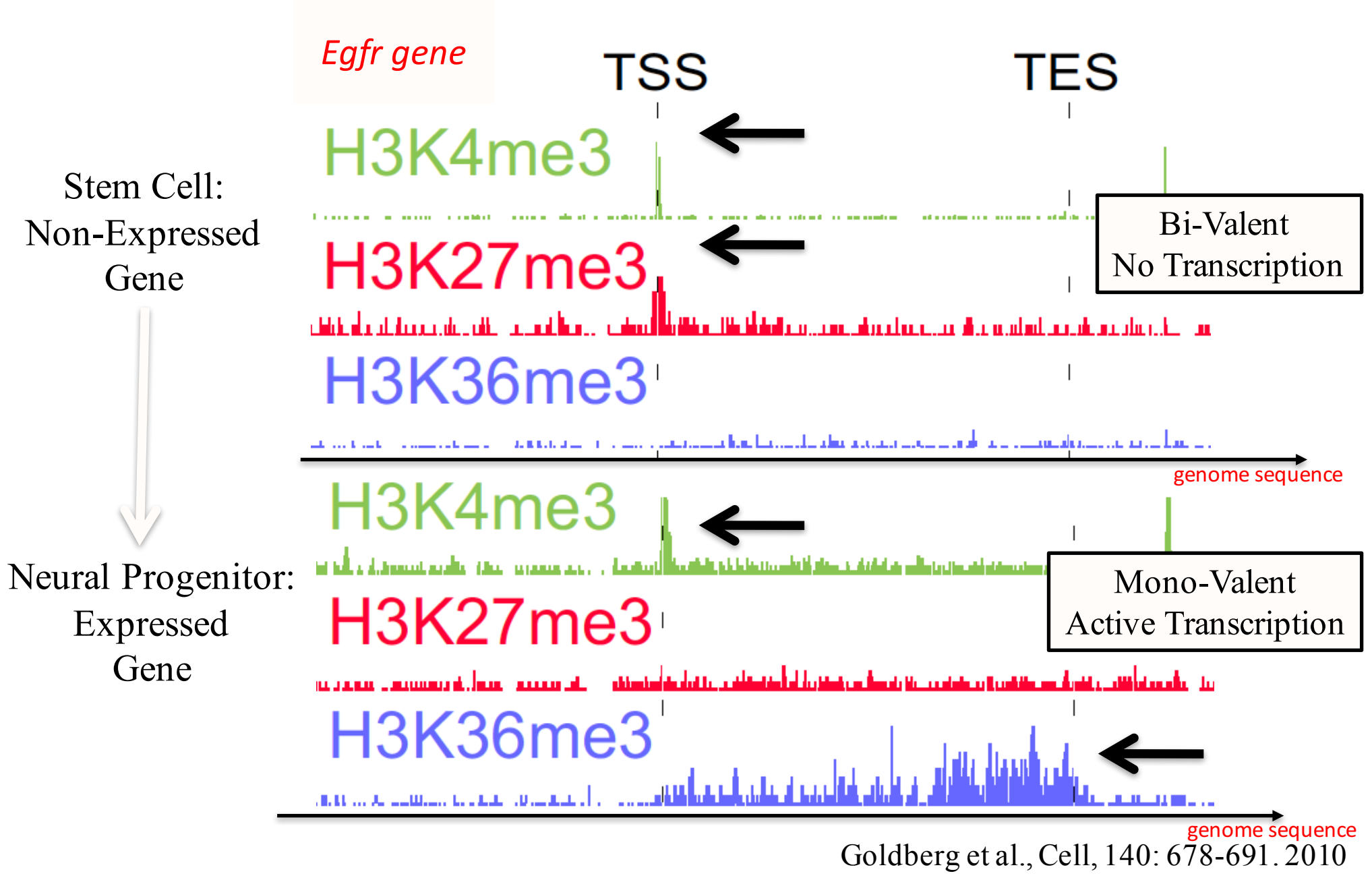
Sequence derived from one amplified cluster

Data analysis: obtained sequence reads are aligned along genomic DNA sequence → high number of reads necessary to obtain full sequence coverage



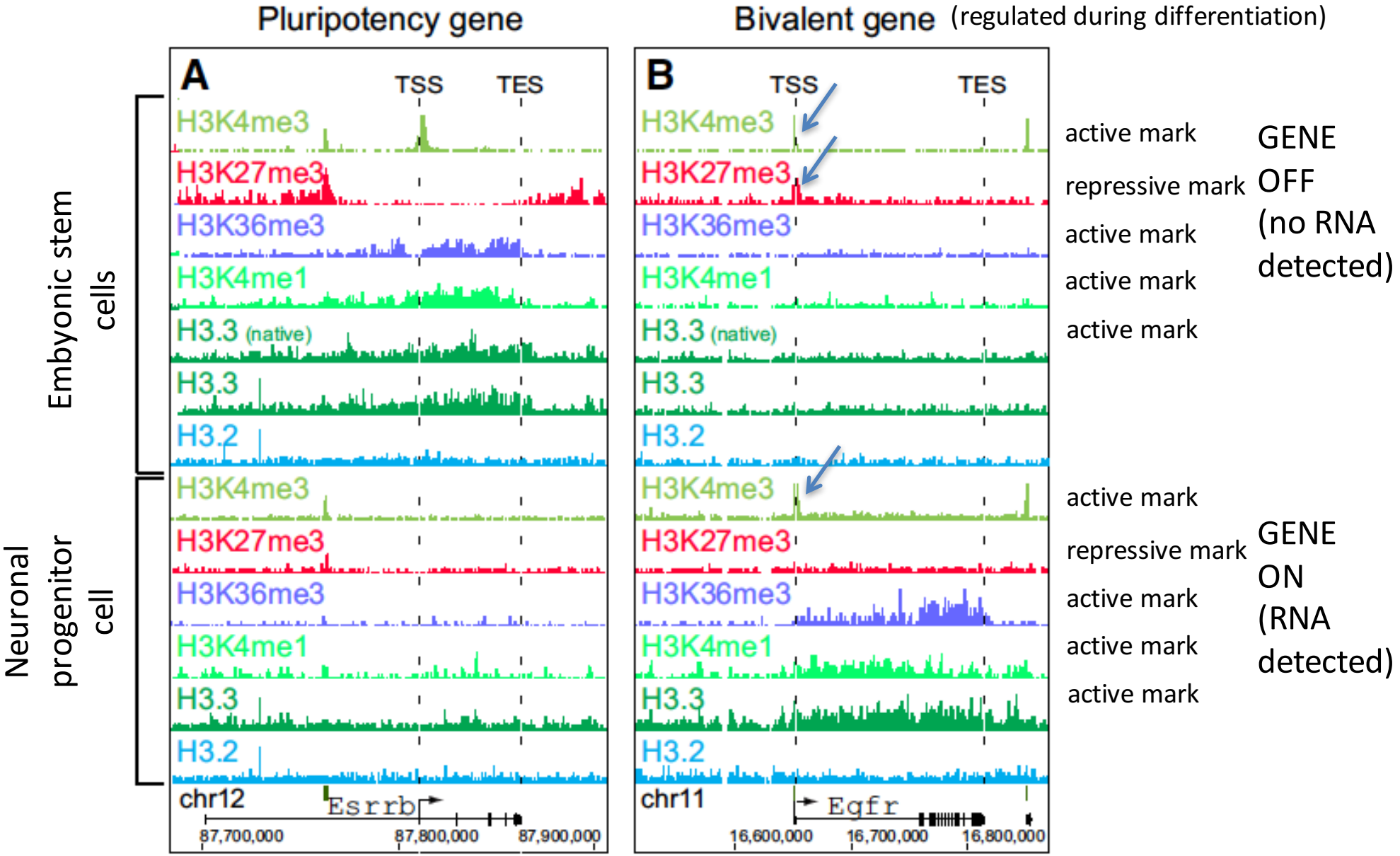
Epigenetic information at single nucleotide level

BIOINFORMATICS ANALYSIS:
Mapping ChIP seq reads agins the human genomic sequence



Goldberg et al., Cell, 140: 678-691. 2010

Mapping the epigenetic landscape enables to define “key rules” to define the epigenetic code of active and silent genes



Goldberg et al., Cell, 140: 678-691. 2010