Epigenetics

MMG 835, SPRING 2016
Eukaryotic Molecular Genetics

George I. Mias

Department of Biochemistry and Molecular Biology
gmias@msu.edu

What is epigenetics

- Twins
  - Same yet different

Martin Schoeller/National Geographic (2012)

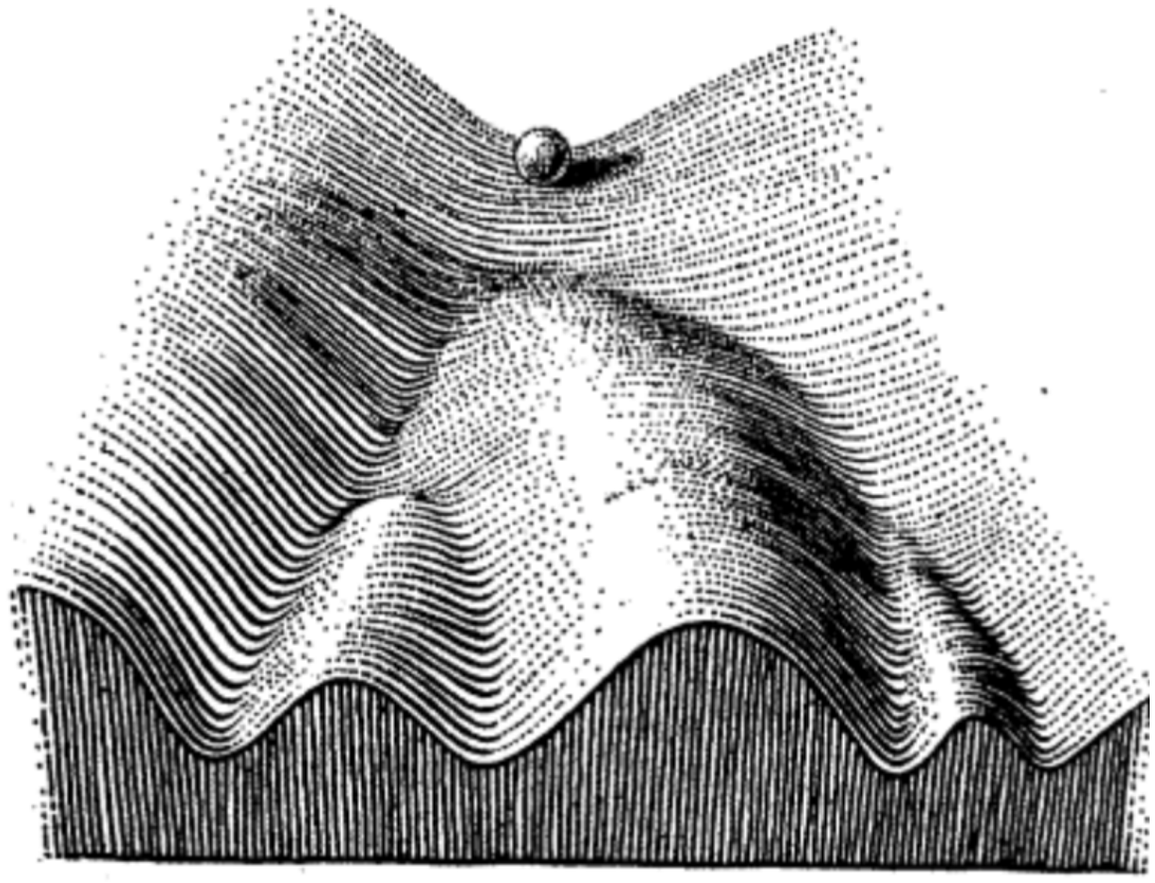
What is epigenetics

- Definitions
  - Changes on top of genetics.
  - Gene expression changes without changes in DNA sequence.
  - Operational Definition: ‘An epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence.
  - Heritability of a phenotype, passed on through either mitosis or meiosis.

Berger et al., Genes & Development 23:781–783 (2009)

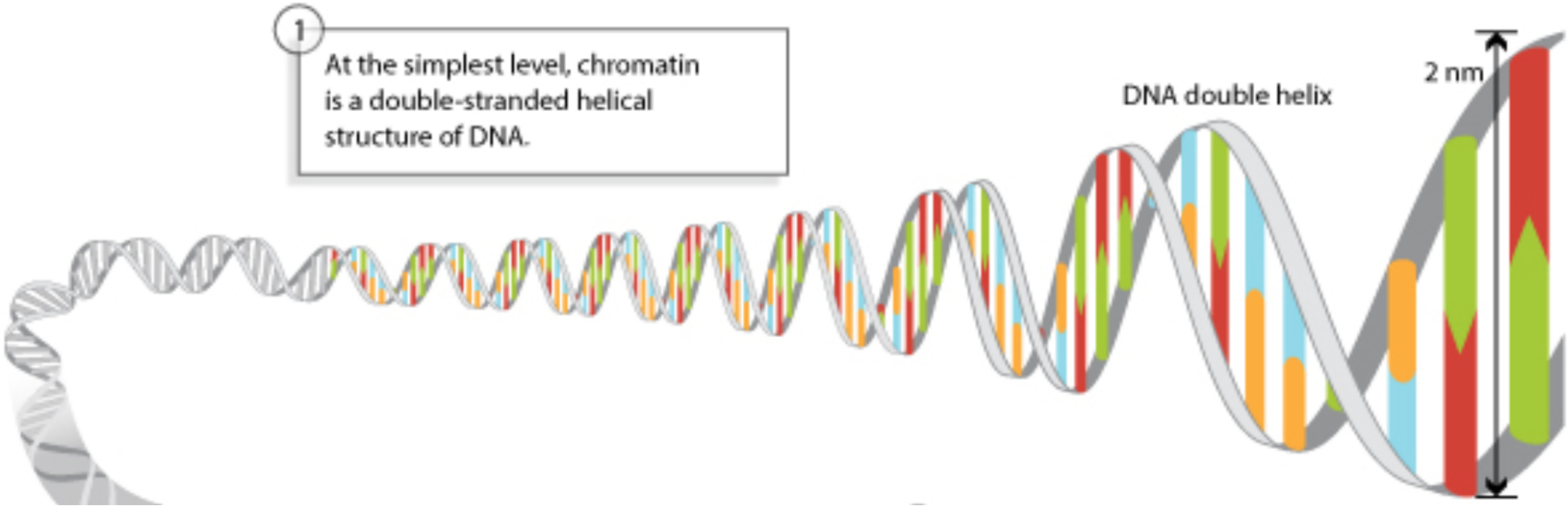
What is epigenetics

- Epigenetic Landscape
  - Differentiation
  - Beyond the genome



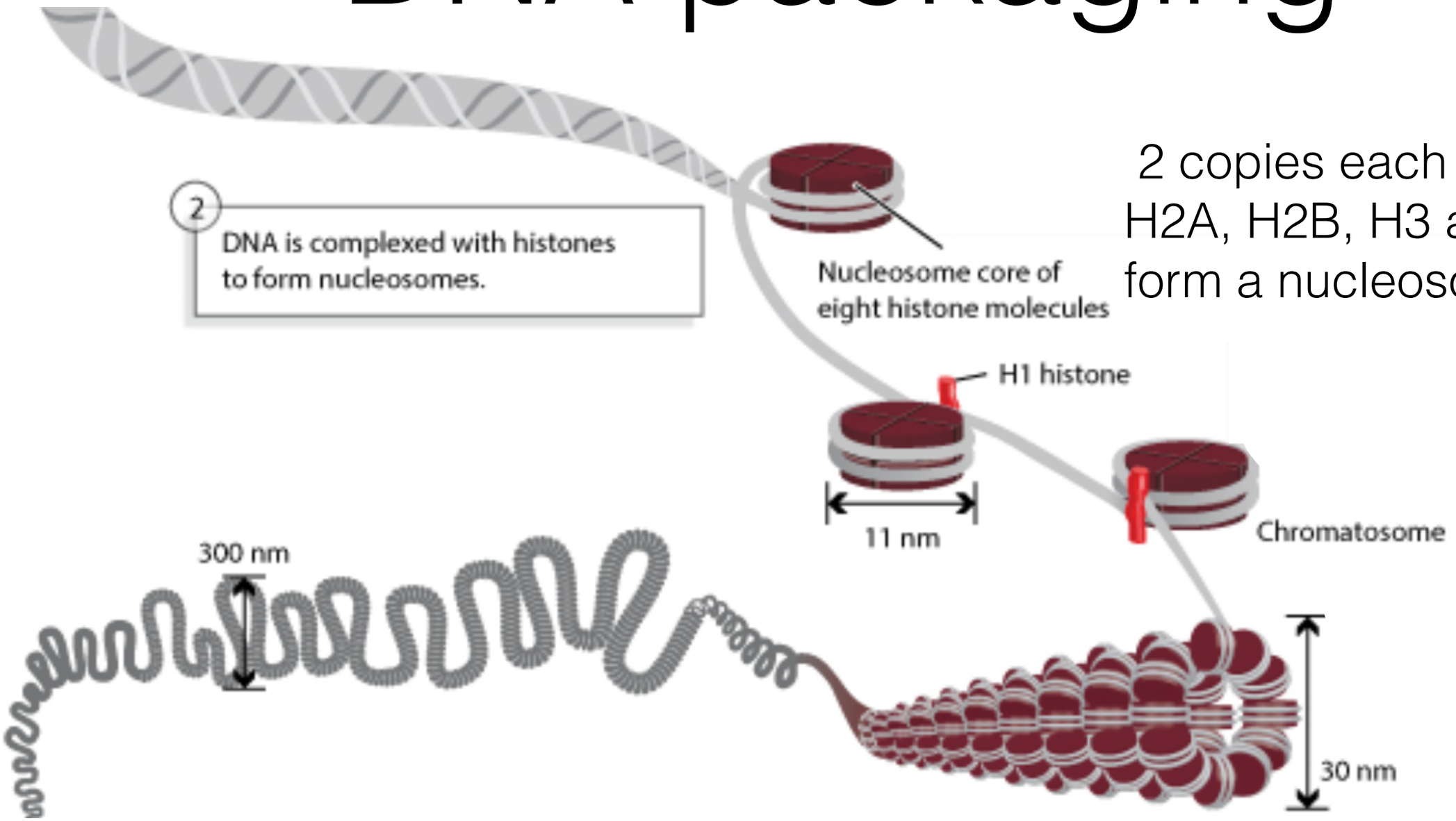
Waddington, C.H. The Strategy of the Genes, Allen & Unwin, London (1957)

DNA packaging



Nature Education (2013) Adapted from Pierce, Benjamin. Genetics: A Conceptual Approach, 2nd ed. http://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310

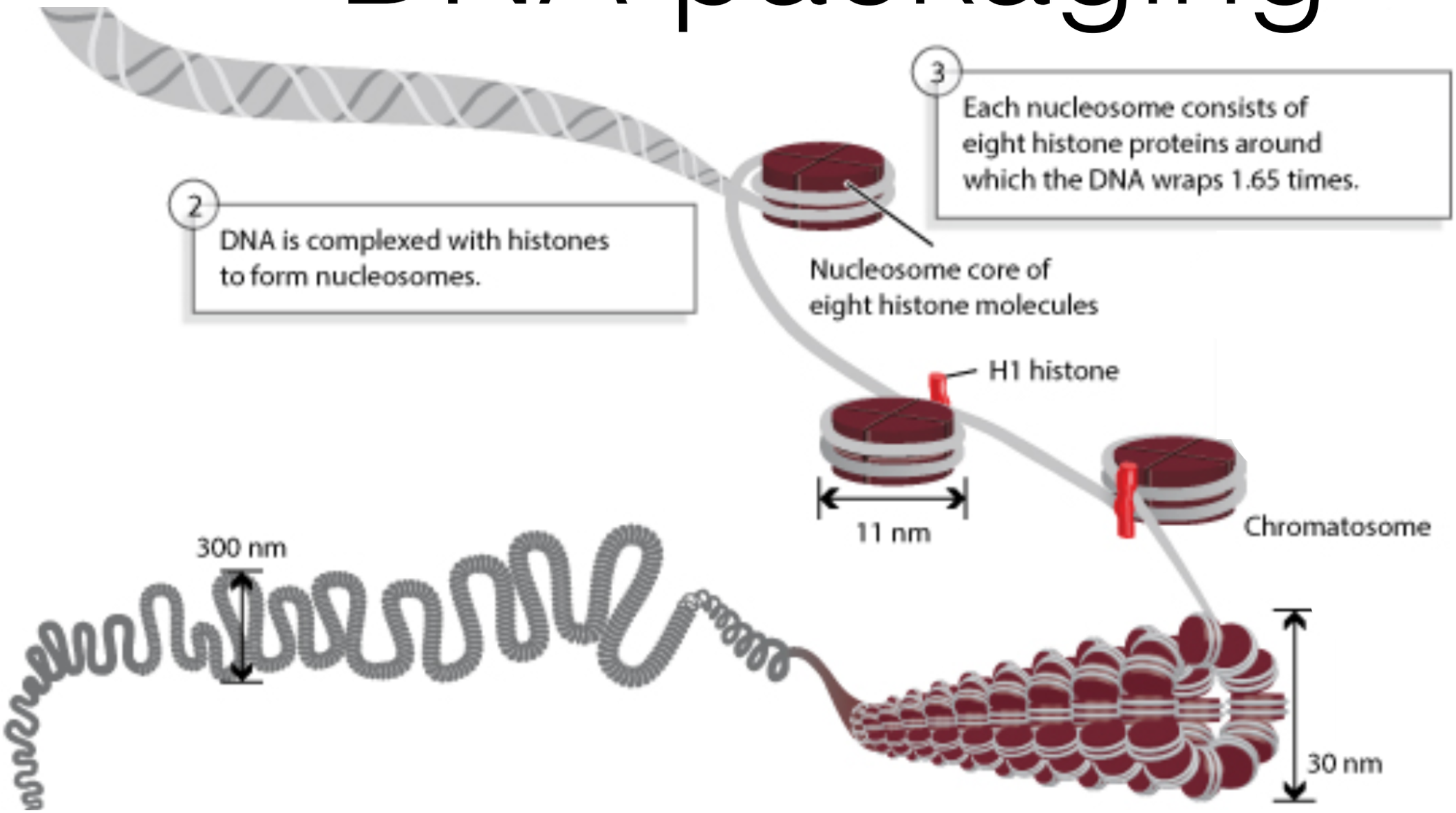
DNA packaging



2 copies each of histones H2A, H2B, H3 and H4 to form a nucleosome.

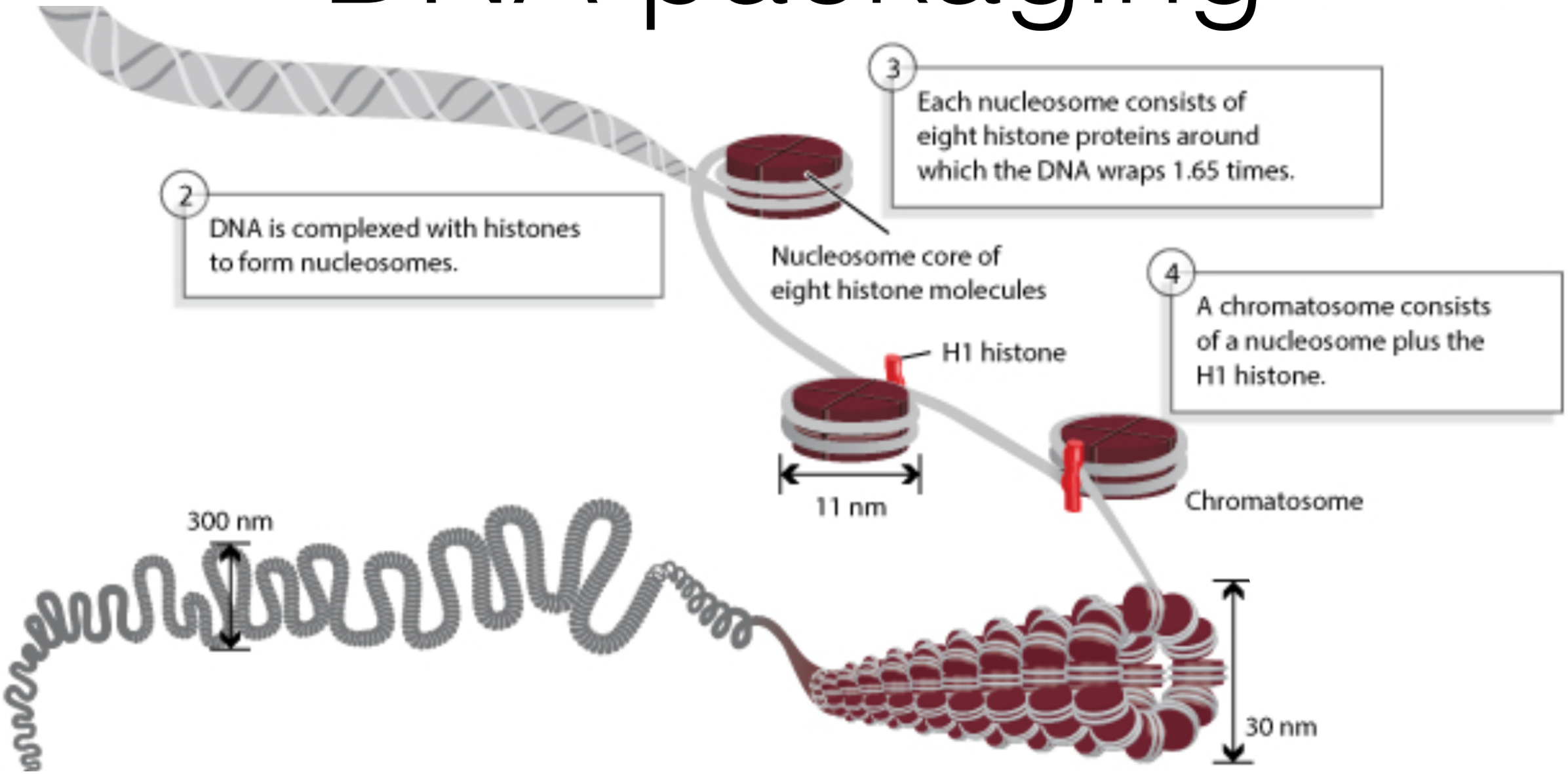
Nature Education (2013) Adapted from Pierce, Benjamin. Genetics: A Conceptual Approach, 2nd ed. http://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310

DNA packaging



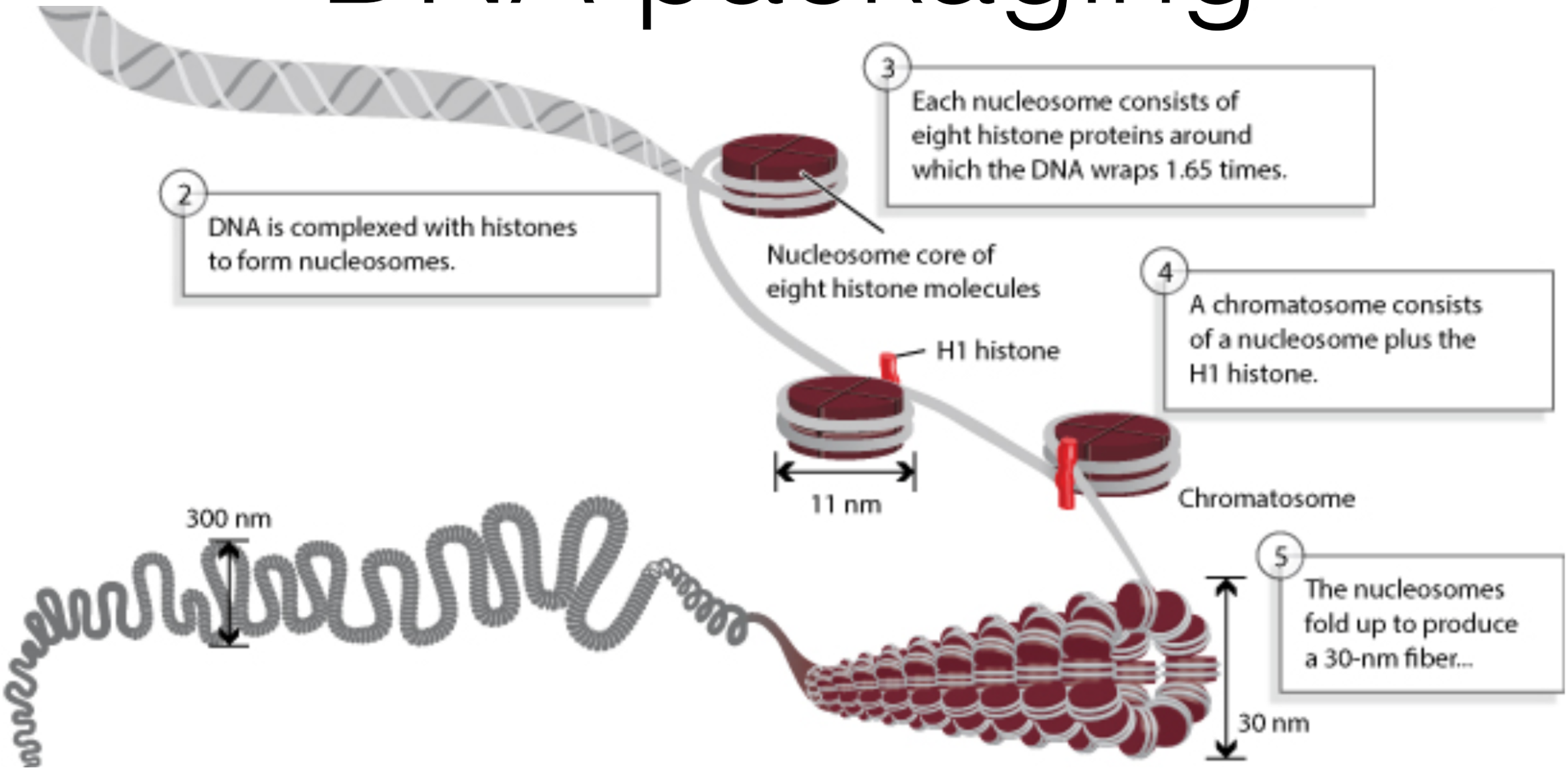
Nature Education (2013) Adapted from Pierce, Benjamin. Genetics: A Conceptual Approach, 2nd ed. http://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310

DNA packaging



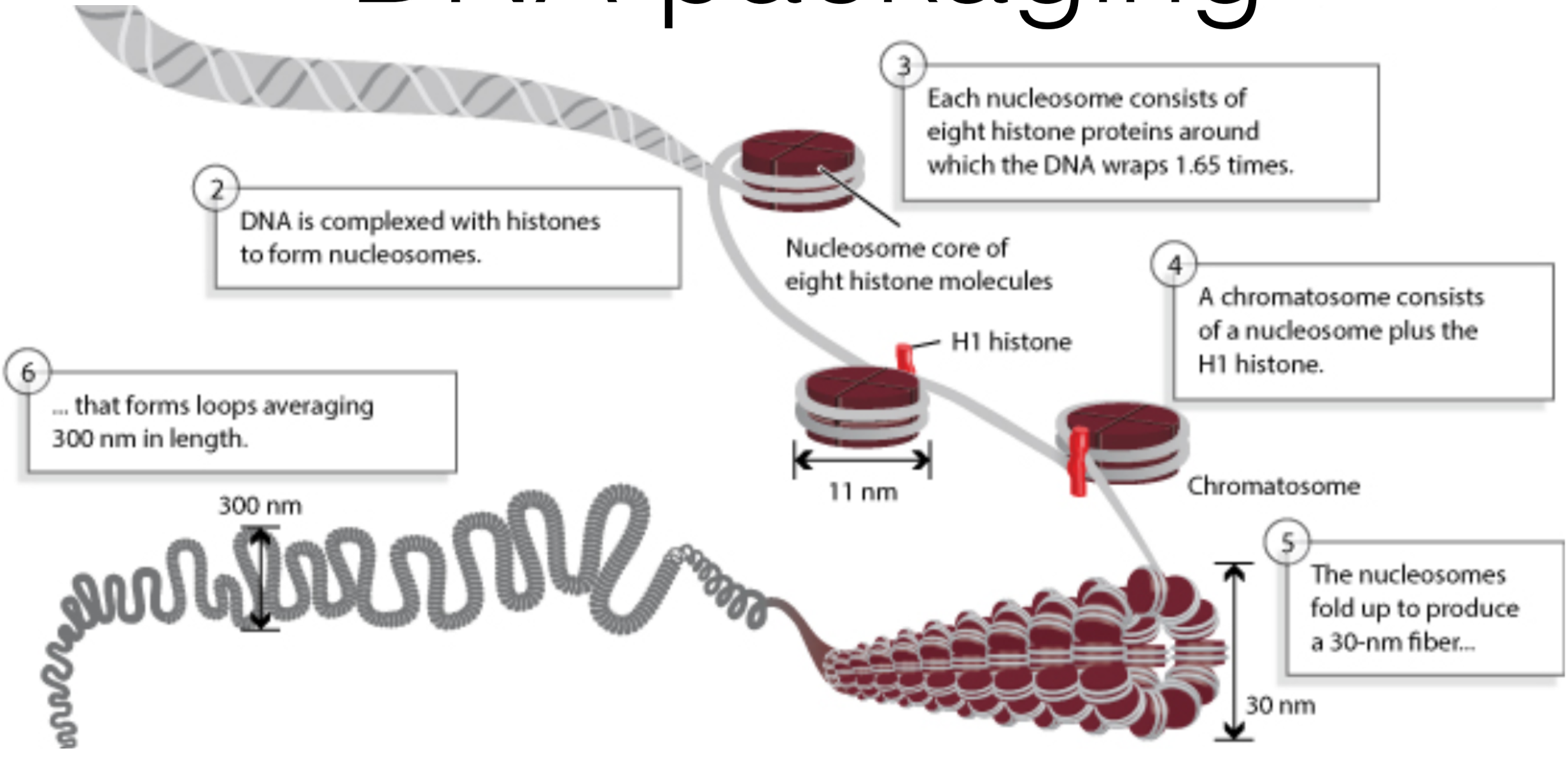
Nature Education (2013) Adapted from Pierce, Benjamin. Genetics: A Conceptual Approach, 2nd ed. http://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310

DNA packaging



Nature Education (2013) Adapted from Pierce, Benjamin. Genetics: A Conceptual Approach, 2nd ed. http://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310

DNA packaging



Nature Education (2013) Adapted from Pierce, Benjamin. Genetics: A Conceptual Approach, 2nd ed. http://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310

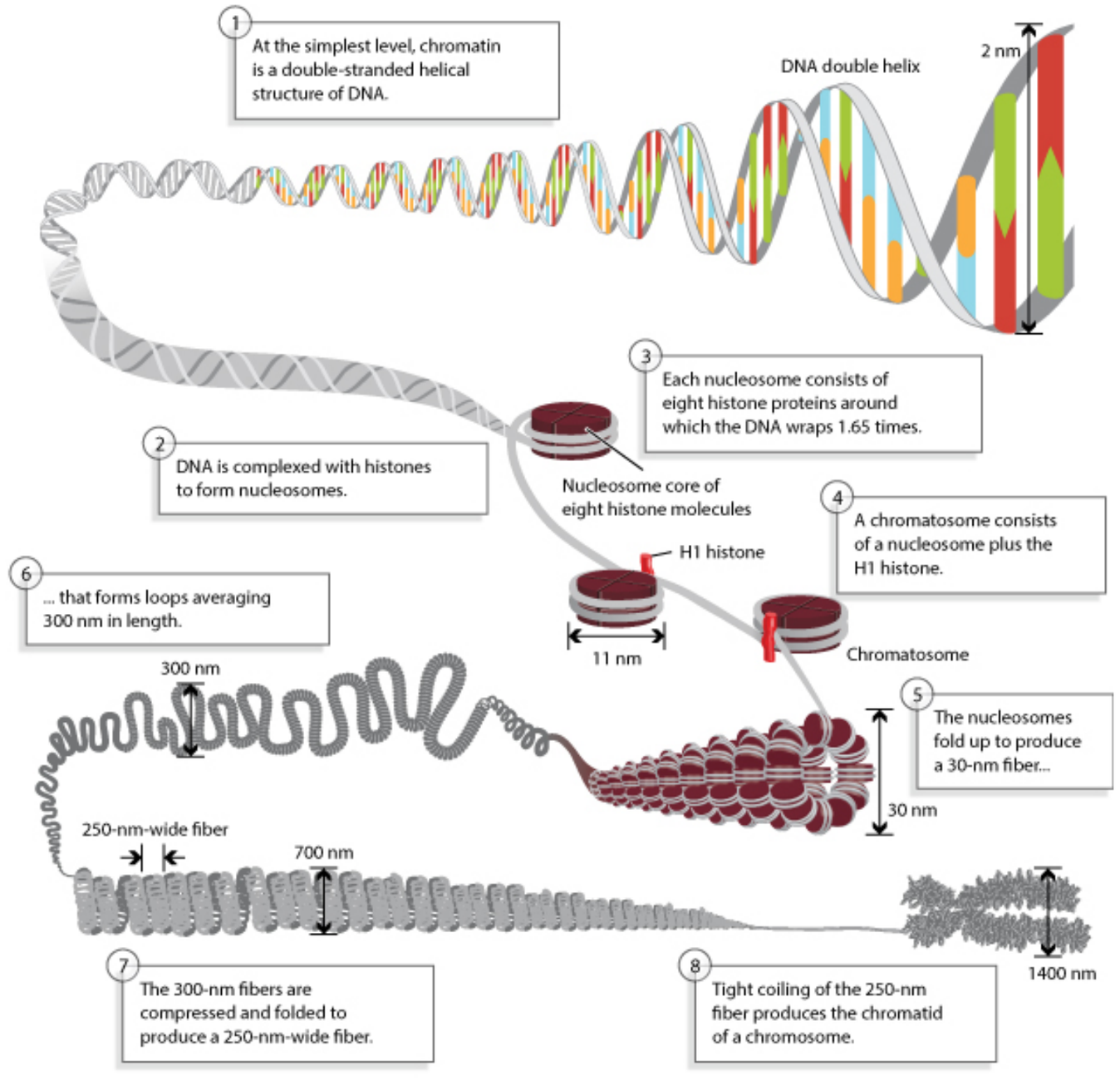
DNA packaging

Nature Education (2013) Adapted from Pierce, Benjamin. Genetics: A Conceptual Approach, 2nd ed. http://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310

DNA packaging

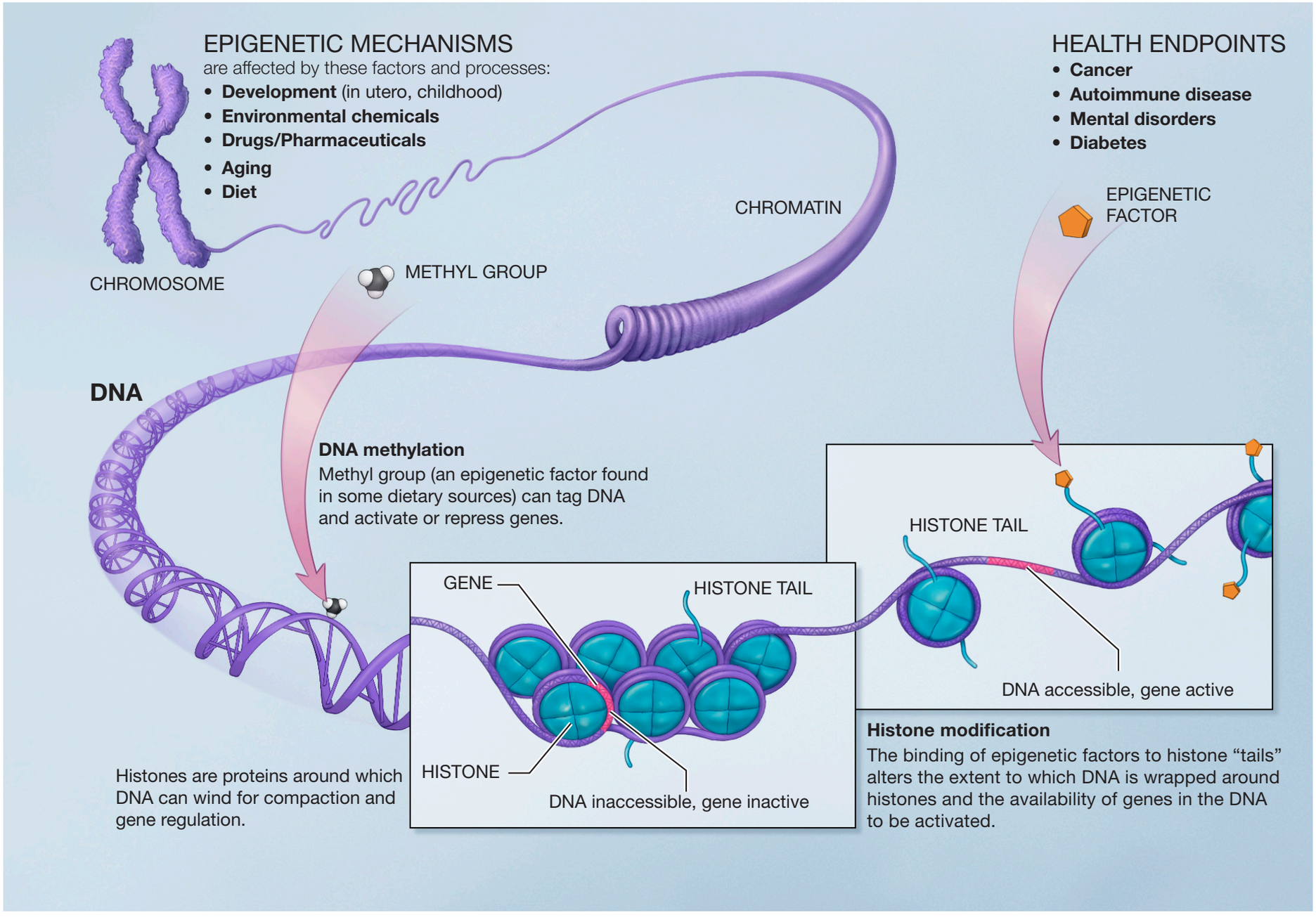
Nature Education (2013) Adapted from Pierce, Benjamin. Genetics: A Conceptual Approach, 2nd ed. http://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310

DNA packaging



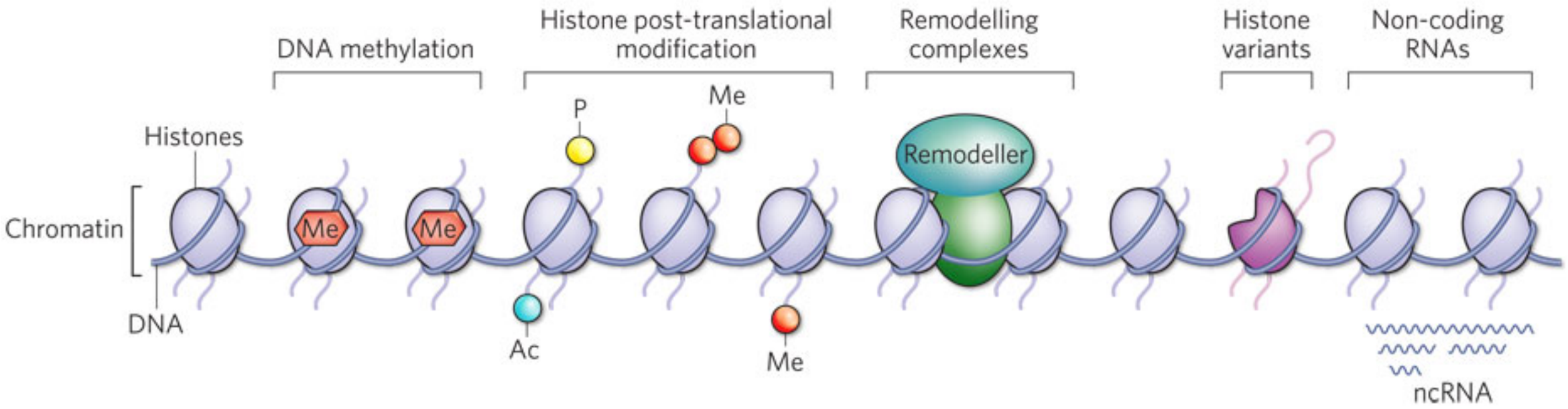
Nature Education (2013) Adapted from Pierce, Benjamin. Genetics: A Conceptual Approach, 2nd ed. http://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310

Epigenetic Mechanisms



http://commonfund.nih.gov/epigenomics/figure (2013)

Mechanisms Affecting Chromatin Structure



Ac, acetyl; Me, methyl; P, phosphate

Dulac, Nature 465, 728–735 (2010)

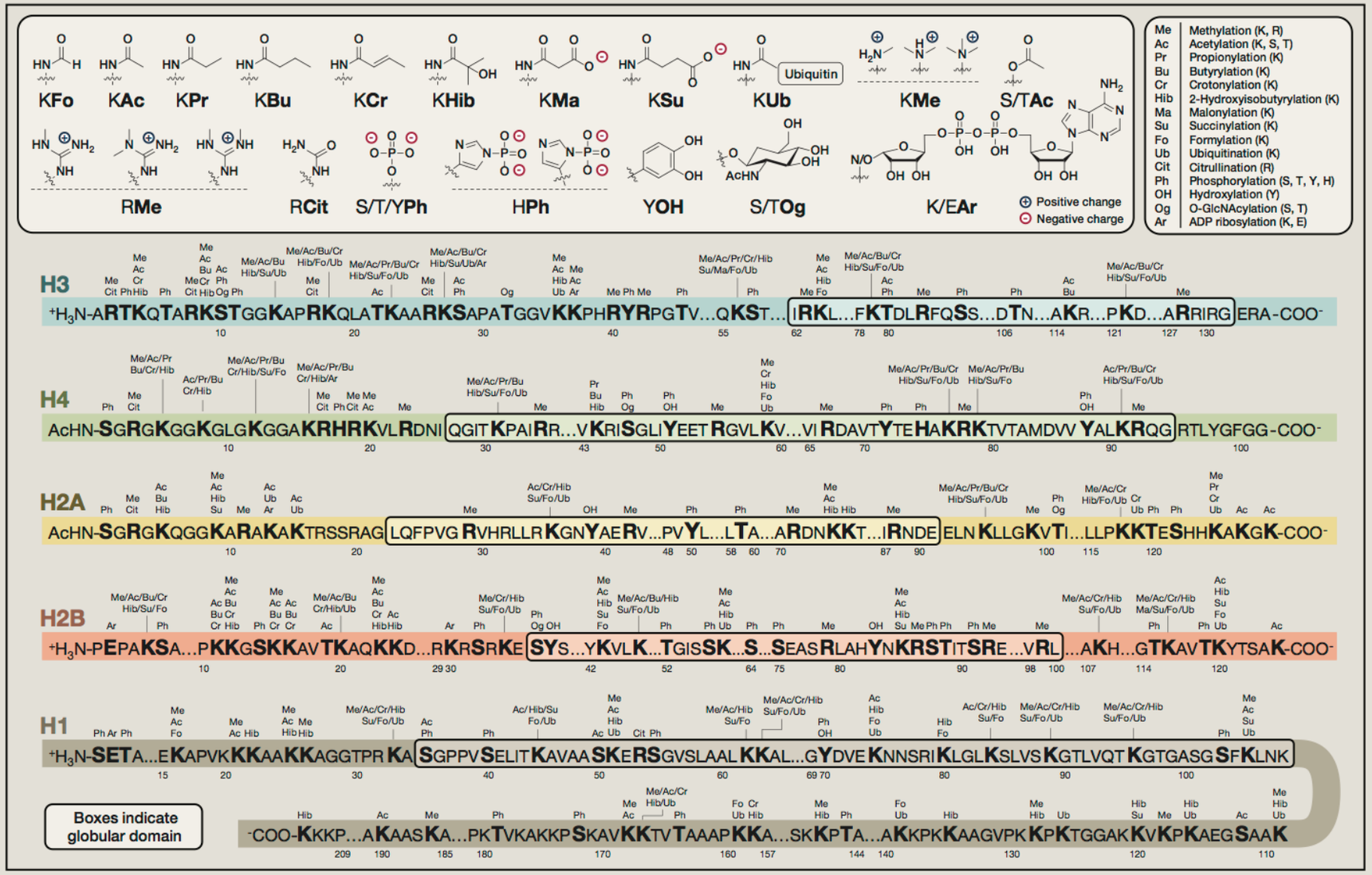
Histone Modifications

458 Cell 159, October 9, 2014 ©2014 Elsevier Inc. DOI http://dx.doi.org/10.1016/j.cell.2014.09.037 See online version for for legend and references.

SnapShot: Histone Modifications

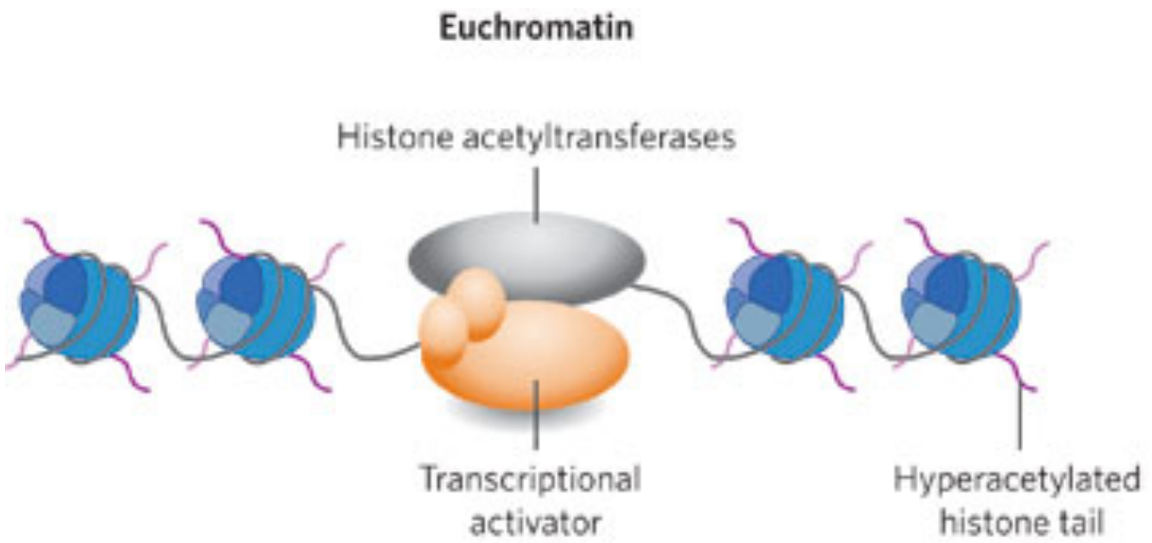
He Huang,[1] Benjamin R. Sabari,[2] Benjamin A. Garcia,[3] C. David Allis,[2] and Yingming Zhao[1]
[1]Ben May Department of Cancer Research, The University of Chicago, Chicago, IL 60637, USA
[2]Laboratory of Chromatin Biology and Epigenetics, The Rockefeller University, New York, NY 10021, USA
[3]Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104, USA

Cell

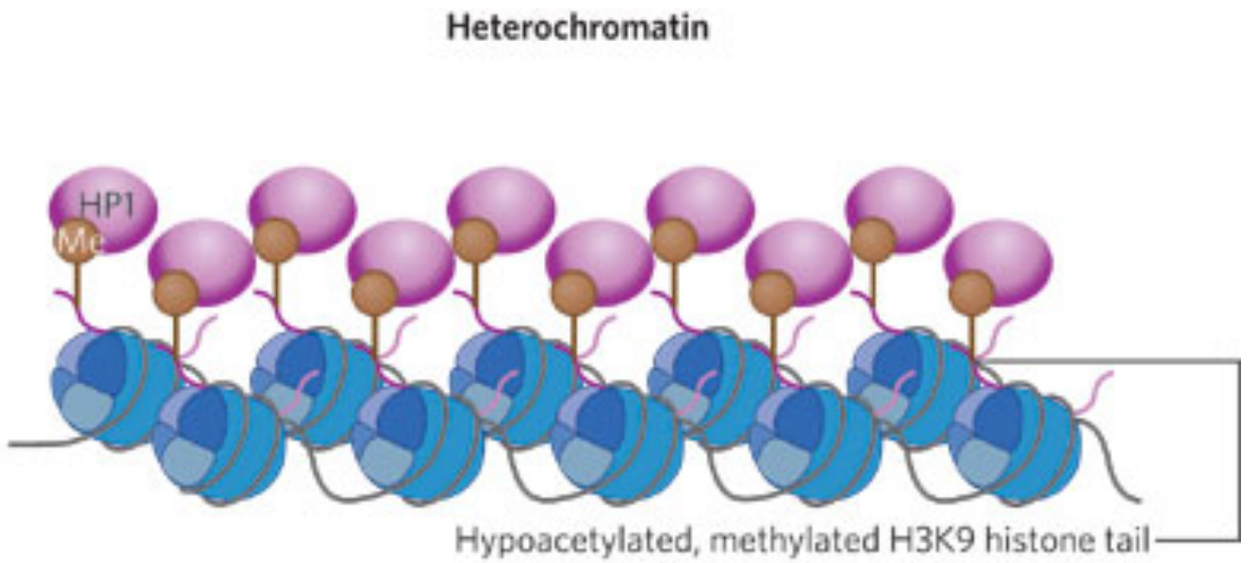


Huang et al., Cell 159 (2014)

Euchromatic Vs heterochromatic regions



- Less condensed
- At chromosome arms
- Contains unique sequences
- Gene-rich
- Replicated throughout S phase
- Recombination during meiosis

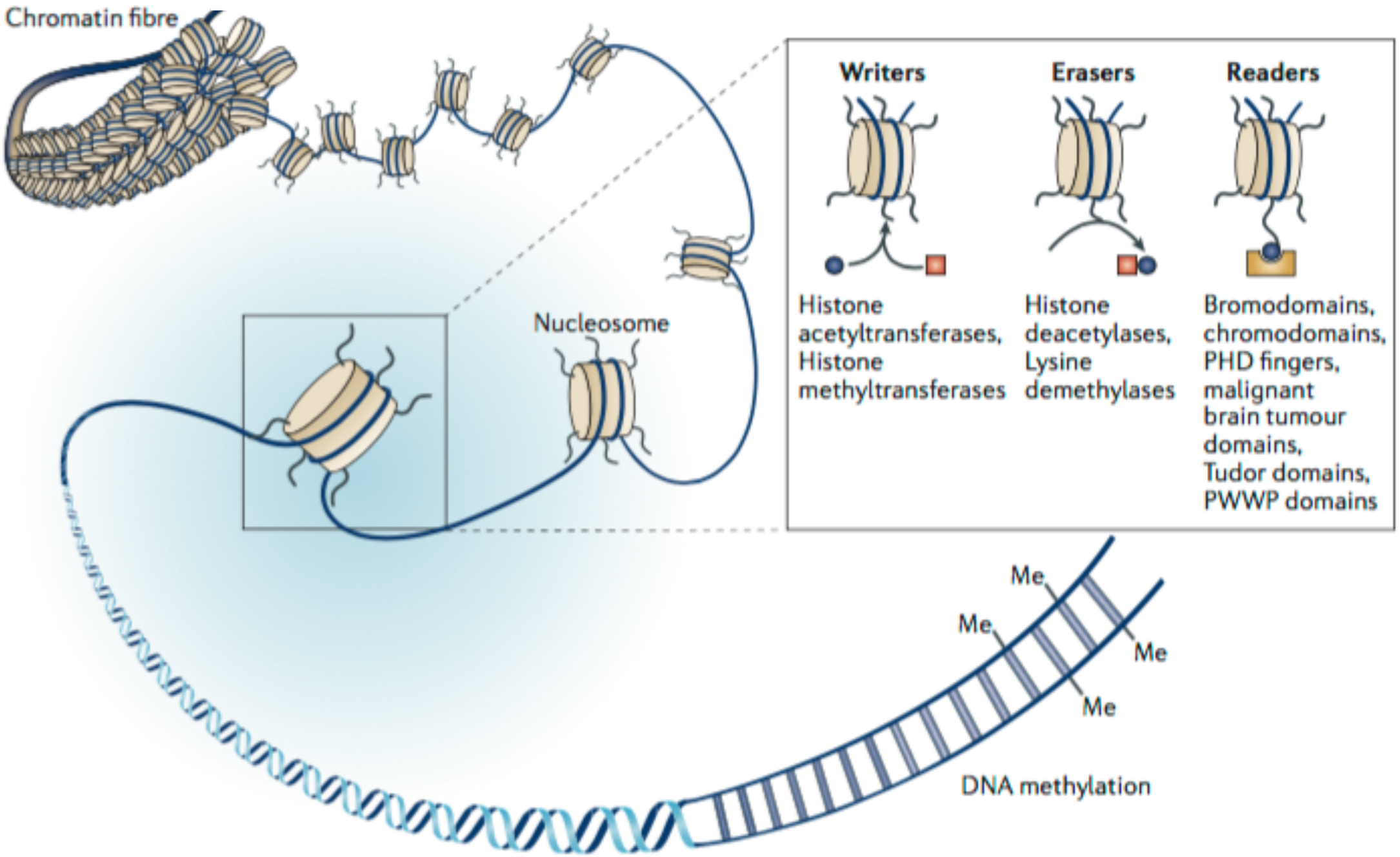


- Highly condensed
- At centromeres and telomeres
- Contains repetitious sequences
- Gene-poor
- Replicated in late S phase
- No meiotic recombination

Less condensed chromatin more accessible to polymerases, transcription factors

Grewal & Elgin, *Nature* **447**, 399-406 (2007)

Histone PTM Mediators



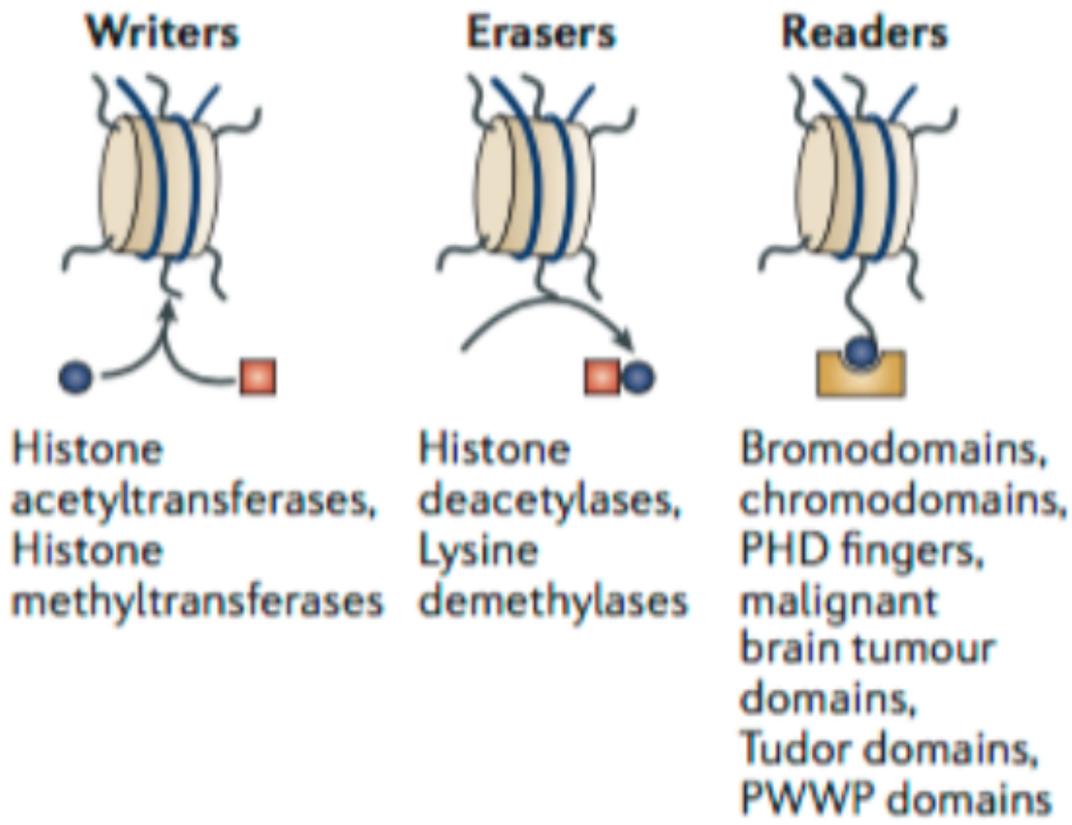
Arrowsmith et al., Nature Reviews Drug Discovery 11, 384-400 (2012)

Protein families

Epigenome protein families

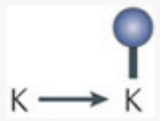
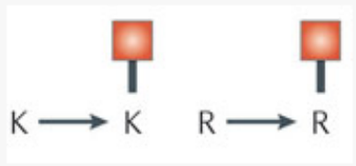
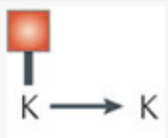
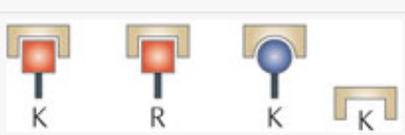
- deposit (‘write')
- bind to (‘read')
- remove (‘erase’)

On specific lysine or arginine side chains of histonesmethyl marks (orange squares) or acetyl marks (blue circles)



Arrowsmith et al., Nature Reviews Drug Discovery 11, 384-400 (2012)

Protein families

Family	Activity	Number of proteins	Major classes and function
Writers			
Histone acetyltransferases		18	• MYST family (MOZ, SAS2, YBF2/SAS3, TIP60) proteins: involved in DNA damage and oncogenic translocation • GNAT: involved in EGF signalling and cell cycle progression • EP300: promiscuous (involved in a range of cellular events)
Protein methyltransferases		60	• SET domain: methylates both histone and non-histone lysines • PRMTs: methylate both histone and non-histone arginines • PRDMs: SET domain-like tissue-specific factors
Erasers			
Histone deacetylases		17	• Classes I, IIb and IV enzymes: have both histone and non-histone substrates, involved in gene silencing • Class IIa enzymes: scaffolding proteins • Sirtuins (class III): NAD-dependent, have deacetylation and ADP-ribosylation activity
Lysine demethylases		25	• Lysine-specific demethylases: flavin-dependent enzymes that regulate transcription during development • Jumonji domain: 2-oxoglutarate-dependent
Readers			
Bromodomain-containing proteins		61	• Targeting of chromatin-modifying enzymes to specific sites, often physically linked to PHD fingers and the catalytic domain of histone acetyltransferases
Methyl-lysine- and/or methyl-arginine-binding domain-containing proteins (for example, Tudor domains, MBT domains, chromodomains and PWWP domains)		95	• Tudor domains: bind dimethylated lysine, trimethylated lysine and dimethylated arginine • MBT domains: bind monomethylated and dimethylated lysine with low sequence specificity • Chromodomains: bind trimethylated lysine with sequence specificity • PWWP domains: bind to both trimethylated lysine and DNA
PHD-containing proteins		104	• A large and diverse family that acts on multiple substrates

EGF, epidermal growth factor

EP300, E1A-associated protein p300

GNAT, glycine-*N*-acyltransferase-like protein 1

MBT, malignant brain tumour domain

MYST, histone acetyltransferase MYST

PHD, plant homeodomain

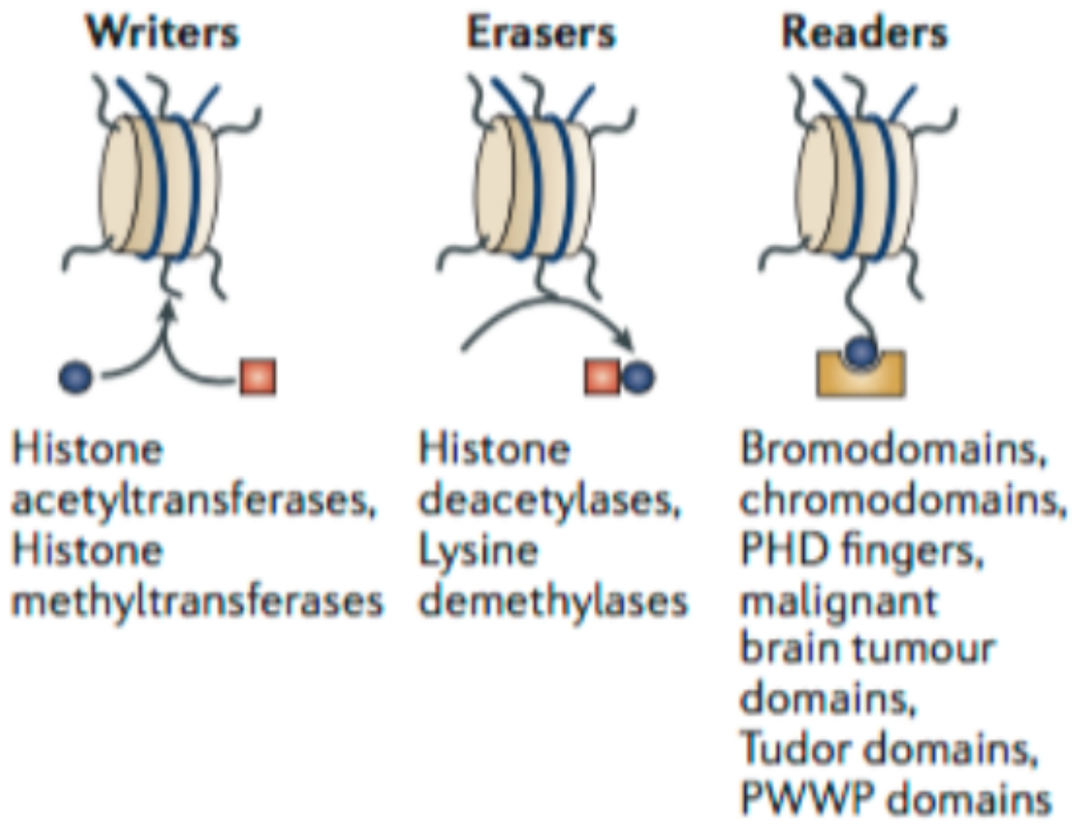
PRDM, PR domain-containing protein

PRMT, protein arginine methyltransferase.

Arrowsmith et al., Nature Reviews Drug Discovery 11, 384-400 (2012)

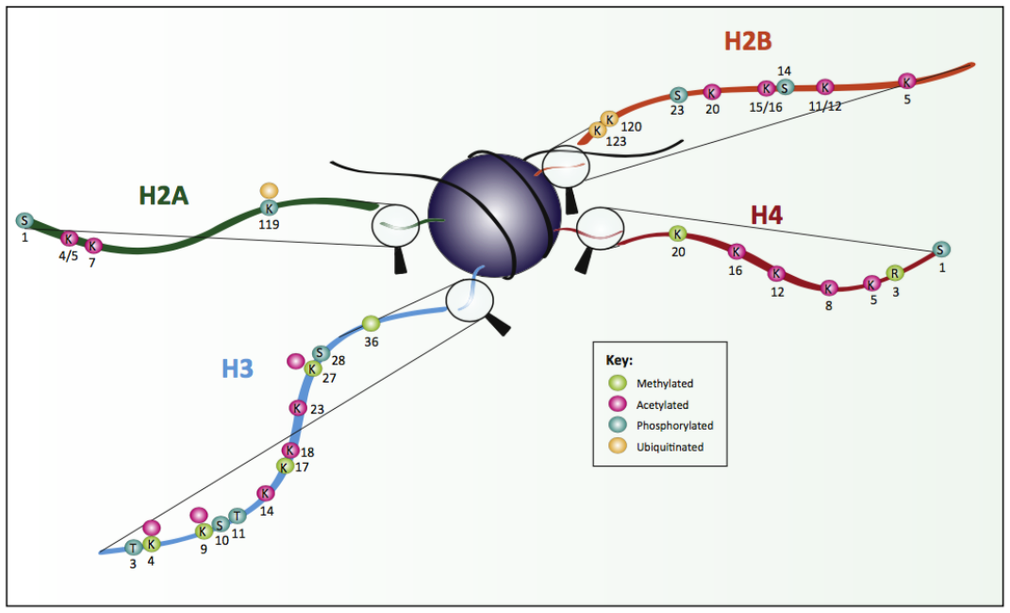
Protein families

- Histone acetyltransferases and protein methyltransferases are the enzymes responsible for writing acetyl and methyl marks, respectively.
- Histone deacetylases and lysine demethylases erase the marks.
- Bromodomains bind acetylated lysines (shown by beige shape).
- Tudor domains, MBT domains, chromodomains and PWWP domains bind methyl marks on lysine or arginine residues (shown by beige shape).
- PHD fingers are present in a large number of proteins and read either methyl or acetyl marks on lysine or arginine side chains, as well as unmodified lysines.



Arrowsmith et al., Nature Reviews Drug Discovery 11, 384-400 (2012)

Histone Tail Modifications



Histone	Modification	Role
H2A	H2AS1P	Mitosis; chromatin assembly
	H2AK4/5ac	Transcriptional activation
	H2AK7ac	Transcriptional activation
	H2AK119P	Spermatogenesis
	H2AK119uq	Transcriptional repression
H2B	H2BS14P	Apoptosis
	H2BS33P	Transcriptional activation
	H2BK5ac	Transcriptional activation
	H2BK11/12ac	Transcriptional activation
	H2BK15/16ac	Transcriptional activation
	H2BK20ac	Transcriptional activation
	H2BK120uq	Spermatogenesis/meiosis
	H2BK123uq	Transcriptional activation

- Modifications on the N-terminal ‘tail’ regions histones
- Accessible on surface

Lawrence, Daujat & Schneider, Trends Genet. 32(1):42-56 (2016)

Histone Tail Modifications

Histone	Modification	Role
H3	H3K4me2	Permissive euchromatin
	H3K4me3	Transcriptional elongation; active euchromatin
	H3K9me3	Transcriptional repression; imprinting; DNA methylation
	H3R17me	Transcriptional activation
	H3K27me3	Transcriptional silencing; X-inactivation; bivalent genes/gene poising
	H3K36me3	Transcriptional elongation
	H3K4ac	Transcriptional activation
	H3K9ac	Histone deposition; transcriptional activation
	H3K14ac	Transcriptional activation; DNA repair
	H1K18ac	Transcriptional activation; DNA repair; DNA replication
	H3K23ac	Transcriptional activation; DNA repair
	H3K27ac	Transcriptional activation
	H3T3P	Mitosis
	H3S10P	Mitosis; meiosis; transcriptional activation
	H3T11/S28P	Mitosis

Histone	Modification	Role
H4	H4R3me	Transcriptional activation
	H4K20me1	Transcriptional silencing
	H4K20me3	Heterochromatin
	H4K5ac	Histone deposition; transcriptional activation; DNA repair
	H4K8ac	Transcriptional activation; DNA repair; transcriptional elongation
	H4K12ac	Histone deposition; telomeric silencing; transcriptional activation; DNA repair
	H4K16ac	Transcriptional activation; DNA repair
	H4S1P	Mitosis

Lawrence, Daujat & Schneider, Trends Genet. 32(1):42-56 (2016)

Histone globular domain Modifications

Central globular domains of histones contain modification sites

Histone	Site	Modification
H2A	H2AK36	Acetylation
	H2AK99	Methylation
	H2AQ105	Methylation
	H2AK119	Acetylation
	H2AK119	Ubiquitylation
H2B	H2BK40	Methylation
	H2BK82	Acetylation
	H2BR96	Methylation
	H2BK105	Acetylation
	H2BK113	Acetylation
	H2BK117	Acetylation

Histone	Site	Modification
H3	H3Y41	Phosphorylation
	H3R42	Methylation
	H3T45	Phosphorylation
	H3R53	Methylation
	H3K56	Acetylation
	H3K56	Methylation
	H3K64	Acetylation
	H3K64	Methylation
	H3K79	Methylation
	H3K115	Acetylation
	H3T118	Phosphorylation
	H3K122	Acetylation
H4	H4K31	Acetylation
	H4S47	Phosphorylation
	H4K59	Methylation
	H4K77	Acetylation
	H4K79	Acetylation
	H4K91	Acetylation
	H4R92	Methylation

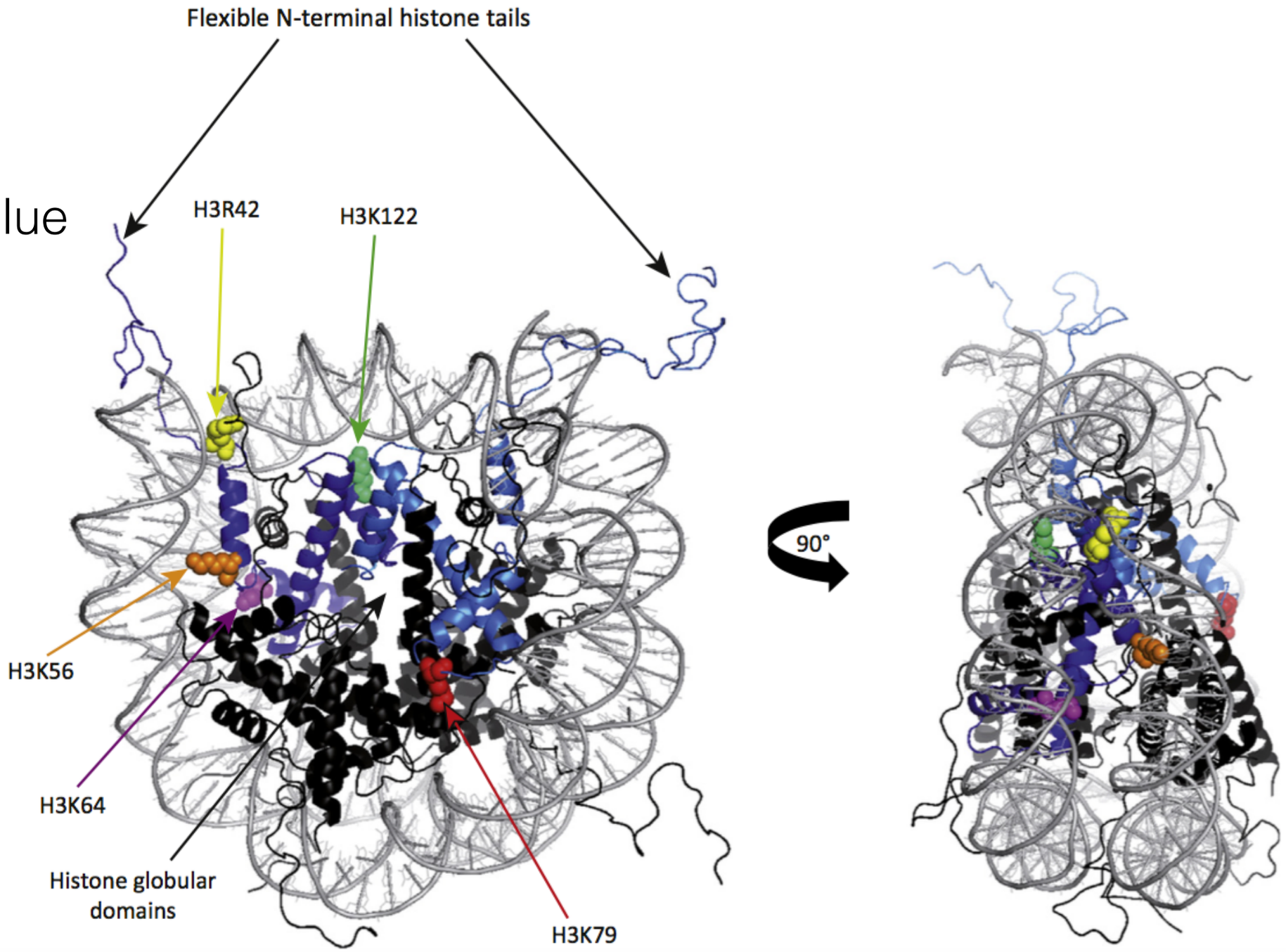
Lawrence, Daujat & Schneider, Trends Genet. 32(1):42-56 (2016)

Histone globular domain Modifications

Crystal structure of a histone octamer.

H3 histones in light blue and purple.

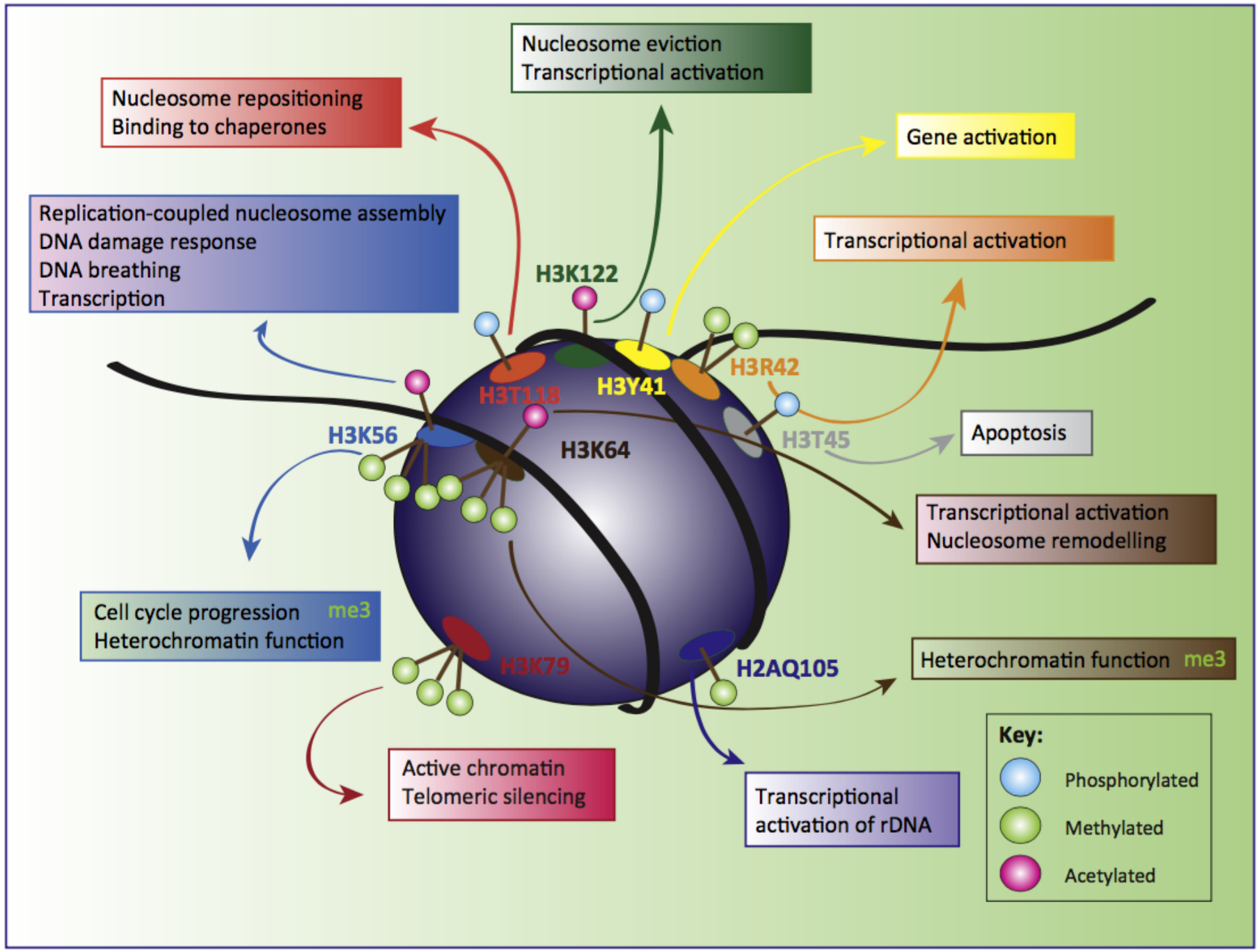
DNA in grey.



Lawrence, Daujat & Schneider, Trends Genet. 32(1):42-56 (2016)

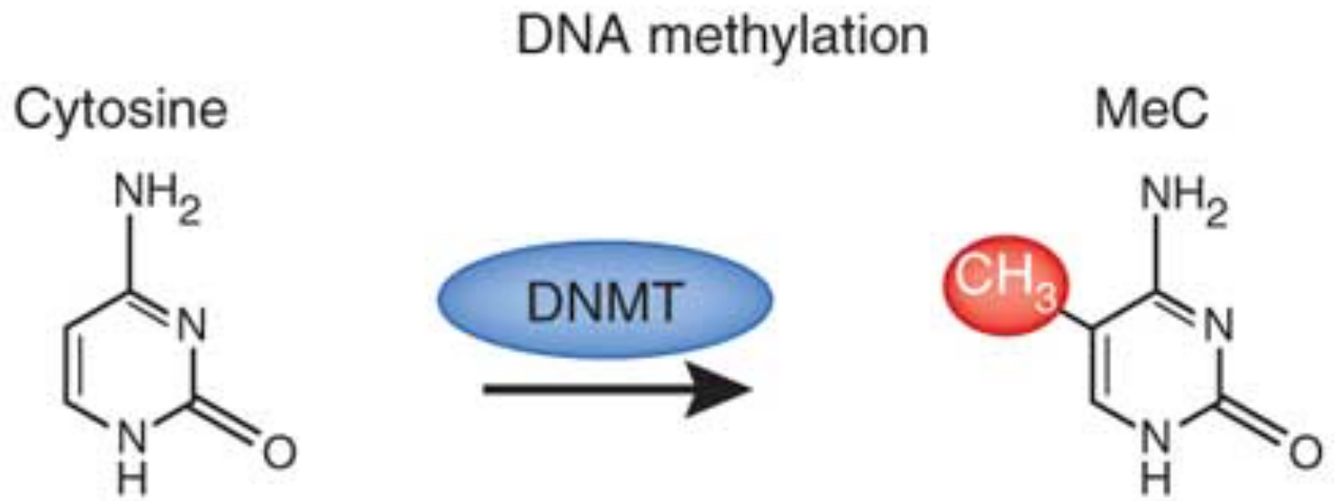
Modifications within the Globular Domains

Schematic of Locations & Functions of Key Modifications



Lawrence, Daujat & Schneider, Trends Genet. 32(1):42-56 (2016)

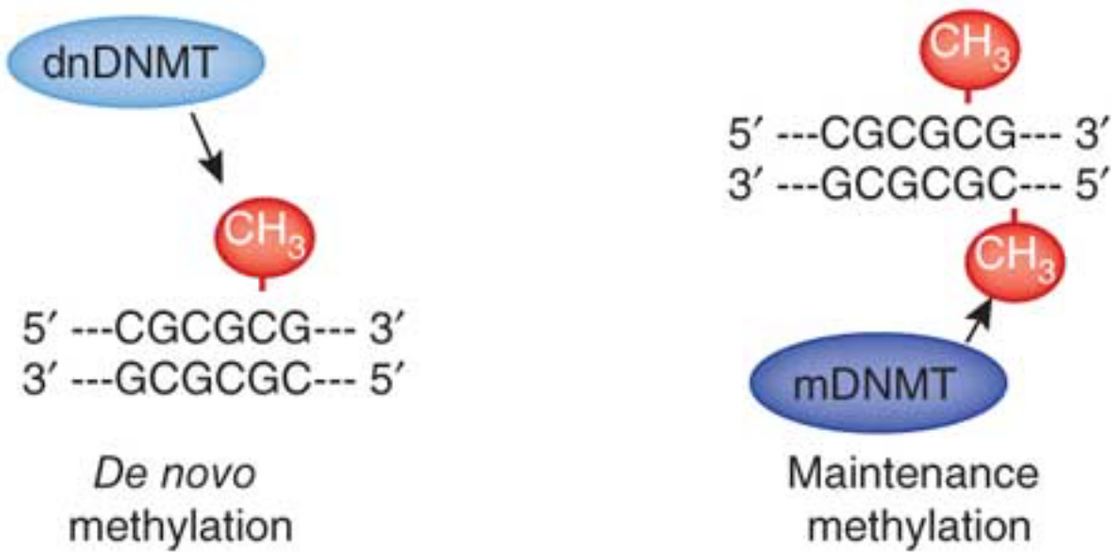
DNA Methylation



- Methylation at cytosine bases
- Methyl group added at the 5′ position on the pyrimidine ring by a
- DNA methyltransferase (DNMT)
- *A fifth nucleotide?*

Day & Sweatt, Nature Neuroscience 13, 1319–1323 (2010)

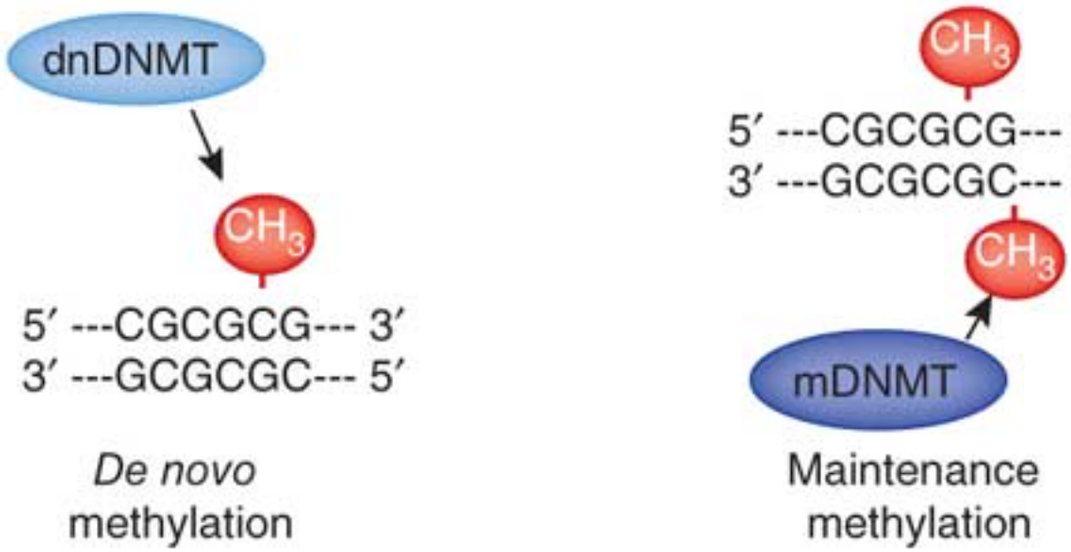
DNA Methylation



- 2 types of DNMTs initiate
- De novo DNMTs methylate previously nonmethylated cytosines
- Maintenance DNMTs methylate hemi-methylated DNA at the complementary strand.

Day & Sweatt, Nature Neuroscience 13, 1319–1323 (2010)

DNA Methylation



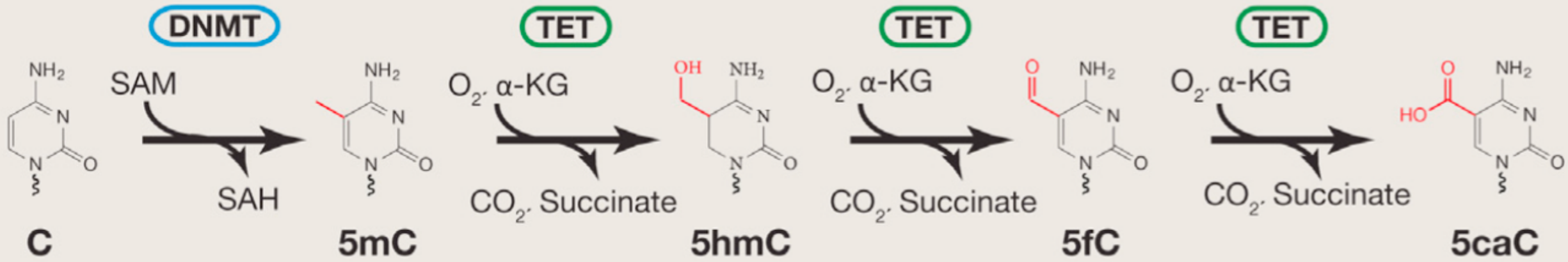
- Predominantly at CpG dinucleotides
- CpG, 5’—C—phosphate—G—3’
- Palindromic

5’ CpG 3’

3’ GpC 5’

Day & Sweatt, Nature Neuroscience 13, 1319–1323 (2010)

DNA Demethylation



TET protein family enzymes

- 5mC
- 5-hydroxymethyl cytosine (5hmC),
- 5-formylcytosine (5fC)
- 5-carboxylcytosine (5caC)

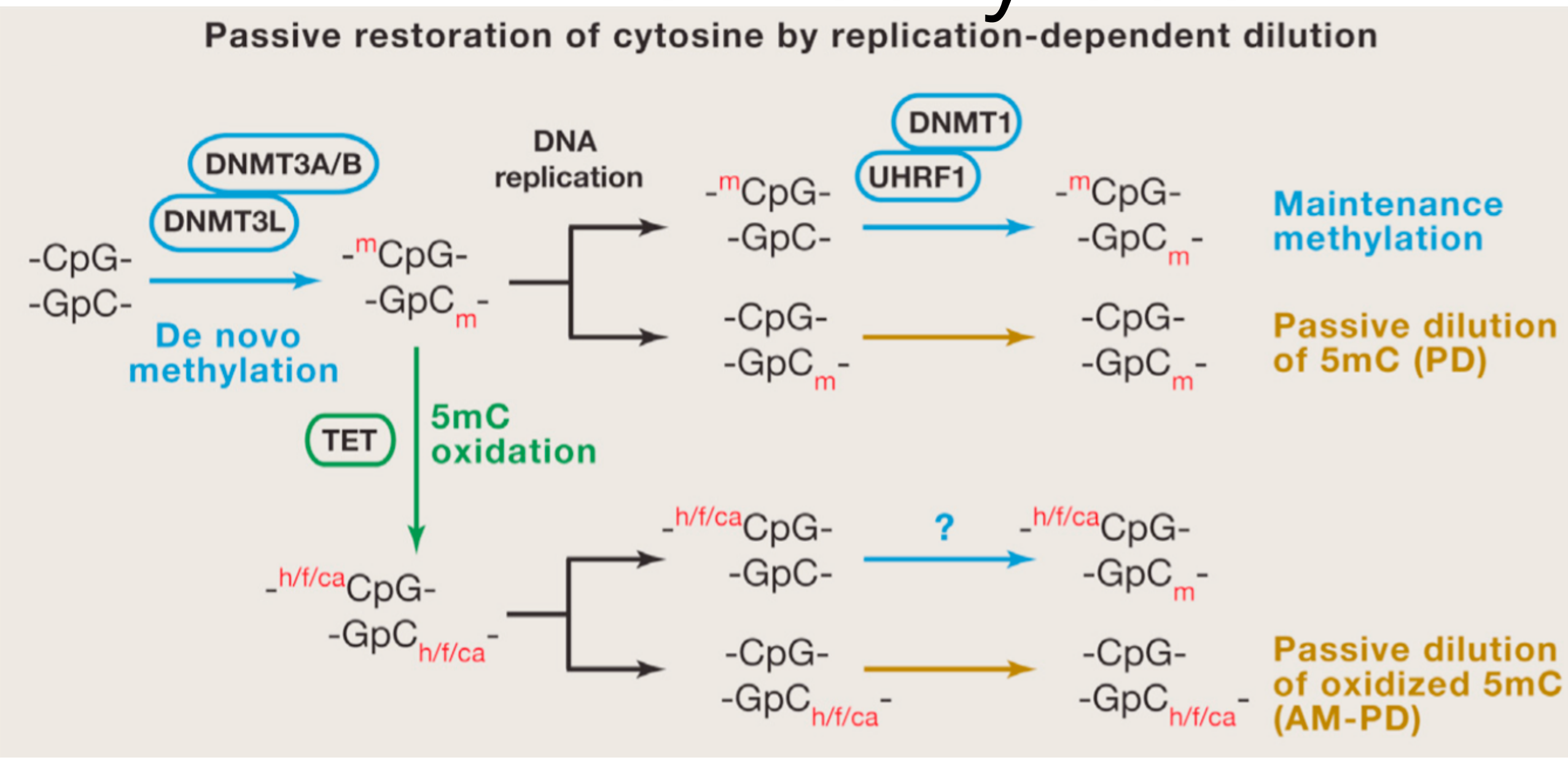
SAM, *S*-adenosylmethionine (methyl donor)

SAH, *S*-adenosylhomocysteine

TET (ten-eleven translocation)

Wu & Zhang Cell 156 p. 45 (2014)

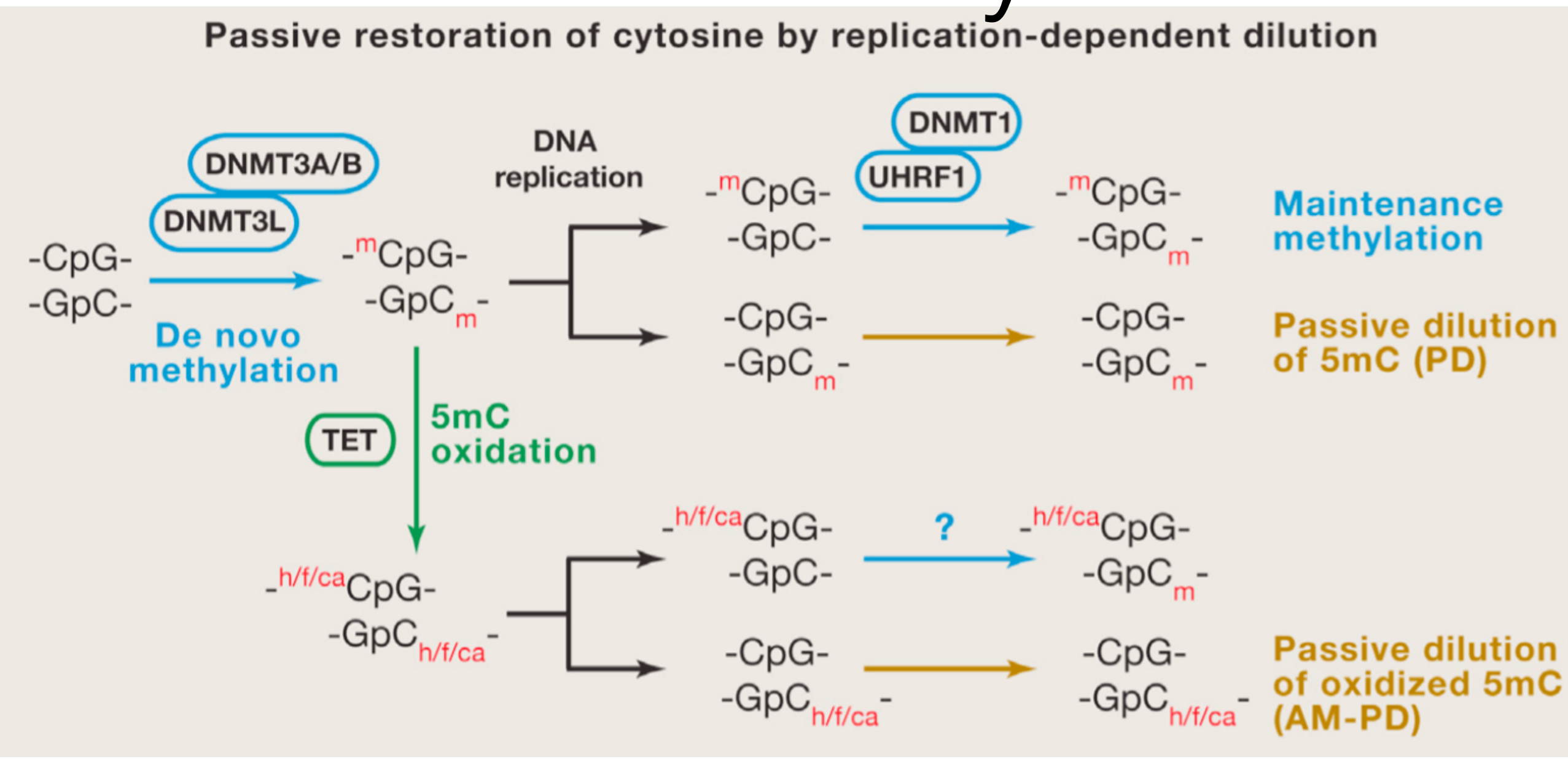
DNA Demethylation



Replication-dependent passive dilution (**PD**) of 5mC occurs in the absence of the DNA methylation maintenance machinery (DNMT1/ UHRF1).

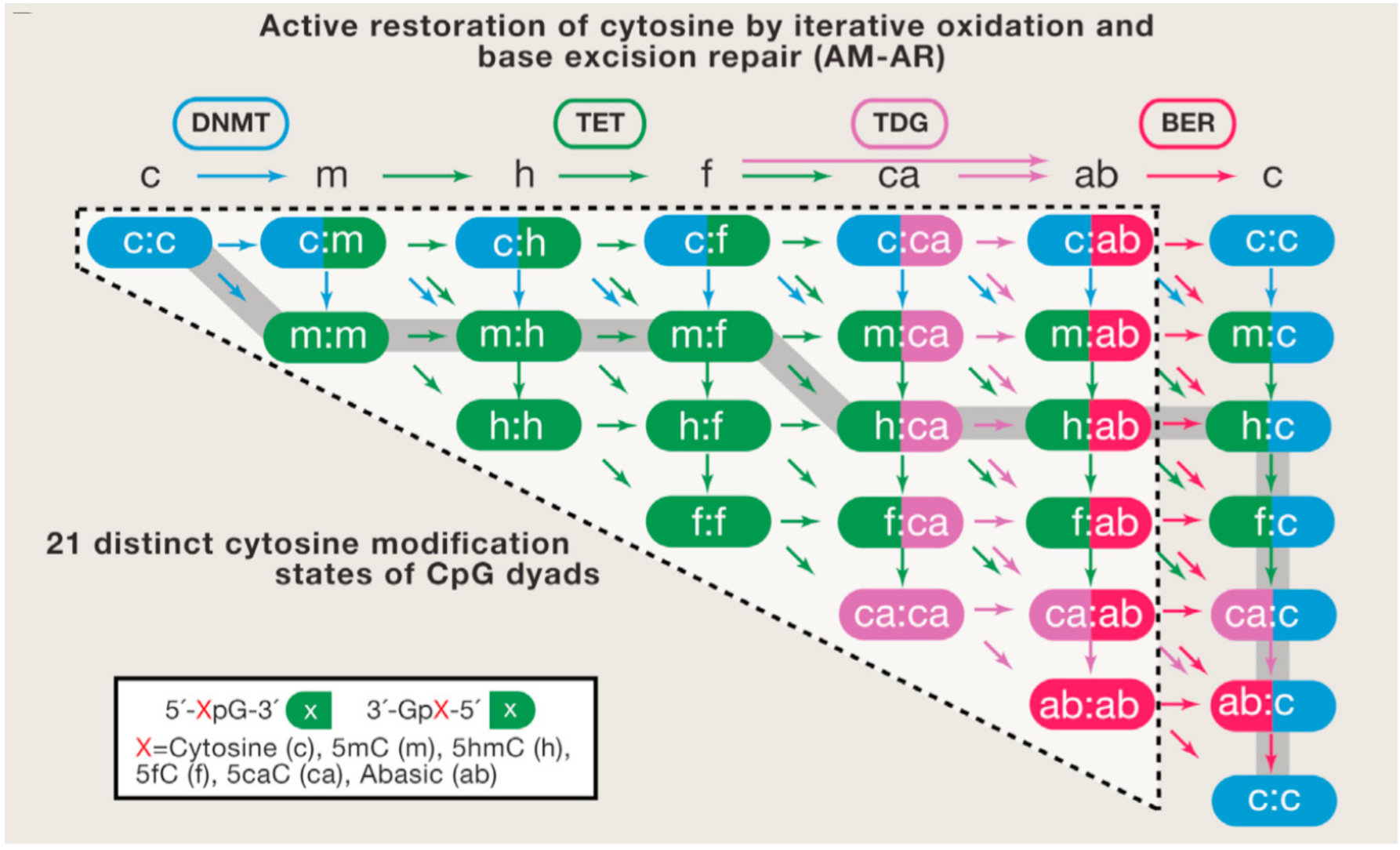
Wu & Zhang Cell 156 p. 45 (2014)

DNA Demethylation



Active modification (AM) followed by passive dilution (AM-PD). 5mC oxidation derivatives, 5hmC (h) (potentially 5fC [f] and 5caC [ca]) may facilitate passive demethylation as hemihydroxymethylated CpGs is an inefficient substrate for DNMT1. Wu & Zhang Cell 156 p. 45 (2014)

DNA Demethylation



- Schematic diagrams of replication-independent DNA demethylation within CpG dyads. TET and TDG mediate sequential 5mC oxidation and 5fC/5caC excision.
- The resulting abasic site is repaired by BER to regenerate unmodified cytosines.
- 21 intermediate states shown.

Wu & Zhang Cell 156 p. 45 (2014)

DNA Demethylation

After fertilization

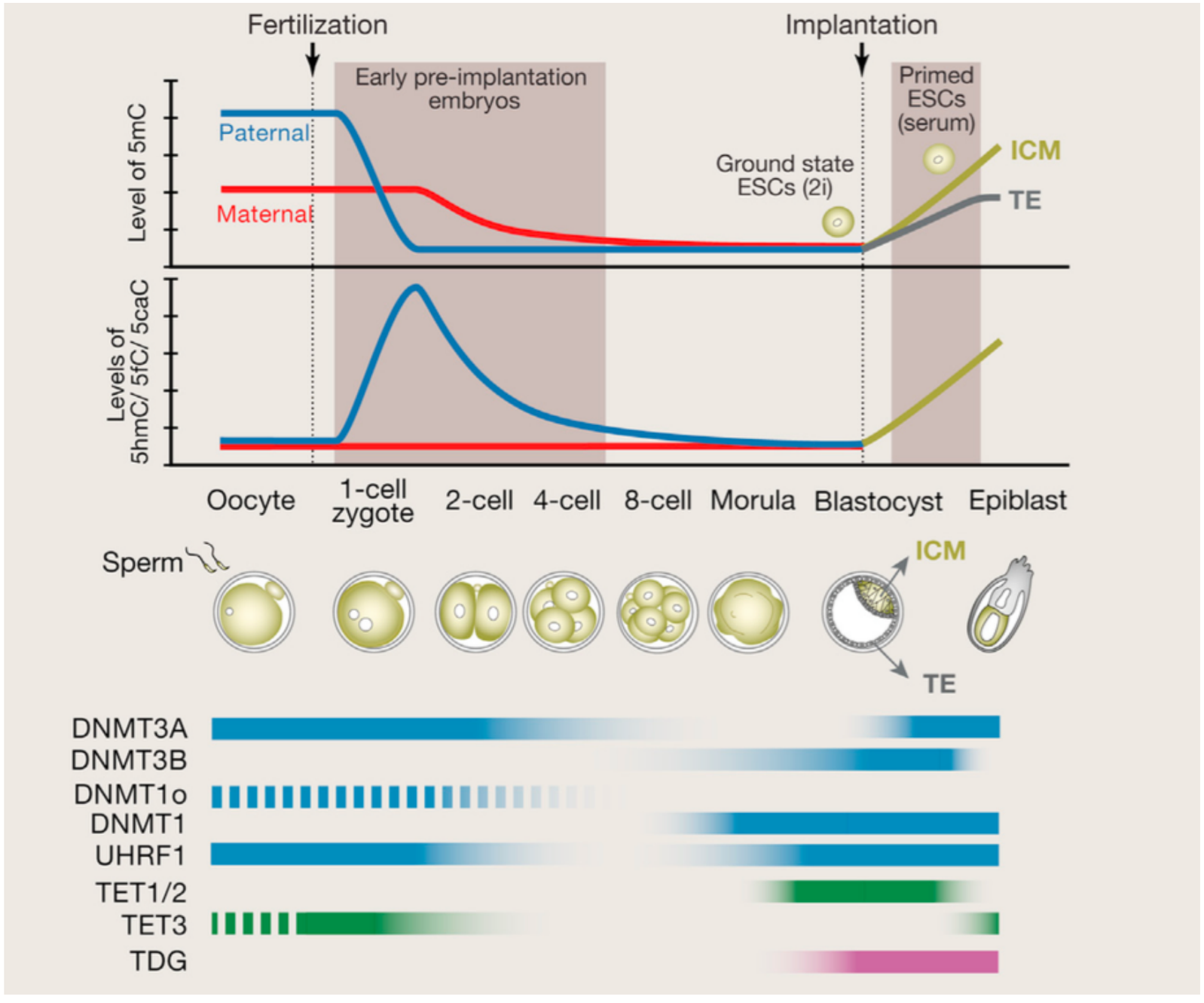
- Paternal 5mC rapidly oxidized to 5hmC/5fC/ 5caC by TET3 proteins.

Early preimplantation embryos

- Oocyte-derived DNMT1o is largely excluded from nucleus (dash line) and consequently maintenance methylation is inefficient.
- Oxidized 5mC bases in the paternal genome and 5mC in the maternal genome are passively diluted.

Wu & Zhang Cell 156 p. 45 (2014)

Changes in cytosine modifications & relevant enzymes during preimplantation development



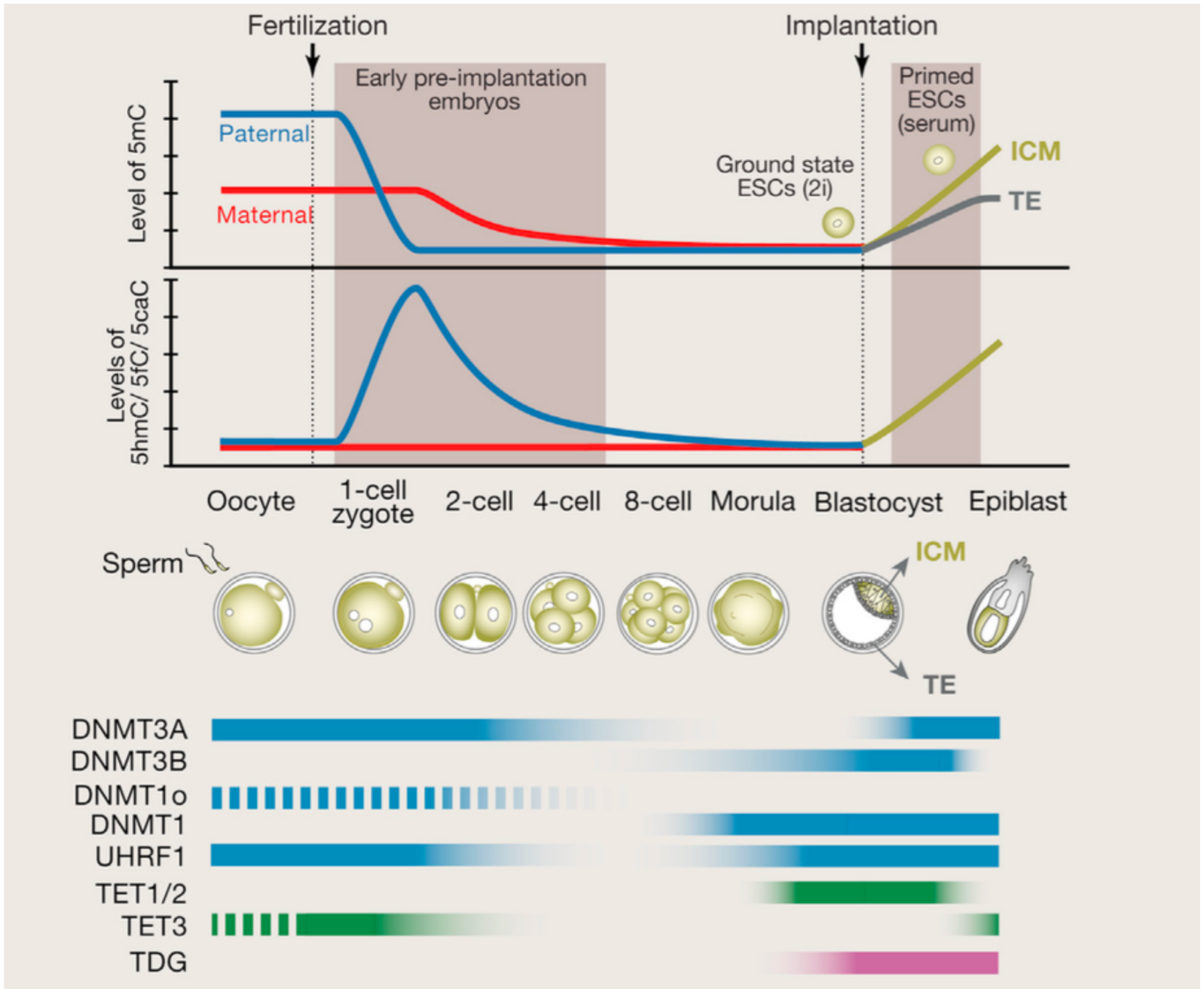
DNA Demethylation

After implantation

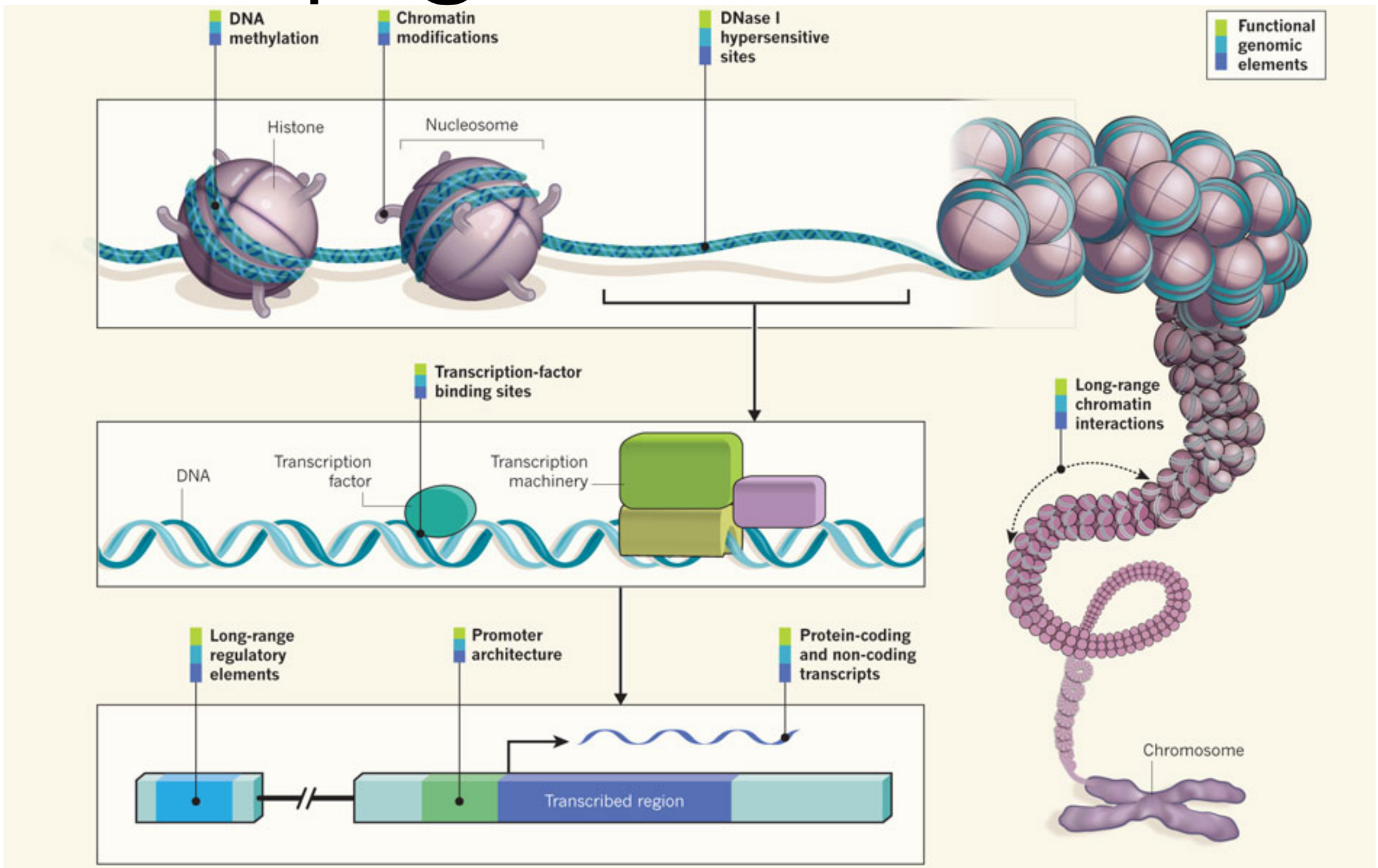
- DNA methylation pattern is reestablished by DNMT3A/ 3B in inner cell mass (ICM) cells, but *not* in trophectoderm (TE) cells.
- Ground-state pluripotent embryonic stem cells (ESCs) genome is hypomethylated (5mC: 1% of all C) & more similar to the methylome of preimplantation ICM cells.
- Primed ESCs (serum) possess a methylome (5mC: 4% of all C) that recapitulates overall methylation pattern in epiblast cells.

Wu & Zhang Cell 156 p. 45 (2014)

Changes in cytosine modifications & relevant enzymes during preimplantation development

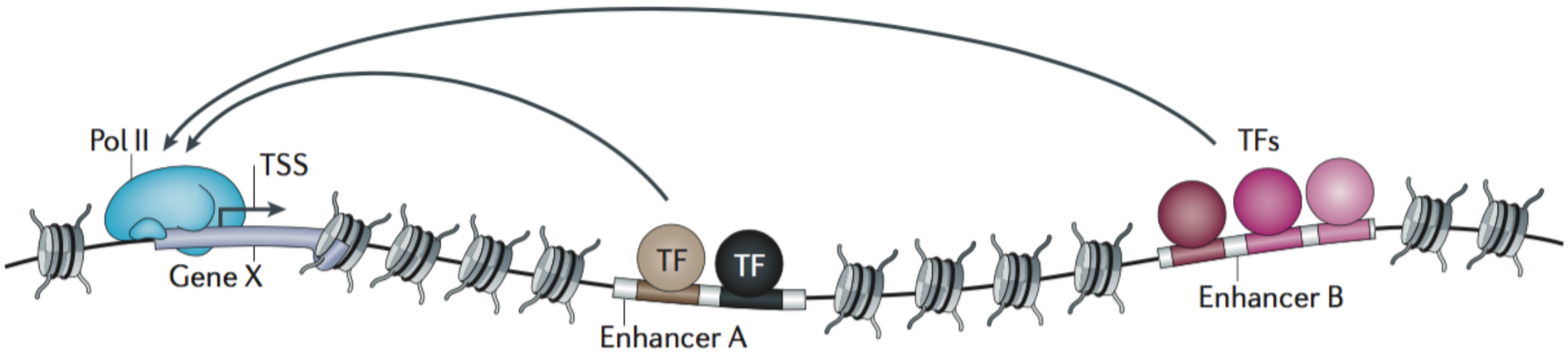


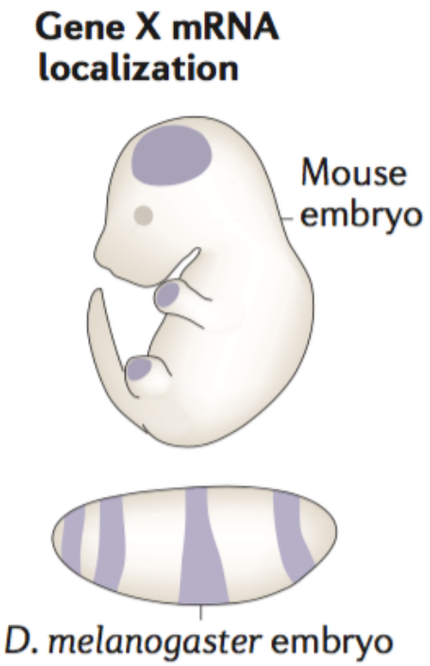
Epigenetics



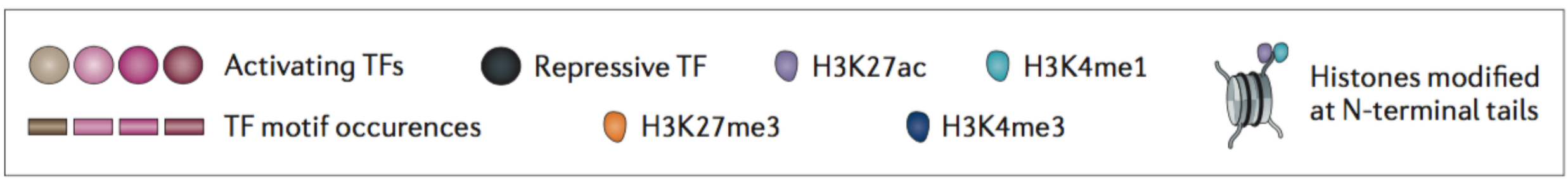
Ecker et al., *Nature* **489**, 52–55 (2012)

Enhancers





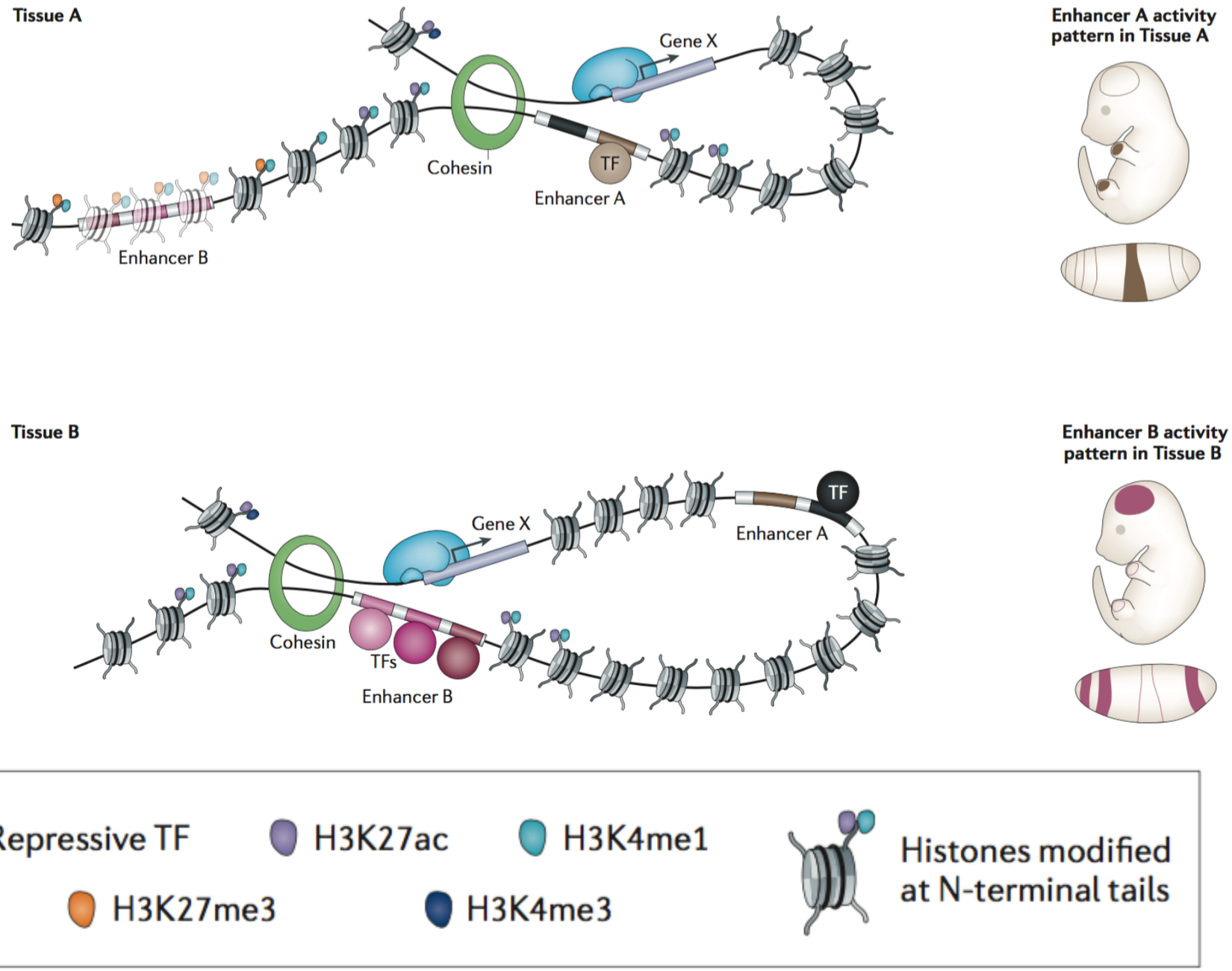
- Enhancers contain binding site sequences for transcription factors (TFs).
- Can upregulate the transcription of a target gene from its transcription start site (TSS).
- Along the linear genomic DNA sequence, enhancers can be located at any distance from their target genes, which makes their identification challenging.



Shlyueva, Stampfel & Stark, Nature Reviews Genetics 15, 272–286 (2014)

Enhancers

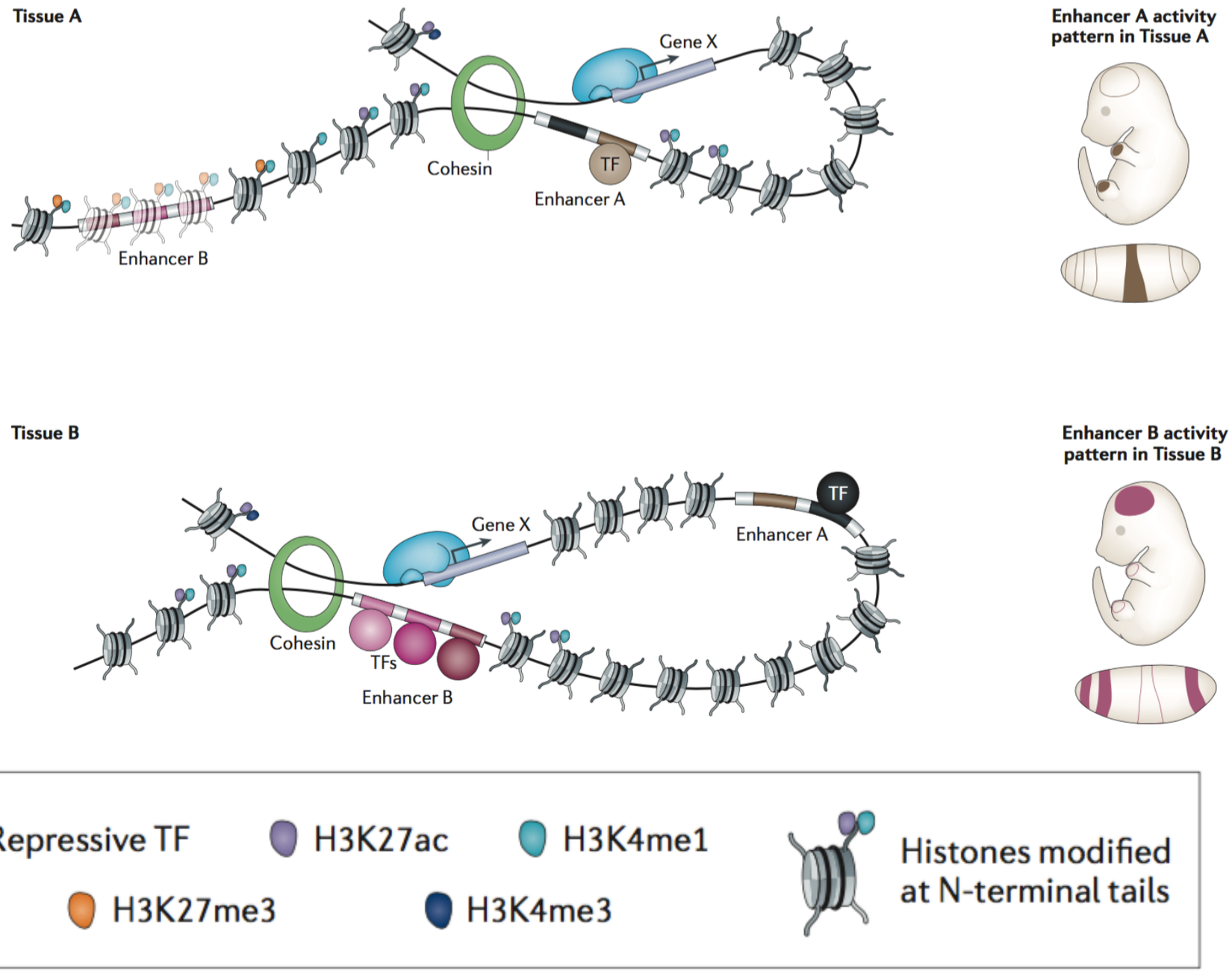
- Enhancers
- Bound by activating TFs
- Brought into proximity of their respective target promoters by looping
- May be mediated by cohesin and other protein complexes.



Shlyueva, Stampfel & Stark, Nature Reviews Genetics 15, 272–286 (2014)

Enhancers

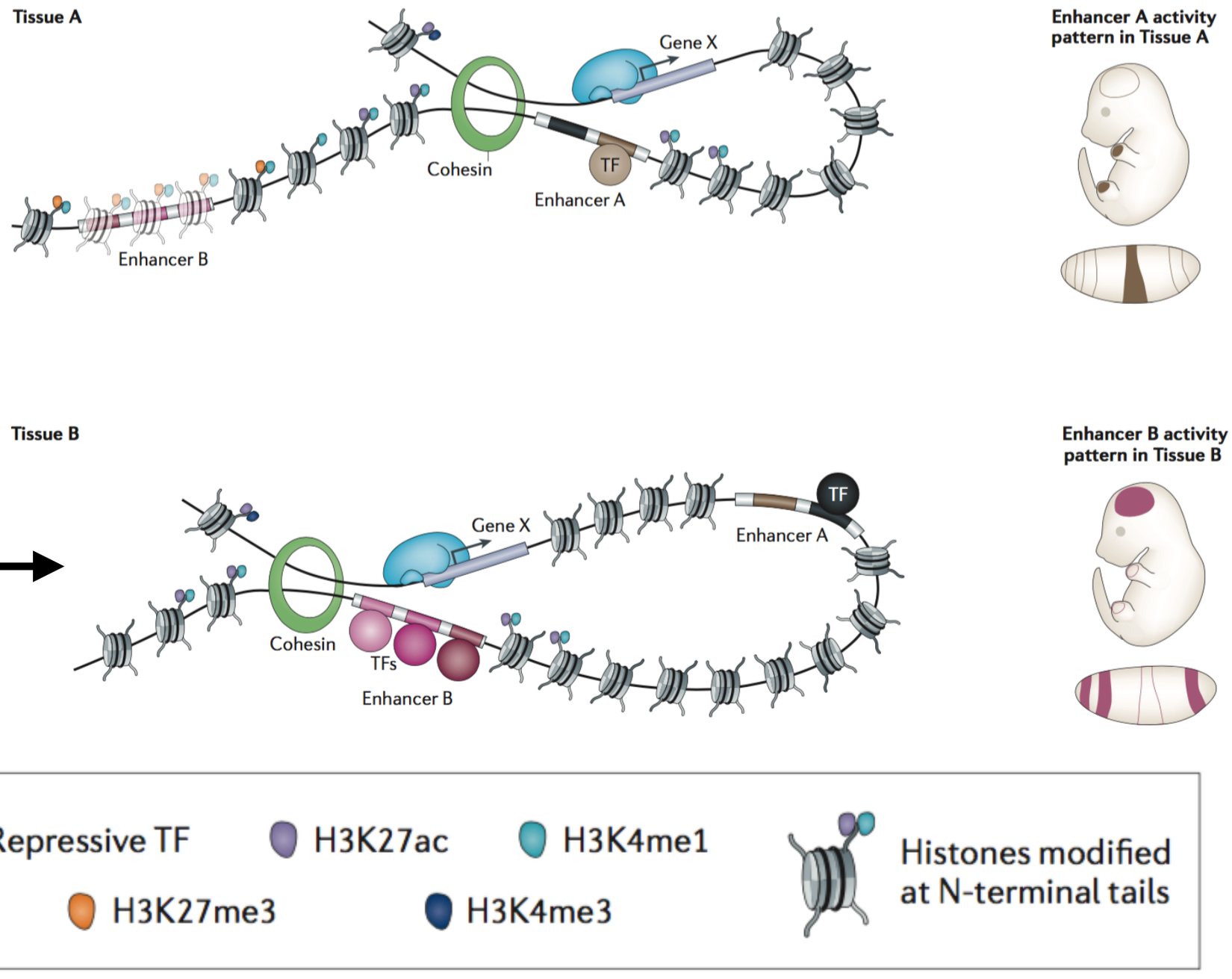
- Active promoters and enhancers
  - Depletion of nucleosomes
  - Nucleosomes that flank active enhancers show specific histone modifications,(eg. H3K4me1) and H3K27 acetylation (H3K27ac).



Shlyueva, Stampfel & Stark, Nature Reviews Genetics 15, 272–286 (2014)

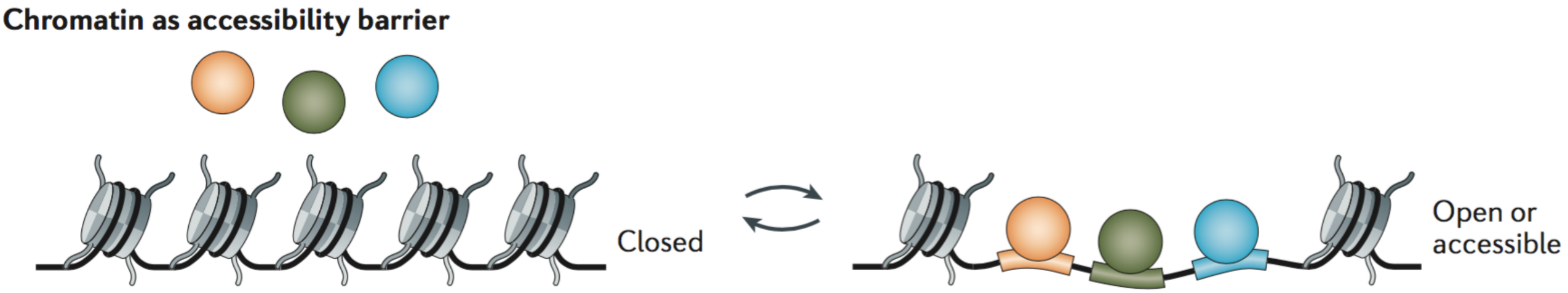
Enhancers

- Inactive enhancers may be silenced by different mechanisms
  - Polycomb protein-associated repressive H3K27me3 mark
  - Repressive TF binding



Shlyueva, Stampfel & Stark, Nature Reviews Genetics 15, 272–286 (2014)

Chromatin accessibility



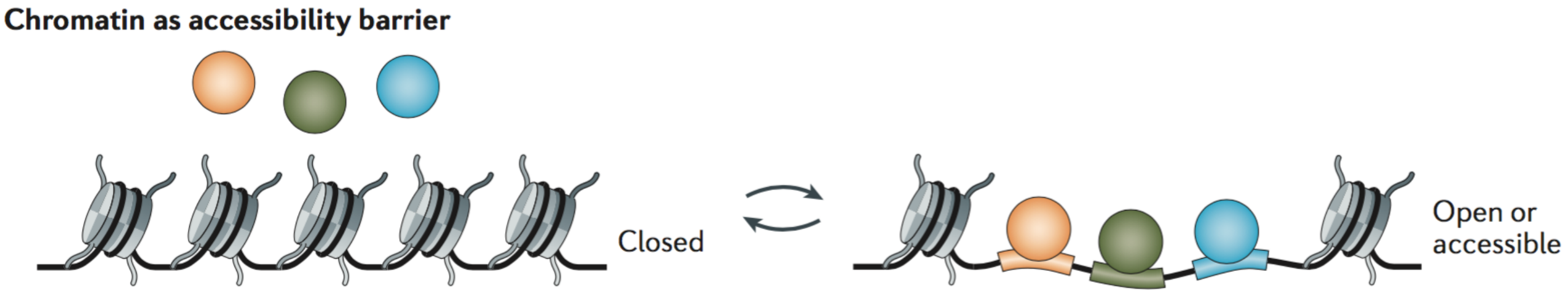
Densely positioned nucleosomes can restrict access for

- transcription factors
- CCCTC-binding factor (CTCF)
- RNA polymerase II (Pol II)
- other proteins.

Nucleosome-free regions can be bound

Shlyueva, Stampfel & Stark, Nature Reviews Genetics 15, 272–286 (2014)

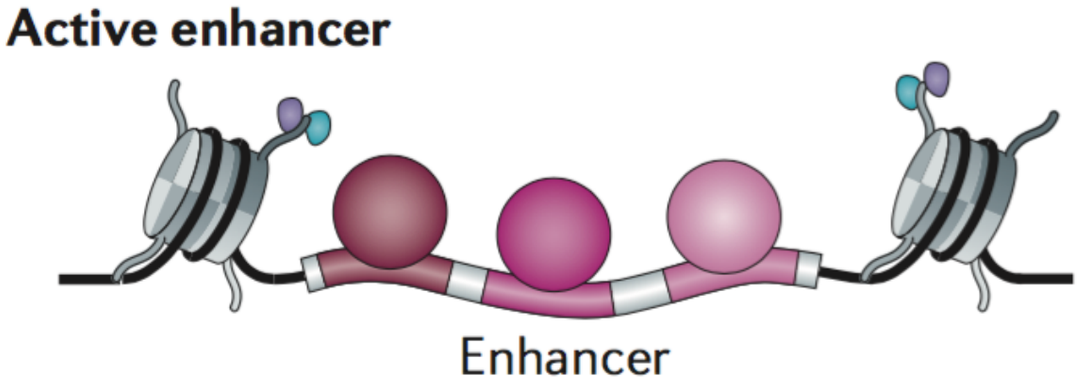
Chromatin accessibility



Transition from ‘open’ to ‘closed’ chromatin, and vice versa determined by regulatory proteins, including pioneer transcription factors. Insulator proteins (for example, CTCF) and other architectural proteins also bind to open regions, and they make up a substantial proportion of sites that are accessible across multiple cell types.

Shlyueva, Stampfel & Stark, Nature Reviews Genetics 15, 272–286 (2014)

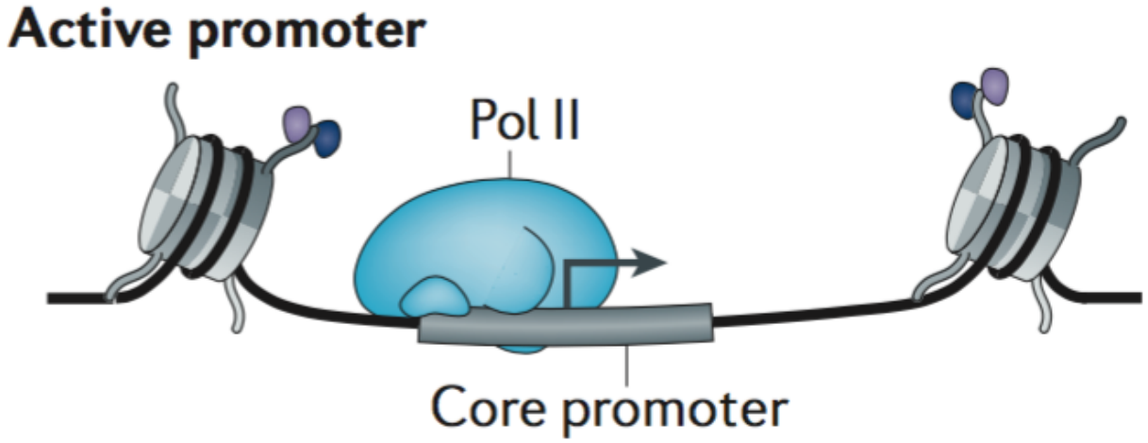
Chromatin accessibility



- Histones flanking active enhancers often marked by
  - H3 acetylated at lysine 27 (H3K27ac)
  - H3 monomethylated at lysine 4 (H3K4me1).

Shlyueva, Stampfel & Stark, Nature Reviews Genetics 15, 272–286 (2014)

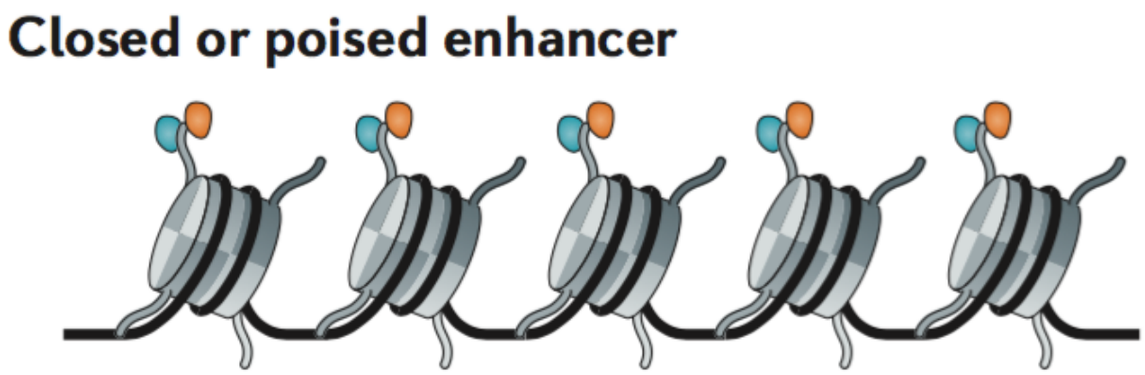
Chromatin accessibility



- Active promoters (depicted as Pol II bound) flanked by nucleosomes with
  - H3K27ac
  - H3K4me3.

Shlyueva, Stampfel & Stark, Nature Reviews Genetics 15, 272–286 (2014)

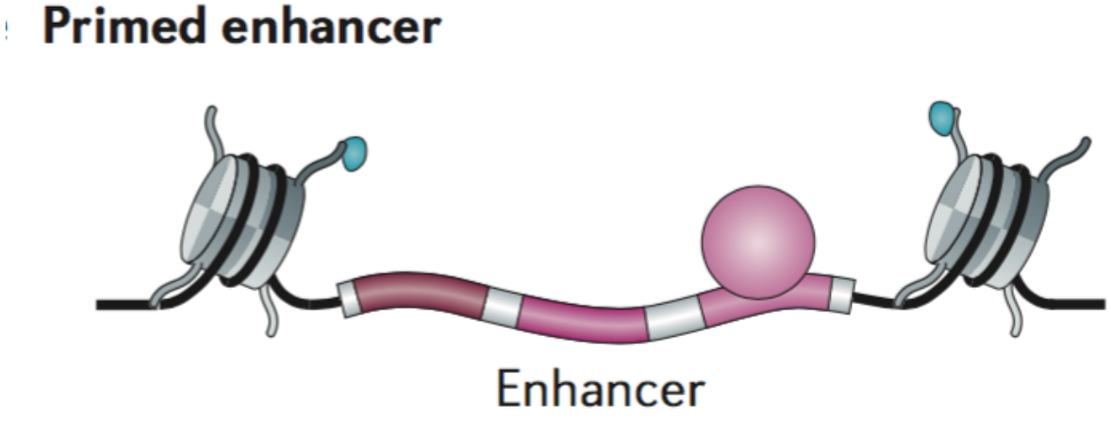
Chromatin accessibility



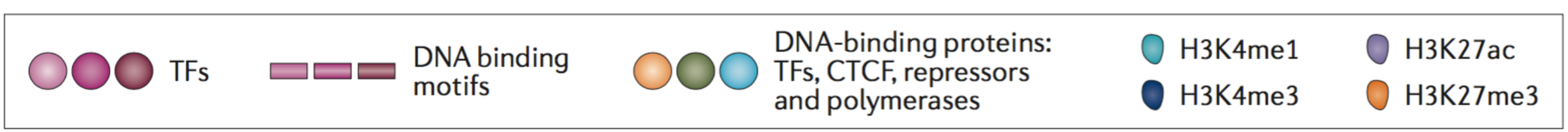
- Active H3K4me1
- Repressive Polycomb protein-associated H3K27me3

Shlyueva, Stampfel & Stark, Nature Reviews Genetics 15, 272–286 (2014)

Chromatin accessibility

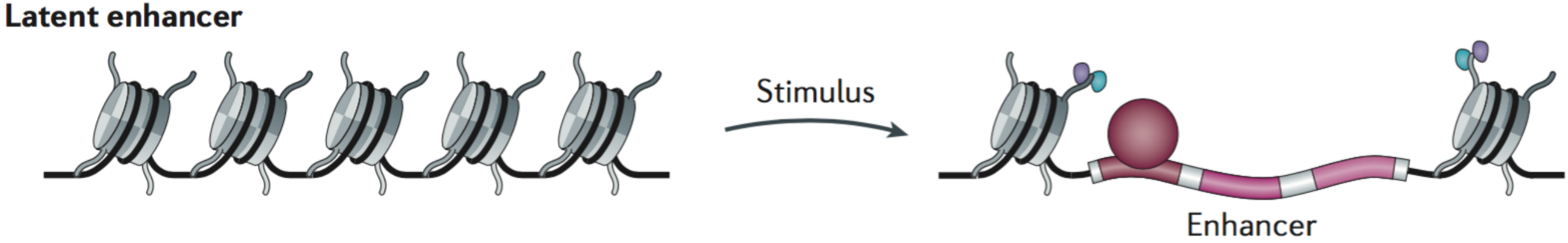


- Enhancers that are not yet active but that are primed for activation
  - at a later developmental time point
  - in response to external stimuli
- Can be pre-marked by H3K4me1.



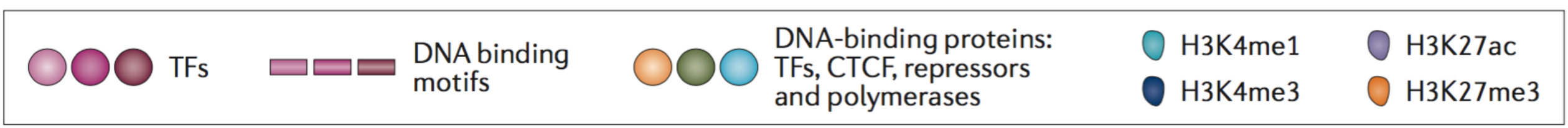
Shlyueva, Stampfel & Stark, Nature Reviews Genetics 15, 272–286 (2014)

Chromatin accessibility



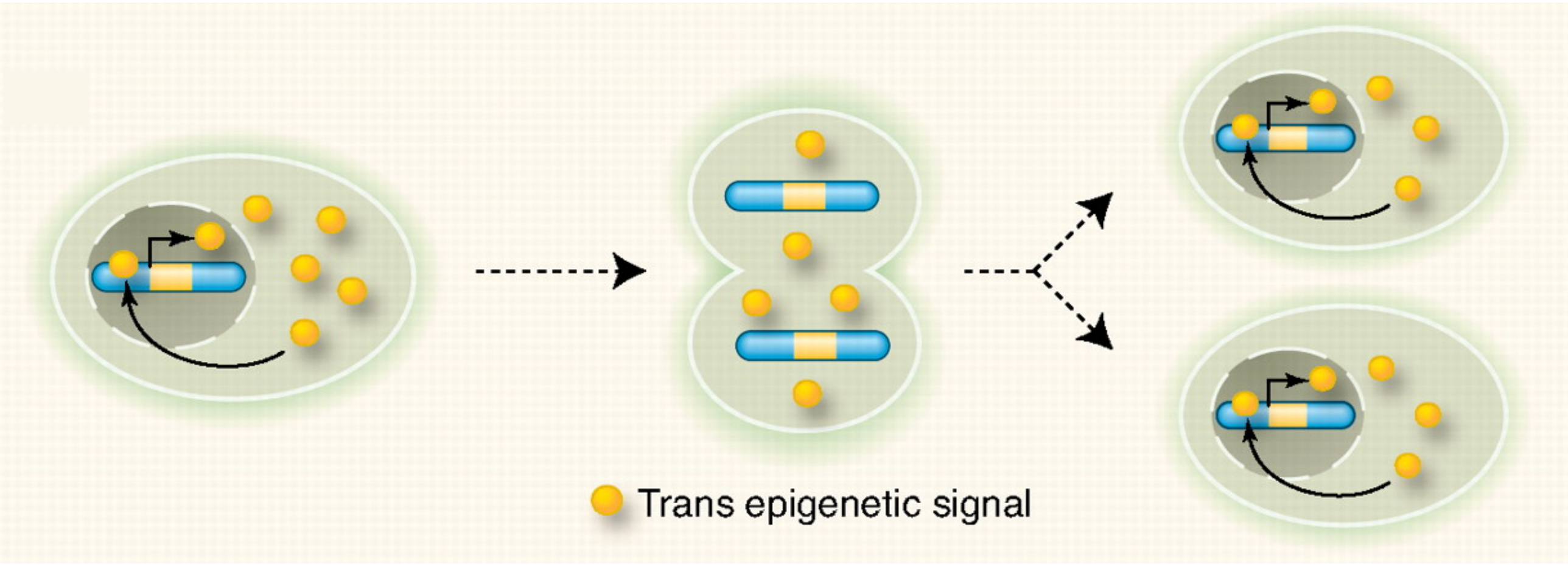
- Latent enhancers
  - located in closed chromatin
  - not pre-marked by known histone modifications.

- DNA becomes accessible
  - flanking nucleosomes acquire marks
    - H3K4me1
    - H3K27ac



Shlyueva, Stampfel & Stark, Nature Reviews Genetics 15, 272–286 (2014)

Epigenetic Signals

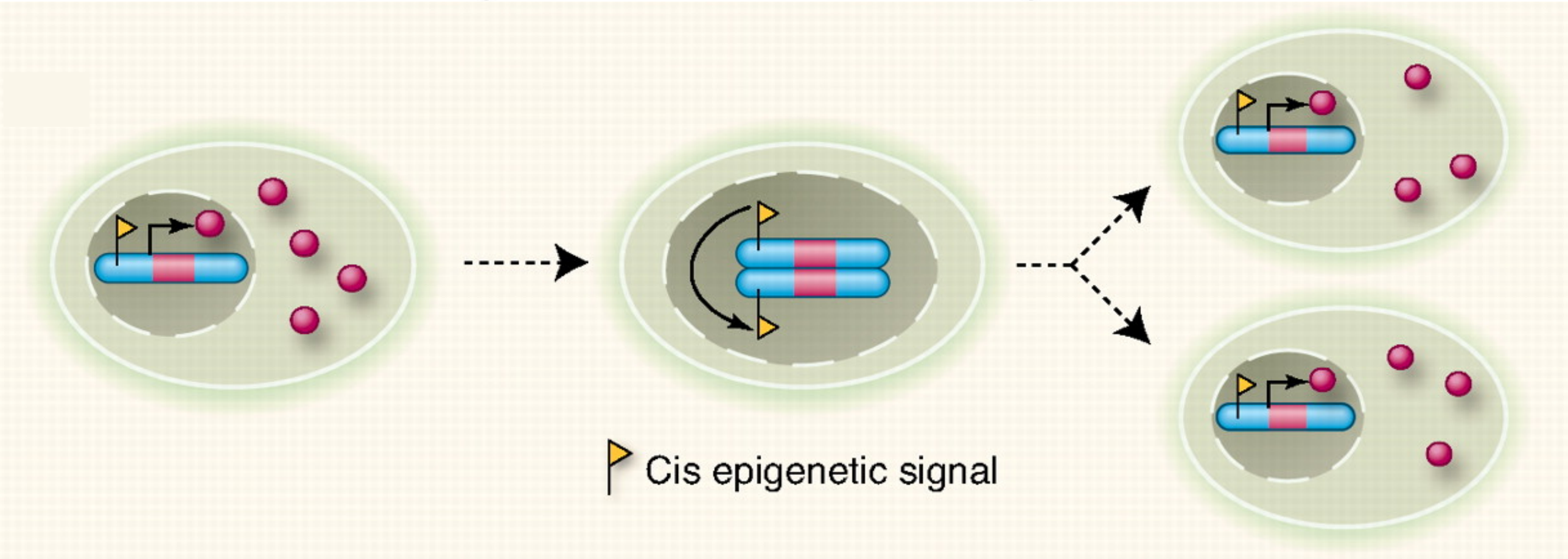


E.g. Regulatory loop, epigenetic signal induces own expression

- Non-sequence-based regulatory signals be inherited across mitosis or/and meiosis. *Trans* signals are unlinked to the DNA.
- Transmitted by partitioning of the cytosol during cell division and maintained by feedback loops.

Bonasio et al., Science 330:612-616 (2010)

Epigenetic Signals



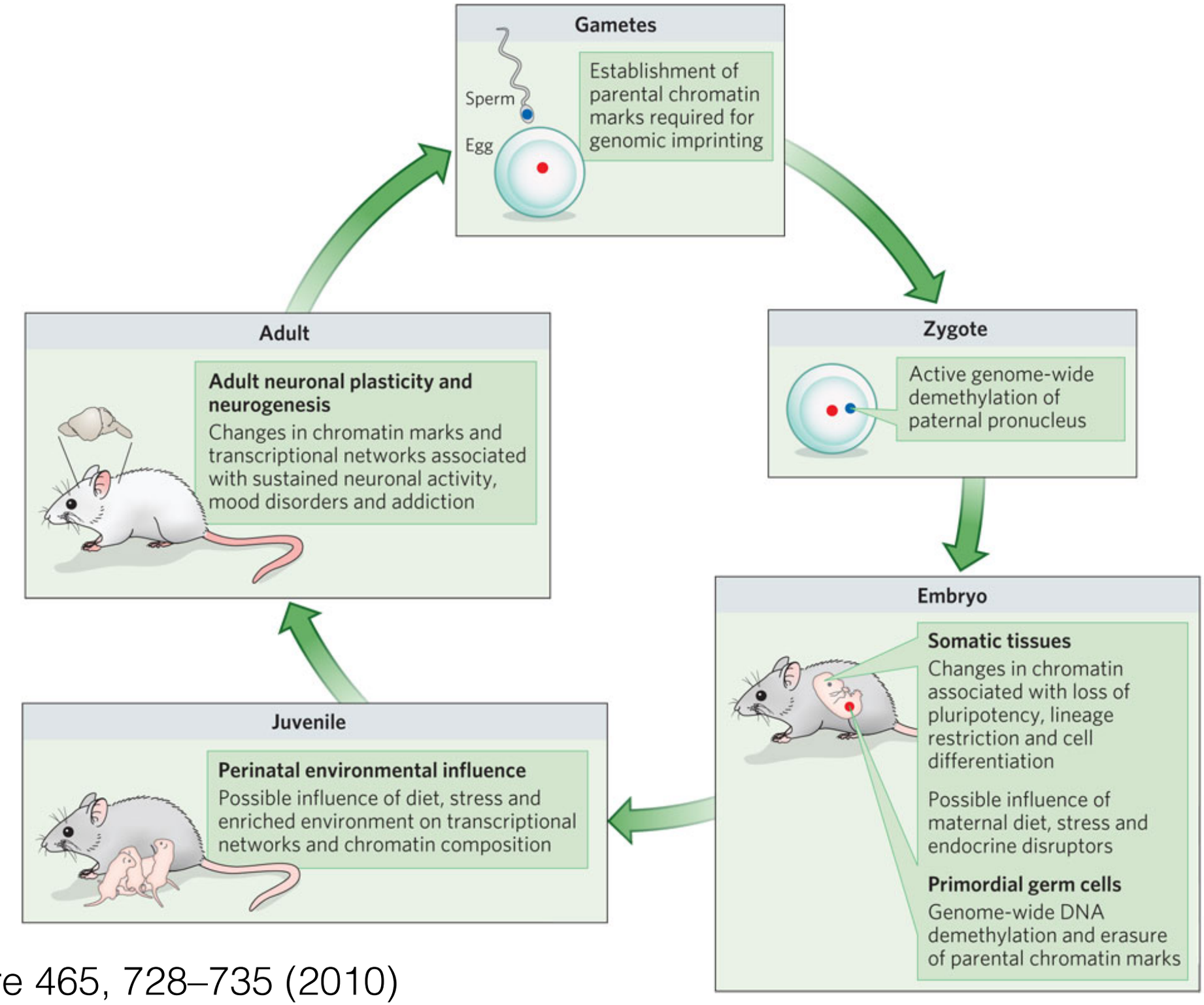
- Cis signals (yellow flags) are molecular signatures physically associated with the DNA
- Inherited via chromosome segregation during cell division.

Bonasio et al., Science 330:612-616 (2010)

DNA Methylation

- Genomic Imprinting (inherited methylation)
- X inactivation (silencing of 1 of 2 X chromosomes in females)
- Gene expression in different tissue
- Methylation changes (cancer cells) Transposition protection
- Silencing

Effects of Chromatin Modifications



Dulac, Nature 465, 728–735 (2010)

Imprinting

Definition

Process that results in diploid cells expressing a small subset of genes from only their maternal- or paternal-inherited chromosome

E.g. in mourse: insulin-like growth factor type 2 receptor (*Igf2r* , cation-independent mannose-6-phosphate receptor, a scavenger receptor for the *Igf2* growth hormone) identified as a maternally expressed imprinted gene using the maternal-effect mutant that mapped to mouse chromosome 17.

In somatic tissues the result is monoallelic expression.

Barlow, Annu. Rev. Genet. 45:379–403 (2011)
Barlow et al., Nature 349:84–87 (1991)

Imprinting

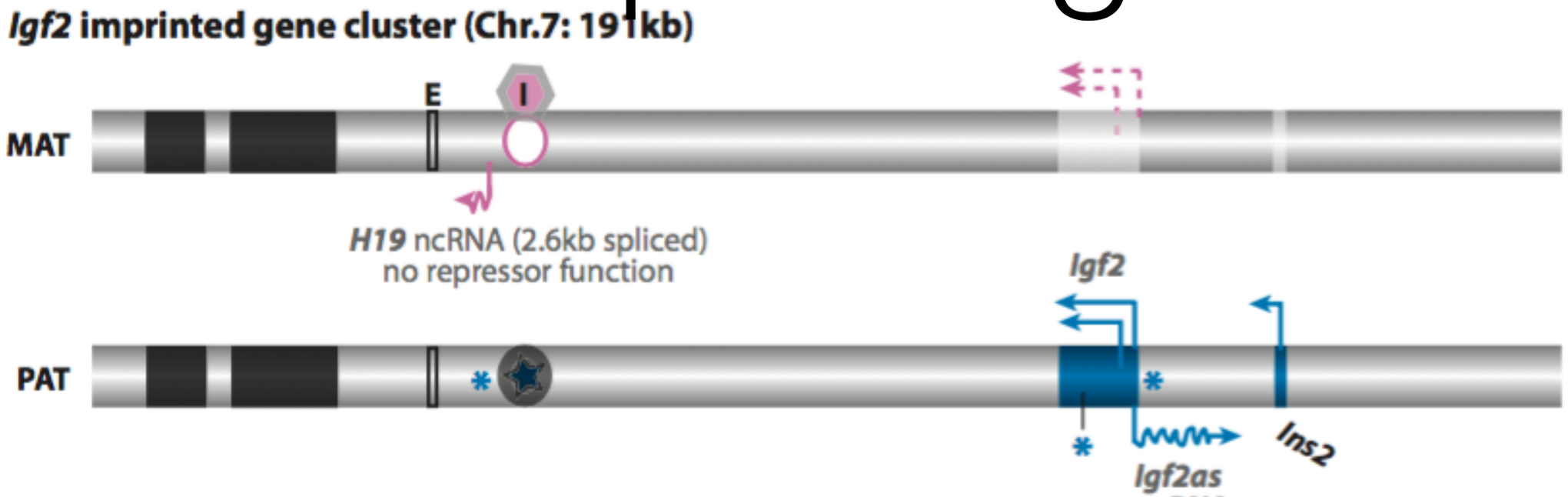
Clusters with imprint control elements (ICE) identified

Imprinted gene cluster[a]	Cluster size (kb)	Gene Nr.	Parental expression M: Mat, P: Pat	Imprint control element (ICE)	Reference	ICE size (kb)	Parental effect of ICE deletion
Igf2r chr17:12,875,272	490	4	**3M**(pc)/1P(nc)	Region2	(153)	3.9	Pat
Kcnq1 chr7:150,293,159	780	12	**11M**(pc)/1P(nc)	KvDMR1	(39)	1.1	Pat
Pws/As chr7:67,127,388	3700	>8[b]	**2M**(pc)/>7P(nc+pc)	*Snrpn*-CGI	(17)	4.8	Pat
Gnas chr2:174,153,609	80	>4[b]	**2M**(pc)/3P(nc)	*Nespas*-DMR	(150)	1.6	Pat
Igf2 chr7:149,836,673	80	3	**1M**(nc)/2P(pc)	*H19*-DMD	(139)	1.6	**Mat**
Dlk1 chr12:110,691,433	830	>5[b]	>**1M**(nc)/4P(pc)	IG-DMR	(75)	4.2	**Mat**
Grb10 chr11:11,830,502	780	4	**2M**(pc)/2P(pc)	*Meg1*/Grb10 DMR	(129)	1.0	Pat

ICE, short DNA sequence whose epigenetic state controls imprinted expression of all genes in one imprinted cluster, also known as ICR, IC

Barlow, Annu. Rev. Genet. 45:379–403 (2011)

Imprinting

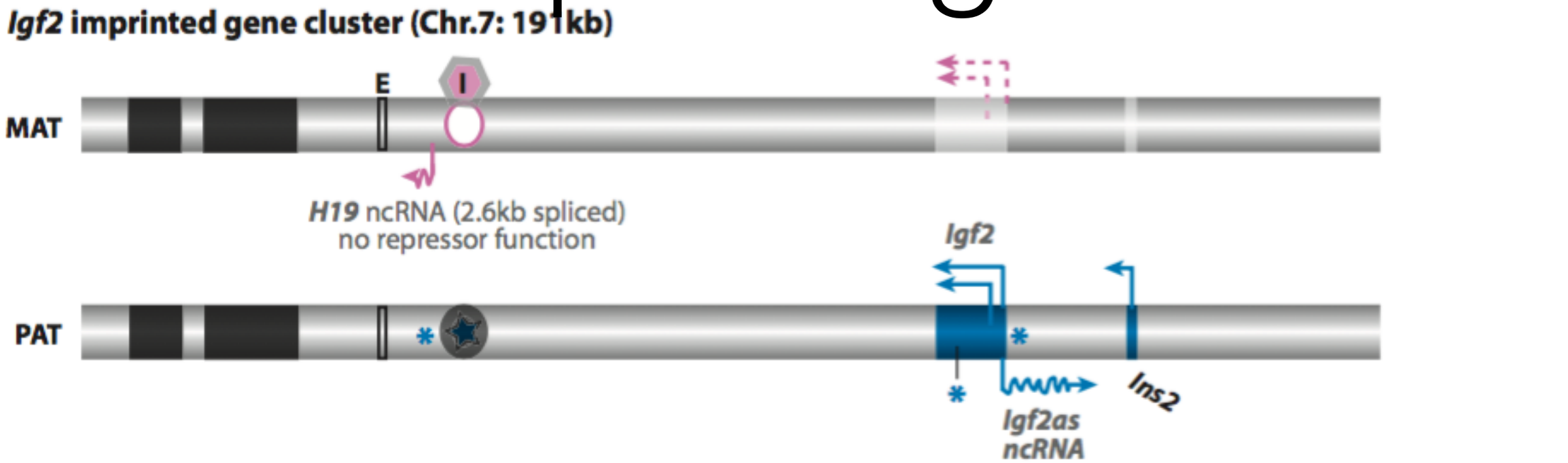


The unmethylated imprint control element (ICE) active (*oval with white center*) on one parental chromosome and required for ncRNA expression and silencing of flanking protein-coding genes

A gametic DNA (gDMR) methylation imprint represses the ICE on the opposite parental chromosome (*gray oval with star*).

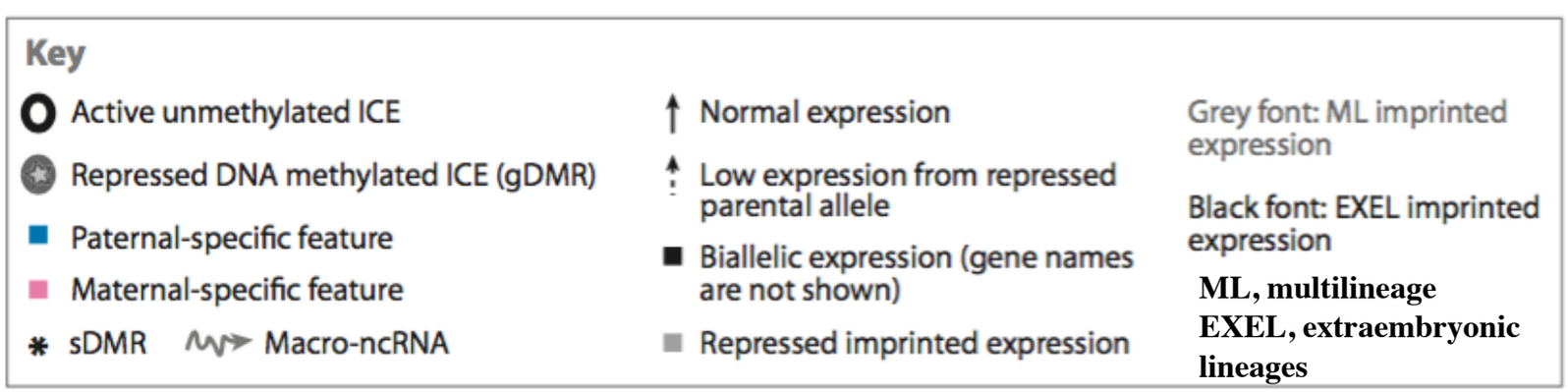
Barlow, Annu. Rev. Genet. 45:379–403 (2011)

Imprinting



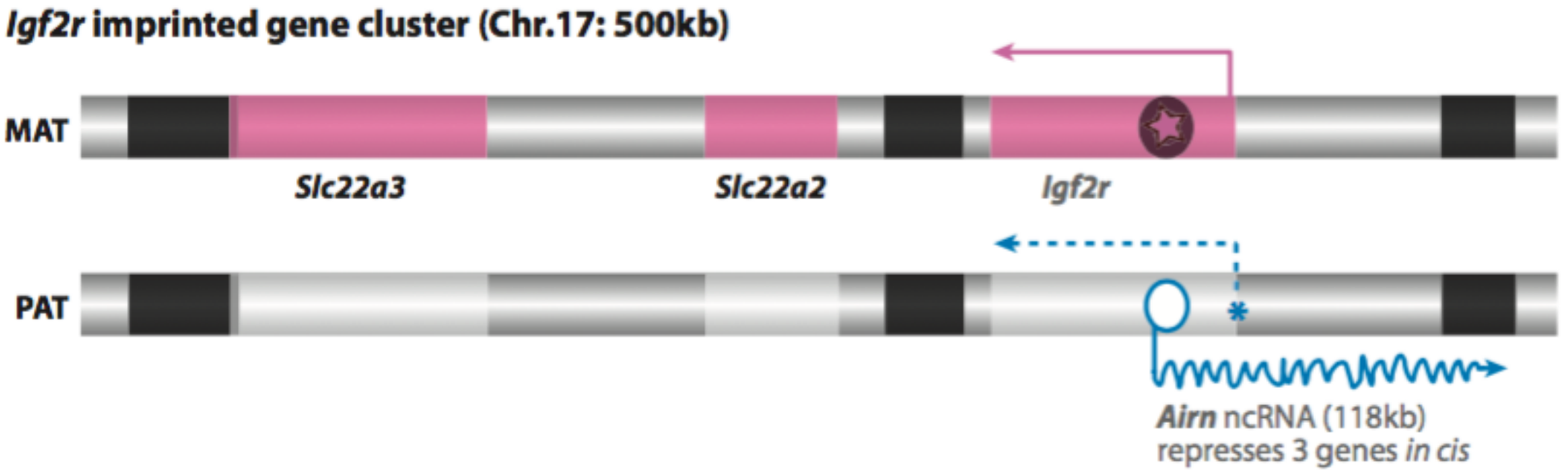
H19 ncRNA controlled by the ICE but not required for silencing

Unmethylated ICE binds insulator proteins (I) that block access of the flanking protein-coding genes to an essential enhancer (E).

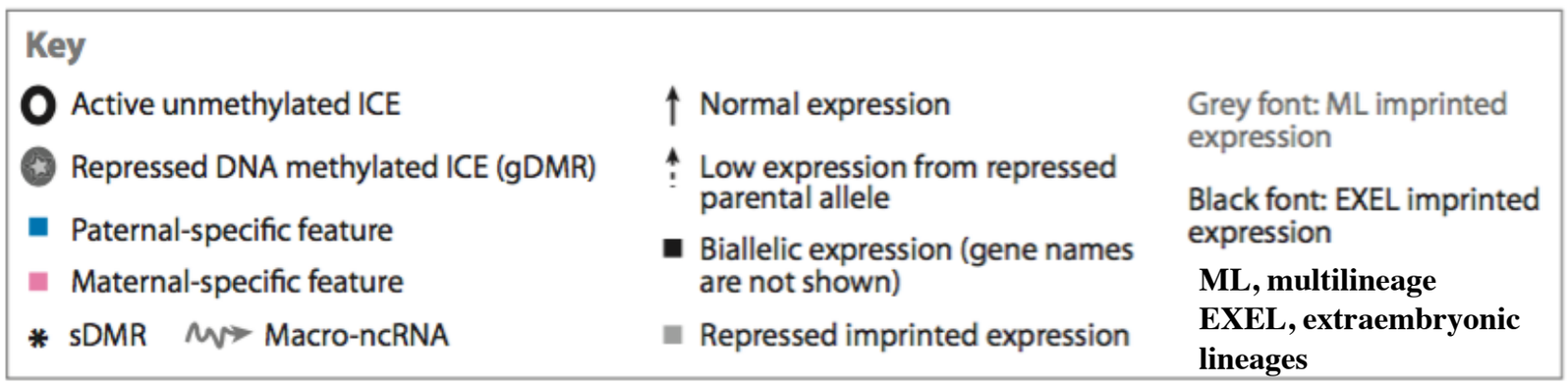


Barlow, Annu. Rev. Genet. 45:379–403 (2011)

Imprinting

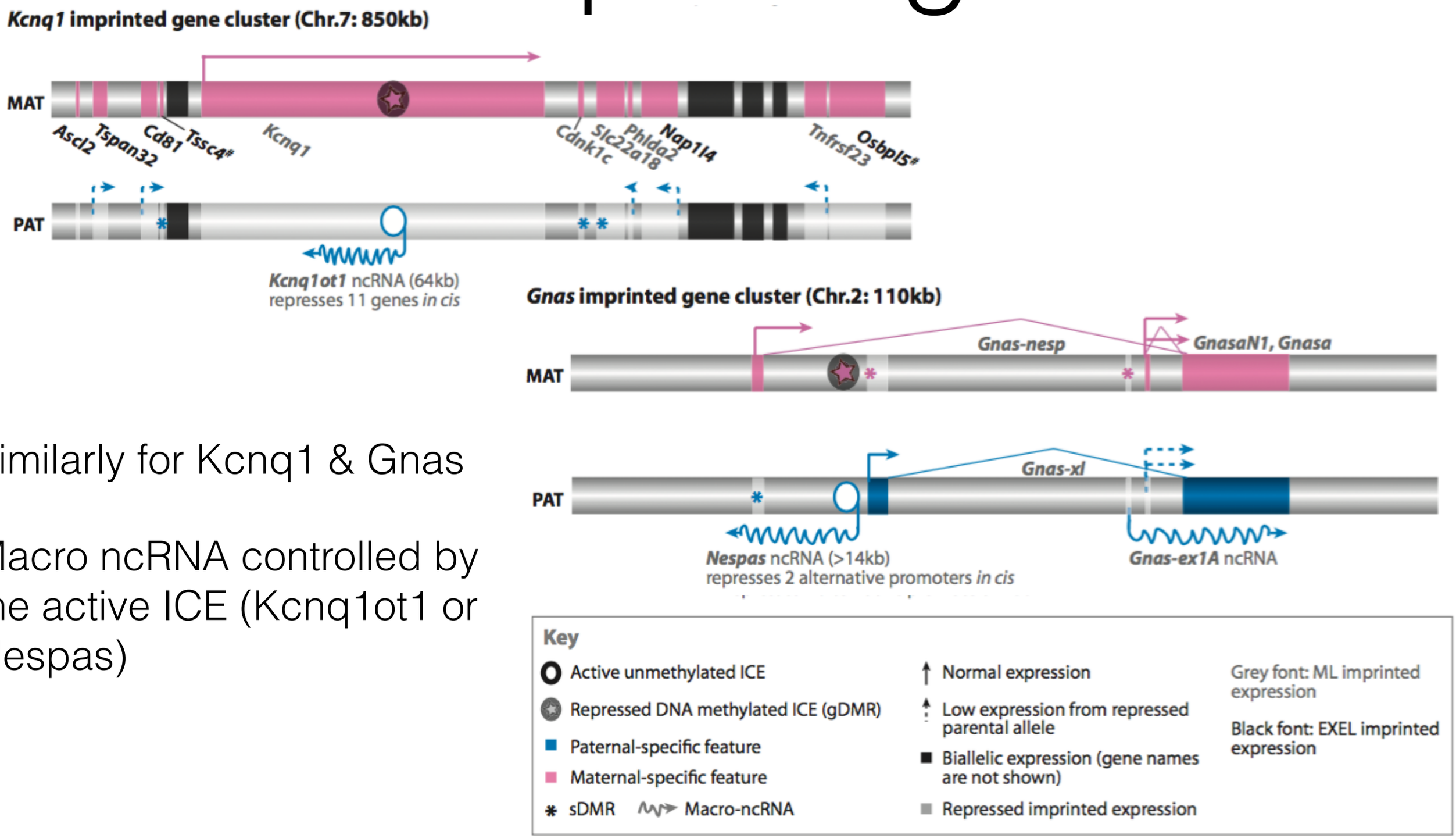


Macro ncRNA controlled by the active ICE (Airn) is responsible for silencing multiple flanking protein-coding genes.



Barlow, Annu. Rev. Genet. 45:379–403 (2011)

Imprinting



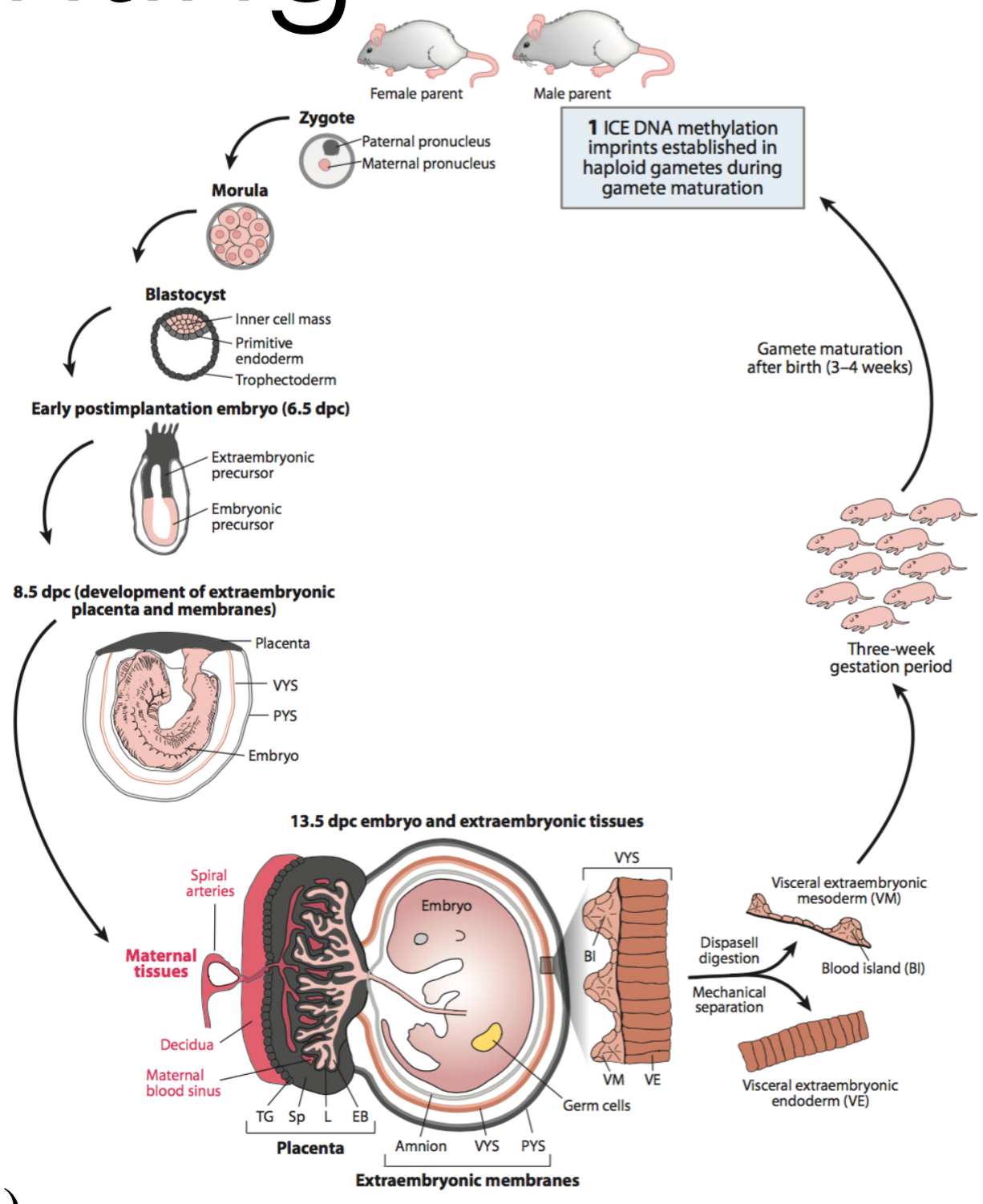
Similarly for Kcnq1 & Gnas

Macro ncRNA controlled by the active ICE (Kcnq1ot1 or Nespas)

Barlow, Annu. Rev. Genet. 45:379–403 (2011)

Imprinting

(*1*) Maternal and paternal parent produce gametes thatcarry imprint control element (ICE) DNA methylation imprints.

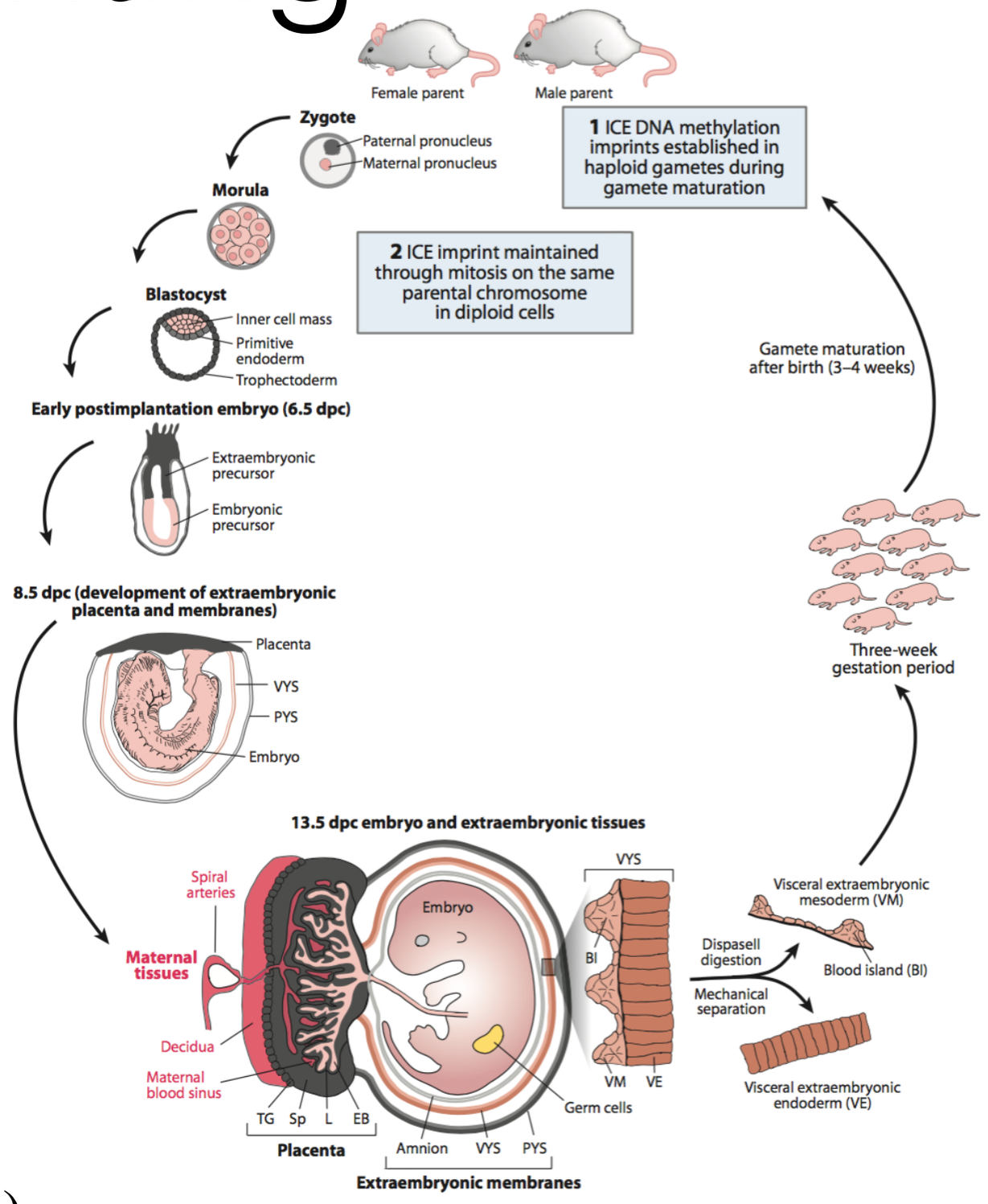


Barlow, Annu. Rev. Genet. 45:379–403 (2011)

Imprinting

2. Imprint maintained on the same parental chromosome during mitosis in diploid cells of the developing embryo.

The unmethylated ICE is active, but its activity requires additional developmentally regulated and tissue-specific factors.



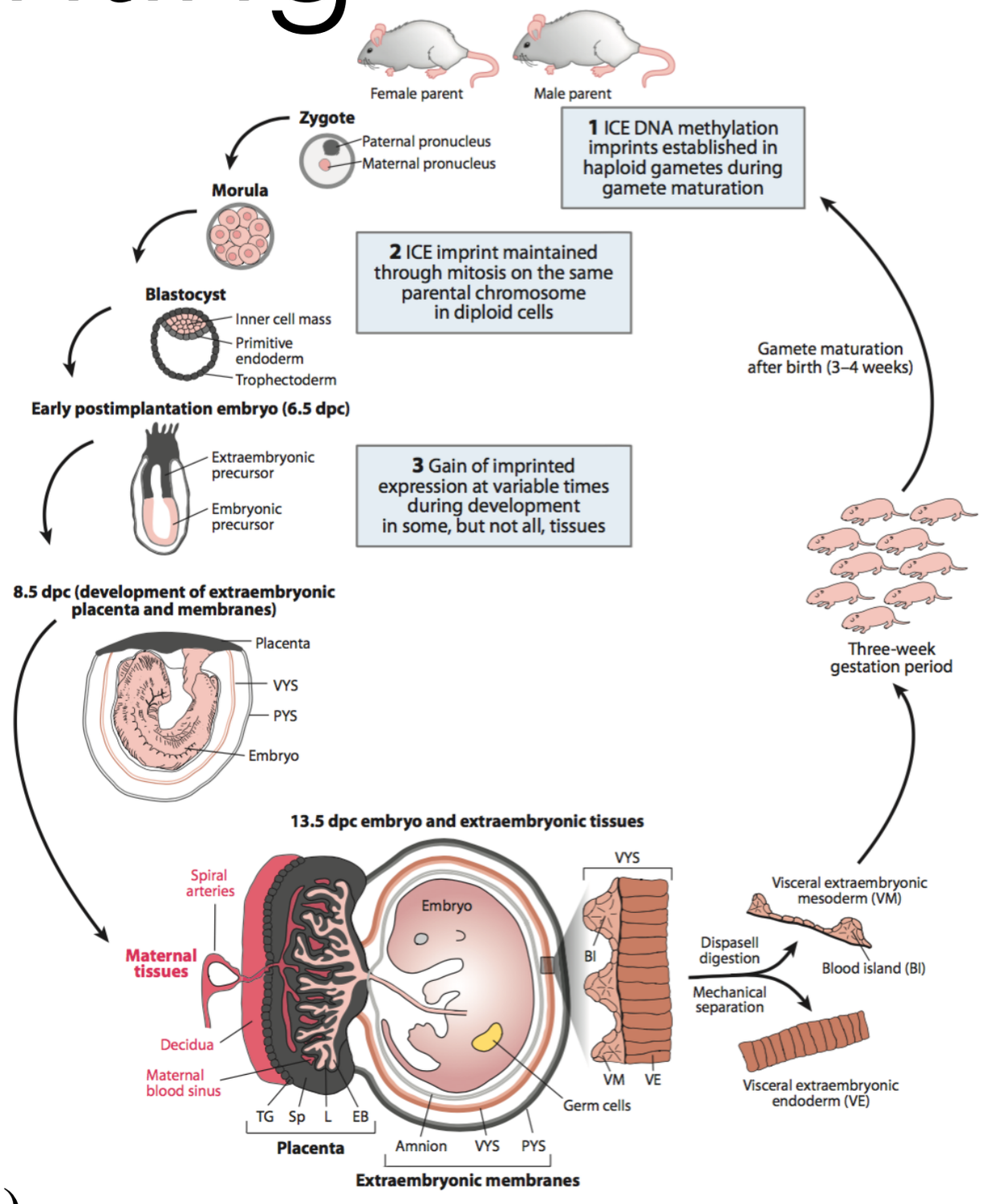
Barlow, Annu. Rev. Genet. 45:379–403 (2011)

Imprinting

(*3*) Imprinted expression is gained during different stages of development and only in some tissues.

Some genes show multilineage (ML) imprinted expression (i.e., imprinted expression in embryonic, extraembryonic, & adult cell lineages).

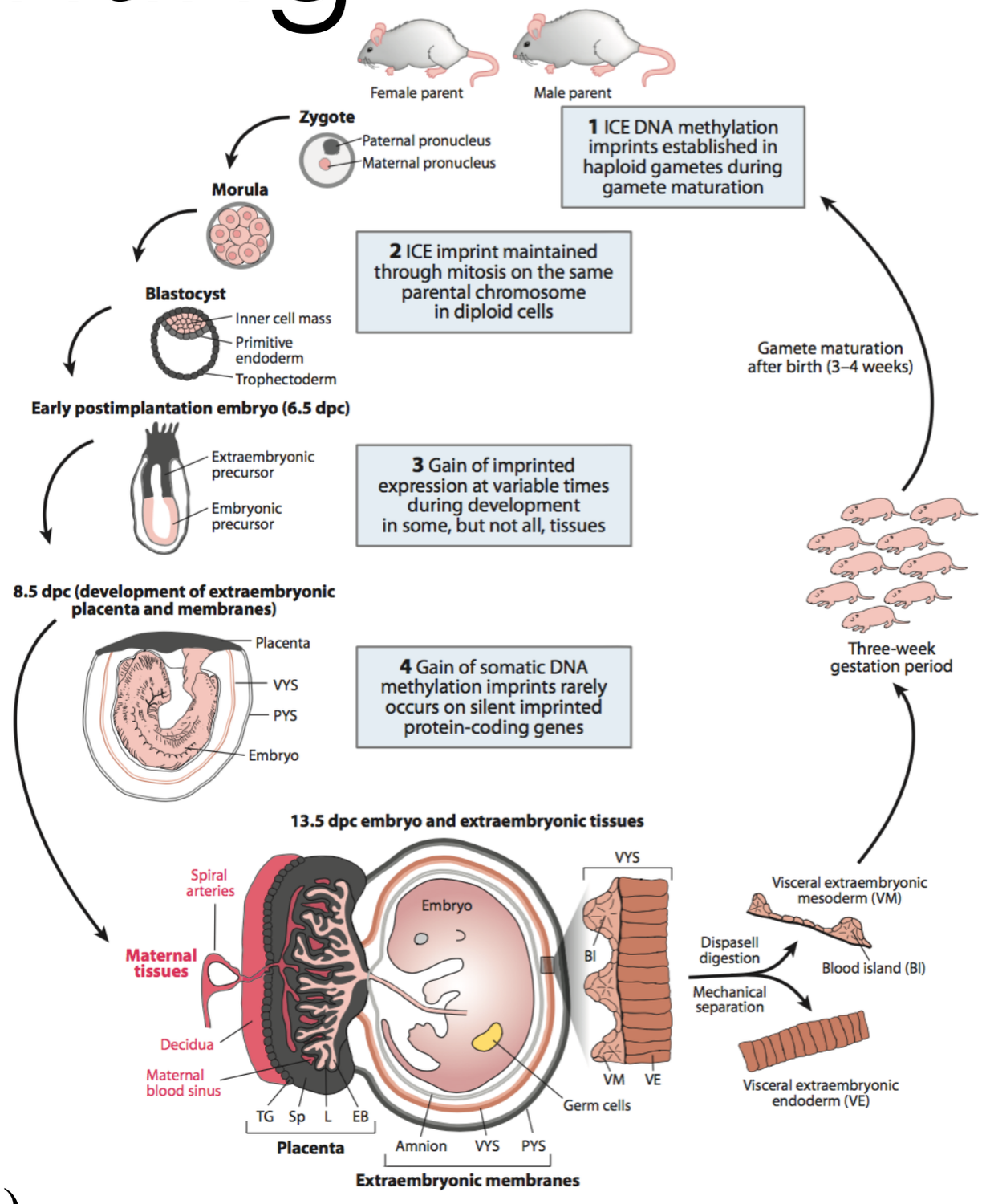
Remainder show imprinted expression in the extraembryonic lineages (EXEL) (i.e., the placenta and membranes), but are expressed from both parental chromosomes in embryonic and adult tissues



Barlow, Annu. Rev. Genet. 45:379–403 (2011)

Imprinting

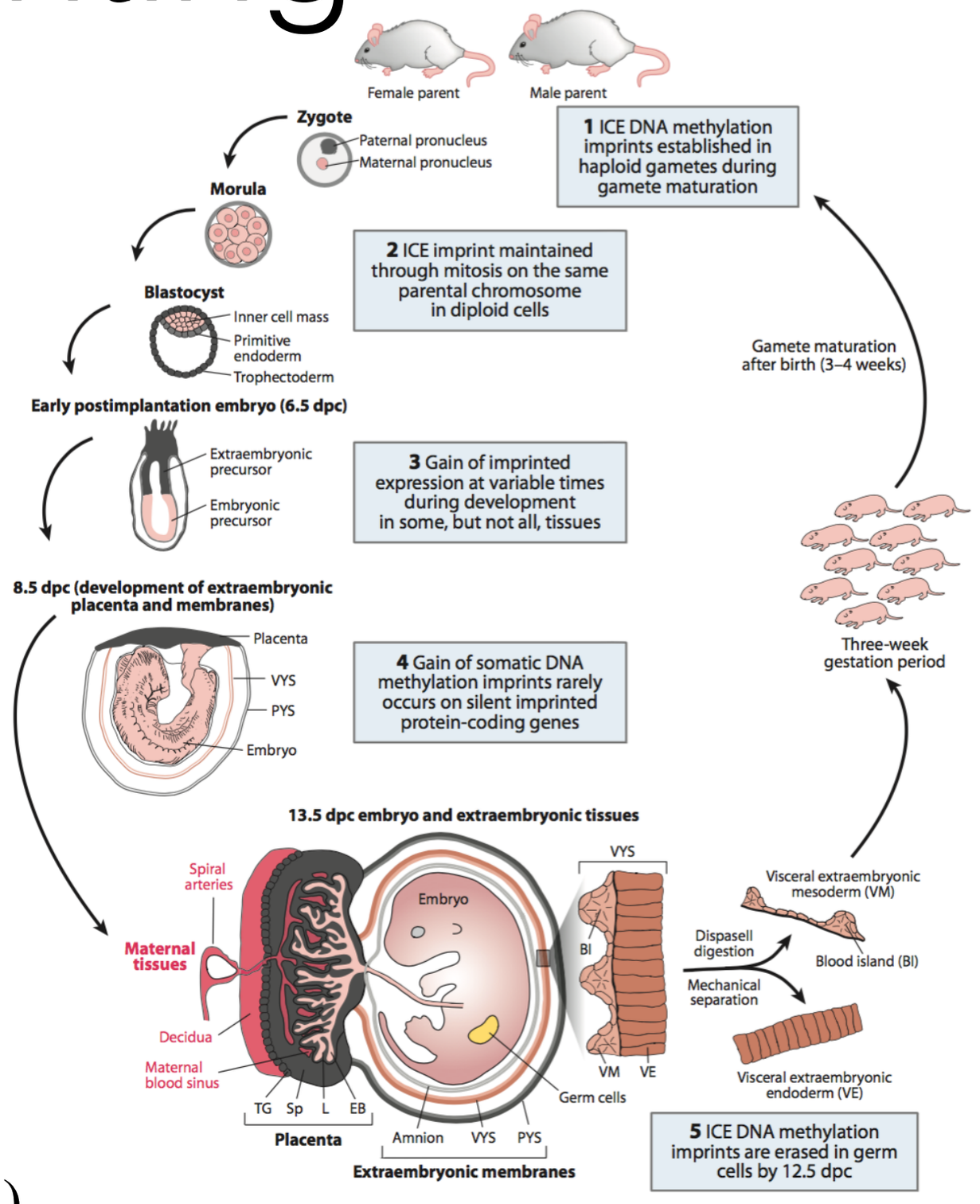
(*4*) Silent imprinted protein-coding genes only rarely gain DNA methylation, and this occurs late in development.



Barlow, Annu. Rev. Genet. 45:379–403 (2011)

Imprinting

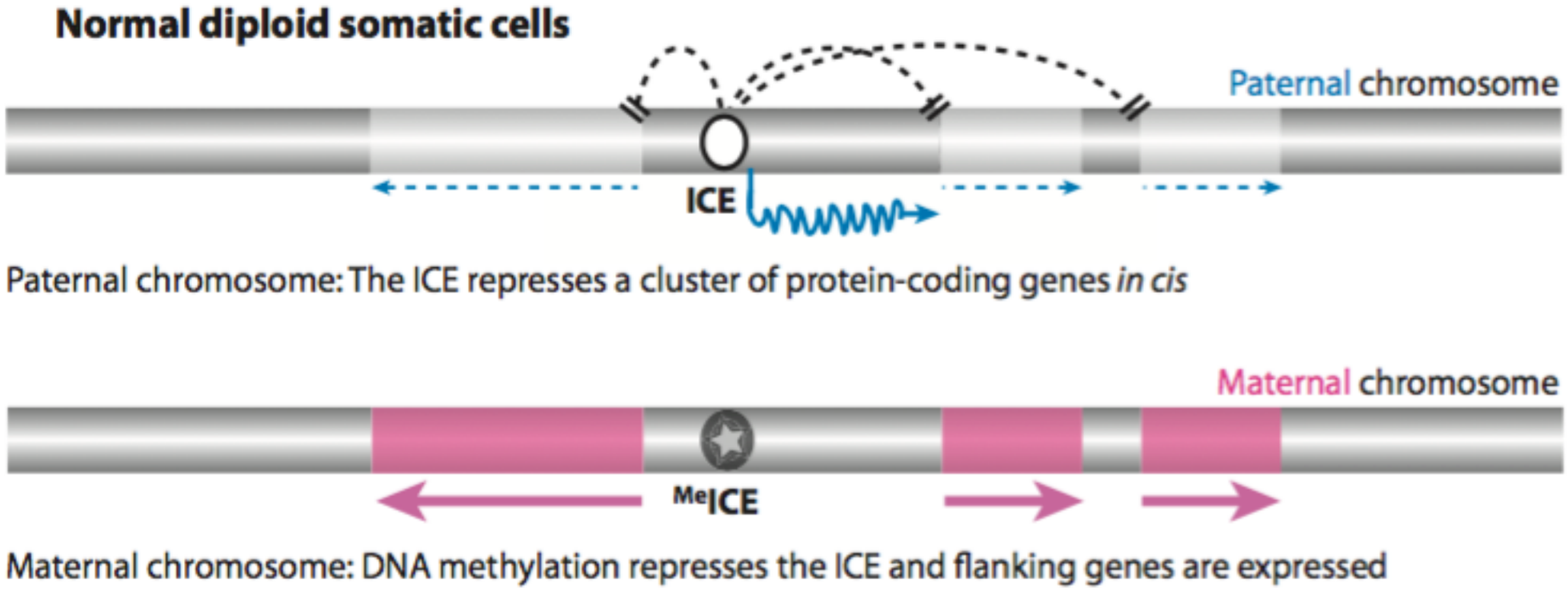
(5) Embryonic germ cells are indicated where the ICE DNA methylation imprint is erased by 12.5 dpc.



Barlow, Annu. Rev. Genet. 45:379–403 (2011)

Imprinting

A double-negative mechanism. ICE is methyl-sensitive, *cis*-acting repressor & macro ncRNA activator

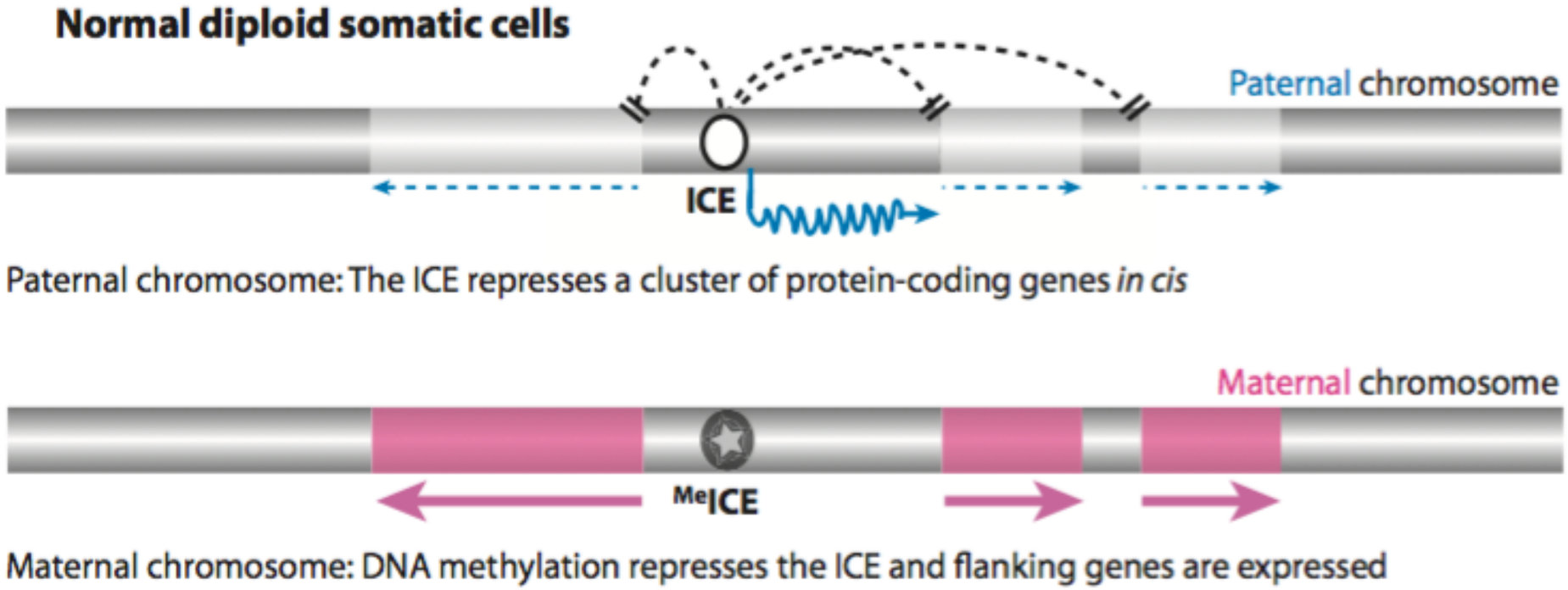


- Paternal chromosome
  - Unmethylated ICE necessary for expression of a macro ncRNA
  - Silencing a cluster of flanking protein-coding genes *in cis*

Barlow, Annu. Rev. Genet. 45:379–403 (2011)

Imprinting

A double-negative mechanism. ICE is methyl-sensitive, *cis*-acting repressor & macro ncRNA activator

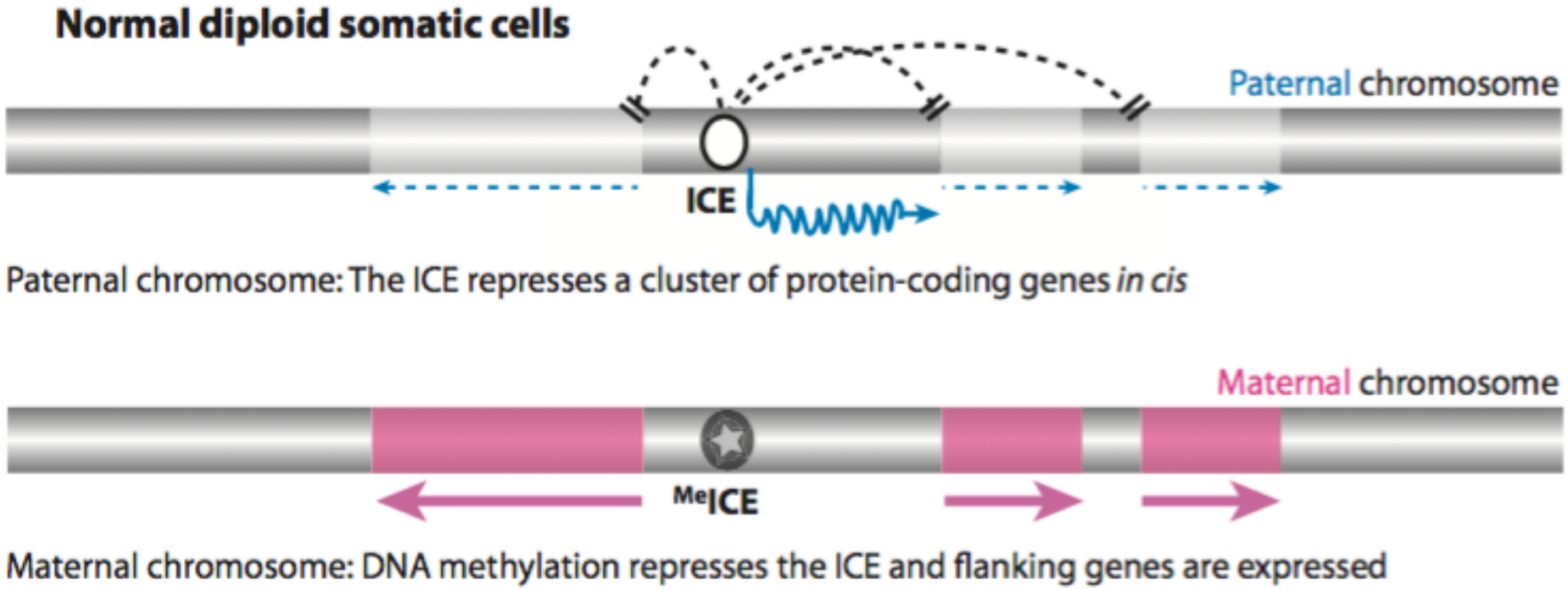


- Maternal chromosome
  - ICE repressed by gametic DNA methylation imprint (MeICE)
  - macro ncRNA is also repressed
  - flanking protein-coding genes are expressed

Barlow, Annu. Rev. Genet. 45:379–403 (2011)

Imprinting

A double-negative mechanism. ICE is methyl-sensitive, *cis*-acting repressor & macro ncRNA activator

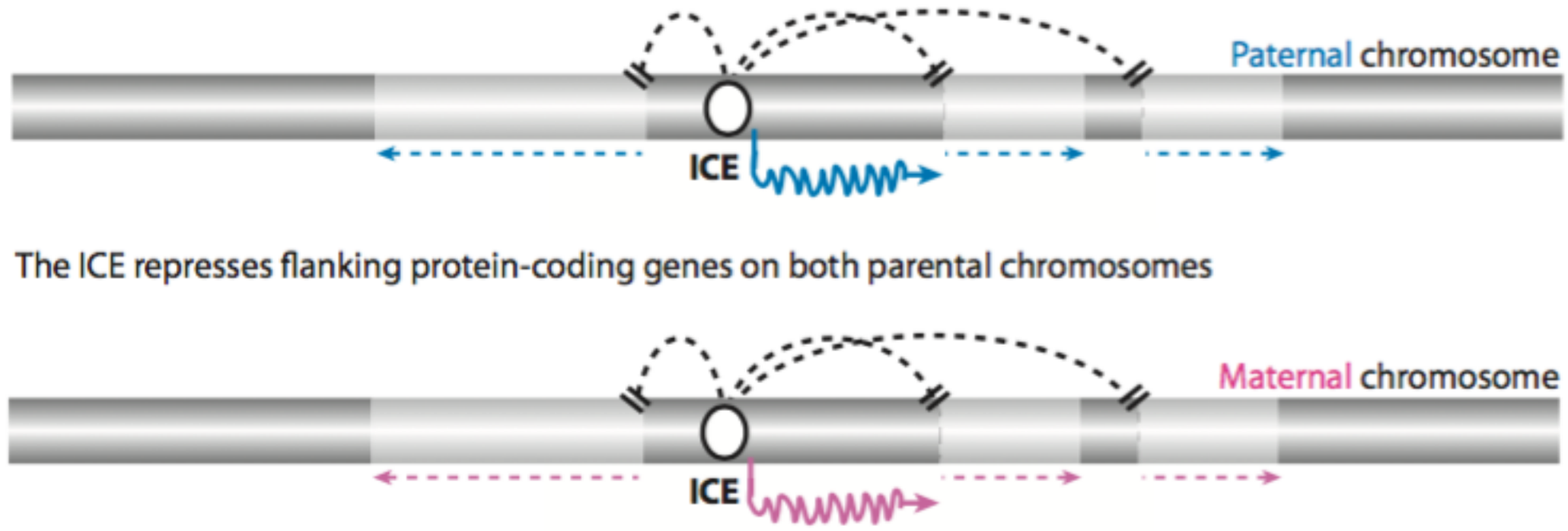


Double negative: DNA methylation represses a repressor

Barlow, Annu. Rev. Genet. 45:379–403 (2011)

Imprinting

A double-negative mechanism. ICE is methyl-sensitive, *cis*-acting repressor & macro ncRNA activator

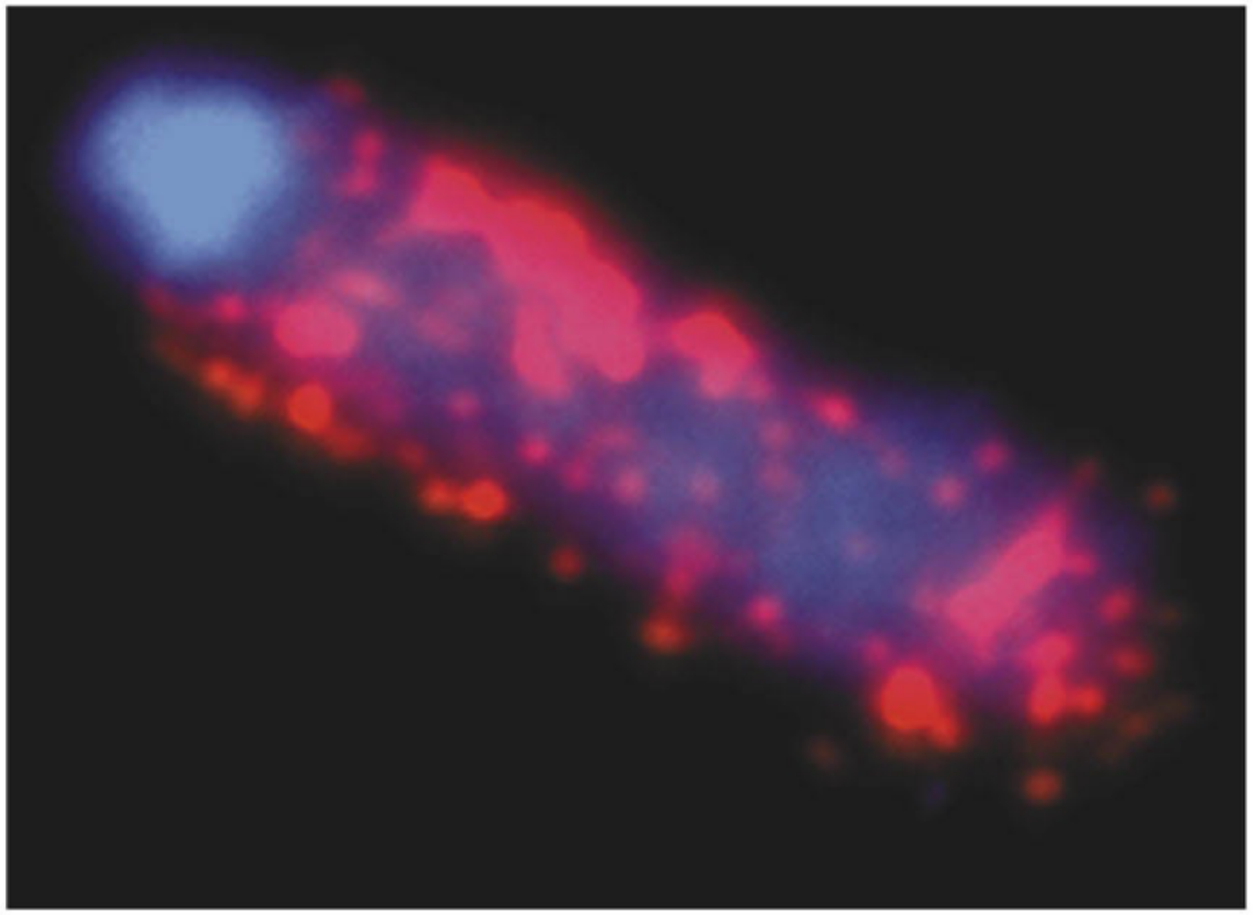


- Absence of DNA methylation
  - ICE is active on each parental chromosome
  - ncRNAs biallelically expressed
  - both ML- and EXEL-protein-coding genes are biallelically repressed.

Barlow, Annu. Rev. Genet. 45:379–403 (2011)

X inactivation

- One X is inactivated (females)
- DNA methylation at all CpG islands



- Inactive X chromosome DNA (blue).
- Red staining is bound *Xist* RNA involved in the heterochromatin formation process.

Cooper & Hausman, The Cell, 3rd edition, ASM Press and Sinauer Associates, Inc (2003)
Figure 6.35

Epigenetics and Cancer

Gene	Tumours
Chromatin remodelling	
SMARCB1	Paediatric malignant rhabdoid tumours
SMARCA4	Lung adenocarcinoma, Burkitt lymphoma, medulloblastoma
PBRM1	Clear cell renal carcinoma
ARID1A	Ovarian clear cell carcinoma, hepatocellular carcinoma, colorectal cancer, lung adenocarcinoma
ARID1B, ARID2	Hepatocellular carcinoma, melanoma, pancreatic cancer, breast cancer
SMARCD1	Breast cancer
SMARCE1	Clear cell meningioma
ATRX	Paediatric glioblastoma, pancreatic neuroendocrine tumours
DAXX	Paediatric glioblastoma, pancreatic neuroendocrine tumours
CHD5	Neuroblastoma, glioma, breast, lung, colon, ovary, prostate cancers
CHD2	Chronic lymphocytic leukaemia
CHD1, CHD3, CHD4, CHD6, CHD7, CHD8	Gastric, colorectal, prostate, breast, bladder, serous endometrial cancers

Feinberg, Koldobskiy & Göndör, Nature Reviews Genetics, doi:10.1038/nrg.2016.13(2016)

Epigenetics and Cancer

Gene	Tumours
DNA methylation	
DNMT3A	T cell lymphoma, myeloid malignancies including acute myeloid leukaemia
DNMT1	Colorectal cancer
TET2	T cell lymphoma, myeloid malignancies including acute myeloid leukaemia
TET1, TET3	Colorectal cancer, chronic lymphocytic leukaemia
MBD1, MBD4	Colorectal cancer, lung adenocarcinoma, breast cancer, melanoma

Feinberg, Koldobskiy & Göndör, Nature Reviews Genetics, doi:10.1038/nrg.2016.13(2016)

Epigenetics and Cancer

Gene	Tumours
Histone acetylation	
EP300	Diffuse large B cell lymphoma, follicular lymphoma, small-cell lung cancer, transitional cell bladder cancer, serous endometrial cancer, pancreatic cancer
CREBBP	Diffuse large B cell lymphoma, follicular lymphoma, small-cell lung cancer, transitional cell bladder cancer, ovarian cancer, relapsed acute lymphoblastic leukaemia
HDAC2	Colorectal cancer
HDAC4	Breast adenocarcinoma
HDAC9	Prostate adenocarcinoma

Feinberg, Koldobskiy & Göndör, Nature Reviews Genetics, doi:10.1038/nrg.2016.13(2016)

Epigenetics and Cancer

Gene	Tumours
Histone methylation	
MLL	Myeloid and lymphoid leukaemias, majority of infant acute lymphoblastic leukaemia, solid tumours (colorectal, lung, bladder, breast)
MLL2	Non-Hodgkin lymphoma (90% of follicular lymphoma, one-third of diffuse large cell lymphoma)
MLL3, MLL4	Solid tumours: bladder, lung, endometrial, hepatocellular
SETD1A	Gastric adenocarcinoma, breast cancer, chronic lymphocytic leukaemia
PRDM9	Head and neck squamous cell carcinoma
MLL	Myeloid and lymphoid leukaemias, majority of infant acute lymphoblastic leukaemia, solid tumours (colorectal, lung, bladder, breast)
NSD1	Acute myeloid leukaemia, head and neck squamous cell carcinoma, endometrial carcinoma, melanoma, colorectal cancer, multiple myeloma
NSD2	Paediatric acute lymphoblastic leukaemia, colorectal cancer, melanoma
SETD2	Renal cell carcinoma, early T cell precursor acute lymphoblastic leukaemia, high-grade glioma
KDM5C (JARID1C)	Renal cell carcinoma
KDM6A (UTX)	Multiple myeloma, oesophageal squamous cell carcinoma, renal cell carcinoma, medulloblastoma, prostate, transitional cell bladder cancer
KDM2B	Diffuse large B cell lymphoma

Feinberg, Koldobskiy & Göndör, Nature Reviews Genetics, doi:10.1038/nrg.2016.13(2016)

Epigenetics and Cancer

Gene	Tumours
Readers	
PHF6	T cell acute lymphoblastic leukaemia, acute myeloid leukaemia
PHF23	Acute myeloid leukaemia
BRD4	NUT midline carcinoma
BRD8	Hepatocellular carcinoma
ING1	Melanoma, oesophageal squamous cell cancer, acute lymphoblastic leukaemia
Histones	
H3F3A	Paediatric glioblastoma, diffuse intrinsic pontine glioma, giant cell tumour of bone
H3F3B	Chondroblastoma
HIST1H3B	Paediatric glioblastoma, diffuse intrinsic pontine glioma
HIST1H1B	Chronic lymphocytic leukaemia, follicular lymphoma, colorectal cancer

Feinberg, Koldobskiy & Göndör, Nature Reviews Genetics, doi:10.1038/nrg.2016.13(2016)

Epigenetics and Cancer

Class	Definition	Examples
Genetic classification		
Oncogene	A gene whose activation by mutation is advantageous to the cancer cell. Acts as dominant	MYC, KRAS, PIK3CA, ABL1, BRAF
Tumour suppressor gene	A gene whose inactivation by mutation is advantageous to the cancer cell. Generally acts as recessive	RB1, TP53, WT1, NF1, NF2, VHL, APC, CDKN2A
Selection classification		
Driver gene	A gene whose mutation or aberrant expression is subject to selection during tumorigenesis	MYC, KRAS, PIK3CA, ABL1, RB1, TP53, WT1
Passenger gene	A gene mutated in cancer that is not a driver	Estimated as 99.9% of all mutational changes in cancer

Feinberg, Koldobskiy & Göndör, Nature Reviews Genetics, doi:10.1038/nrg.2016.13(2016)

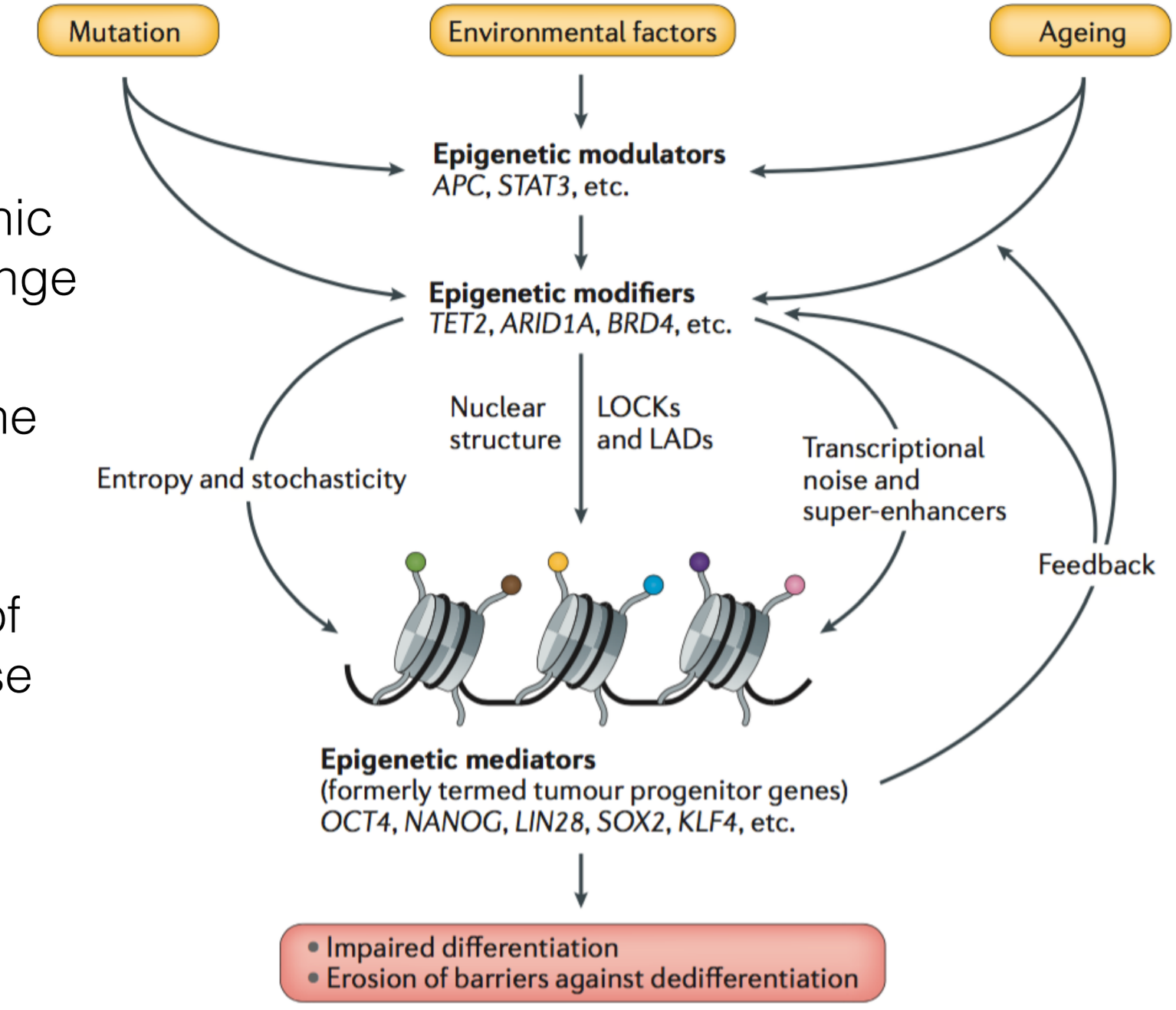
Epigenetics and Cancer

Class	Definition	Examples
Epigenetic functional classification		
Epigenetic modulator	A gene, mutated or not, that activates or represses the epigenetic machinery in cancer	IDH1/2, KRAS, APC, TP53, STAT1/3, YAP1, CTCF
Epigenetic modifier	A gene, mutated or not, that modifies DNA methylation or chromatin structure or its interpretation in cancer	SMARCA4, PBRM1, ARID1A, ARID2, ARID1B, DNMT3A, TET2, MLL1/2/3, NSD1/2, SETD2, EZH2, BRD4
Epigenetic mediator	A gene regulated by an epigenetic modifier in cancer (mutations rare or absent) that increases pluripotency or survival	OCT4, NANOG, LIN28, SOX2, KLF4

Feinberg, Koldobskiy & Göndör, Nature Reviews Genetics, doi:10.1038/nrg.2016.13(2016)

Epigenetics and Cancer

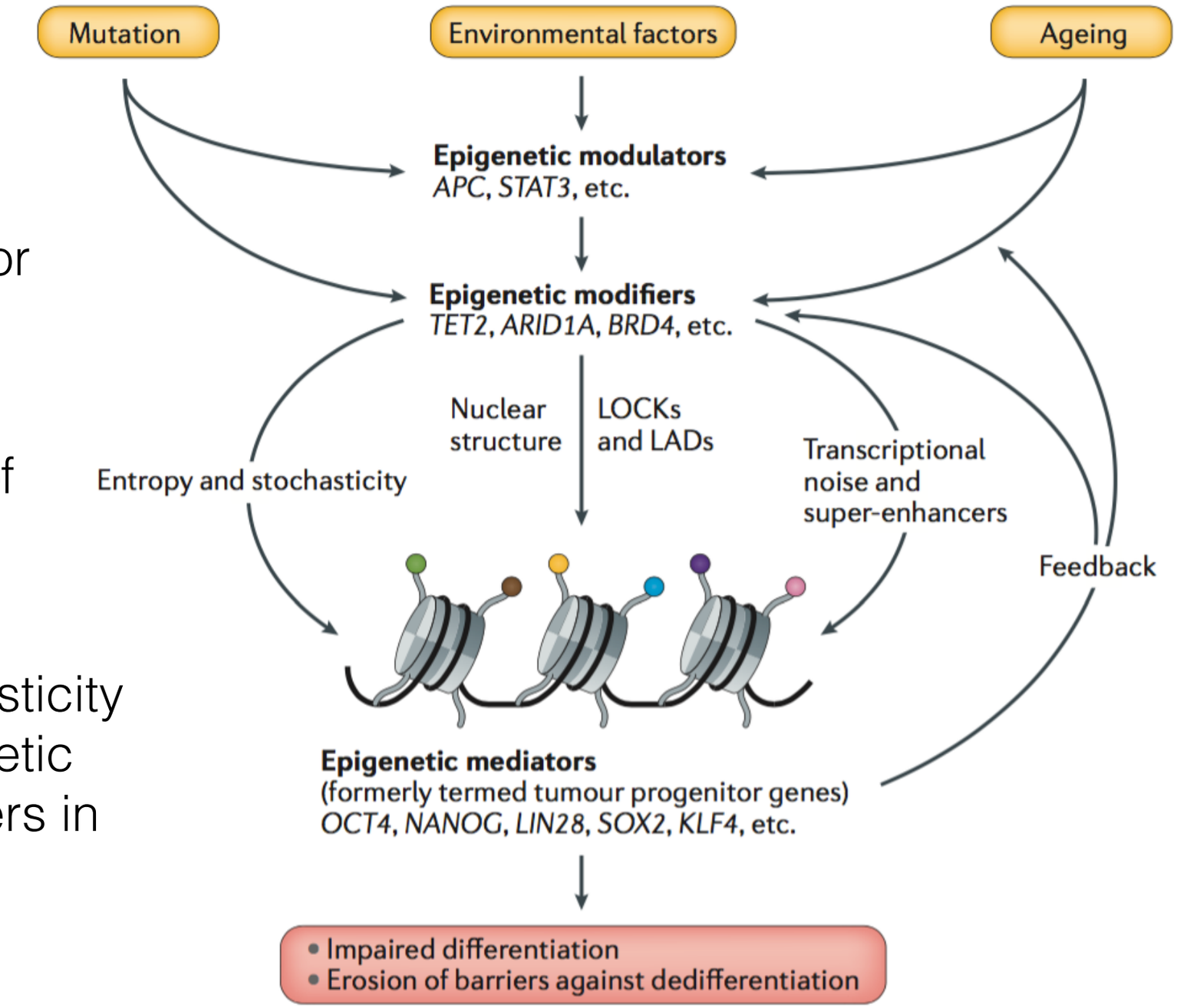
- Aging, inflammation and chronic exposure to carcinogens impinge on epigenetic modulators
- Epigenetic modulators fine tune and regulate the function of epigenetic modifiers
  - change in the expression of epigenetic mediators whose gene products regulate developmental potential.



Feinberg, Koldobskiy & Göndör, Nature Reviews Genetics, doi:10.1038/nrg.2016.13(2016)

Epigenetics and Cancer

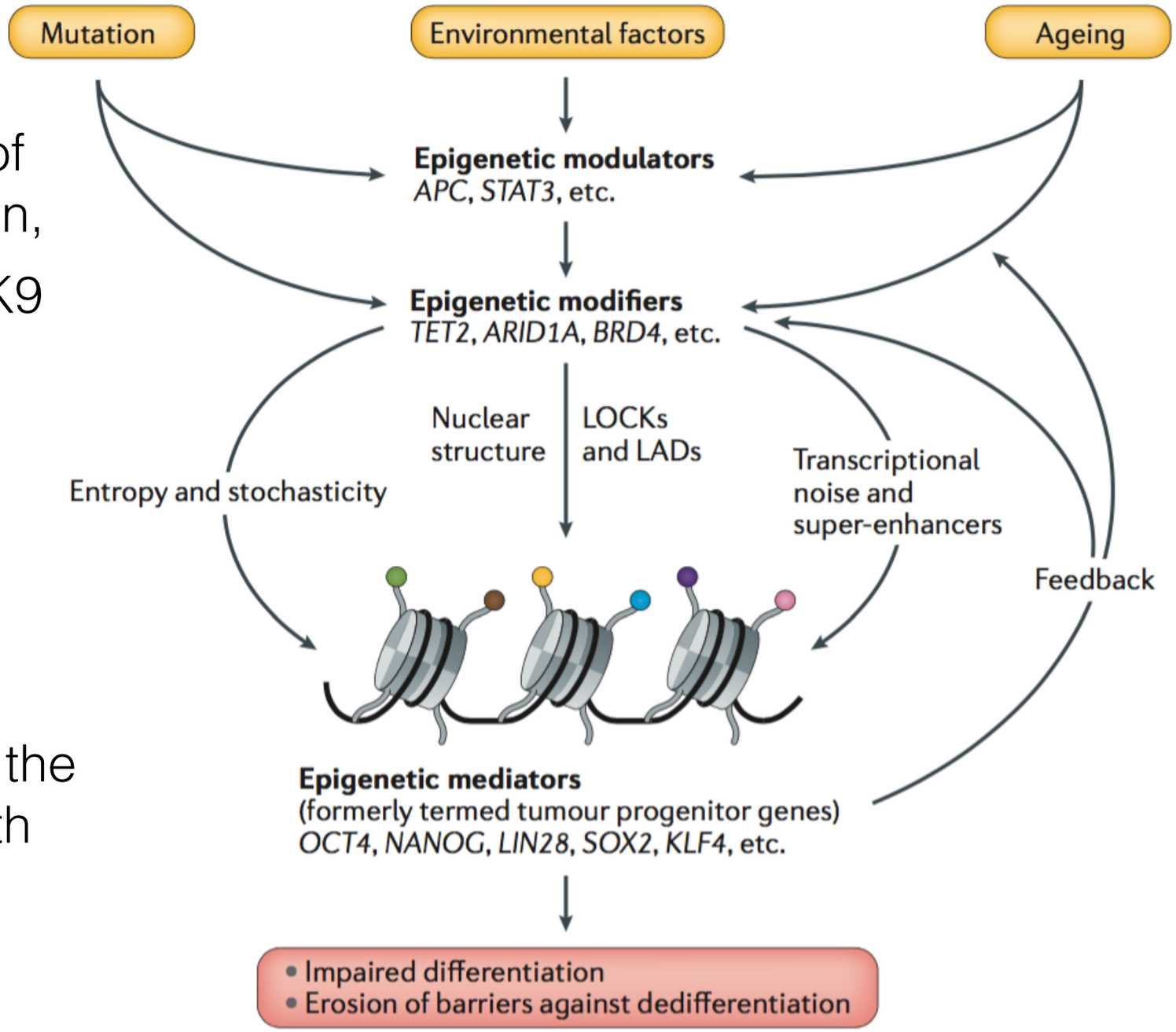
- Mutations in modulators and modifiers are often selected for during cancer development.
  - increased cell proliferation
  - unscheduled expression of mediators
    - inhibit differentiation
    - promote epigenetic plasticity by affecting the epigenetic modulators and modifiers in a feedback loop.



Feinberg, Koldobskiy & Göndör, Nature Reviews Genetics, doi:10.1038/nrg.2016.13(2016)

Epigenetics and Cancer

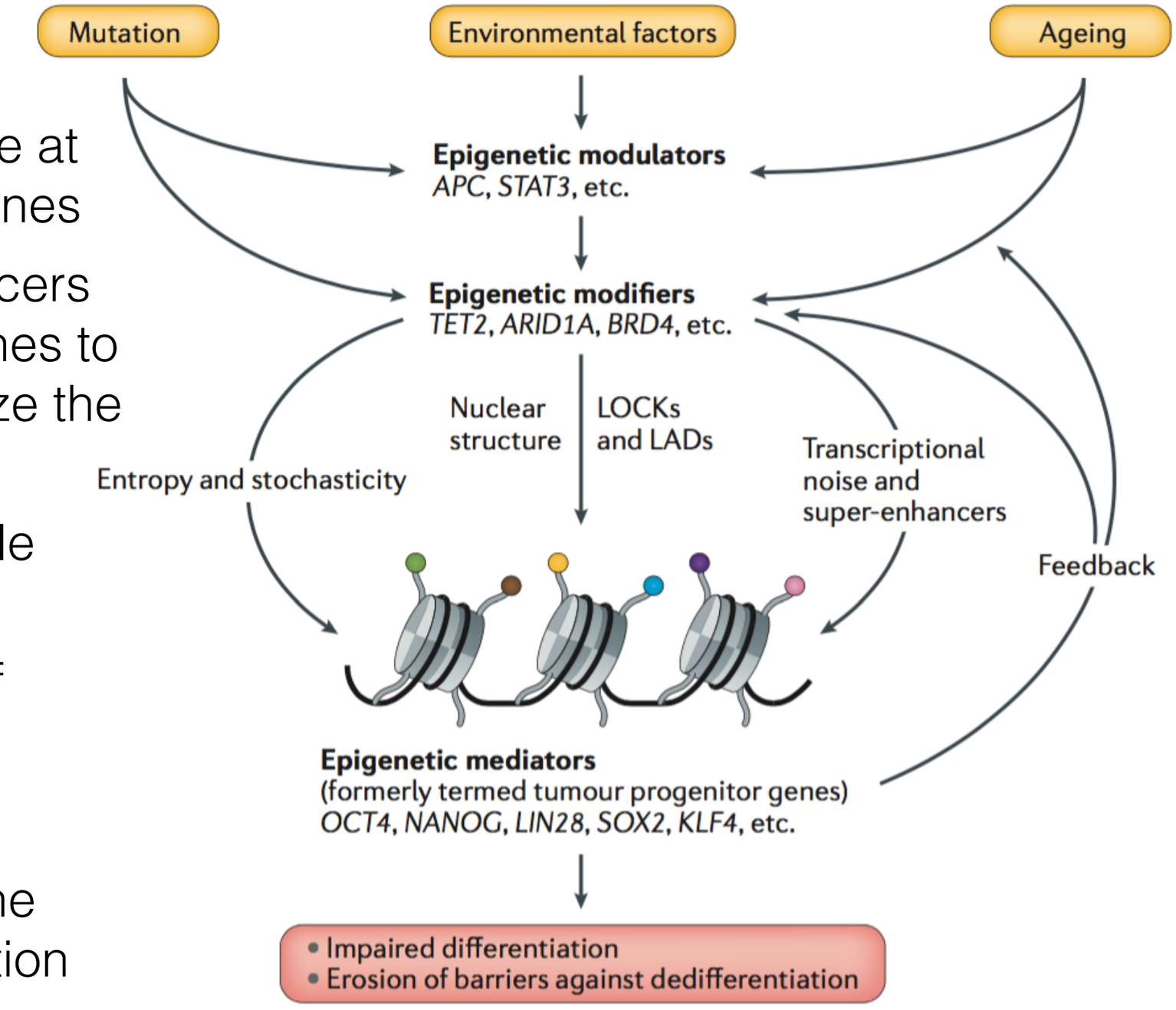
- The mechanism of epigenetic instability involves the erosion of barriers against dedifferentiation,
  - large organized chromatin K9 modifications (LOCKs) overlapping with
    - lamina-associated domains (LADs)
- Emergence of hypomethylated blocks that contain the most variably expressed domains of the tumor genome and interfere with normal differentiation.



Feinberg, Koldobskiy & Göndör, Nature Reviews Genetics, doi:10.1038/nrg.2016.13(2016)

Epigenetics and Cancer

- Increased transcriptional noise at developmentally regulated genes
- Redistribution of super-enhancers from cell-fate-determining genes to oncogenes that further stabilize the cancer cell state.
- Stochastic changes in unstable chromatin states
  - continuous regeneration of epigenetic heterogeneity
  - increased cellular entropy
  - basis for the selection of the fittest during cancer evolution



Feinberg, Koldobskiy & Göndör, Nature Reviews Genetics, doi:10.1038/nrg.2016.13(2016)

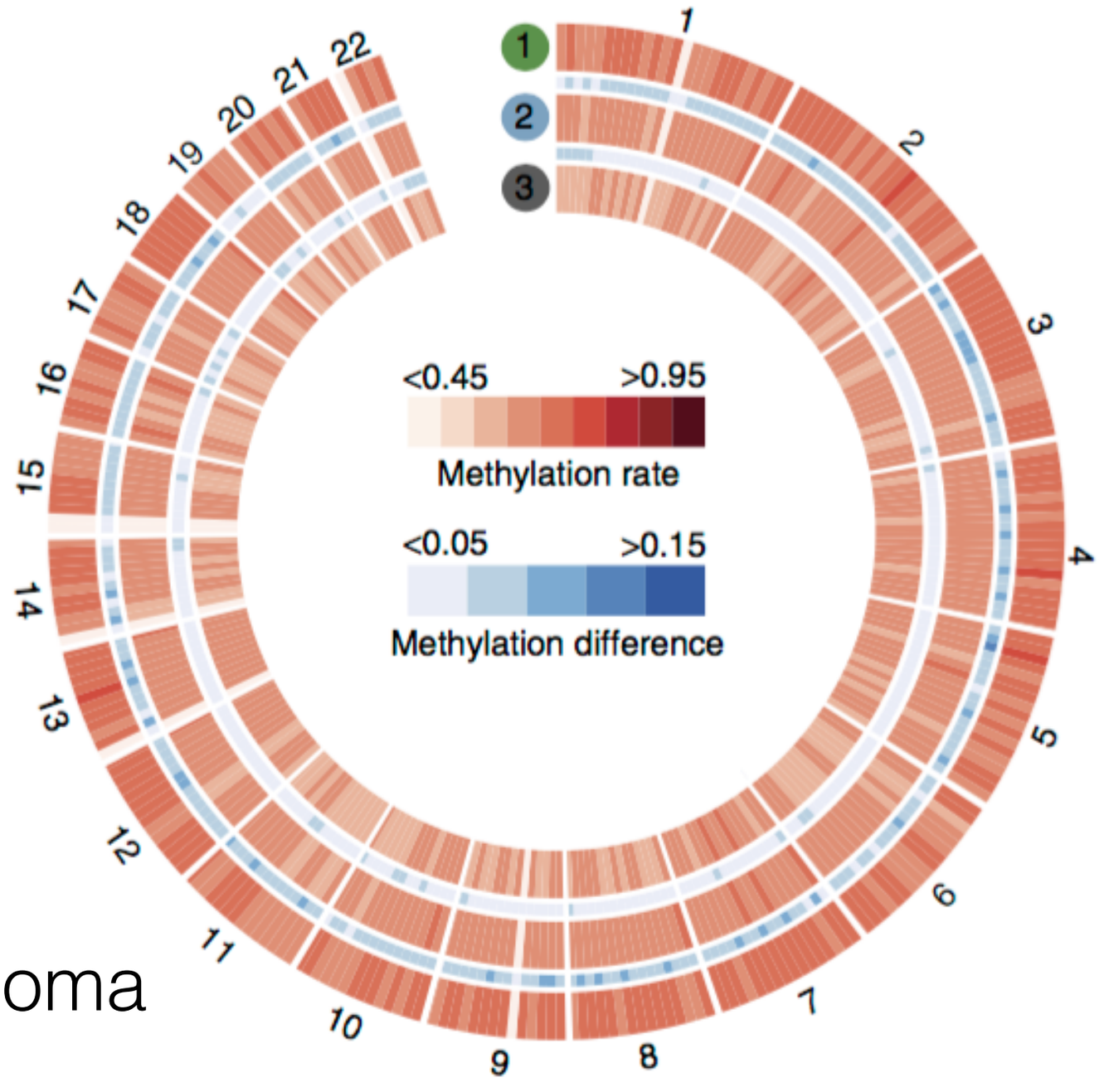
Epigenetics and Cancer

DNA methylome analysis

1: germinal center B cell samples, n = 4

2: follicular lymphomas, n = 9

3: Burkitt lymphomas, n = 13



Loss of methylation in lymphoma

Kretzmer et al., Nature Genetics 47, 1316–1325 (2015)

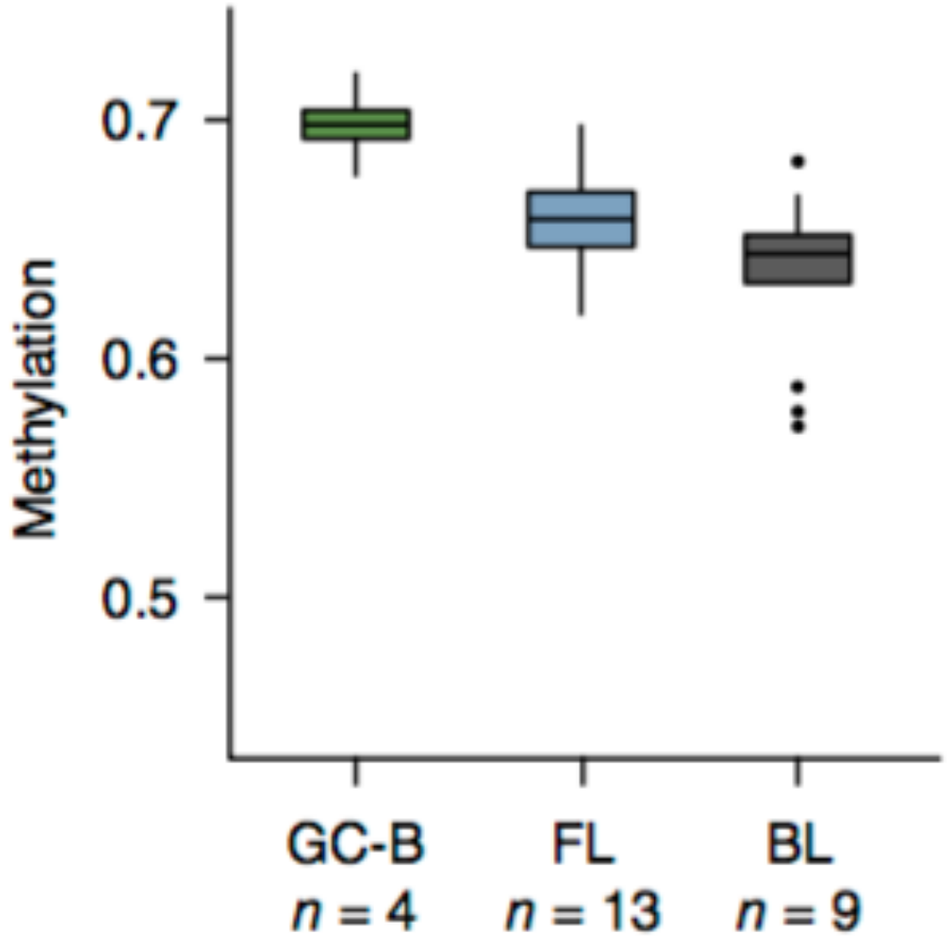
Epigenetics and Cancer

DNA methylome analysis

1: germinal center B cell samples, n = 4

2: follicular lymphomas, n = 9

3: Burkitt lymphomas, n = 13

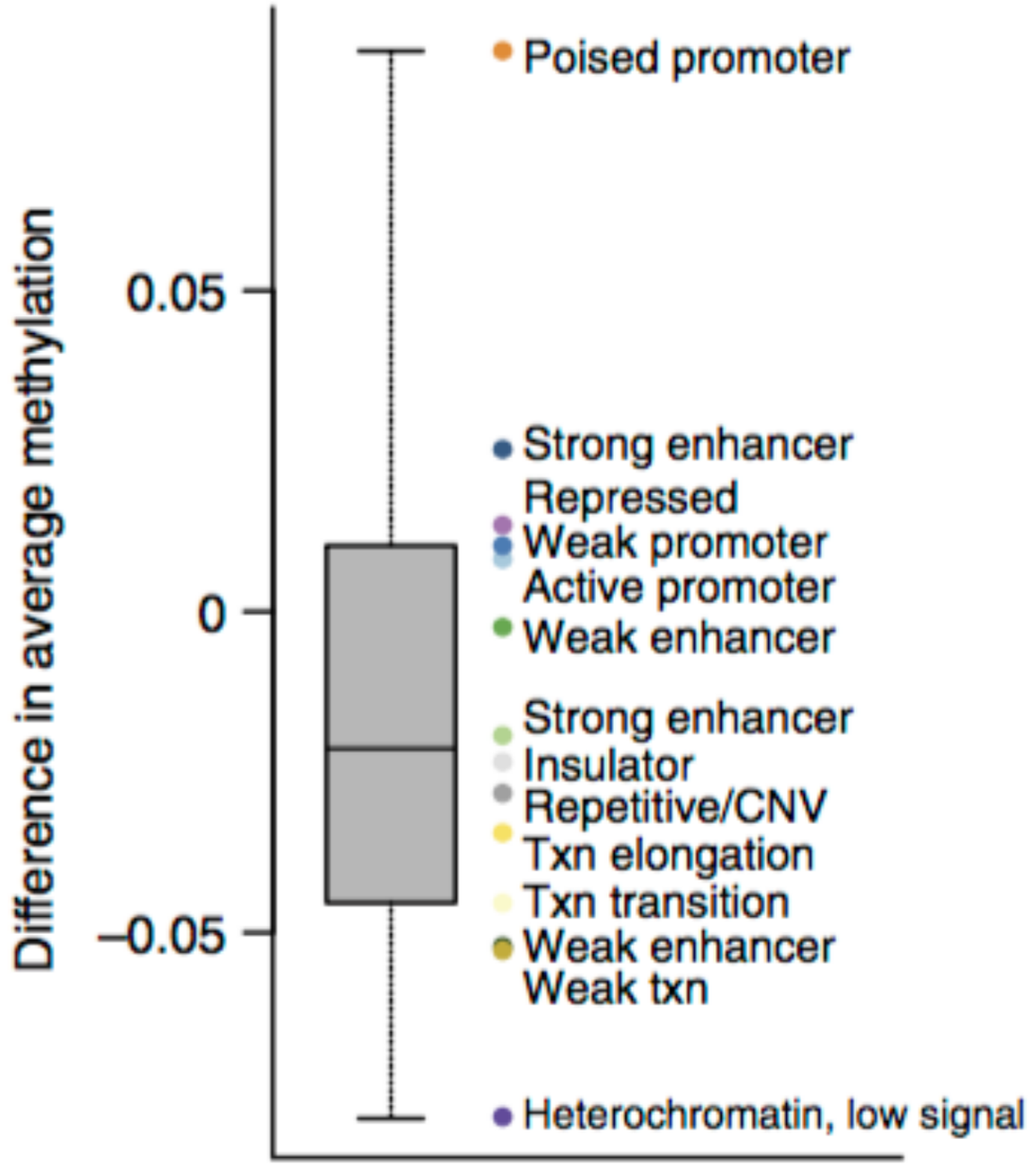


Kretzmer et al., Nature Genetics 47, 1316–1325 (2015)

Epigenetics and Cancer

Average methylation difference (average for Burkitt lymphoma and follicular lymphoma versus germinal center B cells) in GM12878 chromatin segments.

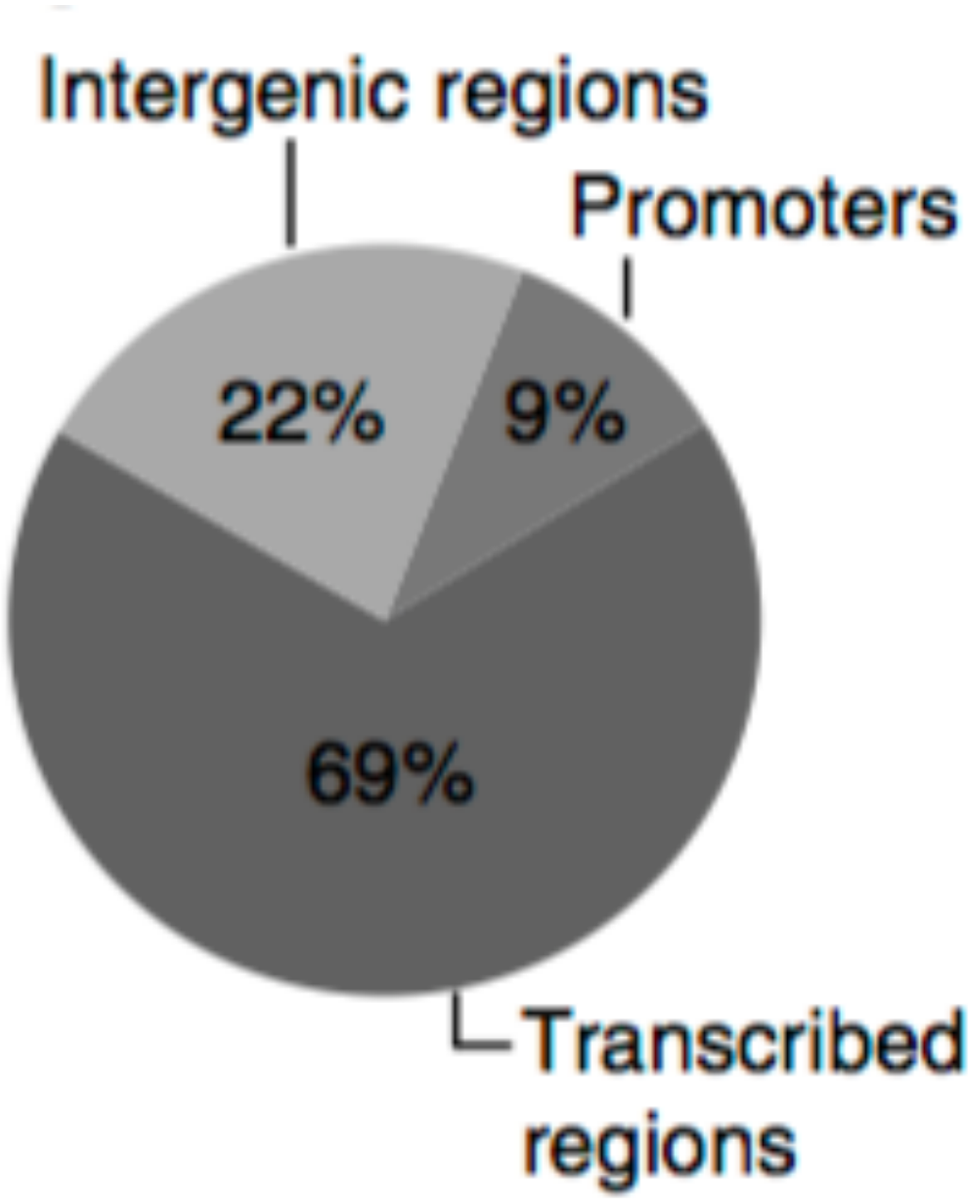
Txn, transcription



Kretzmer et al., Nature Genetics 47, 1316–1325 (2015)

Epigenetics and Cancer

Proportion of DMRs in promoters, transcribed regions & intergenic regions

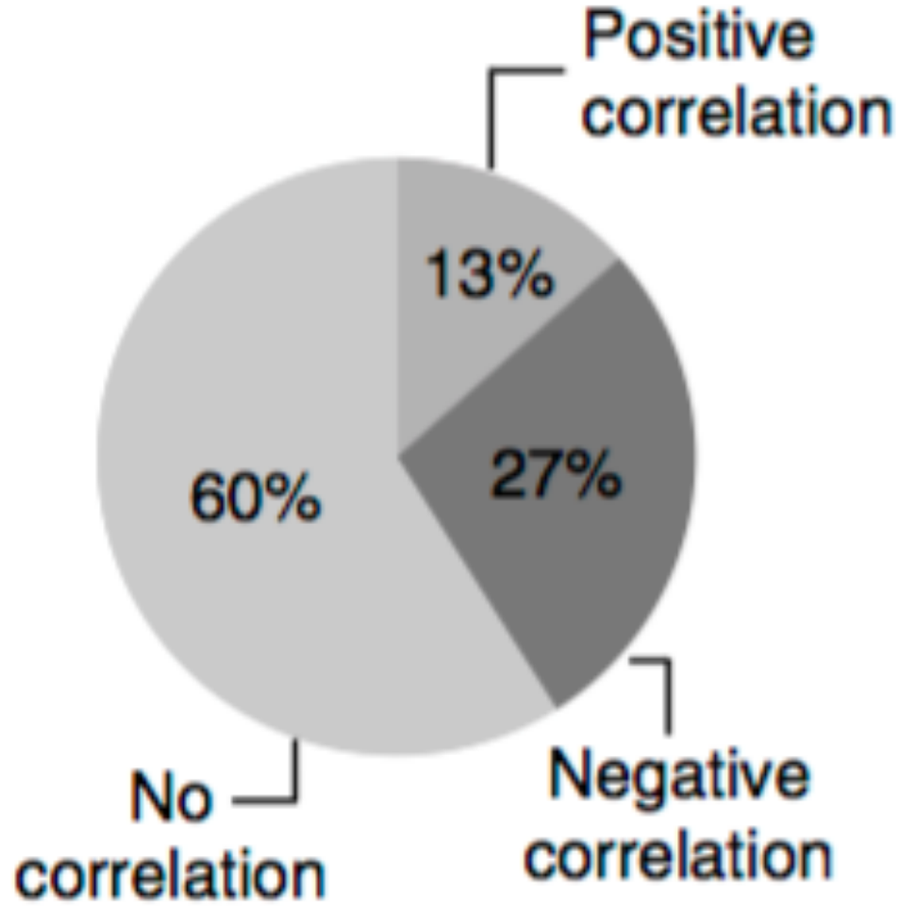


Kretzmer et al., Nature Genetics 47, 1316–1325 (2015)

Epigenetics and Cancer

40% of the intragenic DMRs (differentially methylated regions) showed significant ($P < 0.05$) correlations between methylation and RNA expression.

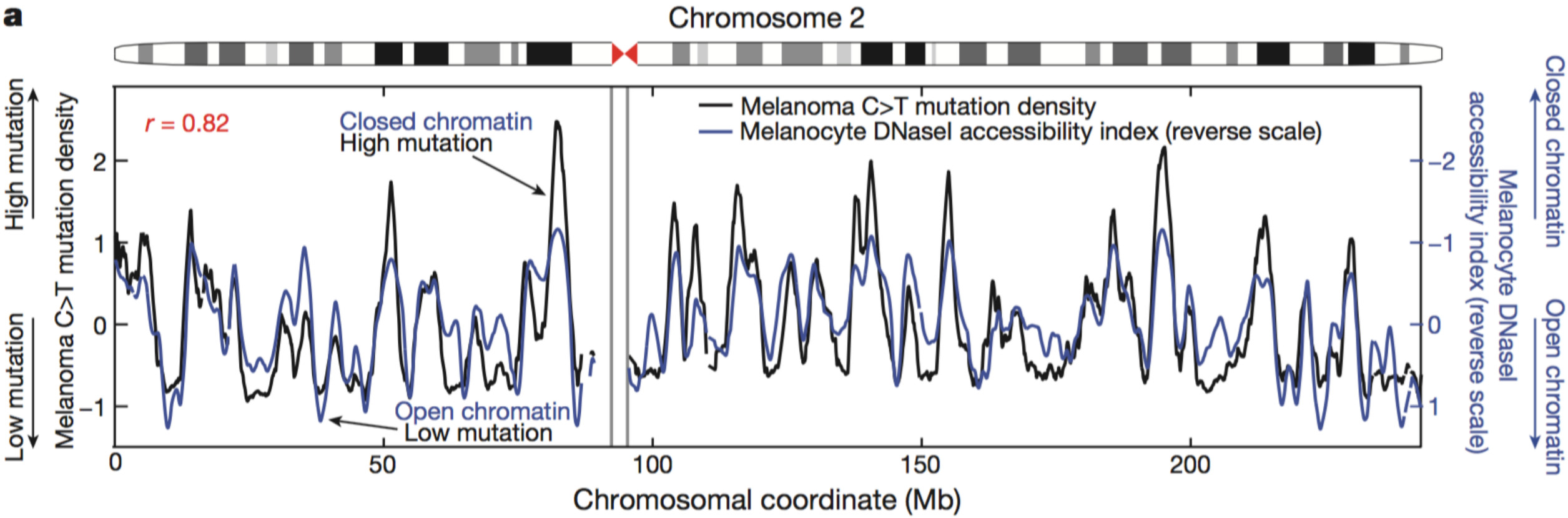
64% of them had negative correlation



Kretzmer et al., Nature Genetics 47, 1316–1325 (2015)

Epigenetics and Cancer

Mutation density in melanoma is associated with individual chromatin features specific to melanocytes.



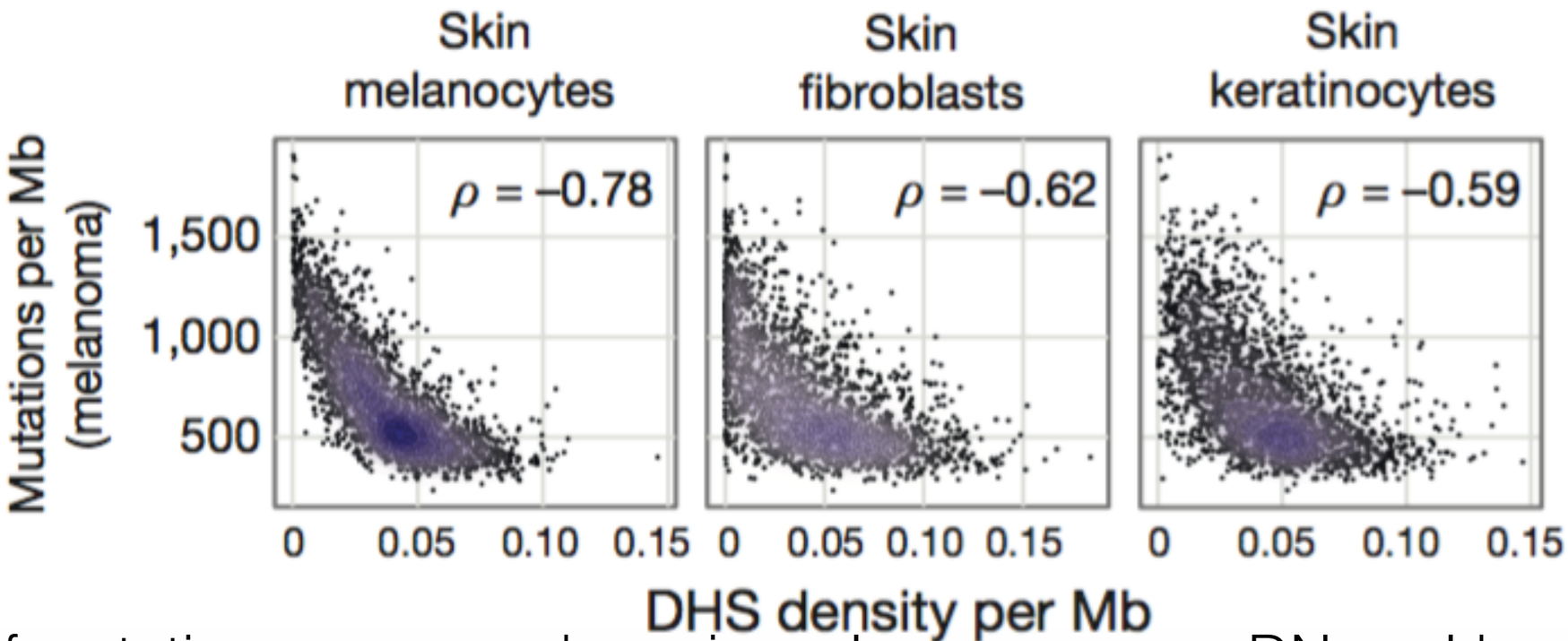
- Active chromatin and transcription associated with low mutation density.
- Repressive chromatin features associated with high mutation density.
- N.B. Not necessarily causal & no specific biological mechanisms.

DNase I hypersensitivity (DHS) as a global measure of chromatin accessibility.

Polak et al., Nature 518, 360–364 (2015)

Epigenetics and Cancer

Mutation density in melanoma is associated with individual chromatin features specific to melanocytes.



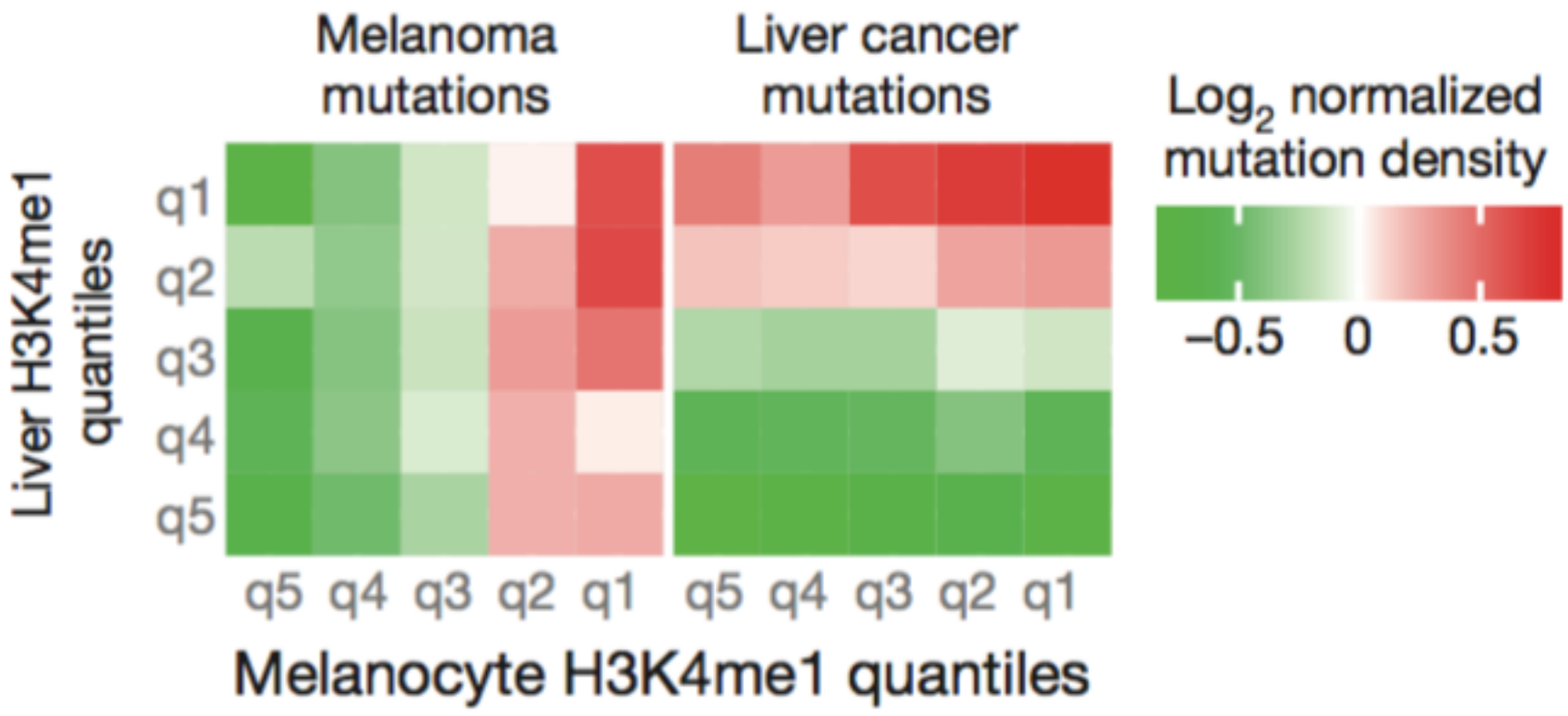
The number of mutations per megabase in melanoma versus DNase I hypersensitive sites (DHS) density.

- DHSs from melanocytes explain larger fraction of the variance in melanoma mutation density than DHSs from other cell types.

Polak et al., Nature 518, 360–364 (2015)

Epigenetics and Cancer

Mutation density in melanoma is associated with individual chromatin features specific to melanocytes.

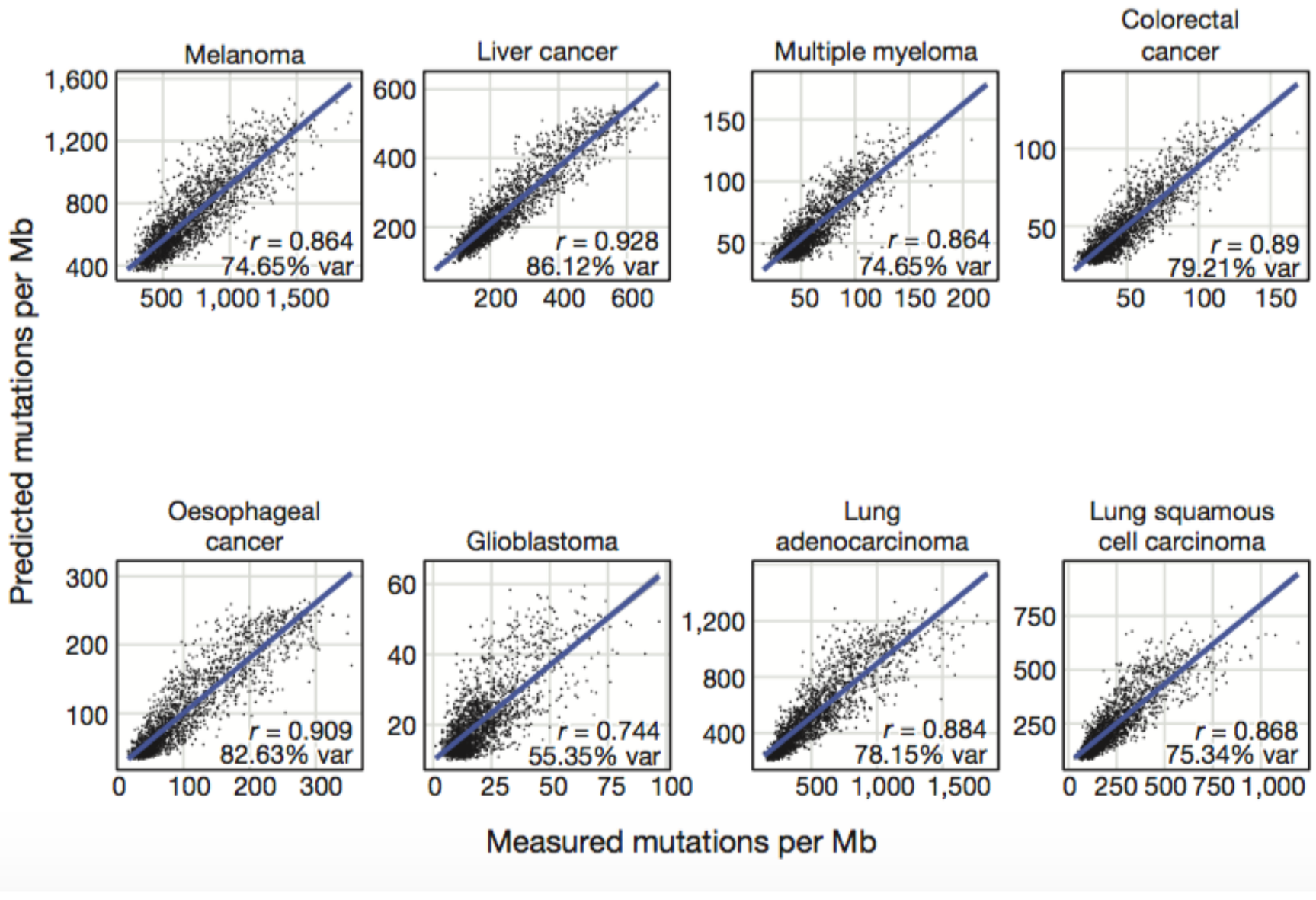


- The normalized density of mutations in liver cancer and melanoma genomes as a function of density quantiles of H3K4me1 marks in liver cells and in melanocytes.
- Mutation density depends only on H3K4me1 marks measured in the cell of origin.

Polak et al., Nature 518, 360–364 (2015)

Epigenetics and Cancer

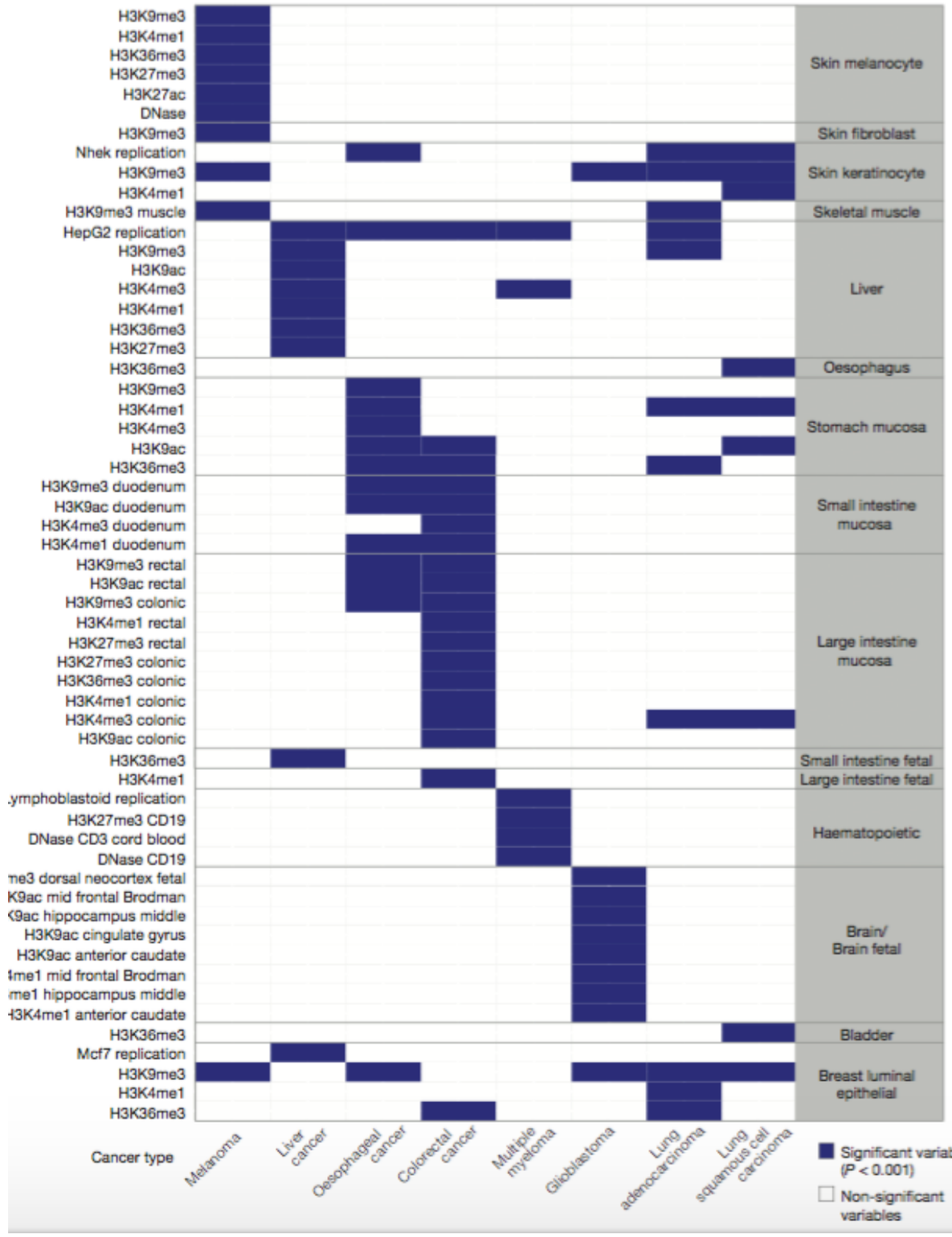
Cell type of origin of a cancer can be determined based on the distribution of mutations along its genome.



Polak et al., Nature 518, 360–364 (2015)

Epigenetics and Cancer

Epigenomic features that significantly contributed to the predictions in at least one cancer type.



Polak et al., Nature 518, 360–364 (2015)

Haplotype-resolved epigenomes across human tissues

- ChIP-seq profiling 6 histone modifications
- 16 human tissue types
- 4 individual donors
- Combined with published data sets (28 cell/tissue types total)
- Developmental states

Leung et al., Nature 518, 350–354 (2015)

Haplotype-resolved epigenomes across human tissues

- Active promoters
  - histone H3 lysine 4 trimethylation (H3K4me3)
  - H3 lysine 27 acetylation (H3K27ac)
- active enhancers
  - H3 lysine 4 monomethylation (H3K4me1)
  - H3K27ac
- transcribed gene bodies
  - H3 lysine 36 trimethylation (H3K36me3)
- silenced regions
  - H3K27 trimethylation (H3K27me3)
  - H3 lysine 9 trimethylation (H3K9me3)

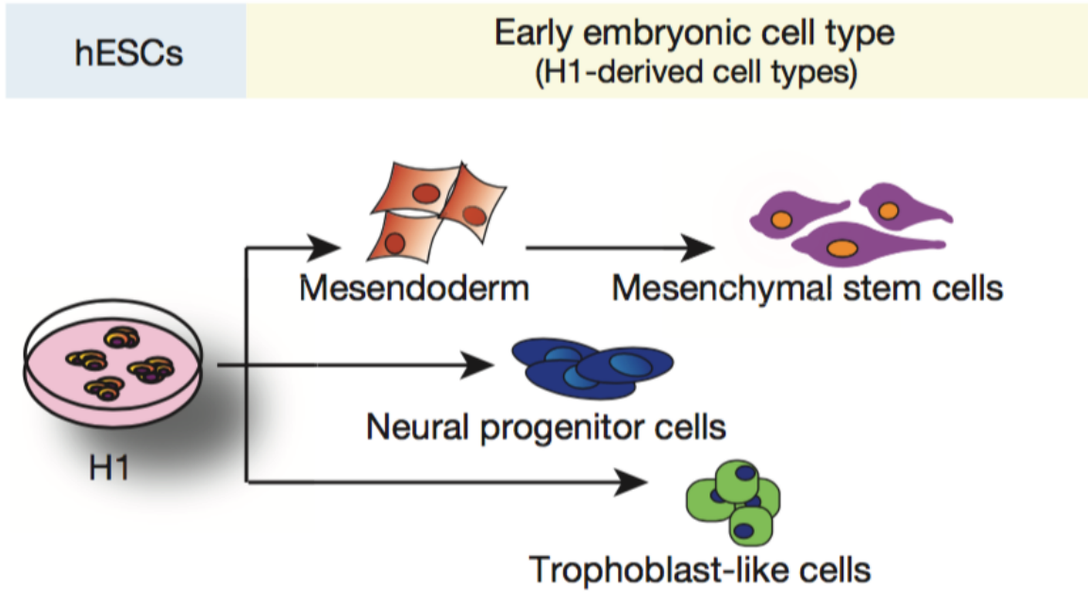
Leung et al., Nature 518, 350–354 (2015)

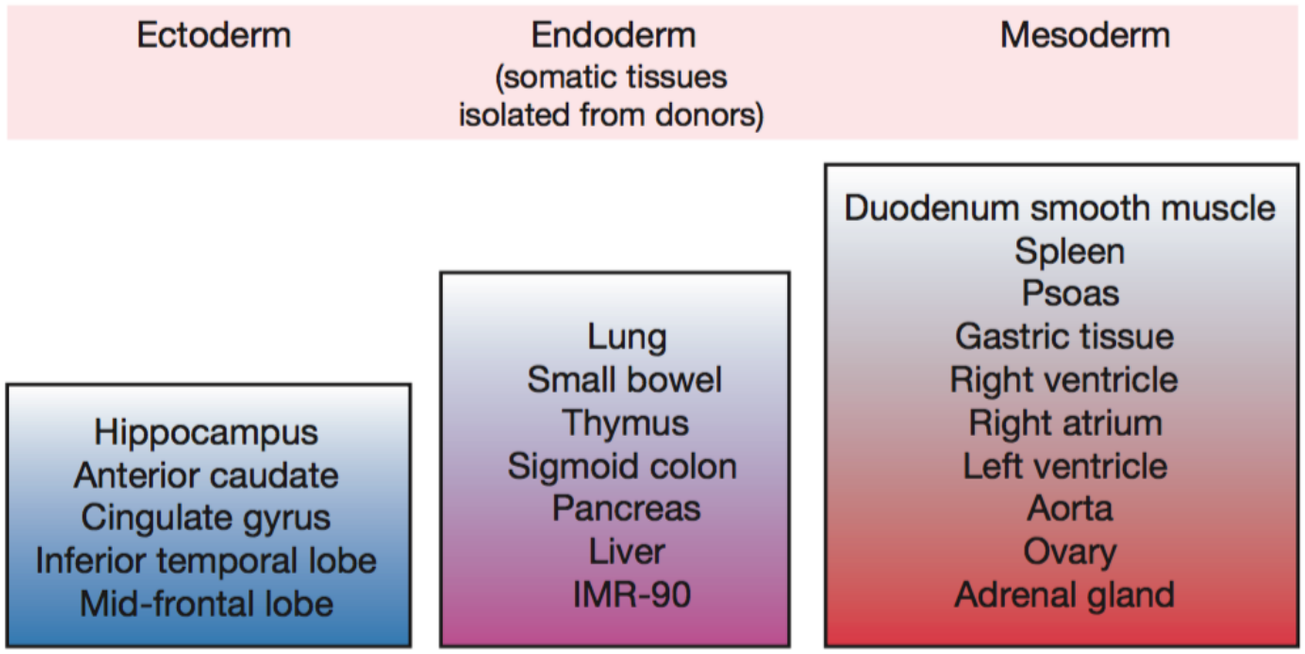
Haplotype-resolved epigenomes across human tissues

Tissues sampled

progress along

developmental lineages

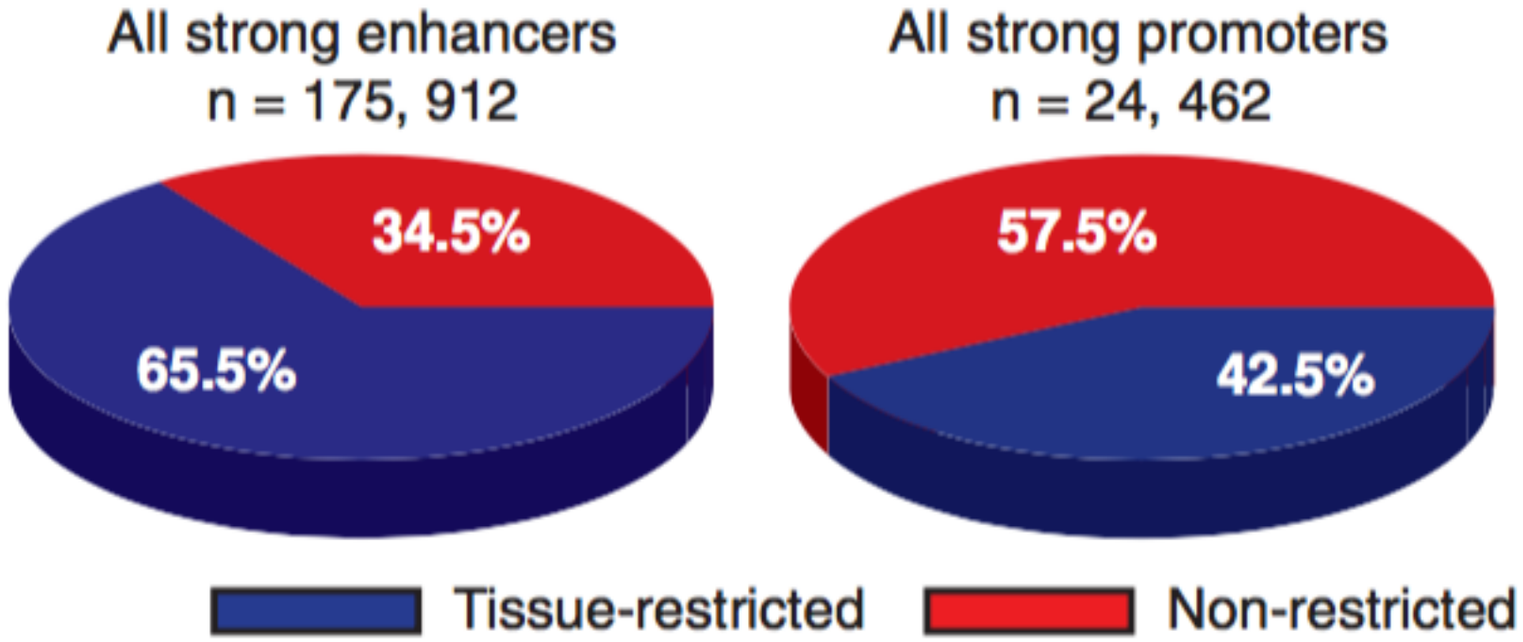




Leung et al., Nature 518, 350–354 (2015)

Haplotype-resolved epigenomes across human tissues

Defined tissue-restricted enhancers (n = 115,222) & promoters (n = 10,396)

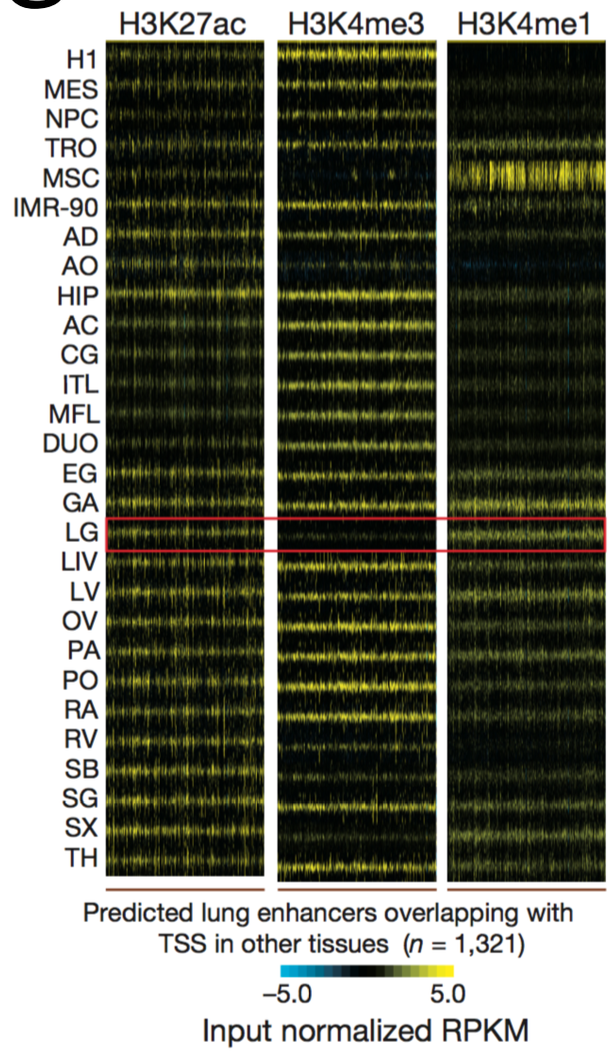


15.2% (n = 3,717) of strong promoters predicted as enhancers in other tissues, (cf mice) where intragenic enhancers act as promoters to produce cell-type-specific transcripts

Leung et al., Nature 518, 350–354 (2015)

Haplotype-resolved epigenomes across human tissues

H3K27ac, H3K4me3 and H3K4me1 enrichment at predicted lung enhancers (n =1,321), which are defined as promoters in other tissues, across 28 samples.



cis-regulatory elements with dynamic signatures (cREDS). Sites that possess histone modification signatures of active enhancers in some tissue/cell types but enriched with active promoter marks in others.

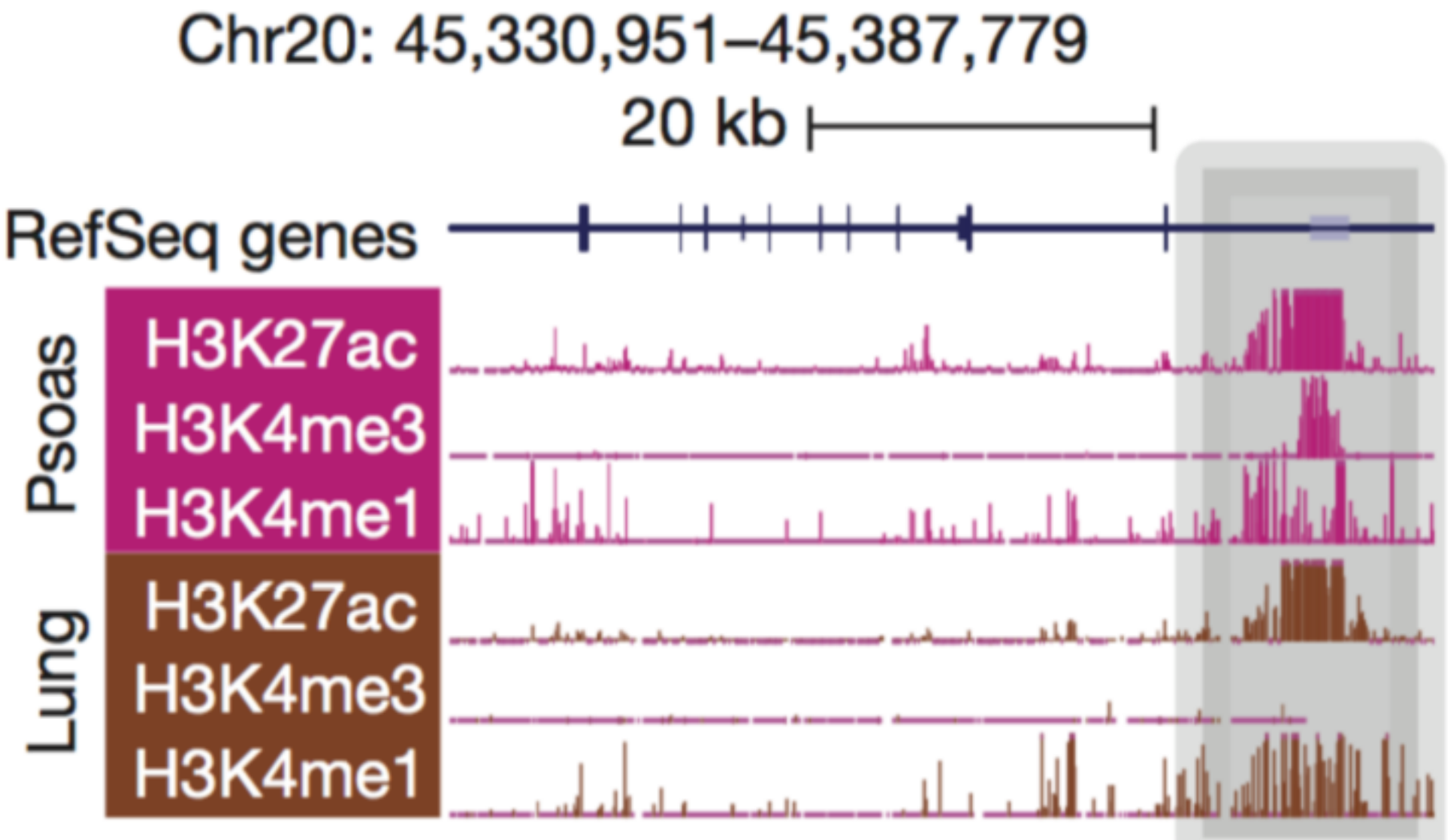
Possibly *fine tuning* of transcriptomes

Embryonic stem cells (H1), early embryonic lineages (mesendoderm cells (MES), neural progenitor cells (NPC), trophoblast-like cells (TRO) and mesenchymal stem cells (MSC)) and somatic primary tissues, representative of all three germ layers (ectoderm: hippocampus (HIP), anterior caudate (AC), cingulate gyrus (CG), inferior temporal lobe (ITL) and mid-frontal lobe (MFL); endoderm: lung (LG), small bowel (SB), thymus (TH), sigmoid colon (SG), pancreas (PA), liver (LIV) and IMR-90 fibroblasts; mesoderm: duodenum smooth muscle (DUO), spleen (SX), psoas (PO), gastric tissue (GA), right heart ventricle (RV), right heart atrium (RA), left heart ventricle (LV), aorta (AO), ovary (OV) and adrenal gland (AD))

Leung et al., Nature 518, 350–354 (2015)

Haplotype-resolved epigenomes across human tissues

Chromatin states of a cREDS element (grey shading) predicted as a promoter in psoas & an enhancer in lung.

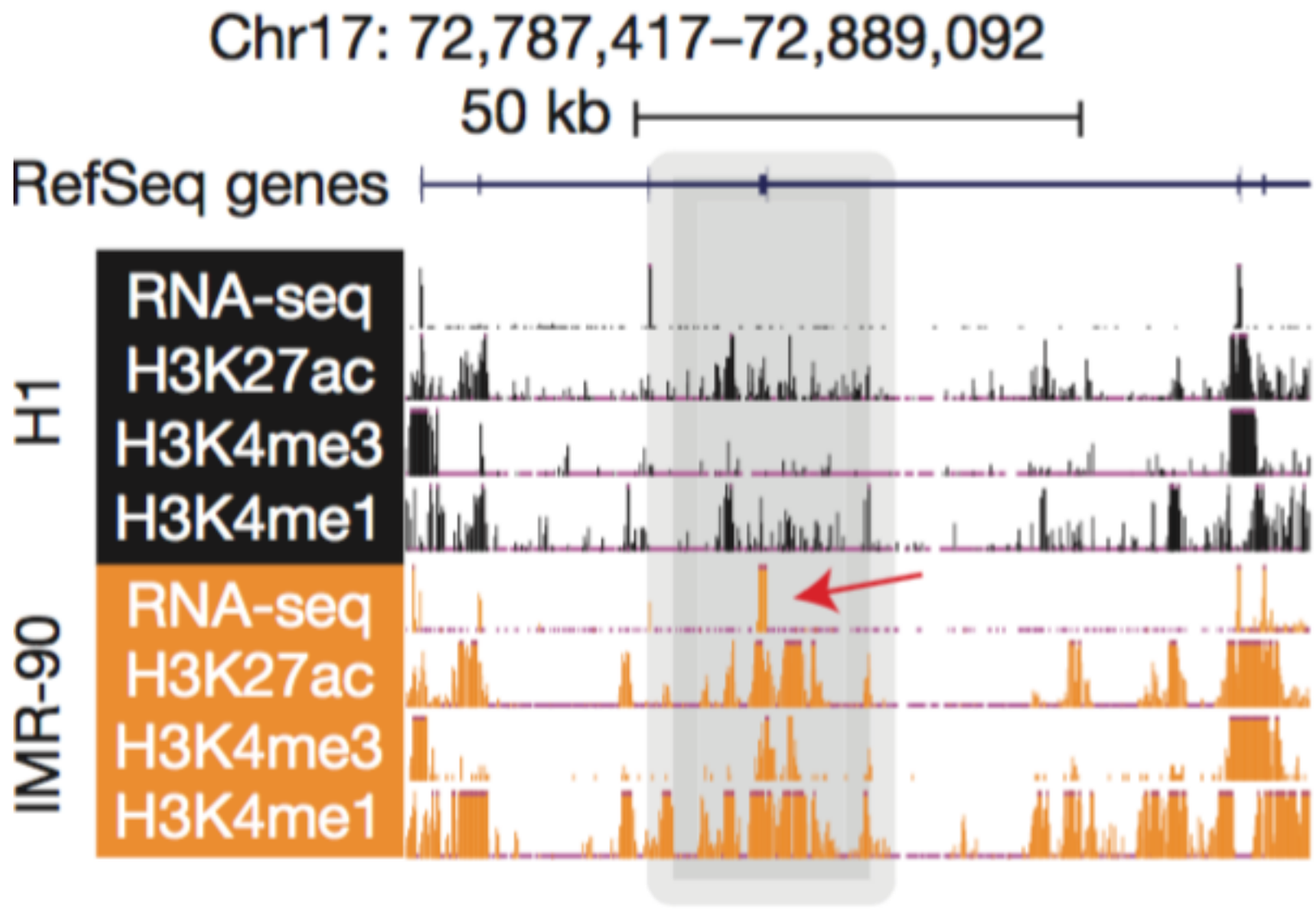


Enrichment of H3K27ac and H3K4me1 & depletion of H3K4me3 in lung

Leung et al., Nature 518, 350–354 (2015)

Haplotype-resolved epigenomes across human tissues

RNA-seq and chromatin states of a cREDS element (grey shading) in H1 and IMR-90

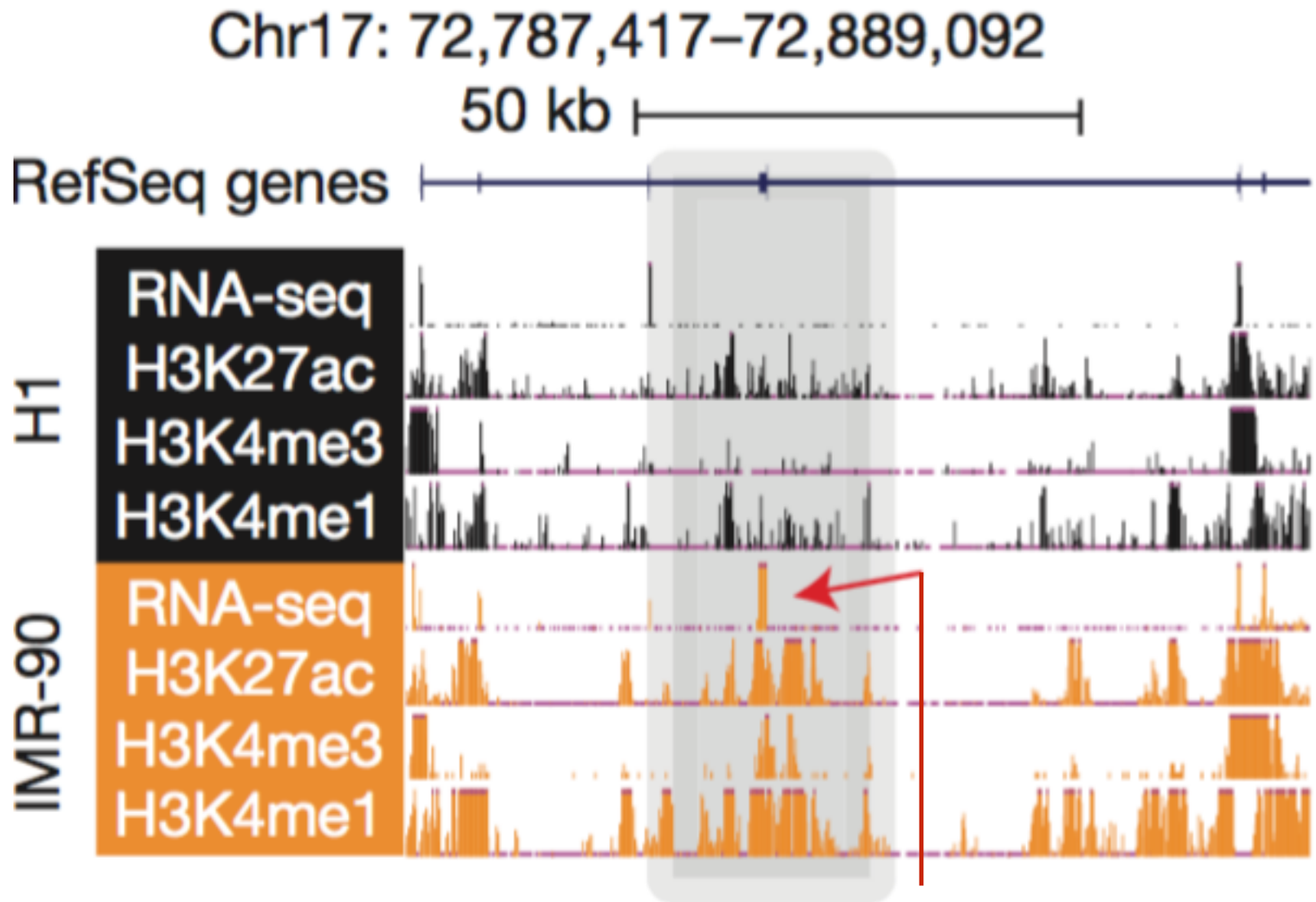


Leung et al., Nature 518, 350–354 (2015)

Haplotype-resolved epigenomes across human tissues

- Subset of cREDS promoters accompanied by creation of new transcripts and/or alternative exon usage (n=99)

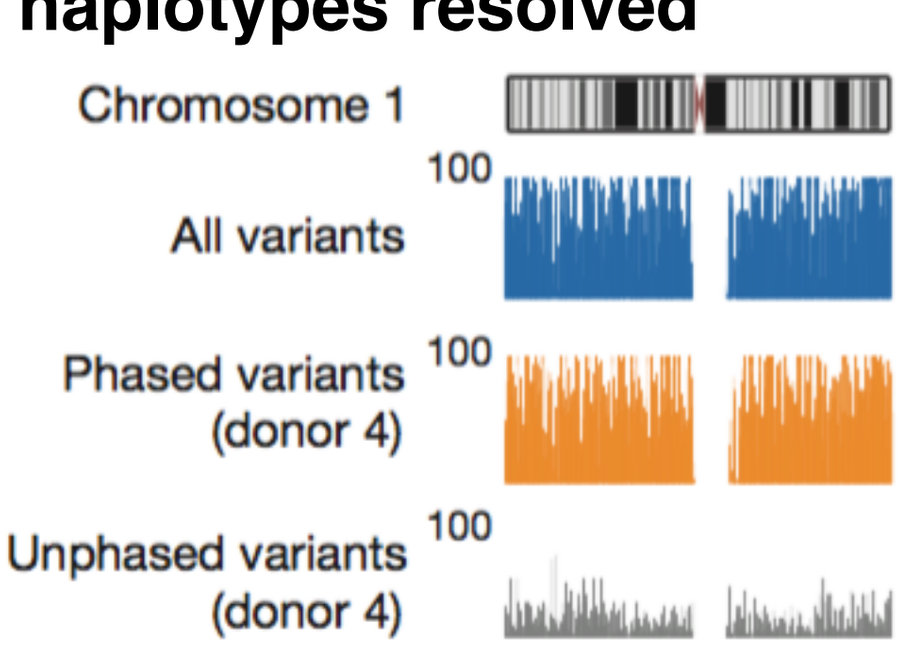
- ? cREDS influence cell/ tissue-specific transcript variants

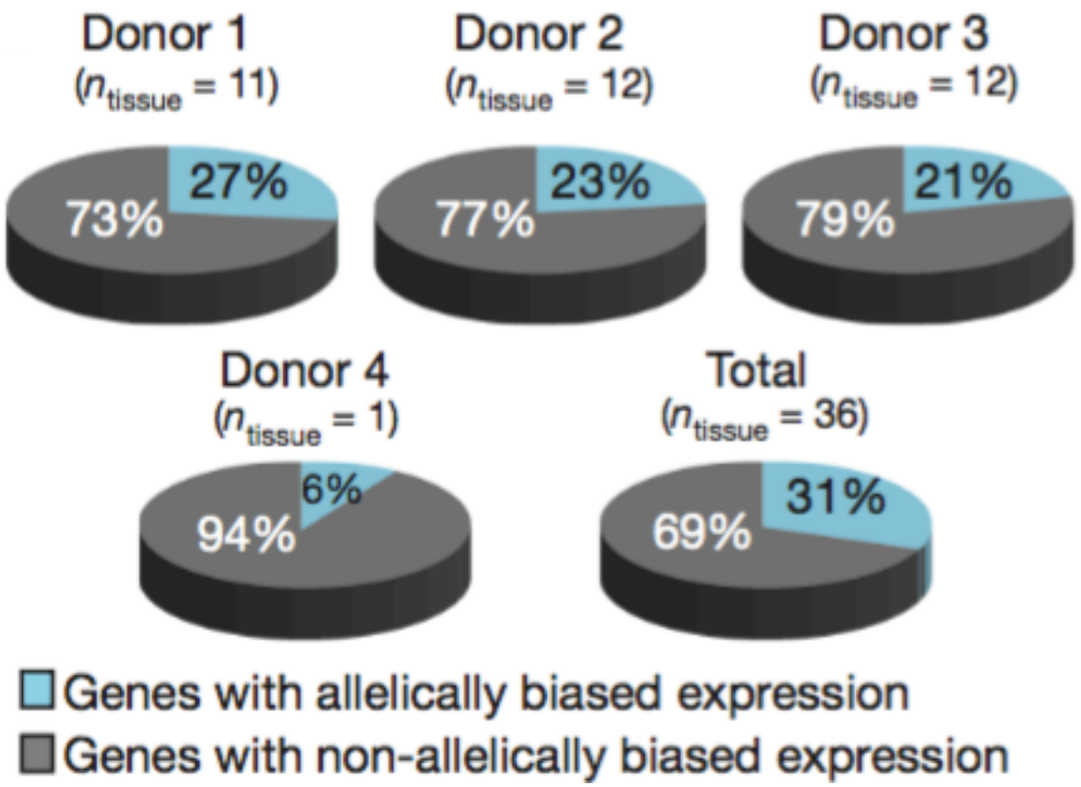


Leung et al., Nature 518, 350–354 (2015)

Haplotype-resolved epigenomes across human tissues

Resolution of haplotypes resolved



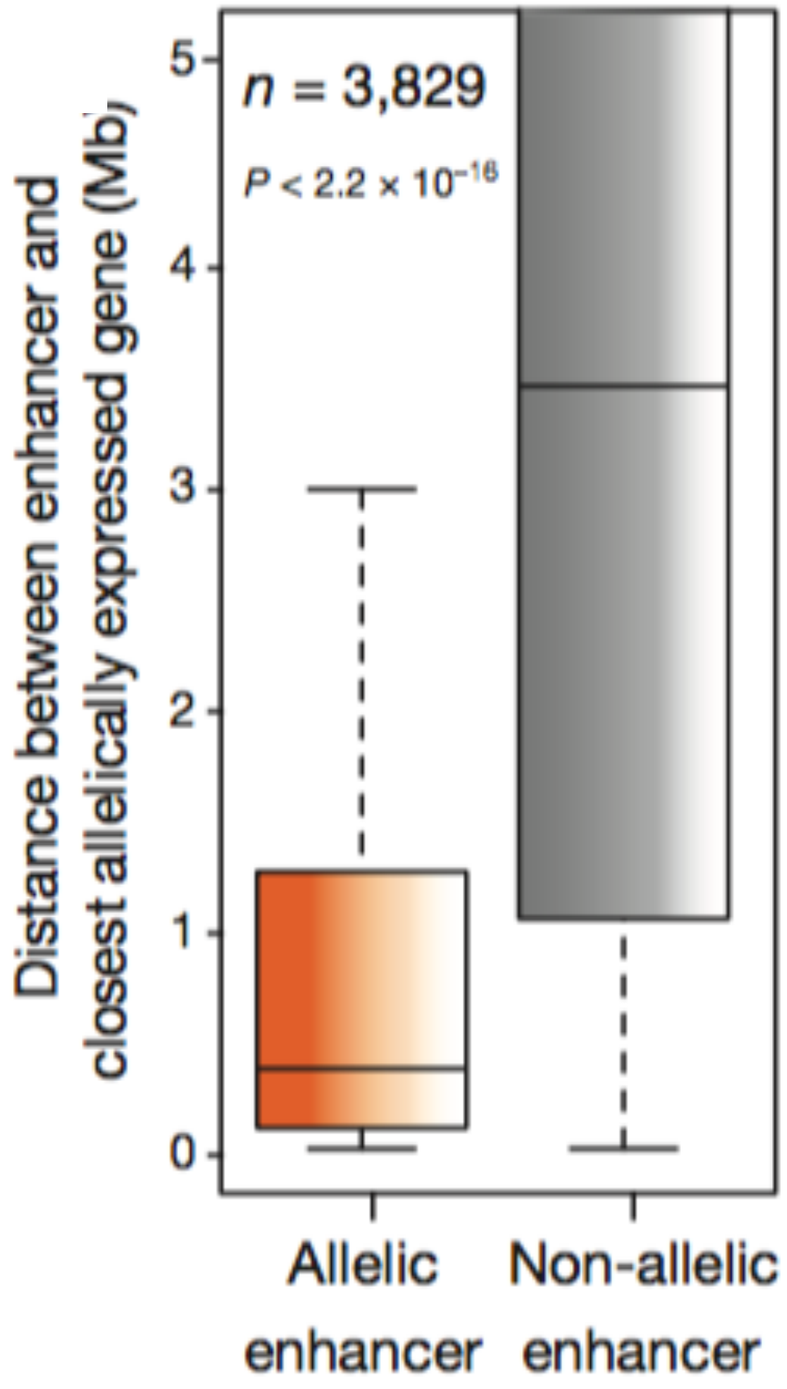


Allelically biased expression among informative genes

Leung et al., Nature 518, 350–354 (2015)

Haplotype-resolved epigenomes across human tissues

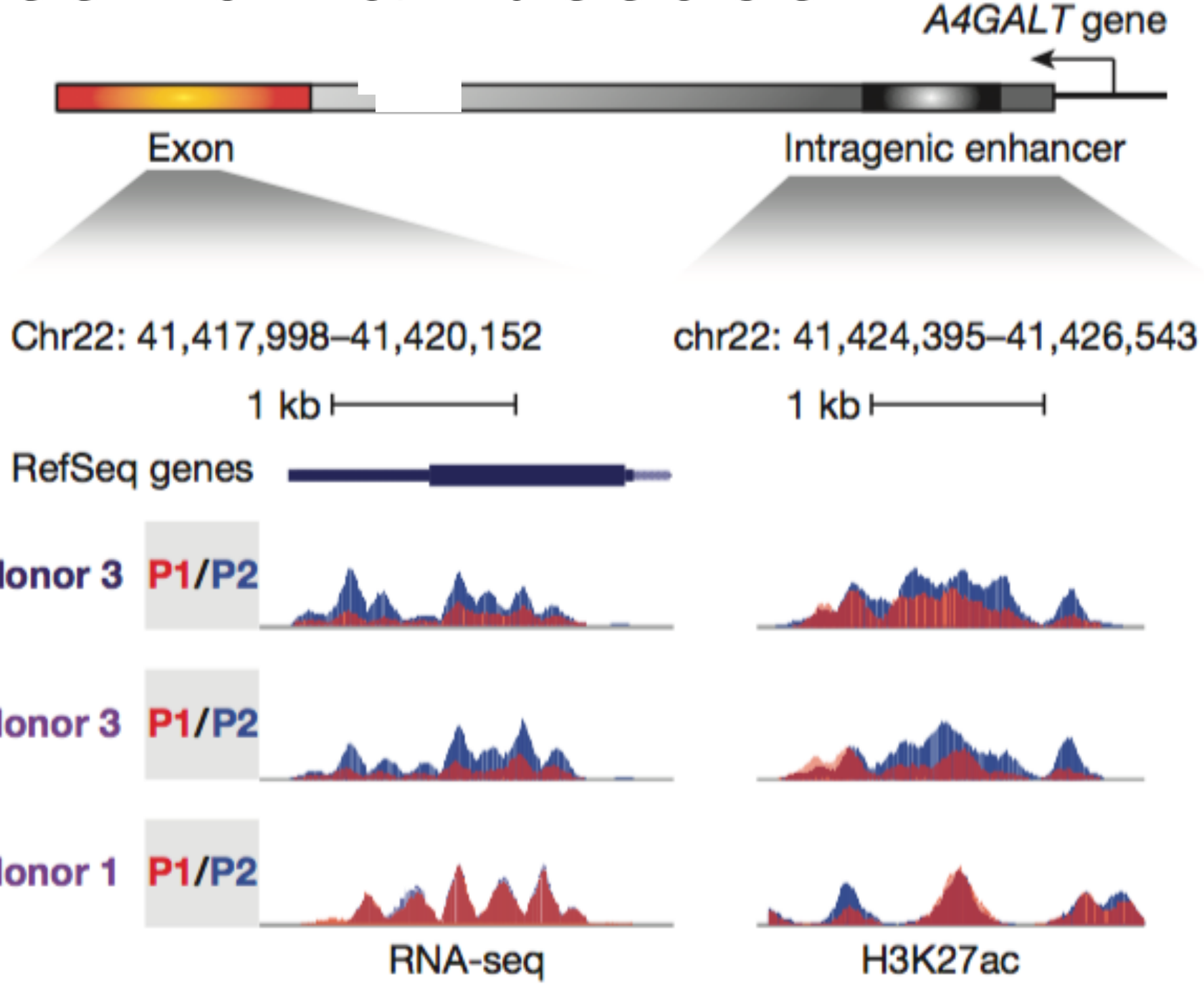
Allelic enhancers resided in closer proximity to genes with allelically biased expression compared to non-allelic enhancers



Leung et al., Nature 518, 350–354 (2015)

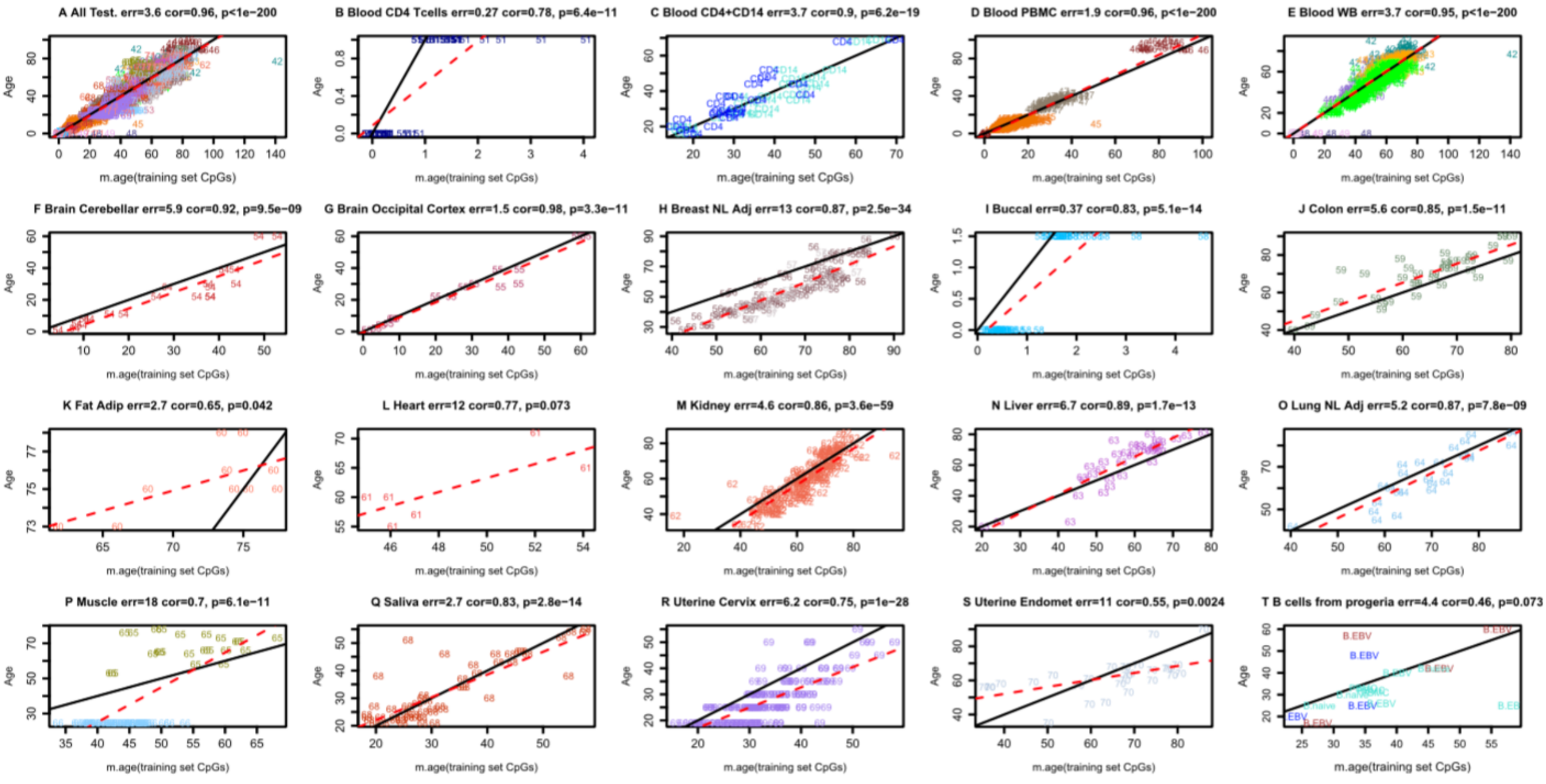
Haplotype-resolved epigenomes across human tissues

Allelic histone acetylation at enhancers associated with allelically biased gene expression.



Leung et al., Nature 518, 350–354 (2015)

Methylation and Age



Horvath, Genome Biology 14:R115 (2013)

Epigenetics and Aging

Environmental inputs	Effects on chromatin	Effects on healthspan and lifespan
Diet (dietary restriction)	• Modulation of chromatin modifiers • Heterochromatin maintenance • rDNA chromatin structure • Inhibition of recombination • Nucleosome positioning	Increased
Circadian cycle (regular)	Circadian epigenome	Increased
Circadian cycle (perturbed)	Modulation of chromatin modifiers	Decreased
Exercise	• Modulation of chromatin modifiers • Chromatin modifications	Increased
Pheromones	Signalling through chromatin modifiers	Increased
Systemic factors (sex steroid hormones)	• Chromatin structure • Chromatin modifications	Increased

Mechanistic link ?

Benayoun, Pollina & Brunet, Nature Reviews Molecular Cell Biology 16, 593–610 (2015)

ncRNA

Name	Size	Location	Number in humans	Functions	Illustrative examples
Short ncRNAs					
miRNAs	19–24 bp	Encoded at widespread locations	>1,424	Targeting of mRNAs and many others	miR-15/16, miR-124a, miR-34b/c, miR-200
piRNAs	26–31bp	Clusters, intragenic	23,439	Transposon repression, DNA methylation	piRNAs targeting RASGRF1 and LINE1 and IAP elements
tiRNAs	17–18bp	Downstream of TSSs	>5,000	Regulation of transcription?	Associated with the CAP1 gene

Esteller, Nature Reviews Genetics 12, 861-874 (2011)

ncRNA

Name	Size	Location	Number in humans	Functions	Illustrative examples
Mid-size ncRNAs					
snoRNAs	60–300 bp	Intronic	>300	rRNA modifications	U50, SNORD
PASRs	22–200 bp	5′ regions of protein-coding genes	>10,000	Unknown	Half of protein-coding genes
TSSa-RNAs	20–90 bp	–250 and +50 bp of TSSs	>10,000	Maintenance of transcription?	Associated with RNF12 and CCDC52 genes
PROMPTs	<200 bp	–205 bp and –5 kb of TSSs	Unknown	Activation of transcription?	Associated with EXT1 and RBM39 genes

Esteller, Nature Reviews Genetics 12, 861-874 (2011)

ncRNA

Name	Size	Location	Number in humans	Functions	Illustrative examples
Long ncRNAs					
lincRNAs	>200 bp	Widespread loci	>1,000	Examples include scaffold DNA–chromatin complexes	HOTAIR, HOTTIP, lincRNA-p21
T-UCRs	>200 bp	Widespread loci	>350	Regulation of miRNA and mRNA levels?	uc.283+, uc. 338, uc160+
Other lncRNAs	>200 bp	Widespread loci	>3,000	Examples include X-chromosome inactivation, telomere regulation, imprinting	XIST, TSIX, TERRAs, p15AS, H19, HYMAI

Esteller, Nature Reviews Genetics 12, 861-874 (2011)

small silencing RNA

Table 1 | **Types of small silencing RNAs**

Name	Organism	Length (nt)	Proteins	Source of trigger	Function
miRNA	Plants, algae, animals, viruses, protists	20–25	Drosha (animals only) and Dicer	Pol II transcription (pri-miRNAs)	Regulation of mRNA stability, translation
casiRNA	Plants	24	DCL3	Transposons, repeats	Chromatin modification
tasiRNA	Plants	21	DCL4	miRNA-cleaved RNAs from the *TAS* loci	Post-transcriptional regulation
natsiRNA	Plants	22	DCL1	Bidirectional transcripts induced by stress	Regulation of stress-response genes
		24	DCL2		
		21	DCL1 and DCL2		
Exo-siRNA	Animals, fungi, protists	~21	Dicer	Transgenic, viral or other exogenous dsRNA	Post-transcriptional regulation, antiviral defense
	Plants	21 and 24			
Endo-siRNA	Plants, algae, animals, fungi, protists	~21	Dicer (except secondary siRNAs in *C. elegans*, which are products of RdRP transcription, and are therefore not technically siRNAs)	Structured loci, convergent and bidirectional transcription, mRNAs paired to antisense pseudogene transcripts	Post-transcriptional regulation of transcripts and transposons; transcriptional gene silencing
piRNA	Metazoans excluding *Trichoplax adhaerens*	24–30	Dicer-independent	Long, primary transcripts?	Transposon regulation, unknown functions
piRNA-like (soma)	*Drosophila melanogaster*	24–30	Dicer-independent	In *ago2* mutants in *Drosophila*	Unknown
21U-RNA piRNAs	*Caenorhabditis elegans*	21	Dicer-independent	Individual transcription of each piRNA?	Transposon regulation, unknown functions
26G RNA	*Caenorhabditis elegans*	26	RdRP?	Enriched in sperm	Unknown

Ghildiyal and Zamore Nat. Rev. Genet. 10:94-108, (2009)

NGS DNA Methylation

Bisulfite Conversion

Step 1

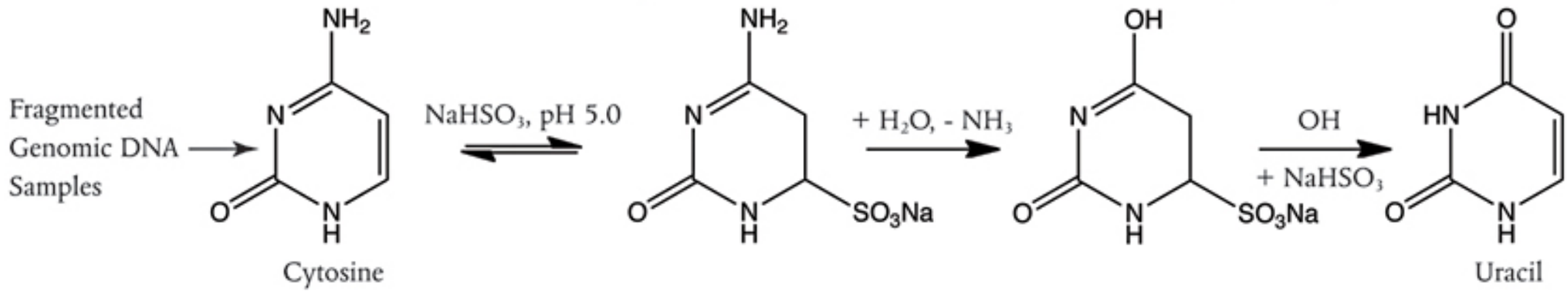
Denaturation
Incubation at 95°C
fragments genomic DNA

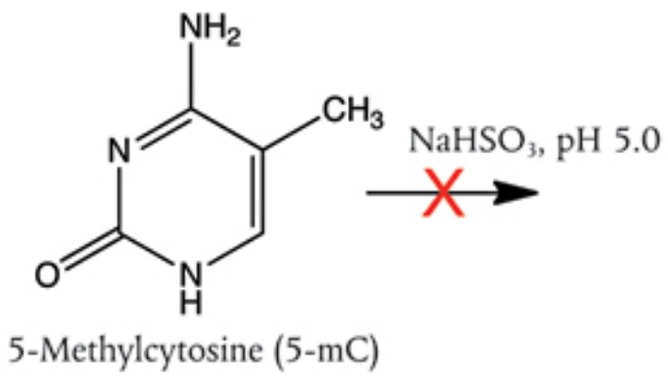
Step 2

Conversion
Incubation with sodium bisulfite
at 65°C and low pH (5-6)
deaminates cytosine residues
in fragmented DNA

Step 3

Desulphonation
Incubation at high pH
at room temperature for 15 min
removes the sulfite moeity,
generating uracil





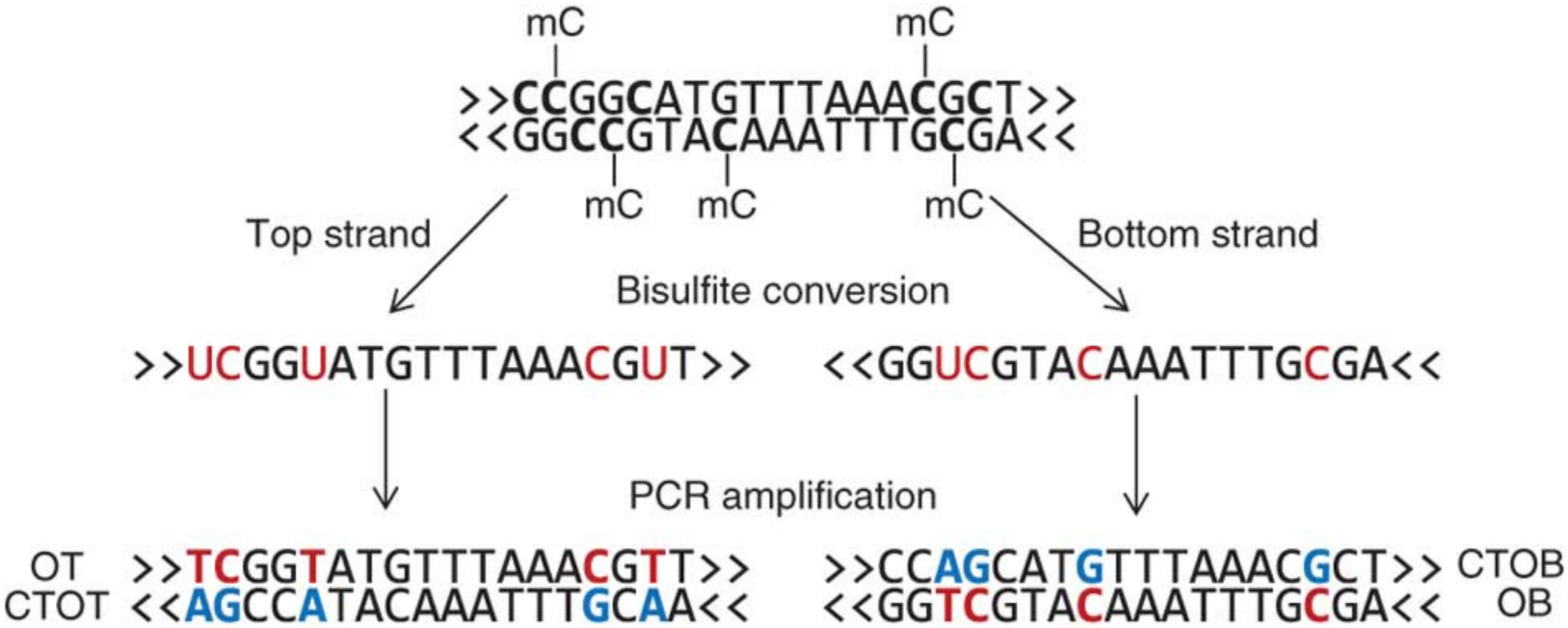
5-Methylcytosine (5-mC)

5-mC and 5-hmC (not shown) are not susceptible to bisulfite conversion and remain intact

neb.com

NGS DNA Methylation

Bisulfite Conversion

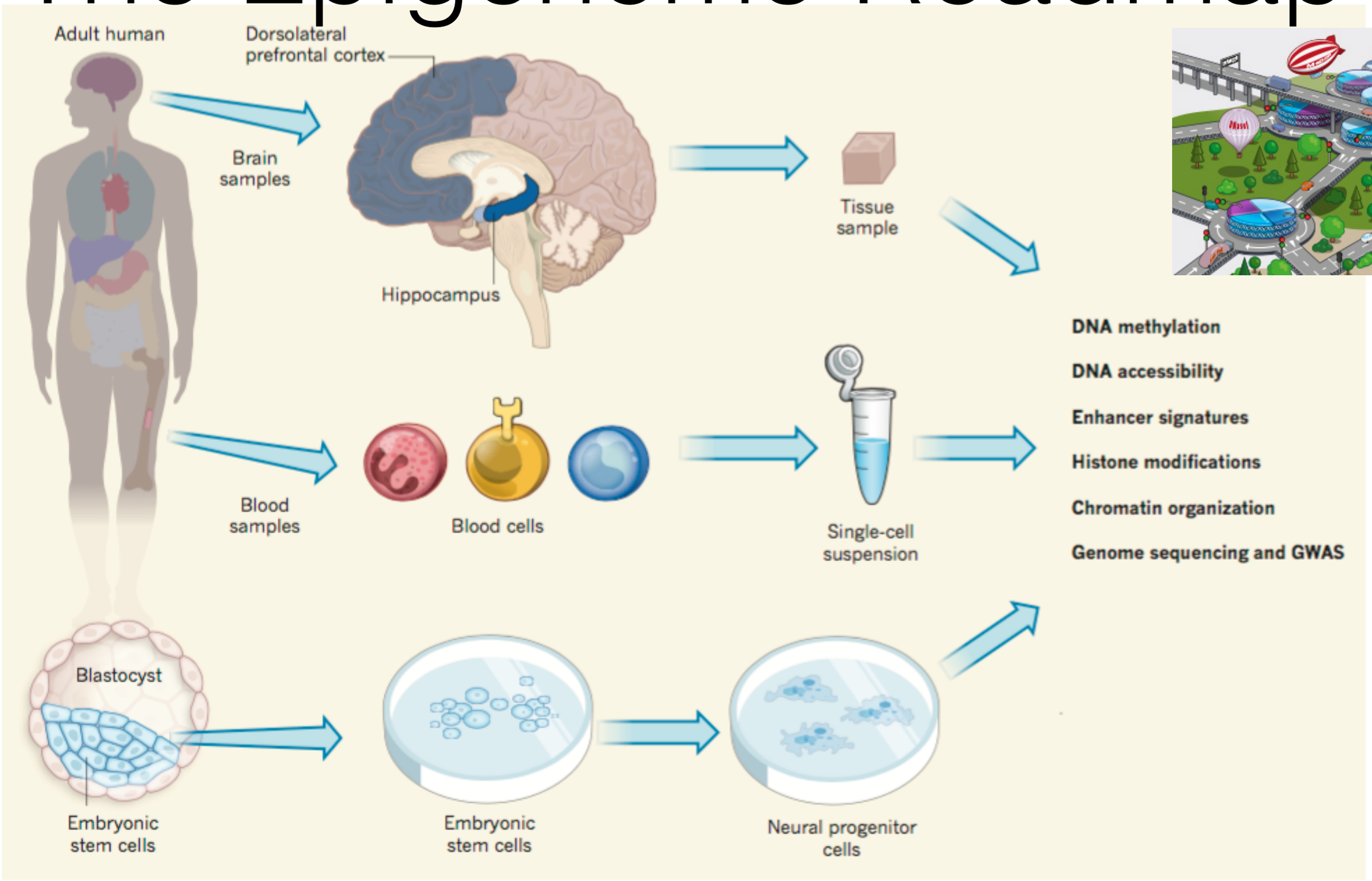


e.g. aligners: BSMAP, Bismark, MAQ, BS Seeker

Krueger et al., Nature Methods 9, 145–151 (2012)

OT, original top strand
CTOT, complementary to original top strand
OB, original bottom strand
CTOB, complementary to original bottom strand.

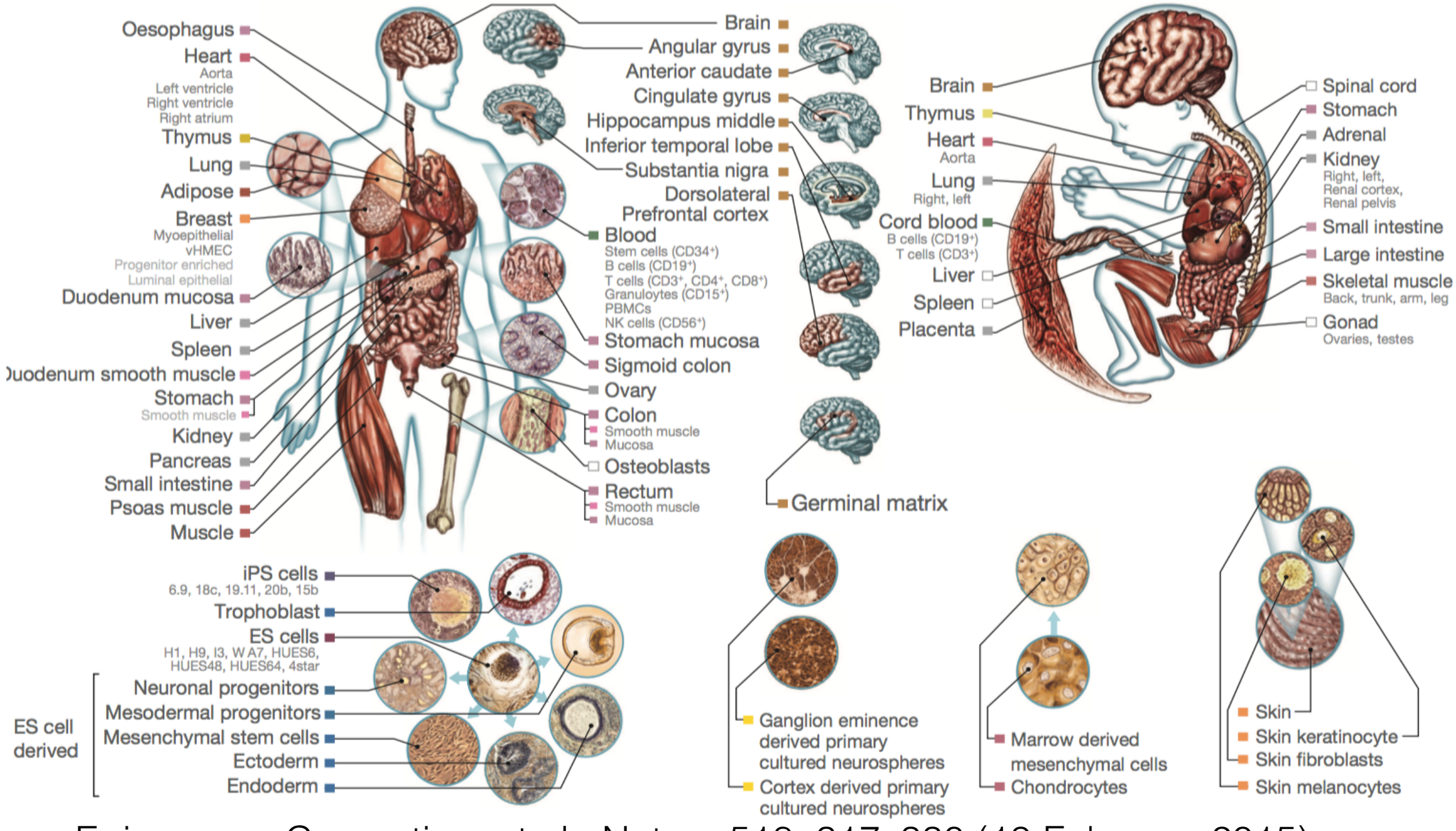
The Epigenome Roadmap



Romanoski et al., Nature 518, 314–316 (2015)
http://www.nature.com/collections/vbqgtr

Human Epigenome

NIH Roadmap Epigenomics Consortium generated the largest collection so far of human epigenomes for primary cells and tissues.



Human Epigenome Consortium et al., Nature 518, 317–330 (19 February 2015)