# L2.4

1. more on DNA methylation
2. tech discussion Research Paper 2

Importance of DNA methylation for epigenetic inheritance (esp. Mammals)

- CGI (CG-rich islands) are frequent at gene promoters

- there are established molecular links between DNA-methylation machinery and the histone repressive complexes

- CpG-methylation profiles display typical cycles accompanying all developmental stages

- Inhibitors of DNA-methylation induce de-differentiation events

- CGI dysmetabolism is frequently linked to disease (esp. cancer)
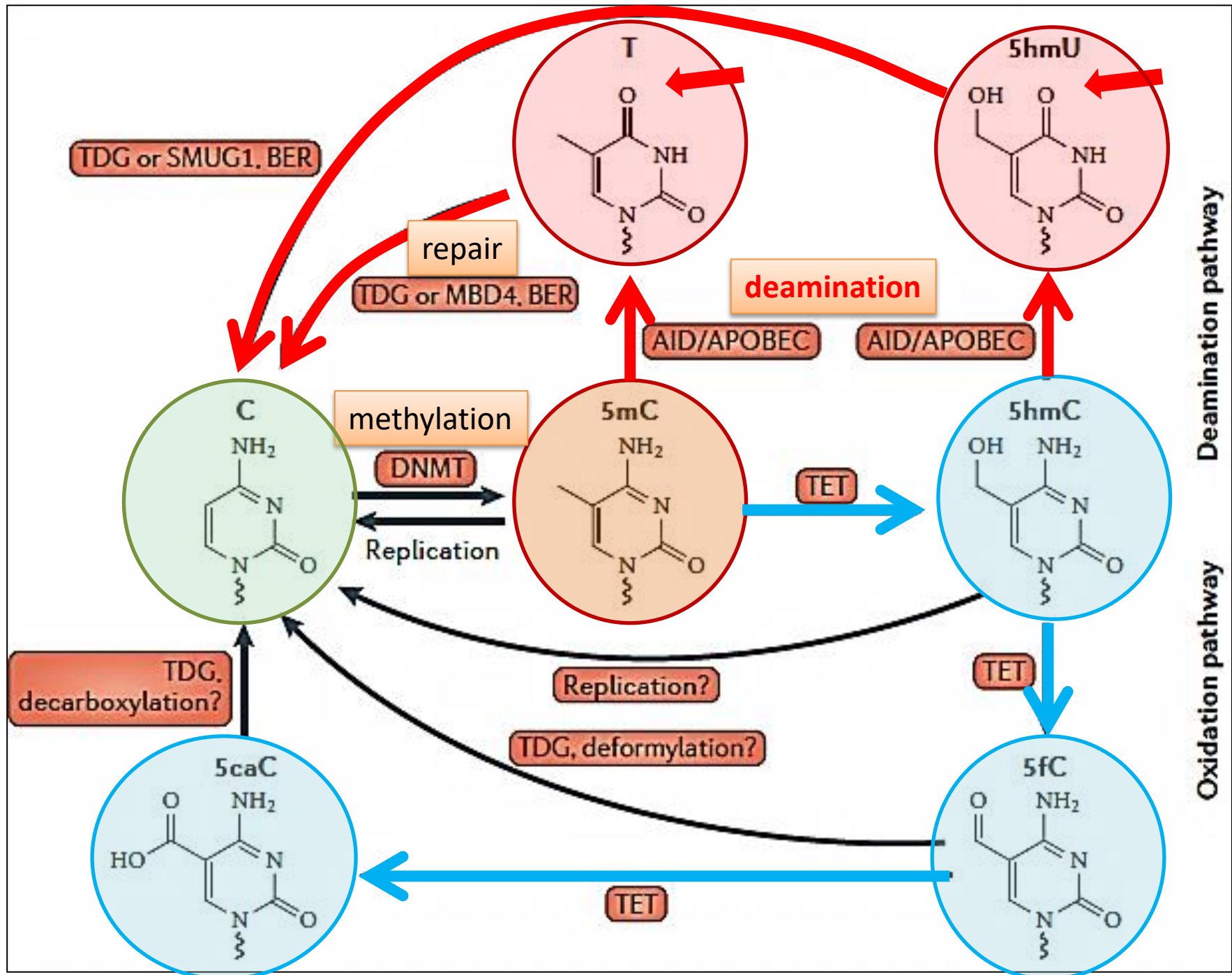
methods for genome-wide DNA methylation analysis

Gupta et al., 2010, Biotechniques 49: iii-xi
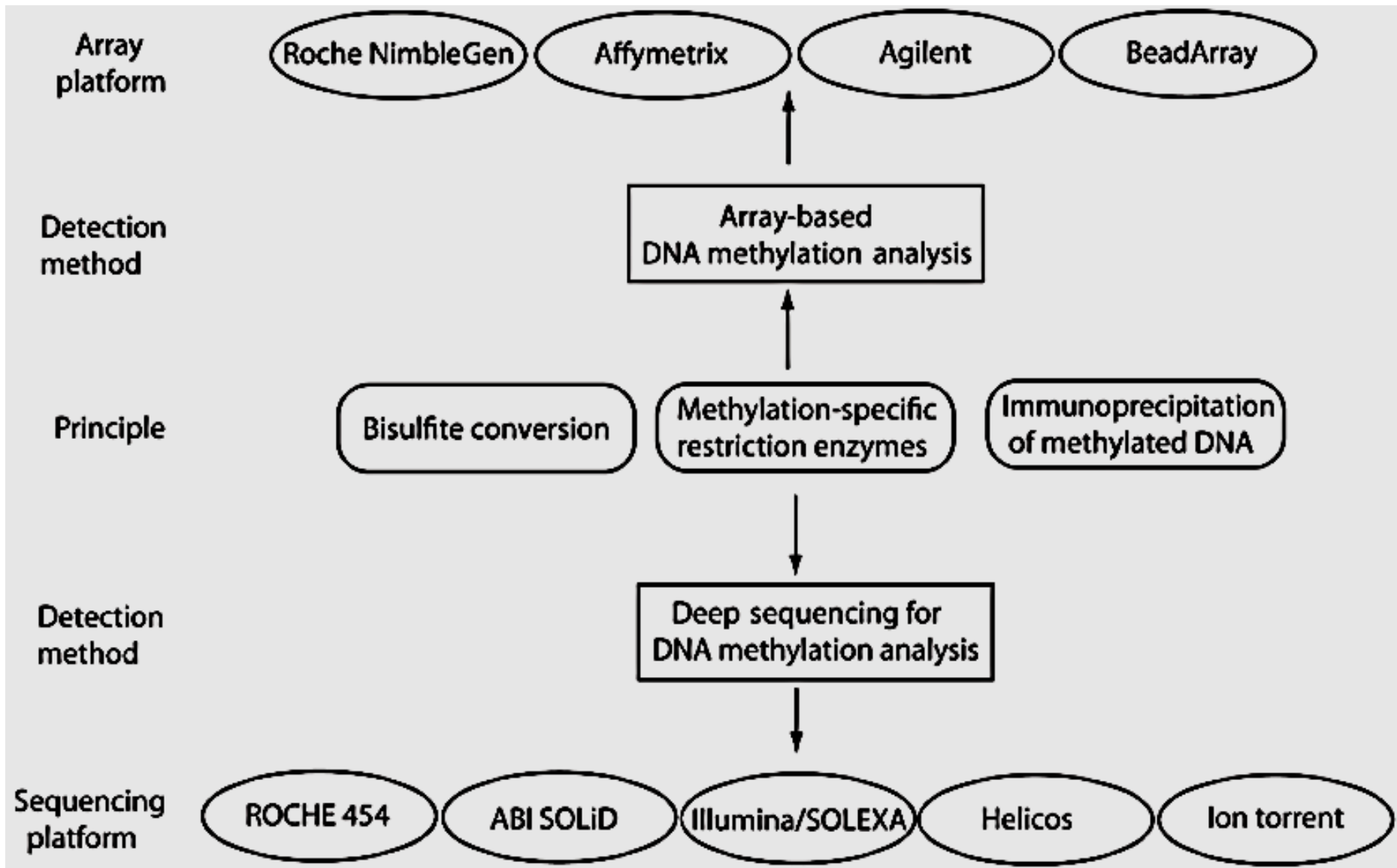
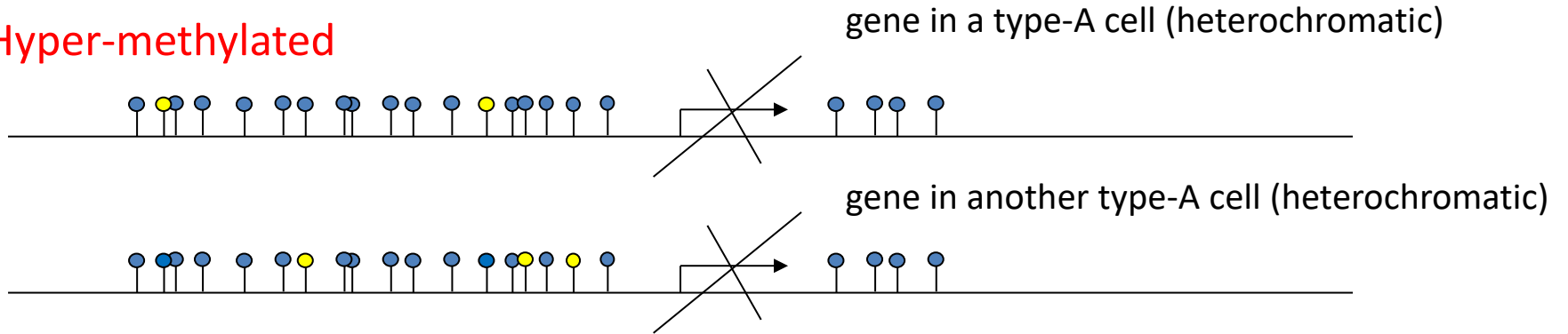Whole genome methylation by bisulfite modification is **very difficult** !

....remember that CpG methylation is stochastic, every cell has its own profile

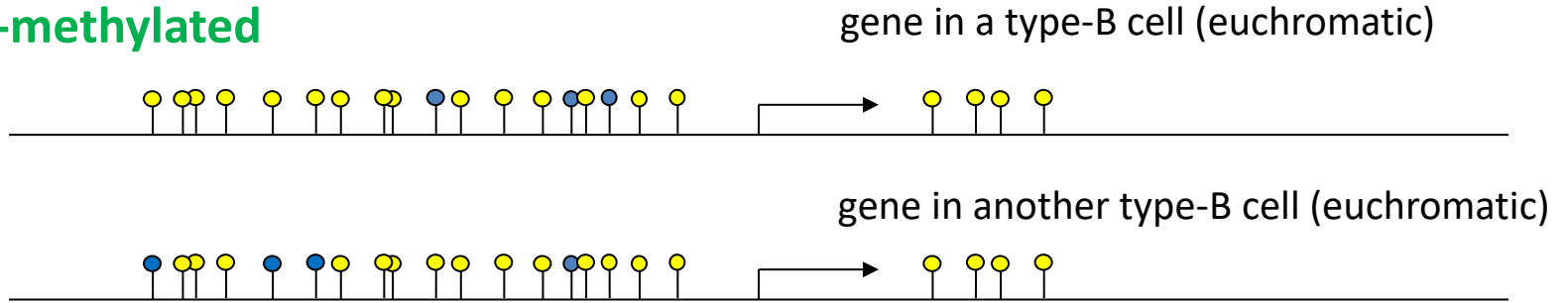Very **redundant** sequencing is needed

Aligning database should contain <u>all possible variations</u>

Warning! – CpG methylation is not site-specific

Hyper-methylated

gene in a type-A cell (heterochromatic)

gene in another type-A cell (heterochromatic)

Hypo-methylated

gene in a type-B cell (euchromatic)

gene in another type-B cell (euchromatic)

unmethylated CpG

methylated CpG

Since CpG methylation of a given DNA fragment is <u>different from cell to cell</u>, sequencing needs either: **cloning + Sanger**   <u>or</u>   **NGS**

e.g. (Sanger sequencing) :

1) extract DNA from cells

2) bisulfite treatment

3) PCR the fragment using "side" primers

4) clone individual fragment in a plasmid vector

5) sequence a <u>representative</u> number of clones (if Sanger)
   today NGS preferred and equal cost
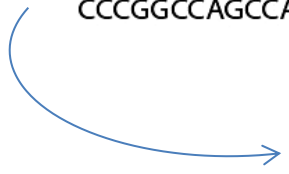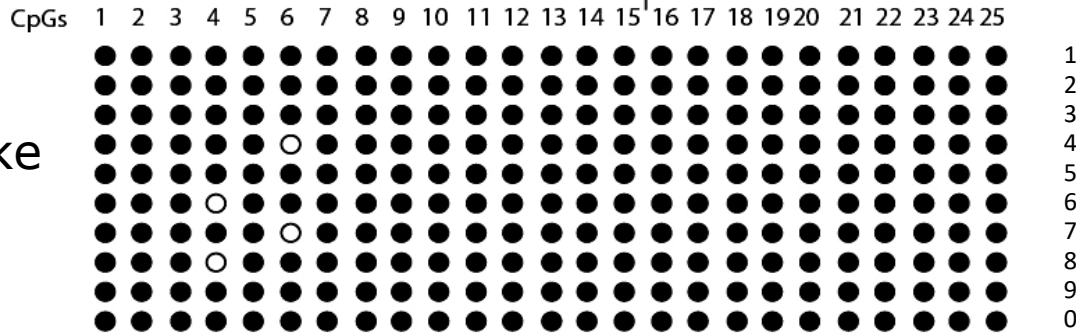
E-Cadherin gene promoter

AGACCCTAGCAACTCCAGGCTAGAGGGTCACCGCGTCTATGCGAGGCCGGGTGGGCGGGCCGTCAGCTCCGCCCTG
GGGAGGGGTCCGCGCTGCTGATTGGCTGTGGCCGGCAGGTGAACCCTCAGCCAATCAGCGGTACGGGGGGCGGTG
CTCCGGGGCTCACCTGGCTGCAGCCACGCACCCCCTCTCAGTGGCGTCGGAACTGCAAAGCACCTGTGAGCTTGCG
GAAGTCAGTTCAGACTCCAGCCCGCTCCAGCCCGGCCCGACCCGACCGCACCCGGCGCCTGCCCTCGCTCGGCGTC
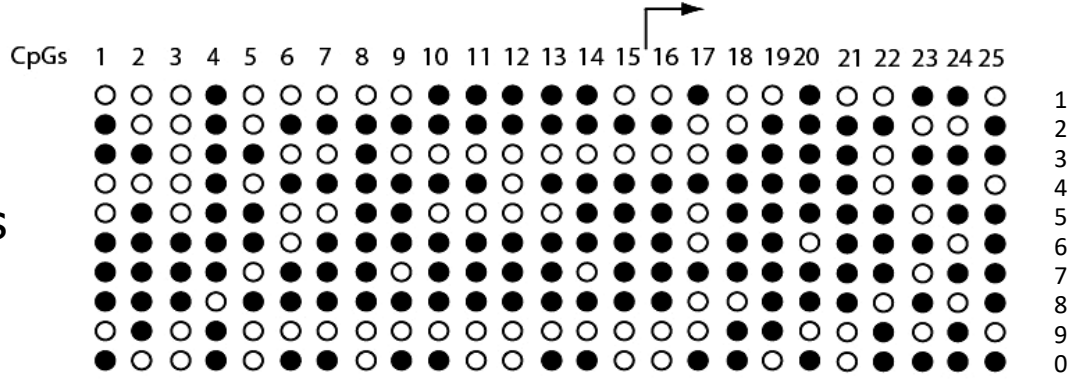CCCGGCCAGCCATGGGCCCTTGGA

CGI

TSS

Epithelial-like

○ methylated

● unmethylated

EMT-cells

CpGs 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

10 clones

1 2 3 4 5 6 7 8 9 0

Human DNA methylomes at base resolution show widespread epigenomic differences

Ryan Lister[1]*, Mattia Pelizzola[1]*, Robert H. Dowen[1], R. David Hawkins[2], Gary Hon[2], Julian Tonti-Filippini[4], Joseph R. Nery[1], Leonard Lee[2], Zhen Ye[2], Que-Minh Ngo[2], Lee Edsall[2], Jessica Antosiewicz-Bourget[5,6], Ron Stewart[5,6], Victor Ruotti[5,6], A. Harvey Millar[4], James A. Thomson[5,6,7,8], Bing Ren[2,3] & Joseph R. Ecker[1]

DNA cytosine methylation is a central epigenetic modification that has essential roles in cellular processes including genome regulation, development and disease. Here we present the first genome-wide, single-base-resolution maps of methylated cytosines in a mammalian genome, from both human embryonic stem cells and fetal fibroblasts, along with comparative analysis of messenger RNA and small RNA components of the transcriptome, several histone modifications, and sites of DNA–protein interaction for several key regulatory factors. Widespread differences were identified in the composition and patterning of cytosine methylation between the two genomes. Nearly one-quarter of all methylation identified in embryonic stem cells was in a non-CG context, suggesting that embryonic stem cells may use different methylation mechanisms to affect gene regulation. Methylation in non-CG contexts showed enrichment in gene bodies and depletion in protein binding sites and enhancers. Non-CG methylation disappeared upon induced differentiation of the embryonic stem cells, and was restored in induced pluripotent stem cells. We identified hundreds of differentially methylated regions proximal to genes involved in pluripotency and differentiation, and widespread reduced methylation levels in fibroblasts associated with lower transcriptional activity. These reference epigenomes provide a foundation for future studies exploring this key epigenetic modification in human disease and development.
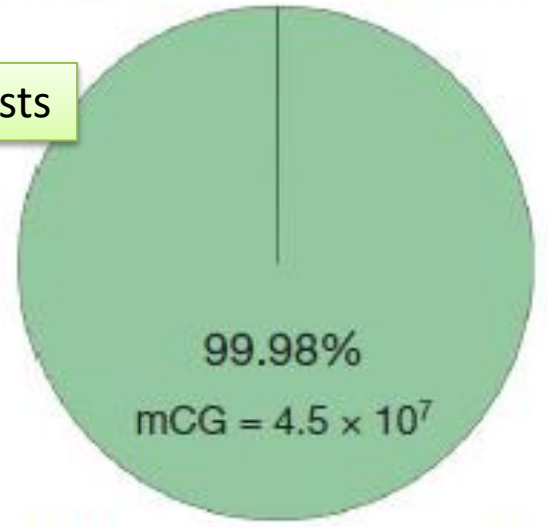
H1 are human embryonic **stem** cells

IMR90 are fetal lung fibroblasts
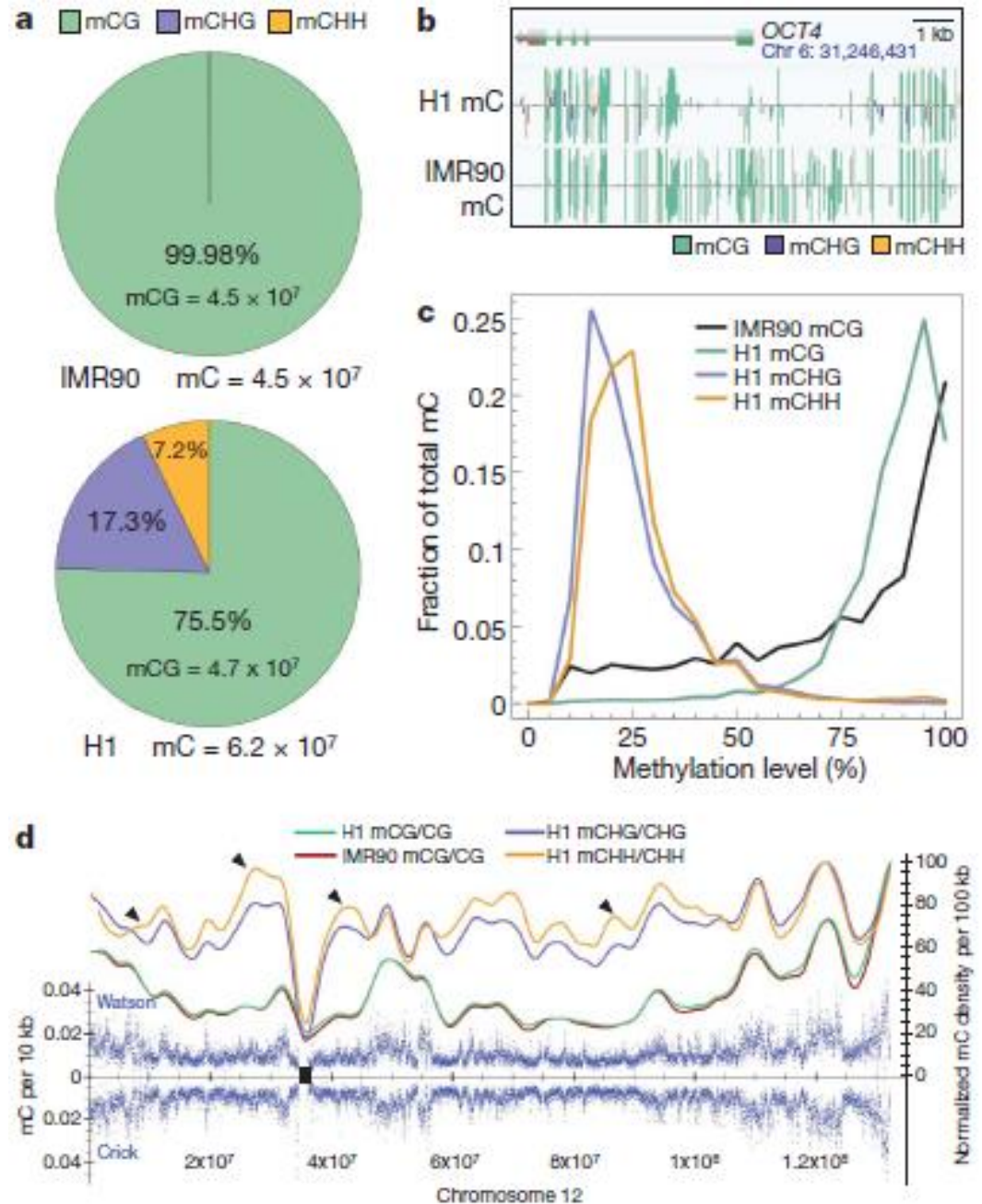
from Lister et al., 2009

a ☐ mCG ☐ mCHG ☐ mCHH
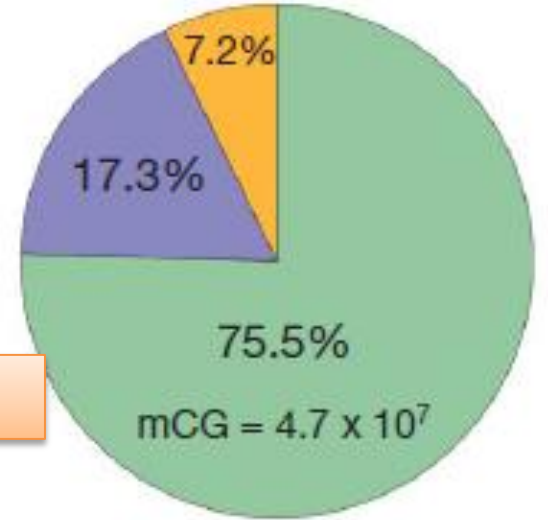
IMR90 are fetal lung fibroblasts

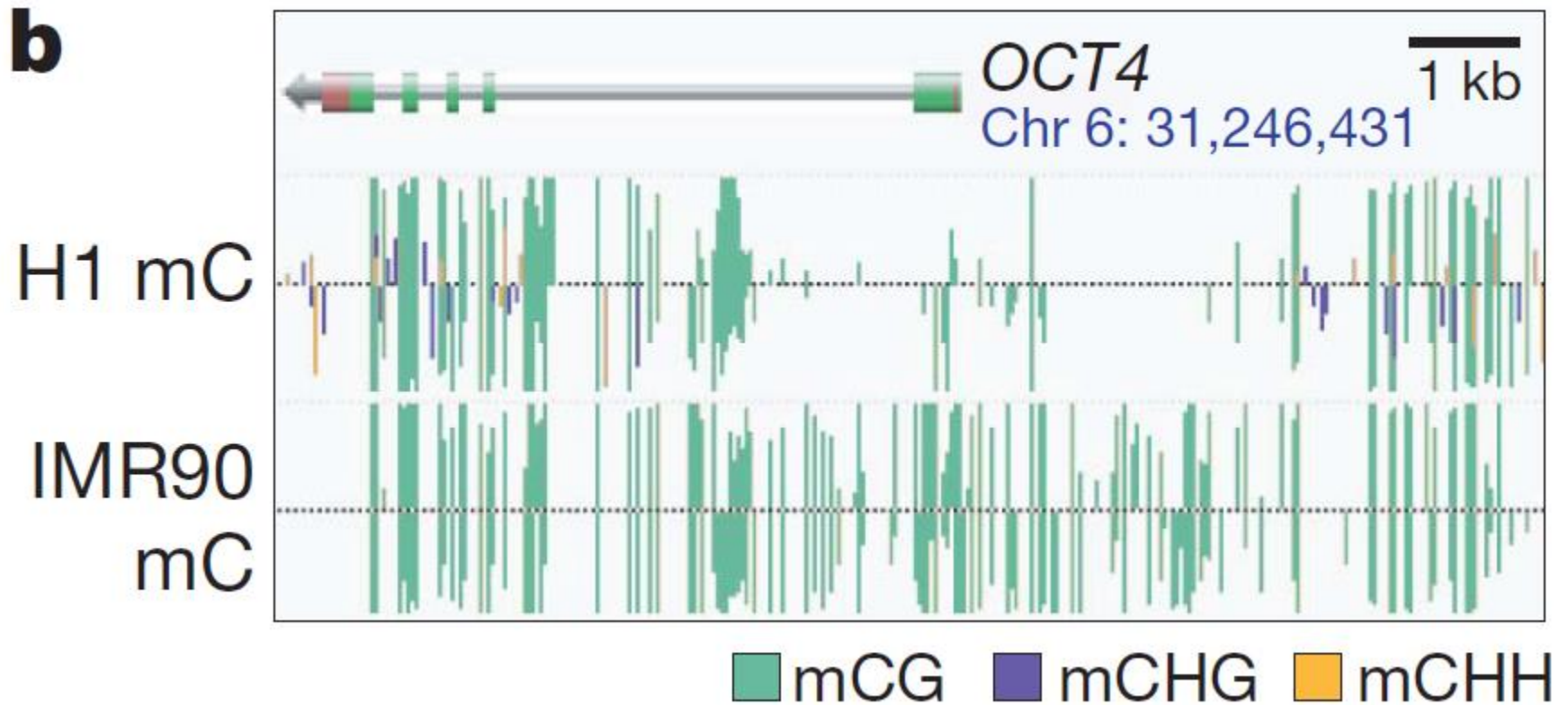Fig 1 Global trends of human DNA methylomes. The percentage of methyl-cytosines identified for H1 and IMR90 cells in each sequence context

99.98%
mCG $= 4.5 \times 10^7$

IMR90   mC $= 4.5 \times 10^7$

7.2%

17.3%

75.5%
mCG $= 4.7 \times 10^7$

H1 are human embryonic stem cells

H1   mC $= 6.2 \times 10^7$

from Lister et al., 2009

b, AnnoJ browser representation of OCT4.

from Lister et al., 2009

from Lister et al., 2009
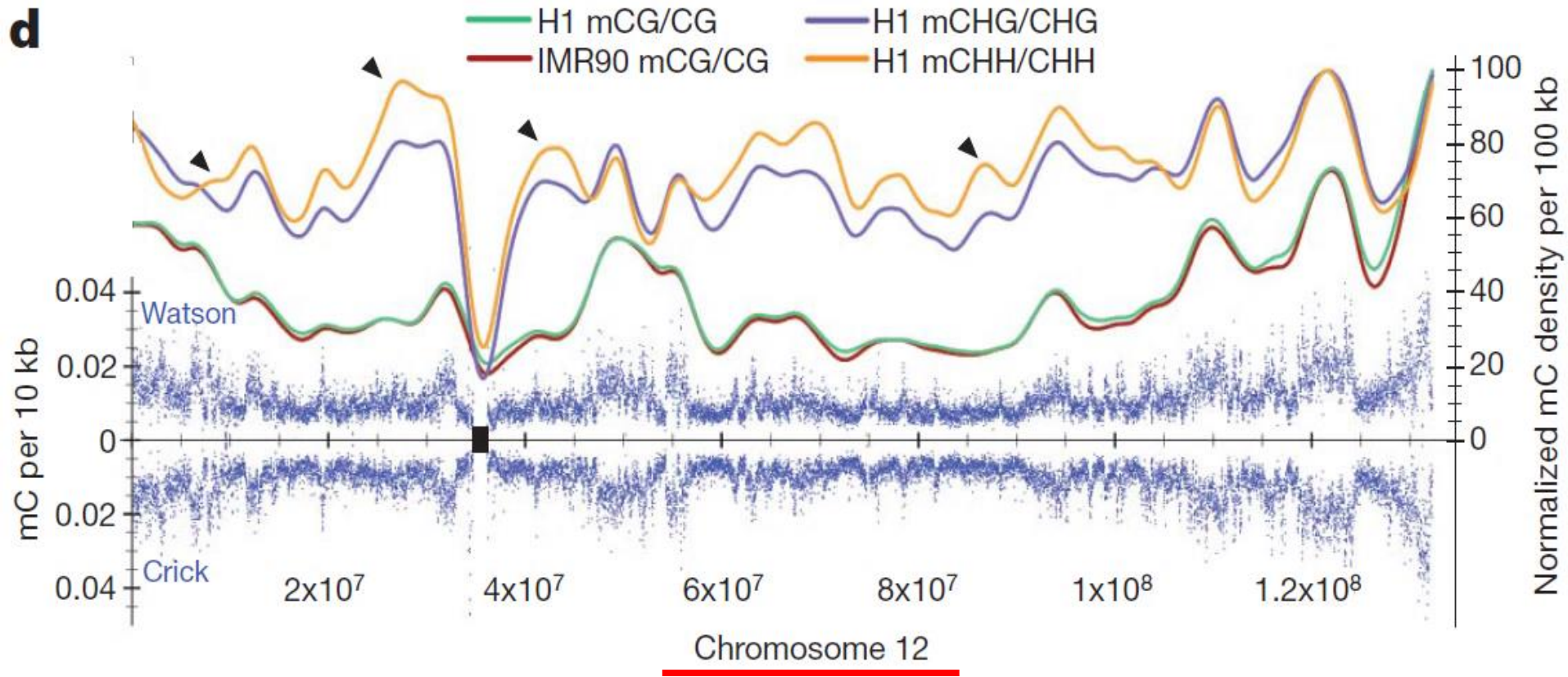
d, Blue dots indicate methyl-cytosine density in H1 cells in 10-kb windows throughout chromosome 12 (black rectangle, centromere). Smoothed lines represent the methyl-cytosine density in each context in H1 and IMR90 cells. Black triangles indicate various regions of contrasting trends in CG and non-CG methylation. mC, methyl-cytosine.

Methylation matters

Dirk Schübeler

Genome-wide maps of methylated cytosine bases at single-base-pair resolution in human cells reveal distinct differences between cell types. These maps provide a starting point to decode the function of this enigmatic mark.
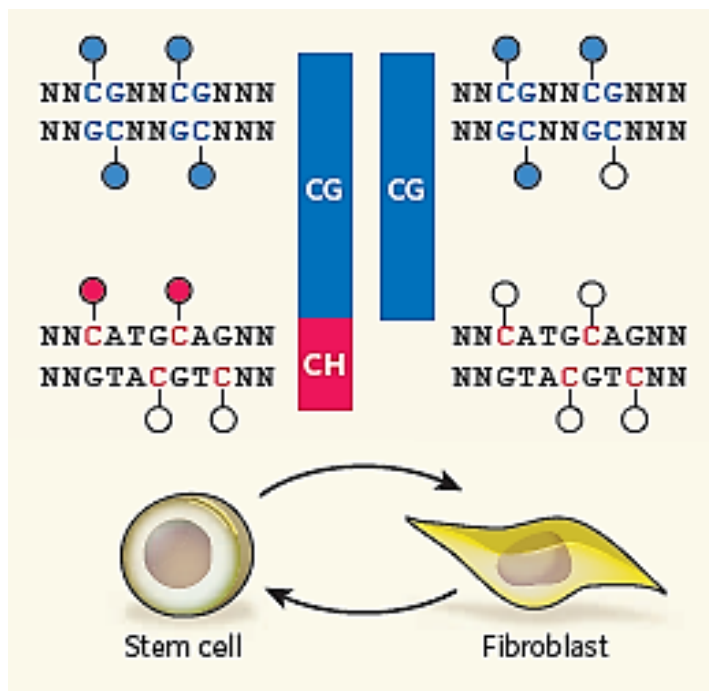
296

Figure 1 | DNA methylation patterns differ between stem cells and differentiated cells[1]. In stem cells, regions of DNA with CpG methylation (blue) are mostly uniformly methylated, whereas this modification is more heterogeneous in fibroblasts. Non-CpG methylation (red), which occurs primarily at CA nucleotides, is detected only in stem cells, yet is asymmetric and more scarce and patchy than CpG methylation. If fibroblasts are converted to induced pluripotent stem cells they regain non-CpG methylation. Filled circles, methylated cytosines; unfilled circles, unmethylated cytosines. H stands for A, C or T; N stands for any nucleotide.

Single-base-resolution maps of DNA methylation for two human cell lines

Single-base DNA methylomes of the flowering plant *Arabidopsis thaliana* were previously achieved using MethylC-Seq[15] or BS-Seq[16]. In this method, genomic DNA is treated with sodium bisulphite (BS) to convert cytosine, but not methylcytosine, to uracil, and subsequent high-throughput sequencing. We performed MethylC-Seq for two human cell lines, H1 human embryonic stem cells[17] and IMR90 fetal lung fibroblasts[18], generating 1.16 and 1.18 billion reads, respectively, that aligned uniquely to the human reference sequence (NCBI build 36/HG18). The total sequence yield was 87.5 and 91.0 gigabases (Gb), with an average read depth of 14.2× and 14.8× per strand for H1 and IMR90, respectively (Supplementary Fig. 1a). In each cell type, over 86% of both strands of the 3.08 Gb human reference sequence are covered by at least one sequence read (Supplementary Fig. 1b), accounting for 94% of the cytosines in the genome.
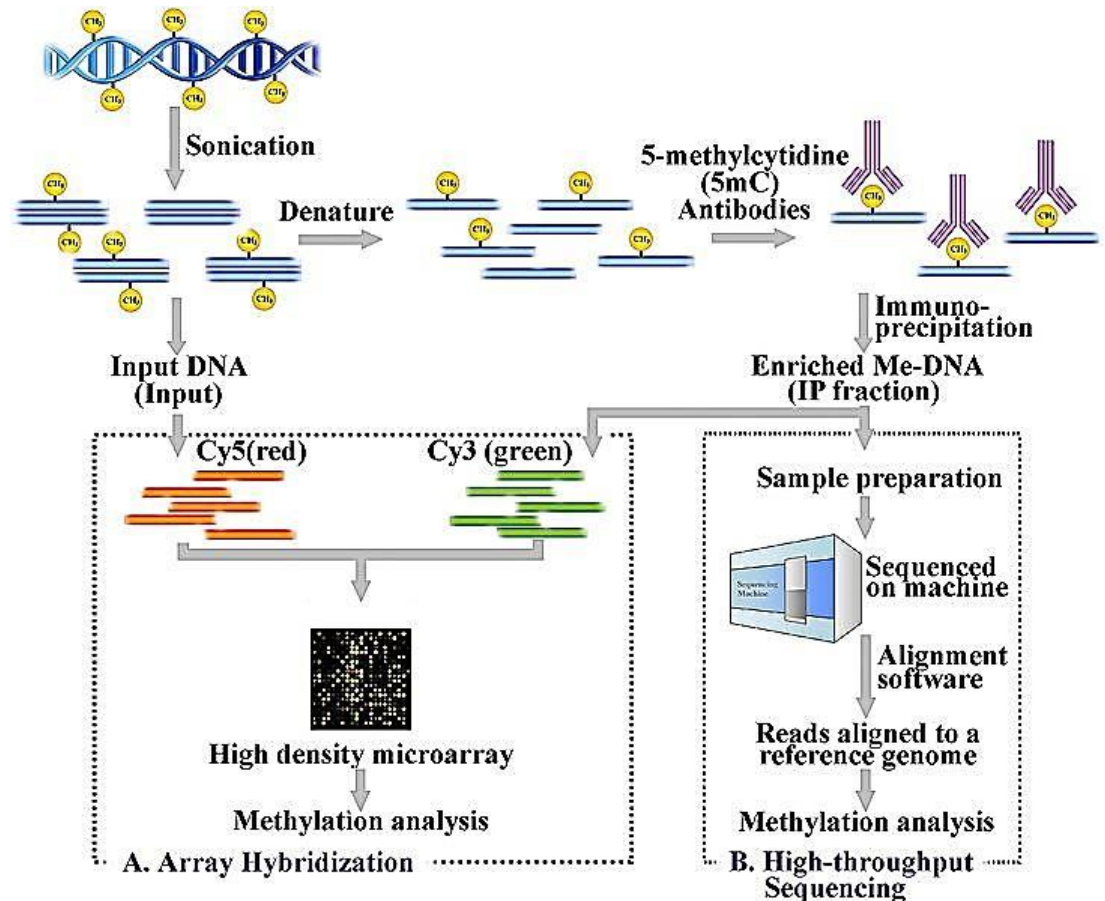
from Lister et al., 2009

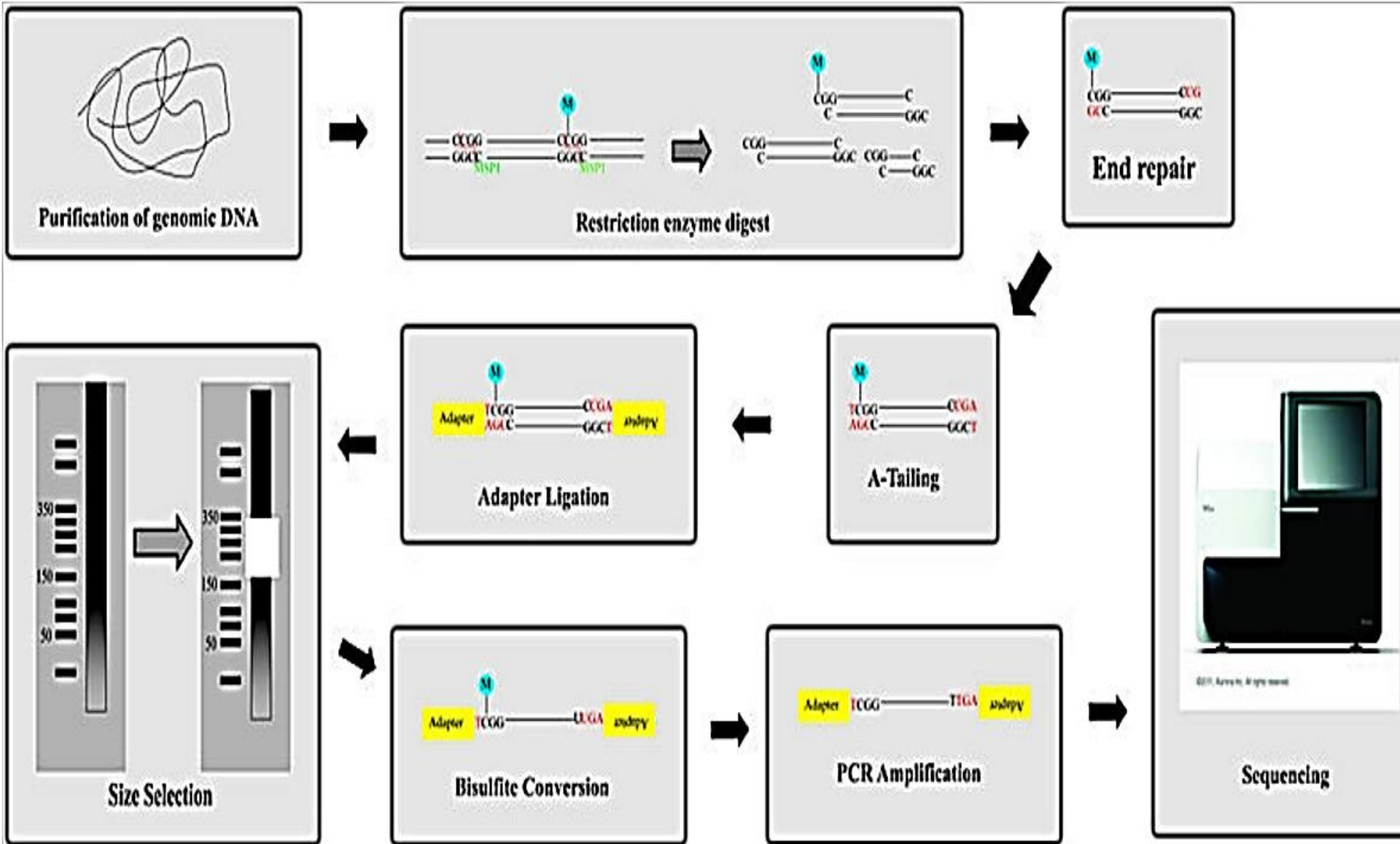**RRBS = Reduced Representation Bisulfite Sequencing**

(1)

DNA cut with CCGG-specific R.E. s will enrich a 200-300 bp fraction of CGI DNA. This fraction is used to produce library → NGS

(2) Or affinity-purified (MeDIP) methylated DNA fragments analyzed:

RRBS workflow

Research Paper 2

Article

CellPress

Capturing the Onset of PRC2-Mediated Repressive Domain Formation

Ozgur Oksuz,[1,2,7] Varun Narendra,[1,2,7] Chul-Hwan Lee,[1,2] Nicolas Descostes,[1,2] Gary LeRoy,[1,2] Ramya Raviram,[5] Lili Blumenberg,[3] Kelly Karch,[4] Pedro P. Rocha,[3,6] Benjamin A. Garcia,[4] Jane A. Skok,[3] and Danny Reinberg[1,2,8,*]

[1]Howard Hughes Medical Institute, New York University School of Medicine, New York, NY 10016, USA
[2]Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY 10016, USA
[3]Department of Pathology, New York University School of Medicine, New York, NY 10016, USA
[4]Epigenetics Program, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[5]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093, USA
[6]Division of Developmental Biology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, Bethesda, MD 20892, USA
[7]These authors contributed equally
[8]Lead Contact
*Correspondence: danny.reinberg@nyumc.org
https://doi.org/10.1016/j.molcel.2018.05.023

Reminder

Research Papers (1 each Ch.) are essential parts of your study.

Deep understanding required.

At the Exam, one question will always be proposed concerning one of the Research Papers; not «general» questions, but going deep into one of the experiments presented in the paper, including methodology.

Summary

Polycomb repressive complex 2 (PRC2) maintains gene silencing by catalyzing methylation of histone H3 at lysine 27 (H3K27me2/3) within chromatin. By designing a system whereby PRC2-mediated repressive domains were collapsed and then reconstructed in an inducible fashion *in vivo*, a two-step mechanism of H3K27me2/3 domain formation became evident. First, PRC2 is stably recruited by the actions of JARID2 and MTF2 to a limited number of spatially interacting "nucleation sites," creating H3K27me3-forming Polycomb foci within the nucleus. Second, PRC2 is allosterically activated via its binding to H3K27me3 and rapidly spreads H3K27me2/3 both in *cis* and in far-*cis* via long-range contacts. As PRC2 proceeds further from the nucleation sites, its stability on chromatin decreases such that domains of H3K27me3 remain proximal, and those of H3K27me2 distal, to the nucleation sites. This study demonstrates the principles of *de novo* establishment of PRC2-mediated repressive domains across the genome.

Summary

Polycomb repressive complex 2 (PRC2) maintains gene silencing by catalyzing methylation of histone H3 at lysine 27 (H3K27me2/3) within chromatin. By designing a system whereby PRC2-mediated repressive domains were collapsed and then reconstructed in an inducible fashion *in vivo*, a two-step mechanism of H3K27me2/3 domain formation became evident.
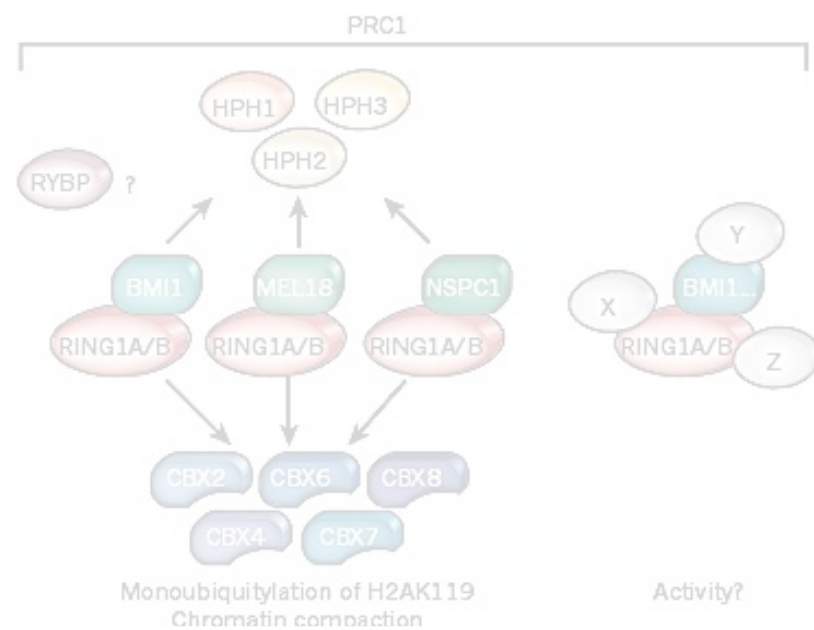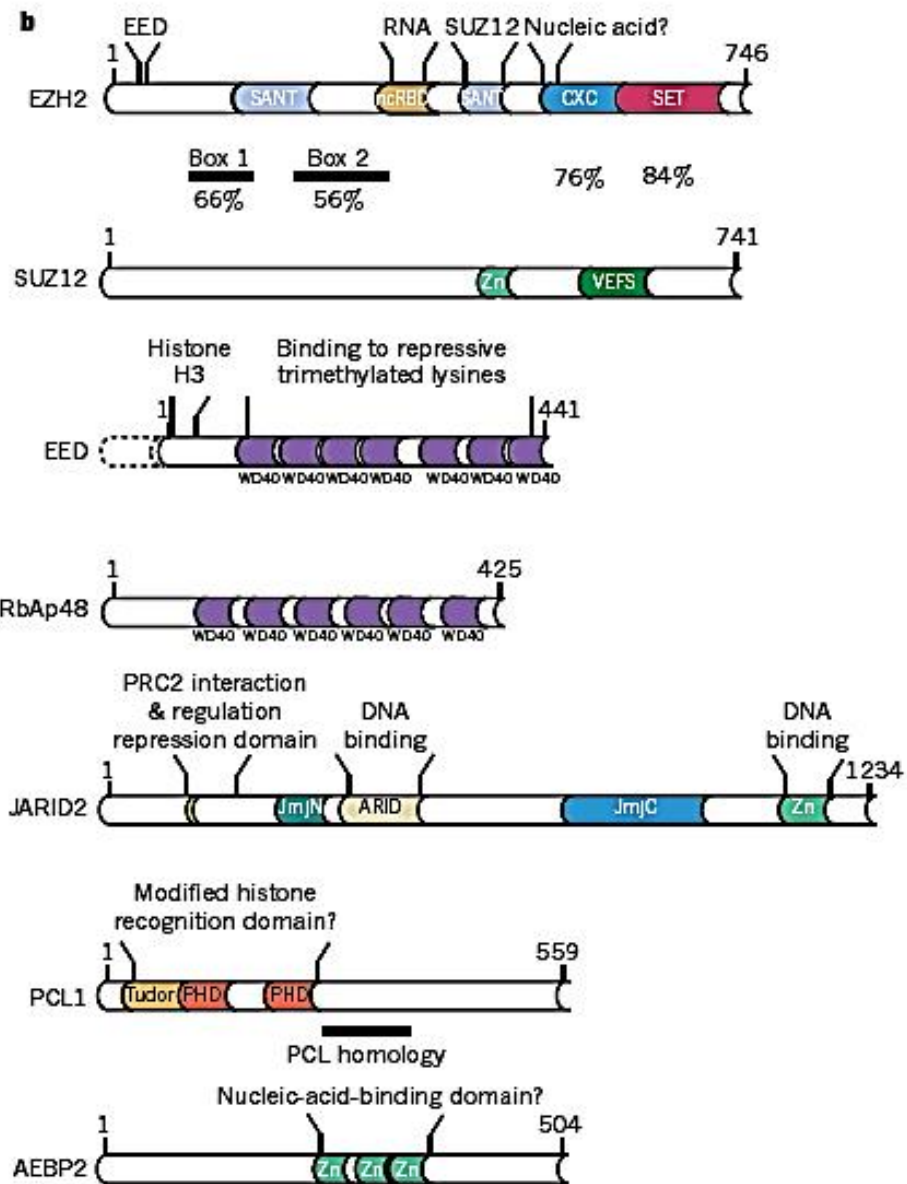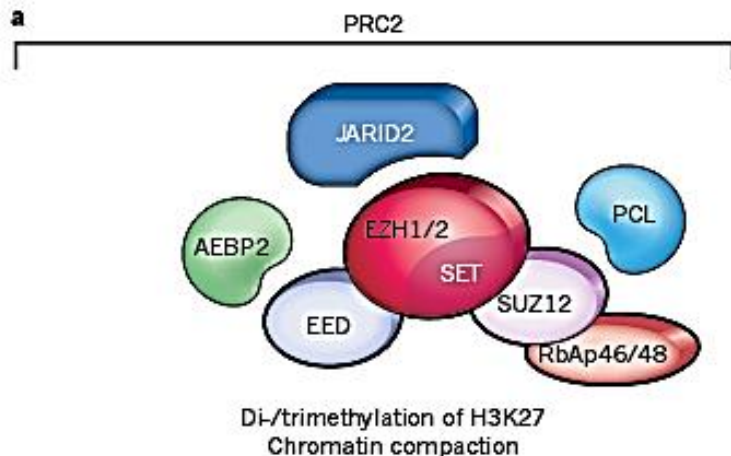
Summary

Polycomb repressive complex 2 (PRC2) maintains gene silencing by catalyzing methylation of histone H3 at lysine 27 (H3K27me2/3) within chromatin. By designing a system whereby PRC2-mediated repressive domains were collapsed and then reconstructed in an inducible fashion *in vivo*, a two-step mechanism of H3K27me2/3 domain formation became evident. First, PRC2 is stably recruited by the actions of JARID2 and MTF2 to a limited number of spatially interacting "nucleation sites," creating H3K27me3-forming Polycomb foci within the nucleus. Second, PRC2 is allosterically activated via its binding to H3K27me3 and rapidly spreads H3K27me2/3 both in *cis* and in far-*cis* via long-range contacts. As PRC2 proceeds further from the nucleation sites, its stability on chromatin decreases such that domains of H3K27me3 remain proximal, and those of H3K27me2 distal, to the nucleation sites. This study

Summary

Polycomb repressive complex 2 (PRC2) maintains gene silencing by catalyzing methylation of histone H3 at lysine 27 (H3K27me2/3) within chromatin. By designing a system whereby PRC2-mediated repressive domains were collapsed and then reconstructed in an inducible fashion *in vivo*, a two-step mechanism of H3K27me2/3 domain formation became evident. First, PRC2 is stably recruited by the actions of JARID2 and MTF2 to a limited number of spatially interacting "nucleation sites," creating H3K27me3-forming Polycomb foci within the nucleus. Second, PRC2 is allosterically activated via its binding to H3K27me3 and rapidly spreads H3K27me2/3 both in *cis* and in far-*cis* via long-range contacts. As PRC2 proceeds further from the nucleation sites, its stability on chromatin decreases such that domains of H3K27me3 remain proximal, and those of H3K27me2 distal, to the nucleation sites. This study demonstrates the principles of *de novo* establishment of PRC2-mediated repressive domains across the genome.

from Margueron & Reinberg, Nature 469:343-349 (2011)

Intro

find the scientific question addressed !

... the mechanism(s) by which PRC2 is initially recruited to specific genomic loci and how PRC2 sets up repressive domains comprising H3K27me2 or H3K27me3 <u>remain unclear</u>.

PRC2 targeting facilitated by JARID2, MTF2 (and other proteins?) for CGI

Thus, we devised a genetic system in mESCs by which we could inducibly reconstruct these domains from scratch. Employing this system, we defined the exact genomic coordinates of PRC2 nucleation and spreading and identified the key factors for its initial recruitment and stability on chromatin

then obtain te necessary background information !

example:

Aa say that << PRC2 targeting is facilitated by JARID2, MTF2 >>

What JARID2, MTF2 are ?

you may use any official database you know
or .....
just google it out, but take care of choosing a good website

then obtain te necessary background information !

example:

Aa say that << PRC2 targeting is facilitated by JARID2, MTF2 >>

What JARID2, MTF2 are ?

NCBI Gene: https://www.ncbi.nlm.nih.gov/gene/3720
                       https://www.ncbi.nlm.nih.gov/gene/22823

GeneCards: https://www.genecards.org/ (Weinzmann Institute of Science)

RESULTS

Disruption of EED-H3K27me3 Interaction Reduced H3K27me3 Levels in mESCs

The crystal structure of EED shows that H3K27me3 is located within an aromatic cage formed by the WD40 repeats of EED and that three amino acids, Phe97, Tyr148, and Tyr365, directly contact the tri-methylated lysine residue (Margueron et al., 2009). To test whether this interaction was necessary for the maintenance of repressive chromatin domains *in vivo*, we generated EED cage mutants having Phe97 or Tyr365 substituted with alanine (F97A or Y365A) in mESCs using the CRISPR/Cas9 system. As a control, we mutated Tyr358 (Y358A), which does not contact the tri-methylated lysine. A western blot (WB) of whole-cell extract preparations from these cage-mutant lines showed a significant reduction in H3K27me3 levels (Figure 1A), while those of the control Y358A mutant were unaffected.

EED is the PRC2 component that binds to H3K27me2/3

The EED domain specifically involved in di-/tri-methylated Lysine binding is called «aromatic Cage» and is formed by <u>WD40 repeats</u>

Where can we get information on this protein ?

NCBI Gene

NCBI Protein

[RCSB Protein Data Bank](#)   see [3JZG](#)

SwissProt

but also Wikipedia

WD40 repeat

The **WD40 repeat** (also known as the **WD** or **beta-transducin repeat**) is a short structural motif of approximately 40 amino acids, often terminating in a tryptophan-aspartic acid (W-D) dipeptide.[2] Tandem copies of these repeats typically fold together to form a type of circular solenoid protein domain called the **WD40 domain**.
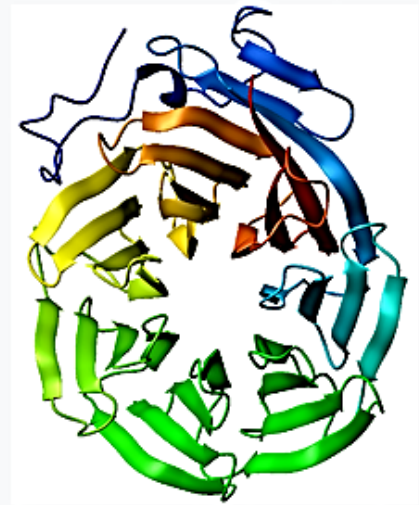
Contents [hide]

WD domain, G-beta repeat

Ribbon diagram of the C-terminal WD40 domain of Tup1 (a transcriptional corepressor in yeast), which adopts a 7-bladed beta-propeller fold. Ribbon is colored from blue (N-terminus) to red (C-terminus).[1]

Identifiers	
Symbol	WD40

Structure [edit]

WD40 domain-containing proteins have 4 to 16 repeating units, all of which are thought to form a circularised beta-propeller structure (see figure to the right).[3][4] The WD40 domain is composed of several repeats, a variable region of around 20 residues at the beginning followed by a more common repeated set of residues. These repeats typically form a four stranded anti-parallel beta sheet or blade. These blades come together to form a propeller with the most common being a 7 bladed beta propeller. The blades interlock so that the last beta strand of one repeat forms with the first three of the next repeat to form the 3D blade structure.

https://en.wikipedia.org/wiki/WD40_repeat

EED Aromatic cage mutants: F97A or Y365A

control mutant: Y358A

Engineered in mESCs using <u>CRISPR/Cas9 system</u>

- cage muts did not disrupt PRC2   (EED KO does)

- cage-muts mESCs have WT phenotype

<u>CRISPR/Cas9 system</u>
Aa replaced endogenous copies of EED gene with the mutated versions.

Genome Editing

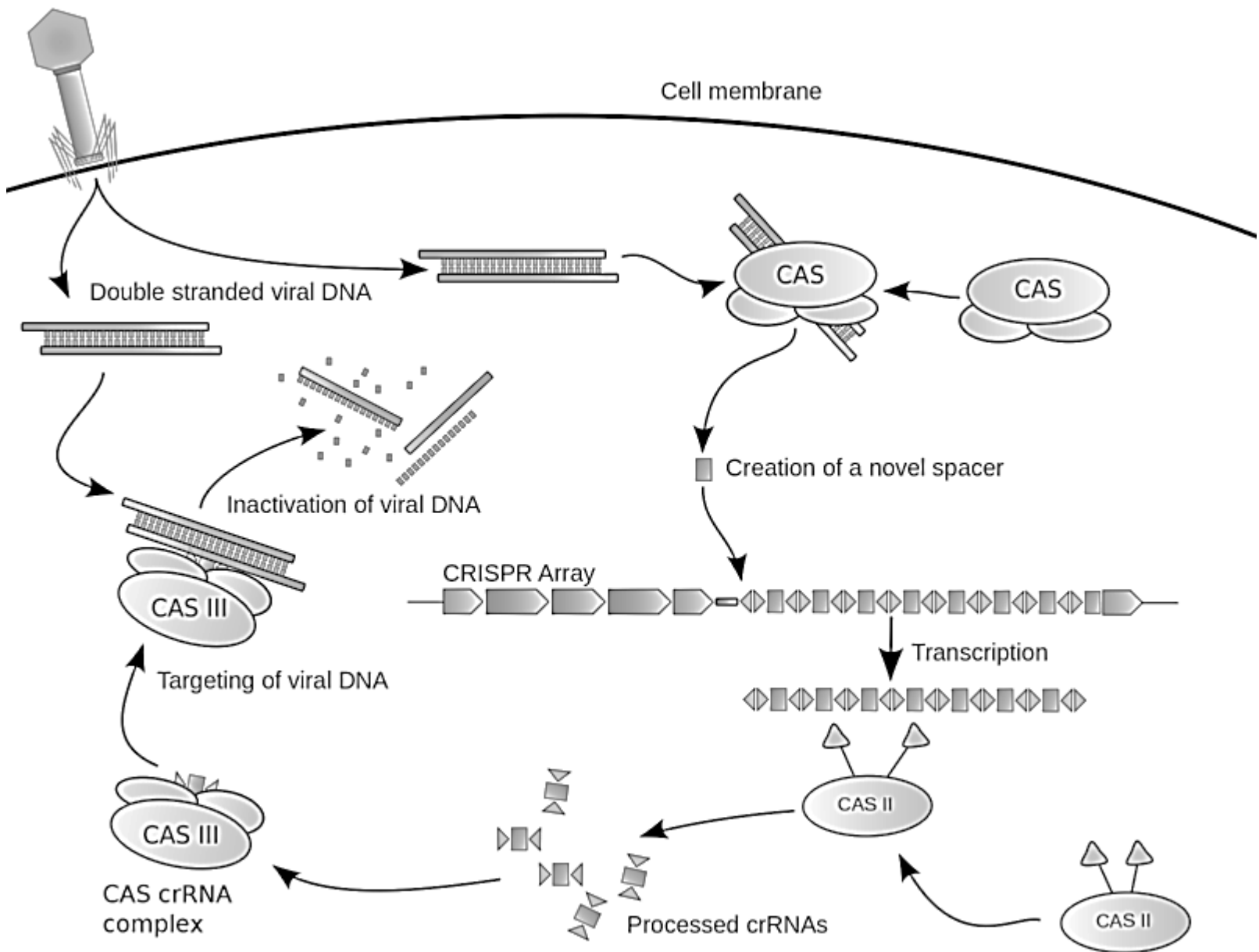gene substitution using **CRISPR-cas9**

with homologous recombination-dependent repair

CRISPR (clustered regularly interspaced short palindromic repeats) is a family of DNA sequences found within the genomes of prokaryotic organisms such as bacteria and archaea. These sequences are derived from DNA fragments from viruses that have previously infected the prokaryote and are used to detect and destroy DNA from similar viruses during subsequent infections. Hence these sequences play a key role in the antiviral defense system of prokaryotes.

Cas 9 (or "CRISPR-associated protein 9") is an enzyme that uses CRISPR sequences as a guide to recognize and cleave specific strands of DNA that are complementary to the CRISPR sequence. Cas9 enzymes together with CRISPR sequences form the basis of a technology known as CRISPR-Cas9 that can be used to edit genes within organisms.

The CRISPR-Cas system is a prokaryotic immune system that confers resistance to foreign genetic elements such as those present within plasmids and phages that provides a form of acquired immunity. RNAs harboring the spacer sequence help Cas proteins recognizing and cutting foreign pathogenic DNA. *Other RNA-guided Cas proteins cut foreign RNA*
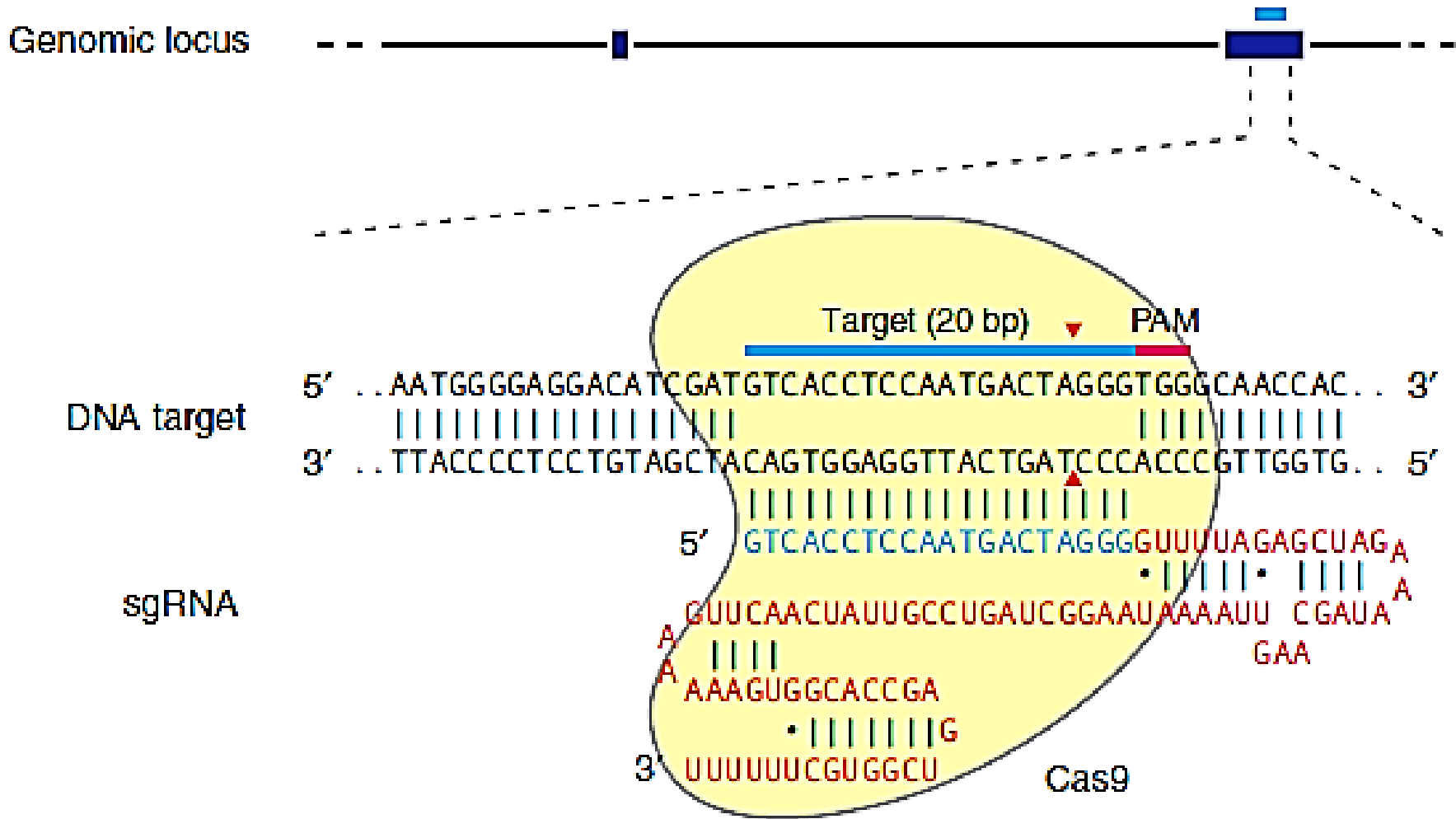
Cell membrane

Double stranded viral DNA

CAS

CAS

Inactivation of viral DNA

Creation of a novel spacer

CRISPR Array

CAS III

Targeting of viral DNA

Transcription

CAS II

CAS crRNA complex

CAS III

Processed crRNAs

CAS II

https://en.wikipedia.org/wiki/CRISPR

Genome engineering using the CRISPR-Cas9 system

F Ann Ran[1–5,8], Patrick D Hsu[1–5,8], Jason Wright[1], Vineeta Agarwala[1,6,7], David A Scott[1–4] & Feng Zhang[1–4]

[1]Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard, Cambridge, Massachusetts, USA. [2]McGovern Institute for Brain Research, Cambridge, Massachusetts, USA. [3]Department of Brain and Cognitive Sciences, MIT, Cambridge, Massachusetts, USA. [4]Department of Biological Engineering, MIT, Cambridge, Massachusetts, USA. [5]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, USA. [6]Program in Biophysics, Harvard University, MIT, Cambridge, Massachusetts, USA. [7]Harvard-MIT Division of Health Sciences and Technology, MIT, Cambridge, Massachusetts, USA. [8]These authors contributed equally to this work. Correspondence should be addressed to F.Z. (zhang@broadinstitute.org).

Targeted nucleases are powerful tools for mediating genome alteration with high precision. The RNA-guided Cas9 nuclease from the microbial clustered regularly interspaced short palindromic repeats (CRISPR) adaptive immune system can be used to facilitate efficient genome engineering in eukaryotic cells by simply specifying a 20-nt targeting sequence within its guide RNA. Here we describe a set of tools for Cas9-mediated genome editing via nonhomologous end joining (NHEJ) or homology-directed repair (HDR) in mammalian cells, as well as generation of modified cell lines for downstream functional studies. To minimize off-target cleavage, we further describe a double-nicking strategy using the Cas9 nickase mutant with paired guide RNAs. This protocol provides experimentally derived guidelines for the selection of target sites, evaluation of cleavage efficiency and analysis of off-target activity. Beginning with target design, gene modifications can be achieved within as little as 1–2 weeks, and modified clonal cell lines can be derived within 2–3 weeks.

Genomic locus

Target (20 bp) PAM

DNA target

5' ..AATGGGGAGGACATCGATGTCACCTCCAATGACTAGGGTGGGCAACCAC.. 3'

3' ..TTACCCCTCCTGTAGCTACAGTGGAGGTTACTGATCCCACCCGTTGGTG.. 5'

sgRNA

5' GTCACCTCCAATGACTAGGGGUUUUAGAGCUAG A

GUUCAACUAUUGCCUGAUCGGAAUAAAAUU CGAUA A

GAA

AAAGUGGCACCGA

3' UUUUUUCGUGGCU

Cas9

Yellow: Cas9 nuclease from *S. pyogenes*

sgRNA : a 20-nt guide sequence (blue) and a scaffold (red)

PAM: a requisite 5'-NGG adjacent motif

Cas9 makes a DSB (double strand break) ~3 bp upstream of the PAM (red triangle).

The RNA-guided nuclease function of CRISPR-Cas is reconstituted in mammalian cells through the heterologous expression of human codon–optimized Cas9 and the requisite RNA components.

Structural RNA + the guide RNA can be fused together → sg RNA
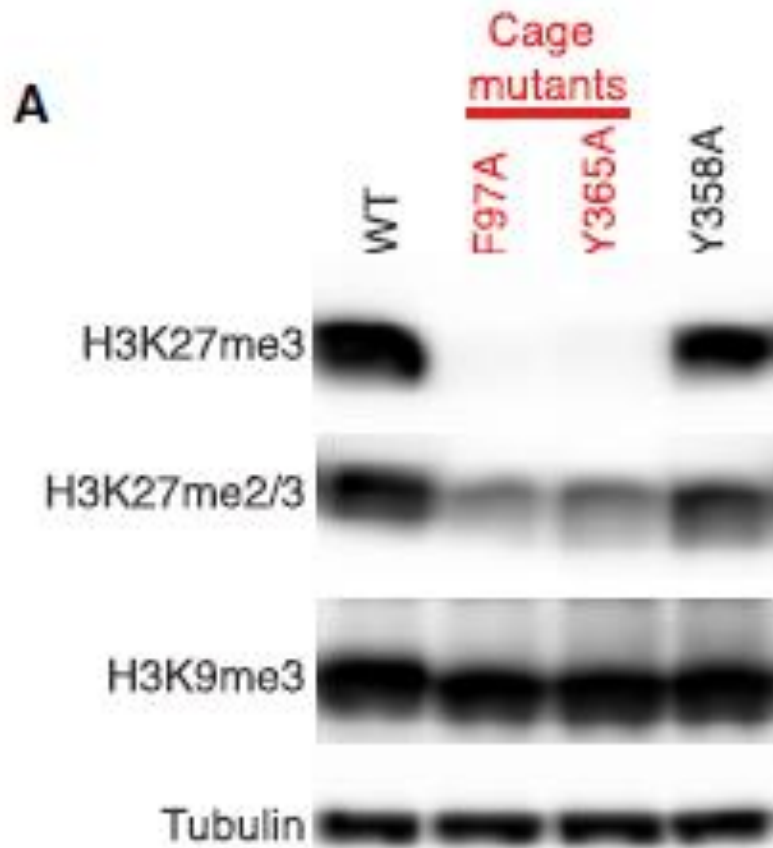   (single guide RNA)
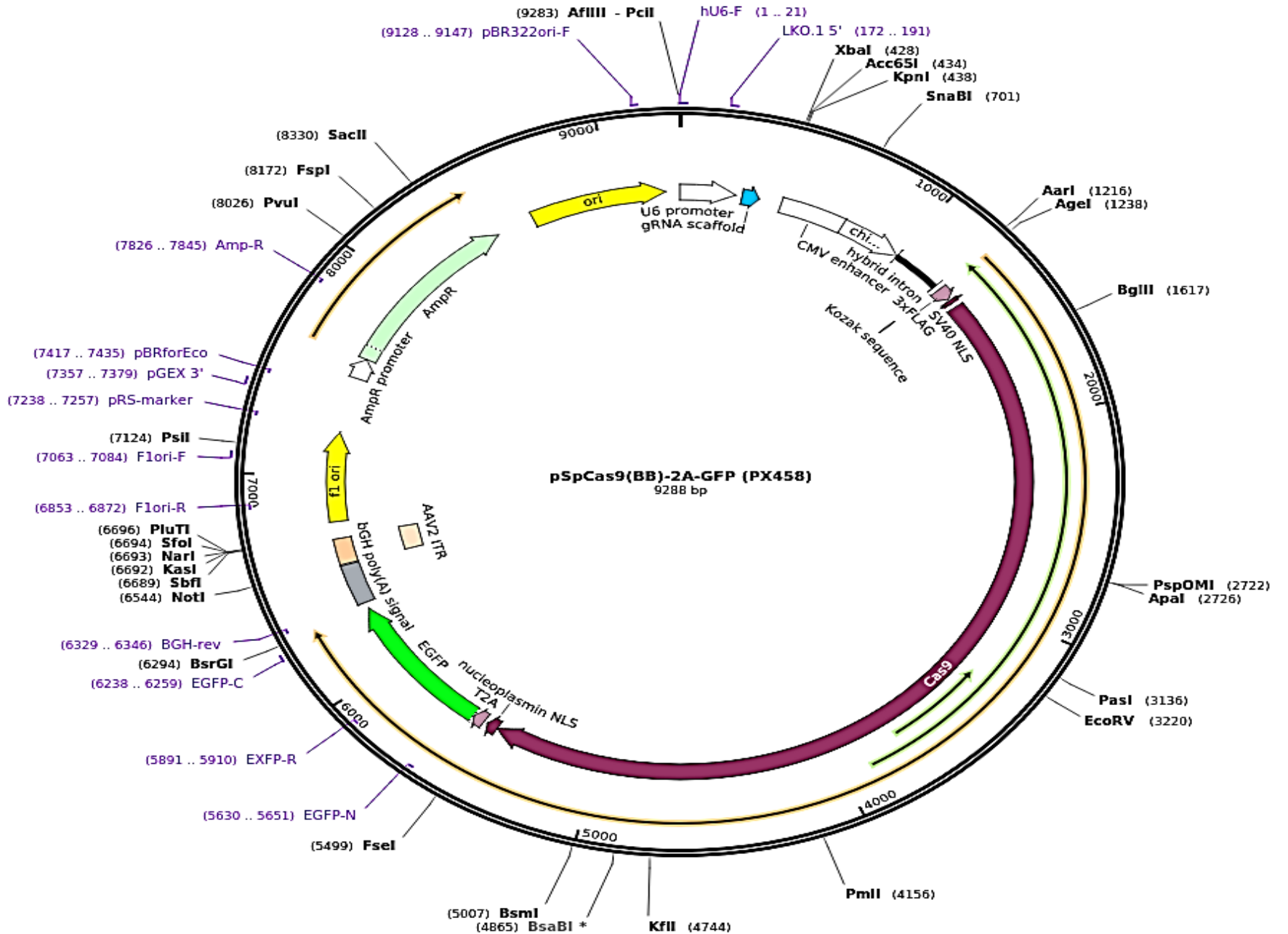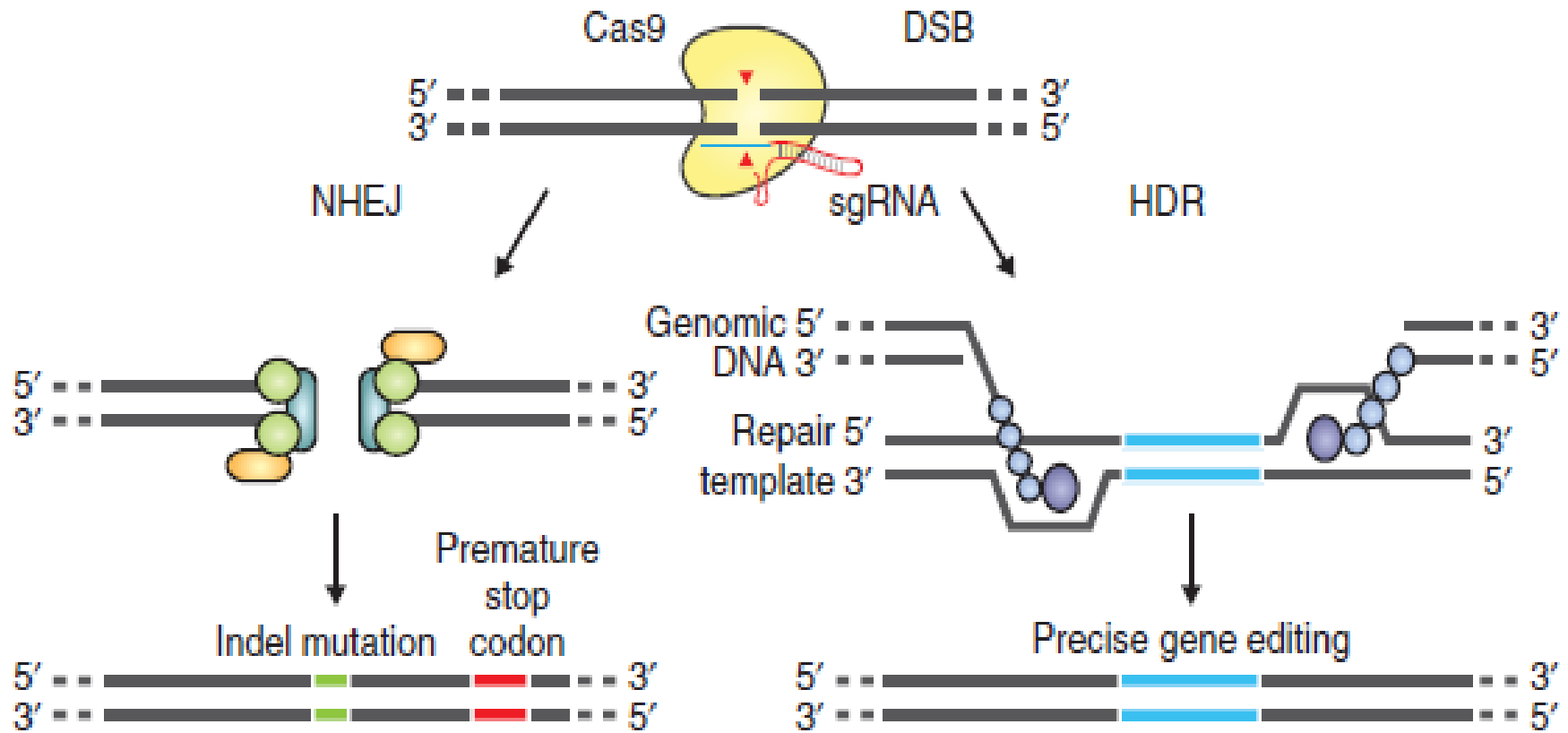
Cells will repair the DSB using two possile pathways:

1. NHEJ – nonhomologous end-joining – introduces errors like small indels
2. HDR - homologous recombination dependent repair
    takes information from homologous DNA, whch can be in cis, i.e. one can
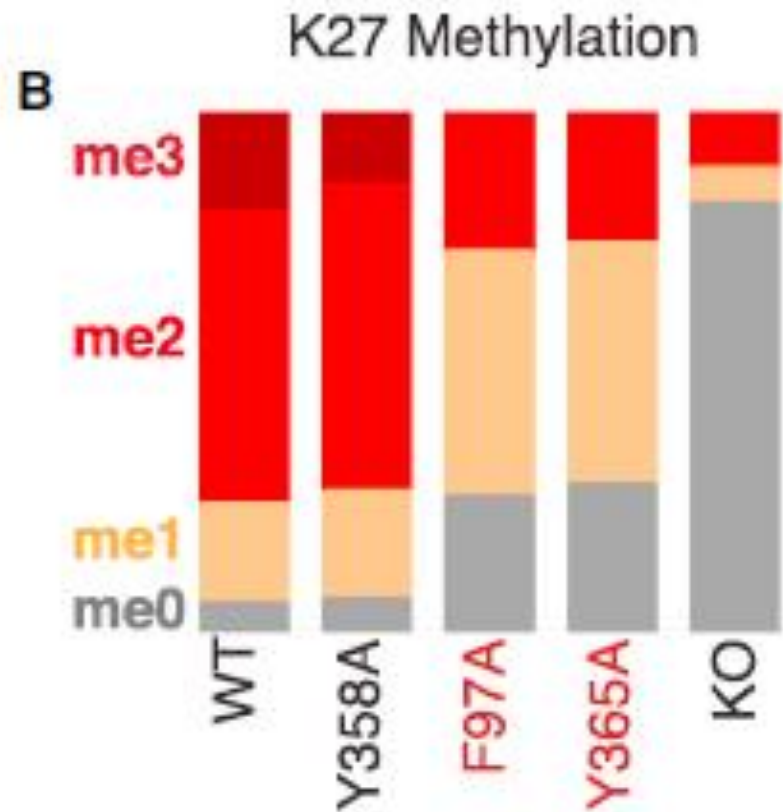   transfect cells with the mutated version of the gene /locus to be substituted.

An aspartate-to-alanine (D10A) mutation in the RuvC catalytic domain of Cas9 allows transformation of wt Cas9 DNA cleaving activity into a nickase: this will yield single-stranded breaks, and the subsequent preferential repair through HDR.

For this kind of gene substitution, Aa used Cas9-nickase

The homologous DNA for recombination was given in *trans* (plasmid) and contained the require mutant EED sequences.
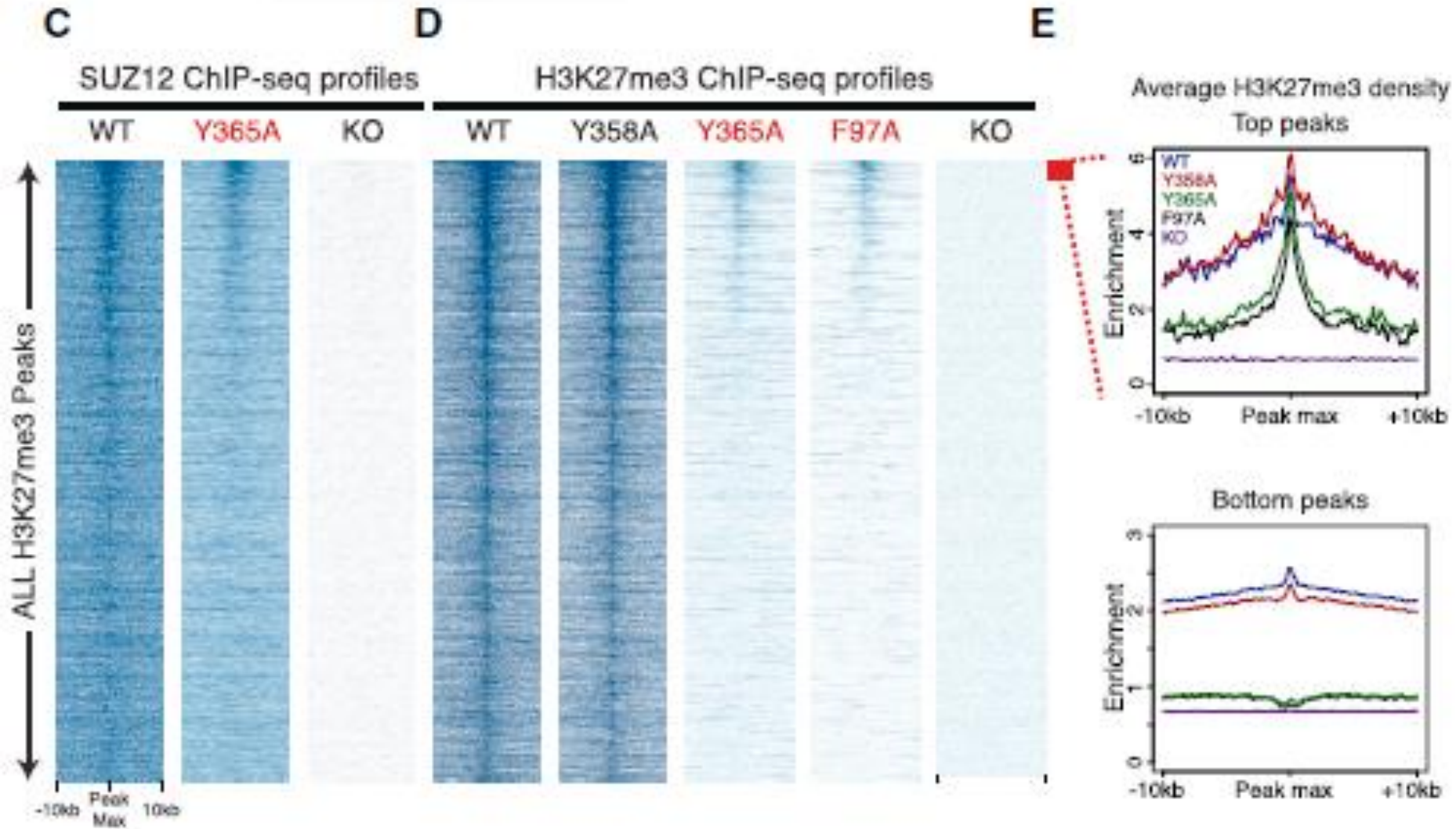
A

Cage mutants

WT | F97A | Y365A | Y358A

H3K27me3

H3K27me2/3

H3K9me3

Tubulin

Total cell lysate WB

B K27 Methylation

me3

me2

me1

me0

WT | Y358A | F97A | Y365A | KO

quantitative histone mass spectrometry

Heatmap

quantitative ChIP-seq

C — SUZ12 ChIP-seq profiles: WT, Y365A, KO

D — H3K27me3 ChIP-seq profiles: WT, Y358A, Y365A, F97A, KO

E — Average H3K27me3 density

Top peaks — WT, Y358A, Y365A, F97A, KO; Enrichment; -10kb, Peak max, +10kb

Bottom peaks — Enrichment; -10kb, Peak max, +10kb

ALL H3K27me3 Peaks

-10kb, Peak Max, 10kb

the window considered

H3K27me3 ChIP-seq profiles

WT   Y358A   Y365A   F97A   KO

Average H3K27me3 density

Top peaks

WT
Y358A
Y365A
F97A
KO

Enrichment

6

4

2

0

-10kb   Peak max   +10kb

Bottom peaks

Enrichment

3

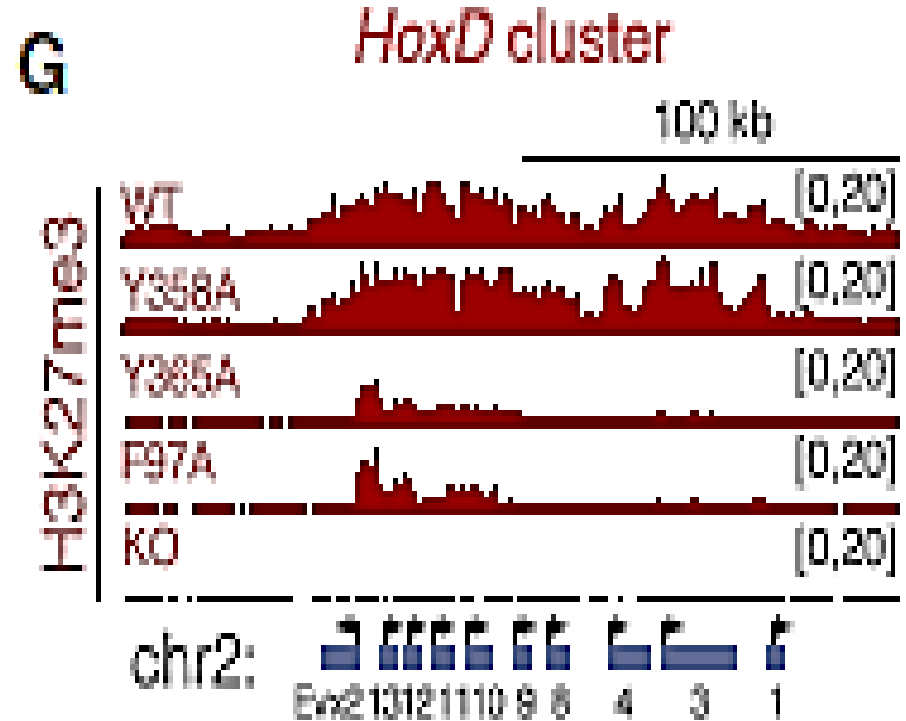2
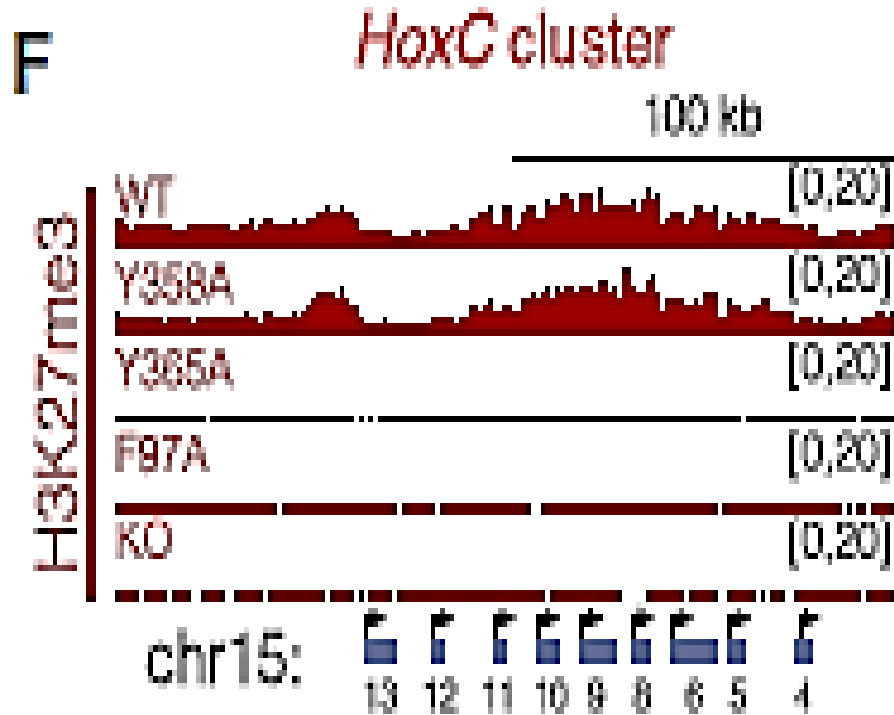
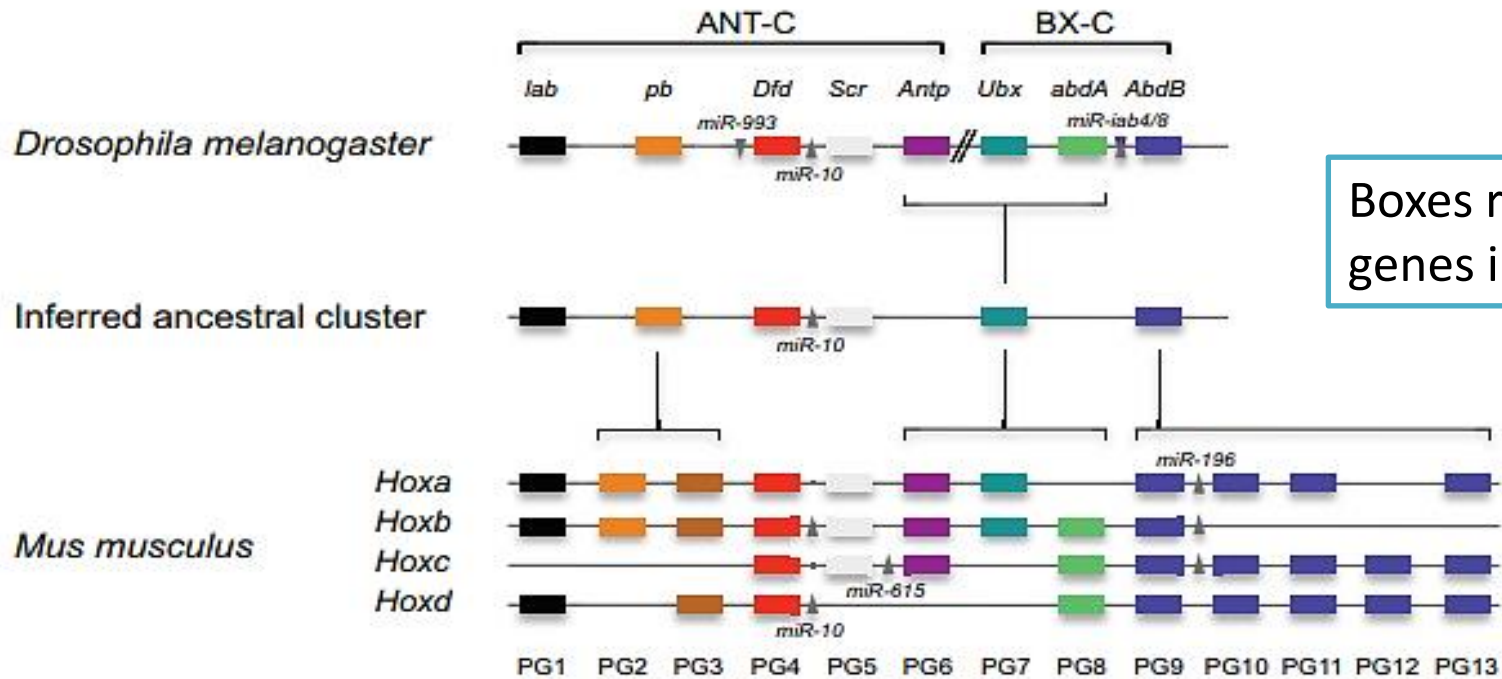1

0

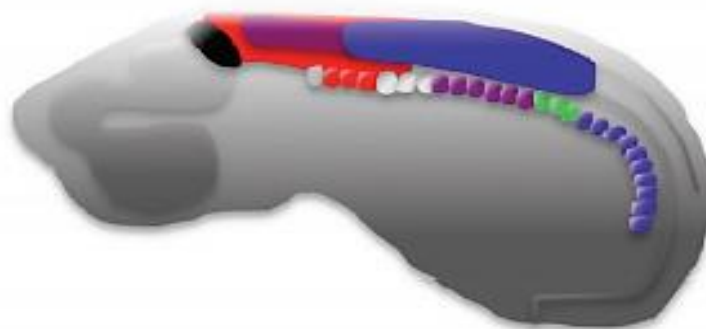-10kb   Peak max   +10kb
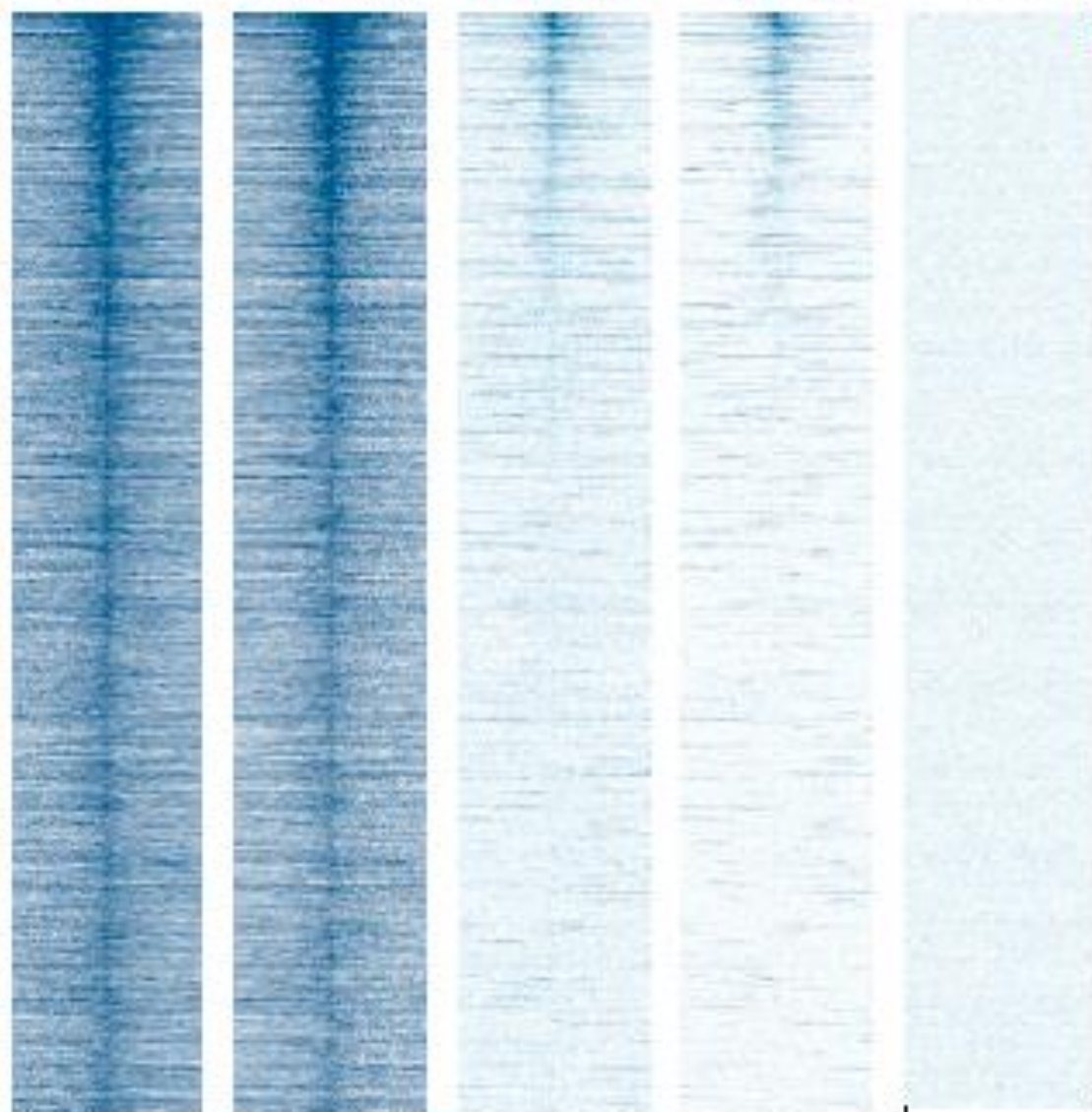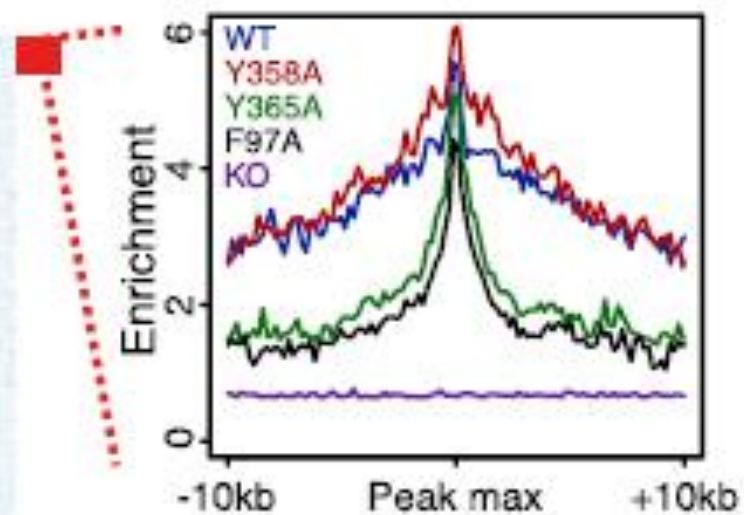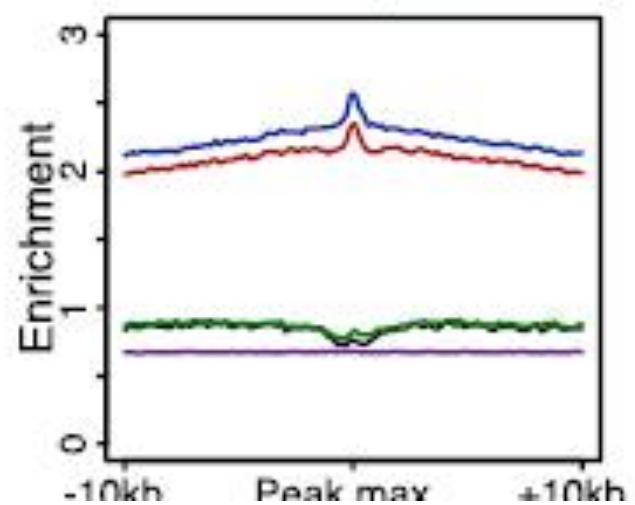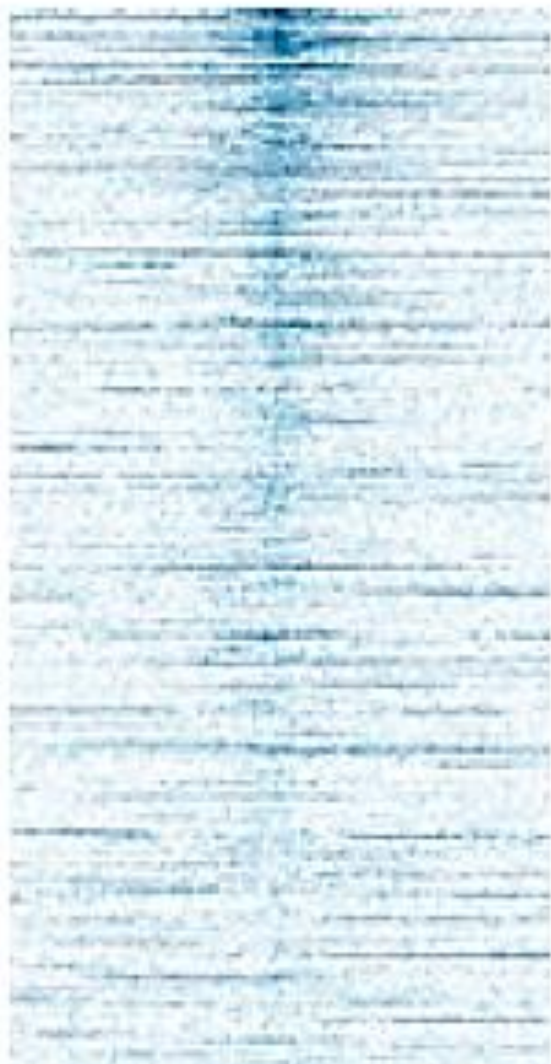
Y365A   F97A   KO

-10kb   +10kB

Details for 2 loci containing the HoxC and the HoxD clusters

Boxes represent single genes in the cluster

...approximate WT levels of H3K27me3 were maintained in cage-mutant mESCs at <u>a limited number of discrete regions</u> within larger Polycomb domains in contrast to other regions in the genome, with a progressive decrease distal to these sites (Figures 1E, cf. top and bottom plots)

At these regions, apparently, Lysine 27 me2/3 binding by EED is not required for PRC2 carrying on methylation

Initial Deposition and Spreading Mechanism of H3K27me2/3 Domains

To test whether regions of active PRC2 recruitment identified above might serve as nucleation sites for the initial establishment of Polycomb chromatin domains, we devised a Cre-lox-based system in which new domains could be established ''de novo'' in an inducible fashion in vivo, and their formation was tracked over time.
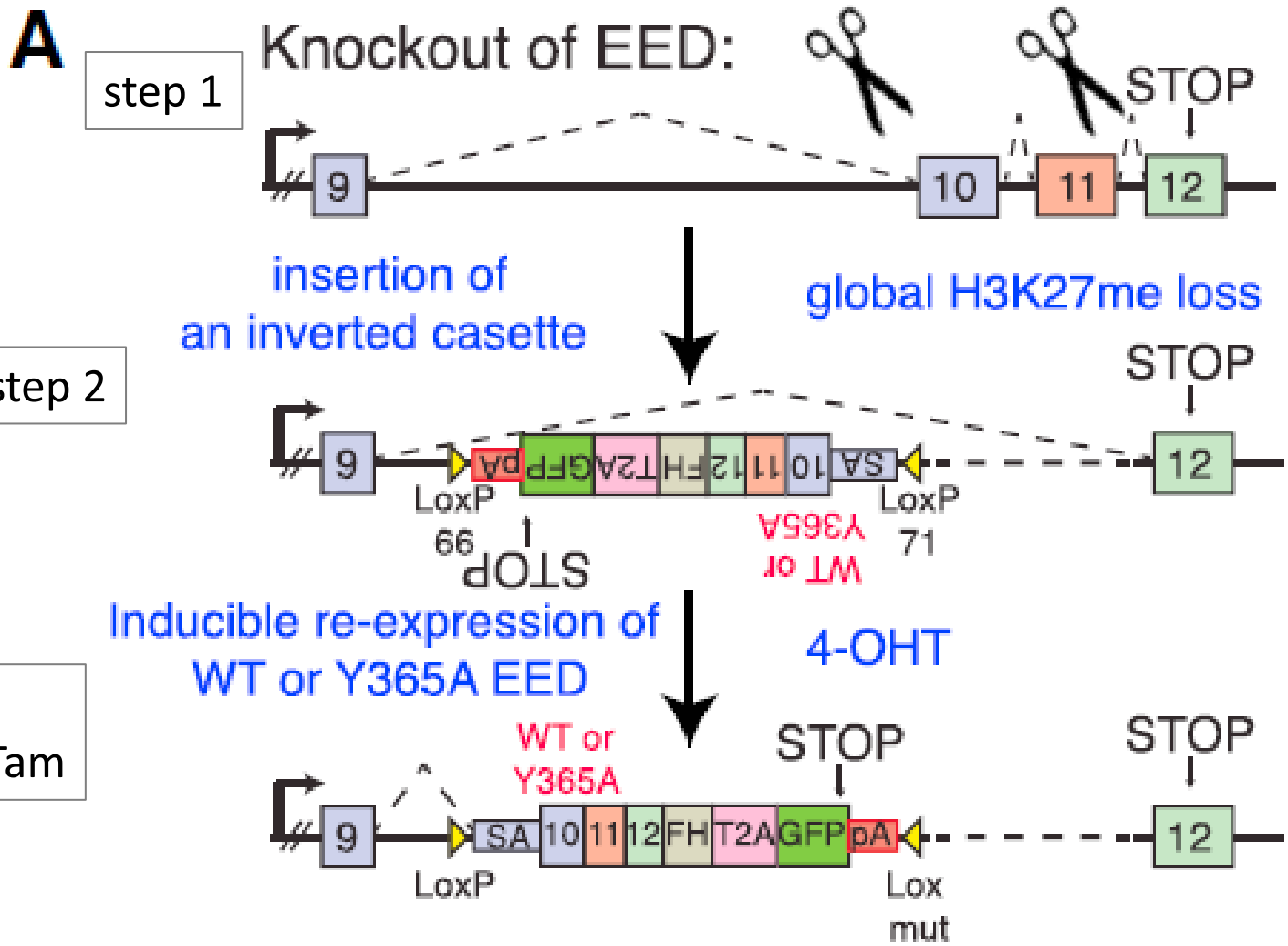
To this end, we deleted the endogenous copy of EED in mESCs possessing a CreERT2 transgene (Figure 2A; see STAR Methods for details).

The cells were propagated until H3K27me2/3 was completely depleted (Figures S3A, see 0 hr, and S3E).

We then expressed FLAG-HA-tagged versions of WT or cage-mutant EED (Y365A) from its endogenous locus in response to 4-Hydroxytamoxifen (4-OHT) to follow de novo H3K27me2/3 deposition.

Targeting scheme to conditionally rescue EED KO mESCs with EED, either WT or Y365A.

Fig 2

A Knockout of EED:

step 1

step 2

insertion of an inverted casette

global H3K27me loss

step 3 + OH-Tam

Inducible re-expression of WT or Y365A EED

4-OHT

Aa. used (C57BL/6) mESCs that contain CreERT2. It encodes a Cre recombinase (Cre) fused to a mutant estrogen ligand-binding domain (ERT2) that requires the presence of tamoxifen for activity.

Fig. 2

CpG Motif 1 (GA-rich): 4.2e-171

CpG Motif 2 (GCN tandem repeat): 9.7e-155

S-Logos