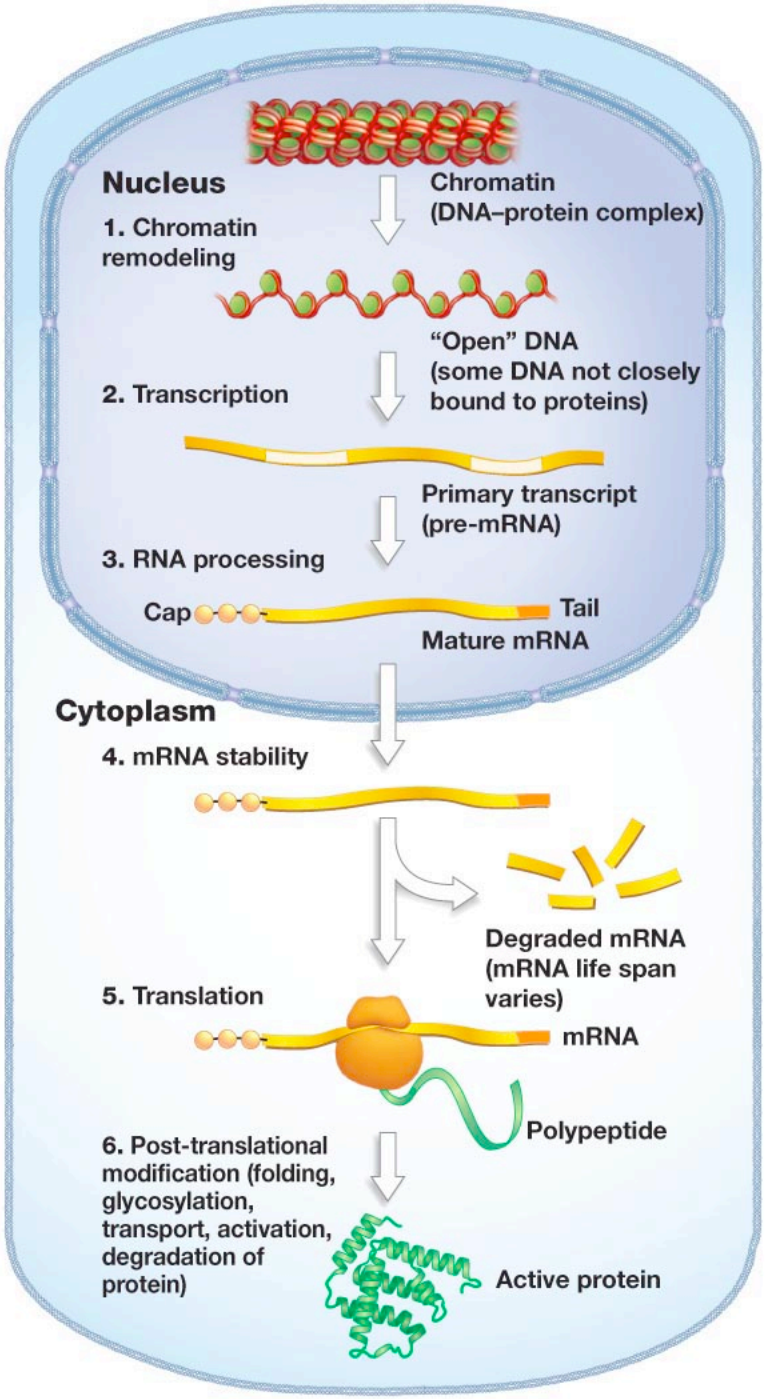
regulation of gene expression in eukaryotes

May 23, 2016

Levels of regulation



© 2011 Pearson Education, Inc.

transcription: chromatin state, interactions of *cis* and *trans* regulators with transcriptional machinery determine if/when/where genes are transcribed

post-transcription: alternative splicing, mRNA stability, RNA interference

translation: differential rates of translation initiation, elongation

post-translation: folding, chemical modifications of amino acids, protein sorting, protein degradation

Induction of transcriptional activity by environmental and biological factors

- heat shock
- hormones

book mentions only these
(comparable to prokaryotic induction)

- light
- development
- metabolism
- immunity
- growth/maintenance
- many other factors

but there are many more factors that affect transcription!

regulation of gene expression in eukaryotes

- overview
- transcriptional regulation
  - *cis* regulatory sequences and *trans* acting factors
  - chromatin organization
- post-transcriptional regulation
- regulation of whole chromosomes

interaction of *cis*-regulatory sequences and *trans*-acting proteins

- *trans*-acting factors-- transcription factors (TFs) are regulatory proteins that:
  - have DNA binding domains
  - bind at transcription factor binding sites in *cis*-regulatory sequences
- *cis*-regulatory sequences come in different types with different properties:
  - promoter
  - enhancers / silencers
- Transcription factors bind *cis*-regulatory sequences to control gene expression

transcription factors

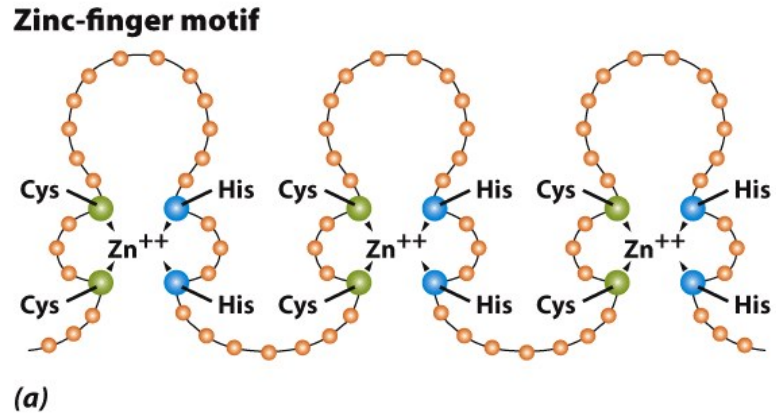
- regulatory proteins that bind to *cis*-regulatory sequences such as enhancers
- TFs that repress gene expression are called repressors
- TFs that promote gene expression are called activators
- interact directly or indirectly with RNA Pol II and associated TFs at the promoter to affect gene expression
- can also interact with other TFs bound at other enhancers

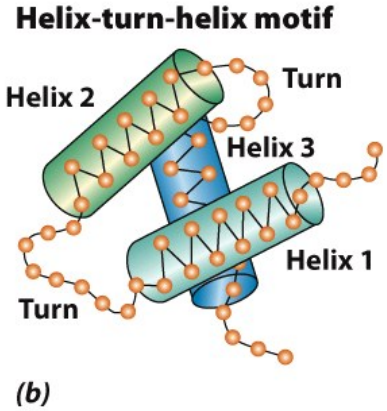
Transcription factors have different protein domains:

1) Structural motifs bind DNA – will bind specific sequences in DNA

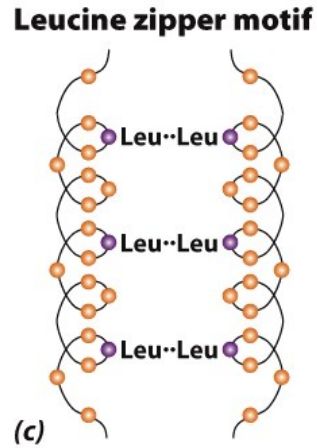
2) Transcriptional activation domain (not pictured) interacts with other TFs/ RNA Pol II at promoter

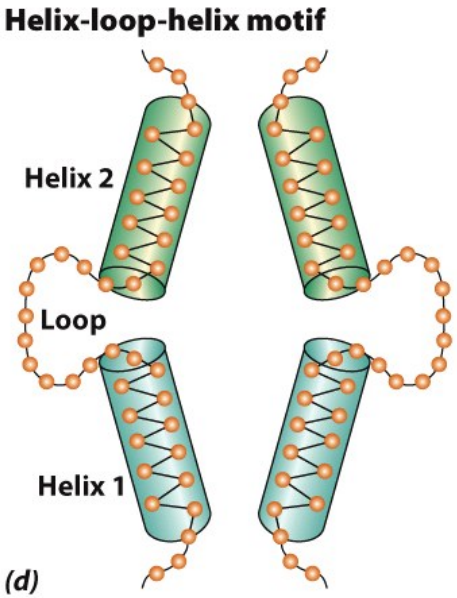
These domains can be in different parts of protein, or be overlapping





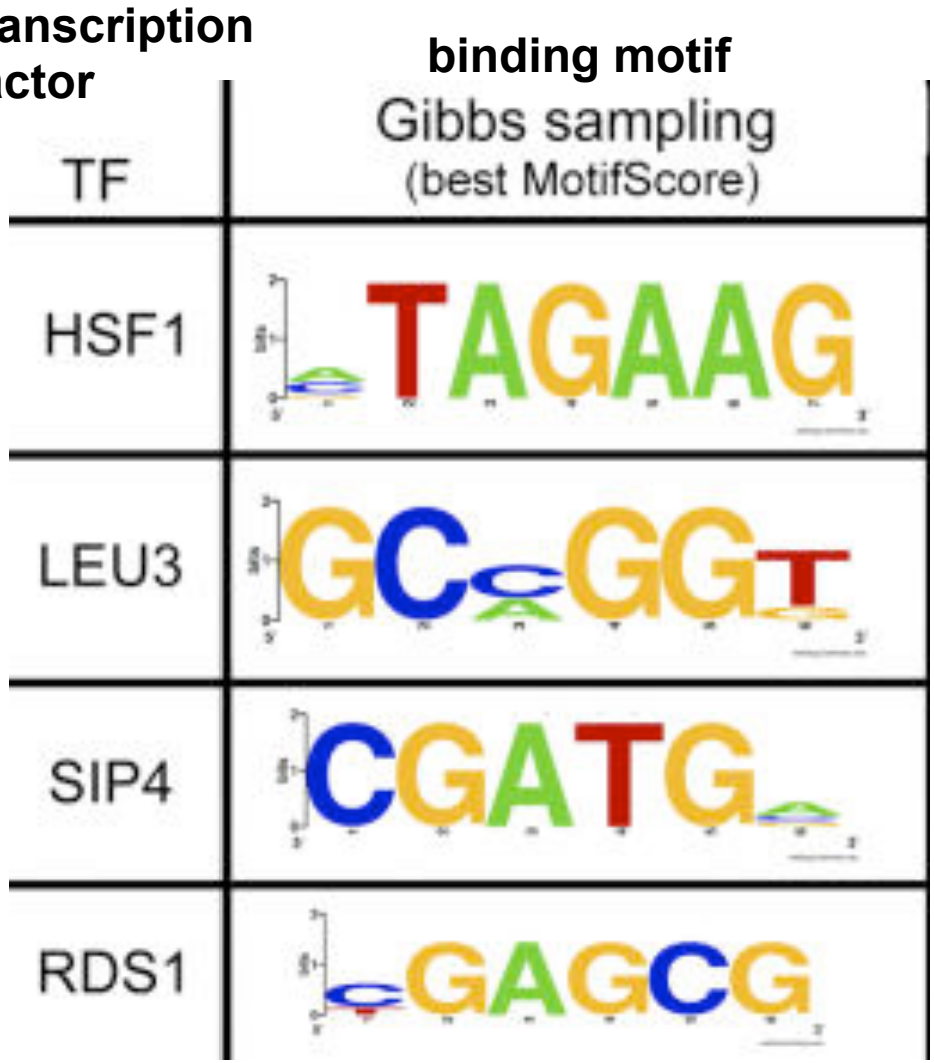
Examples of structural domains in TFs





© 2012 John Wiley & Sons, Inc. All rights reserved.

Transcription factors bind specific DNA sequences in *cis*-regulatory regions



The short sequences where TFs bind are known as **transcription factor binding motifs**

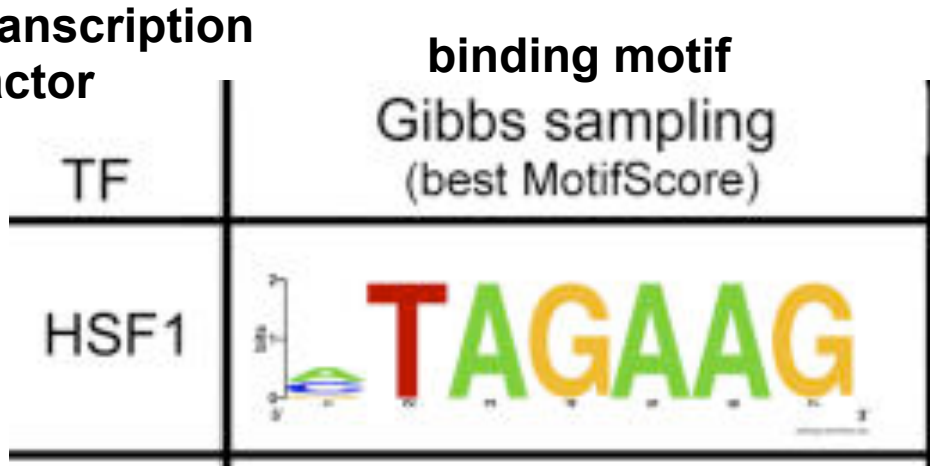
The places in the genome where TFs bind are called **transcription factor binding sites**

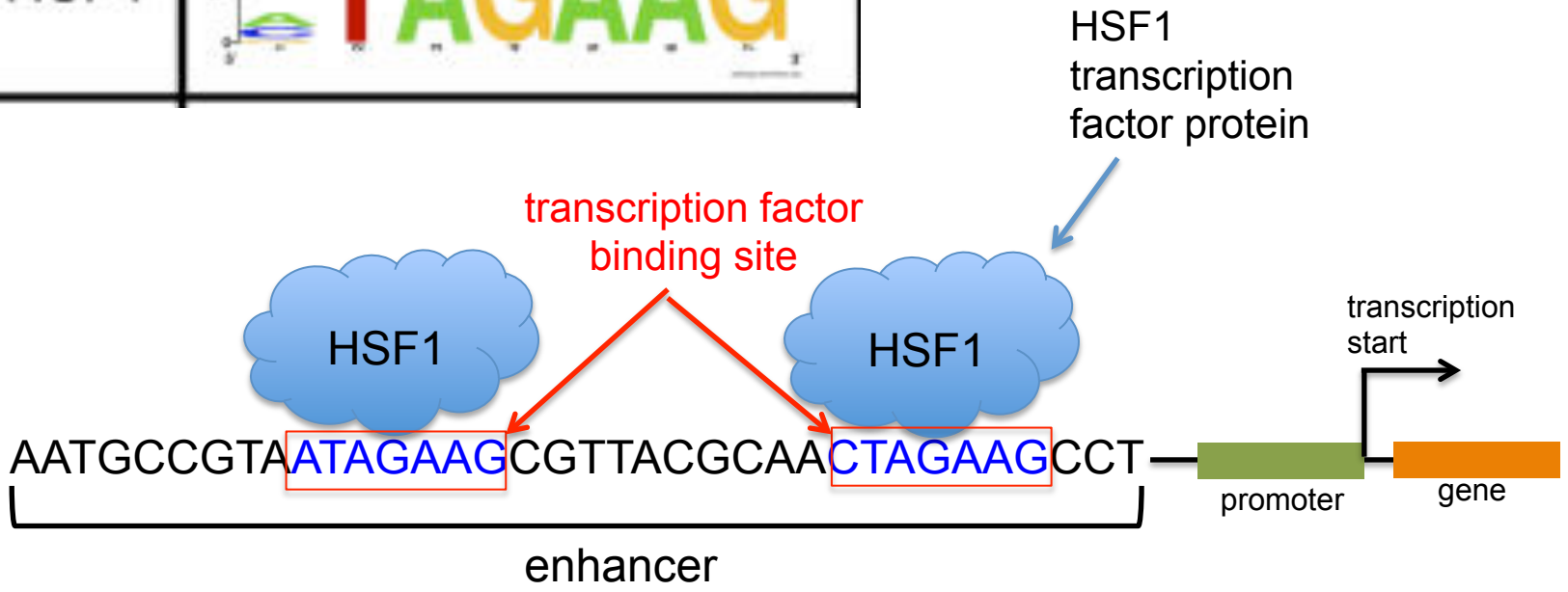
transcription factor binding sites are located in cis-regulatory elements such as enhancers/silencers

This is because of specific DNA sequence in the **DNA binding domain** of the TF

Reddy, et al. (2007) *PLOS Comp. Bio*

Transcription factors bind specific DNA sequences in *cis*-regulatory regions





the HSF1 protein can bind to HSF1 transcription factor binding sites in the enhancer of a gene it will regulate

interaction of *cis*-regulatory sequences and *trans*-acting proteins

- *trans*- acting factors: transcription factors (TFs)
  - DNA binding domains
  - transcription factor binding sites

- types of *cis*-regulatory sequences
  - promoter
  - enhancers / silencers

- these proteins and sequences work together to control gene expression

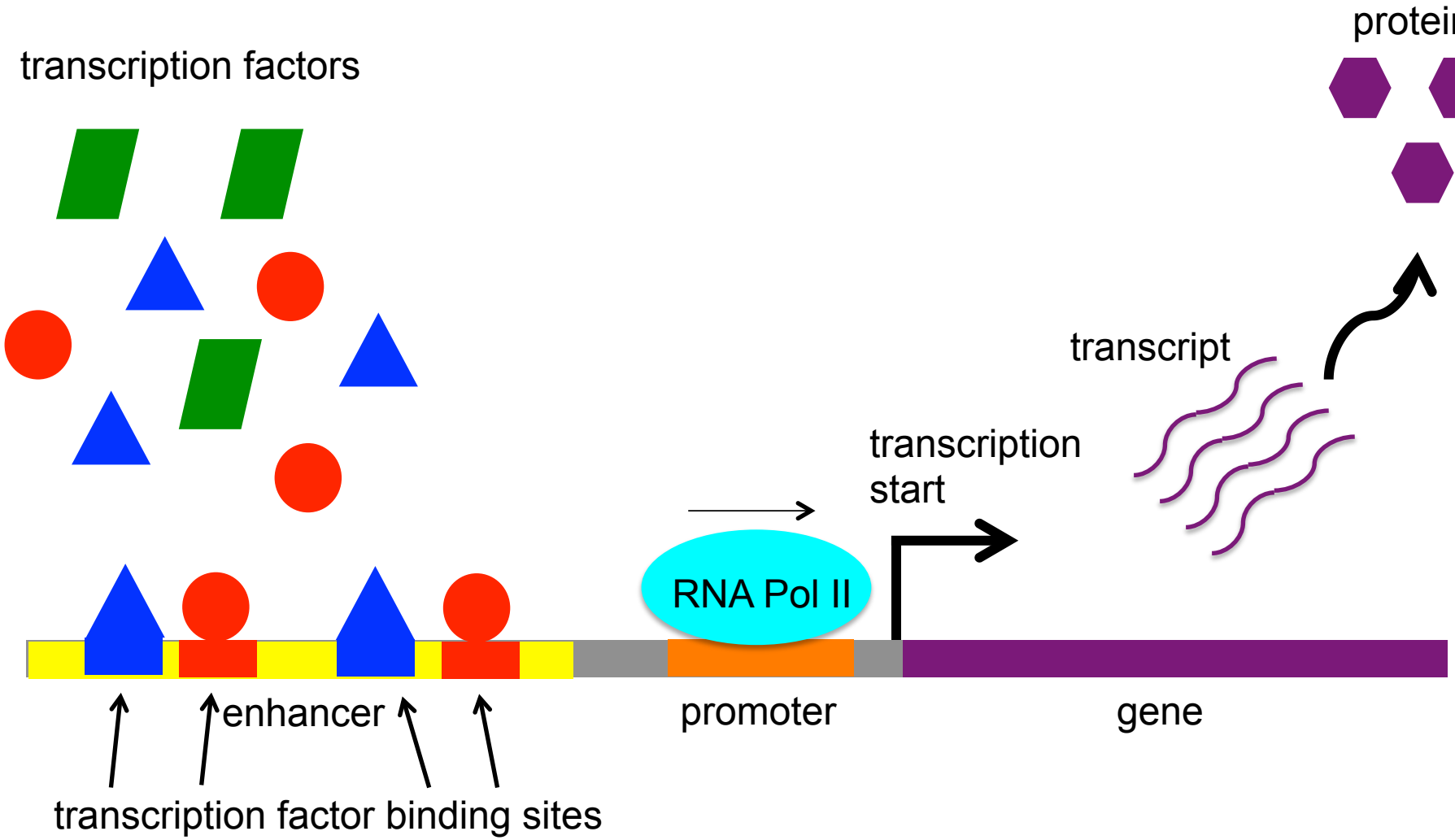
cis-regulatory sequences are DNA sequences involved in regulation of transcription

- promoter
  - right next to transcription start site
  - contains TATA box and other sequences that recruit DNA Pol II

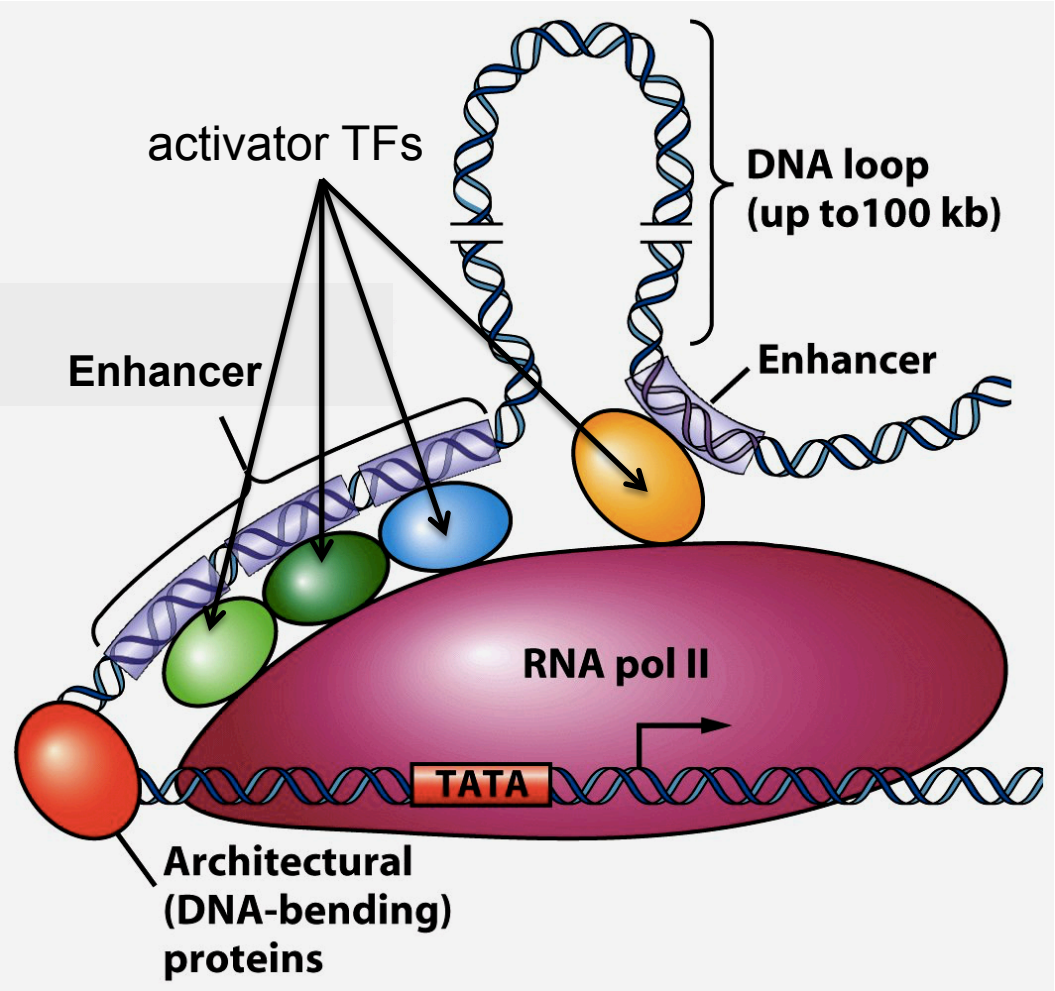
- enhancers/silencers
  - further away from transcription start than promoter, can act from longer distances
  - binds TFs that interact with proteins bound to promoter
  - can be upstream, downstream, in introns
  - can function even when flipped in orientation (5' to 3')

enhancers/silencers are DNA sequences involved in gene regulation

- enhancers promote transcription
- silencers repress transcription
- Some sequences can bind both activator and repressor TFs, and can either promote or repress transcription based on which TFs are bound. Confusingly, these sequences are still called enhancers.
- transcription factors (TFs) bind TF binding sites in enhancers/ silencers
- TFs bound at enhancers/silencers interact, directly or indirectly, with RNA Pol II and associated TFs to affect transcription
- different *cis*-regulatory sequences can affect transcription in different tissues or under different conditions


protein
transcription factors
transcript
transcription
start
RNA Pol II
enhancer
promoter
gene
transcription factor binding sites

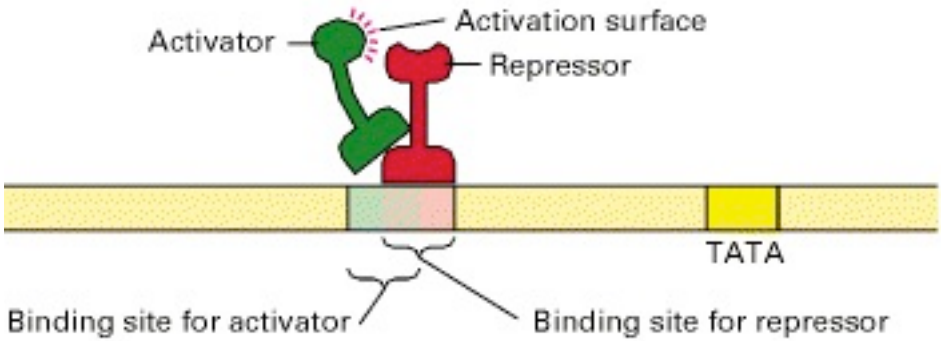
activators bound to enhancers can promote transcription by stabilizing transcription initiation



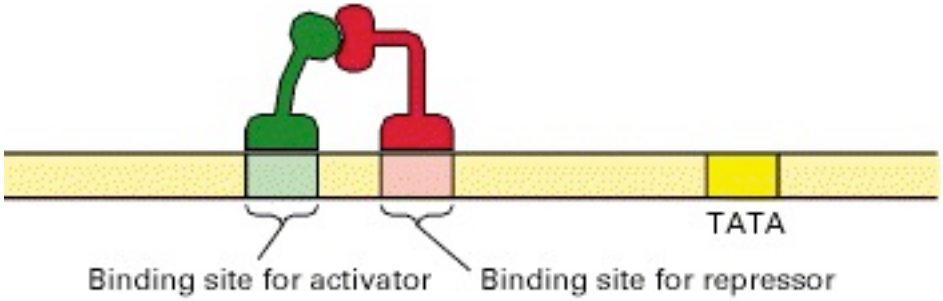
Activators bound to enhancers stabilize the RNA Pol II complex at the TATA box, activating transcription

How do repressors bound to *cis*-regulatory sequences repress?

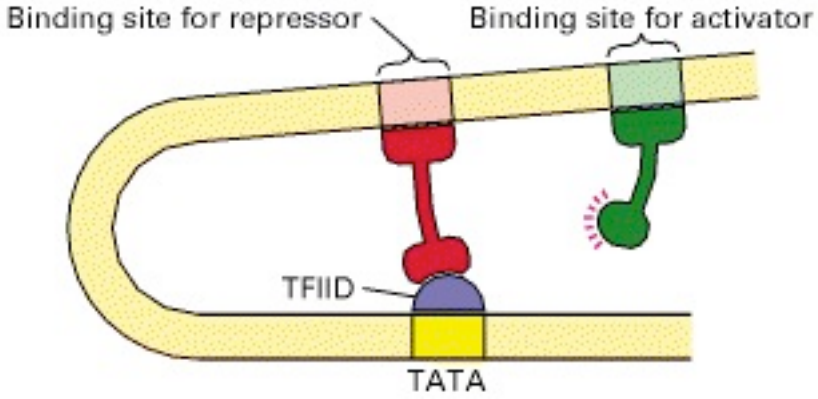
Competitive binding with activator TF



Interaction with bound activator TF



Interaction with RNA Pol or its initiation TFs



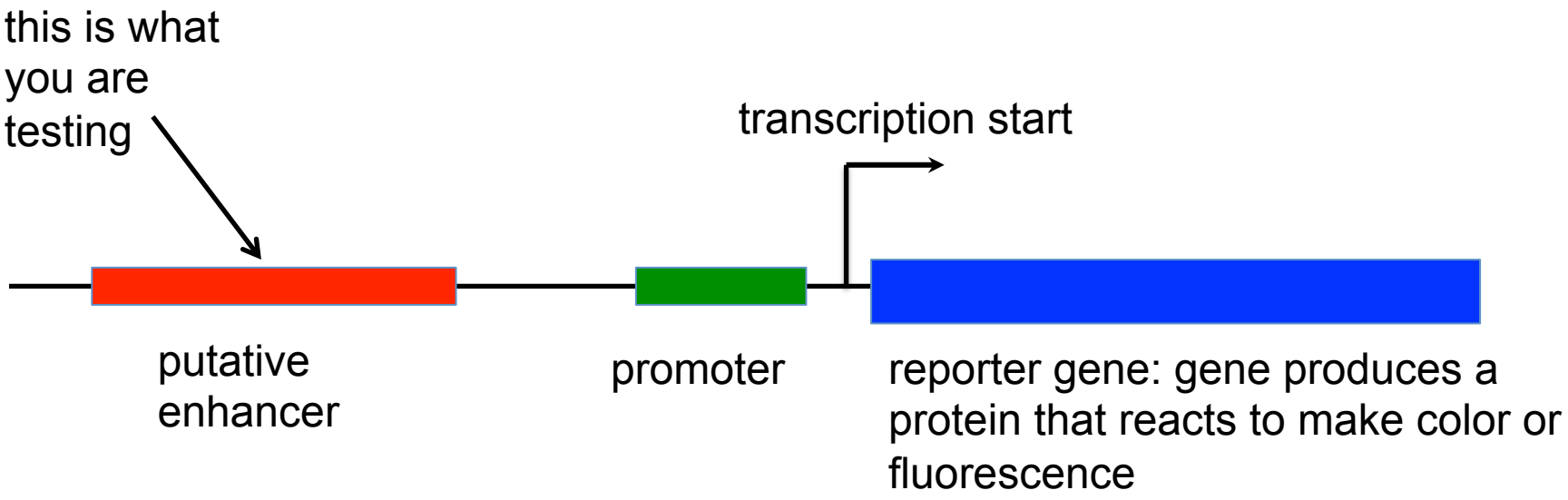
Lodish et al, Molecular Cell Biology, 4th ed. 2000

how to find enhancers? look for:

- computational methods
  - evolutionary conservation of non-coding regions
  - clusters of binding sites for transcription factors

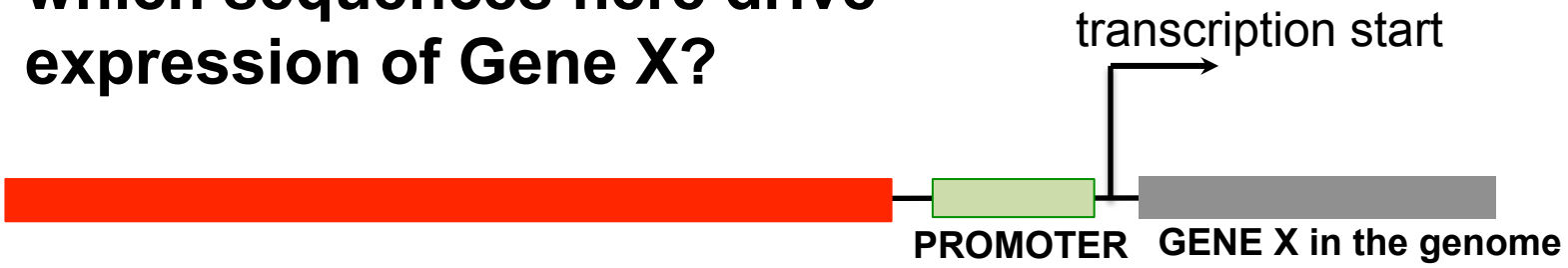
- experimental methods
  - expression phenotypes associated with mutations in particular non-coding regions
  - where TFs bind
  - chromatin marks/state associated with active transcription

use of reporter genes to find enhancers



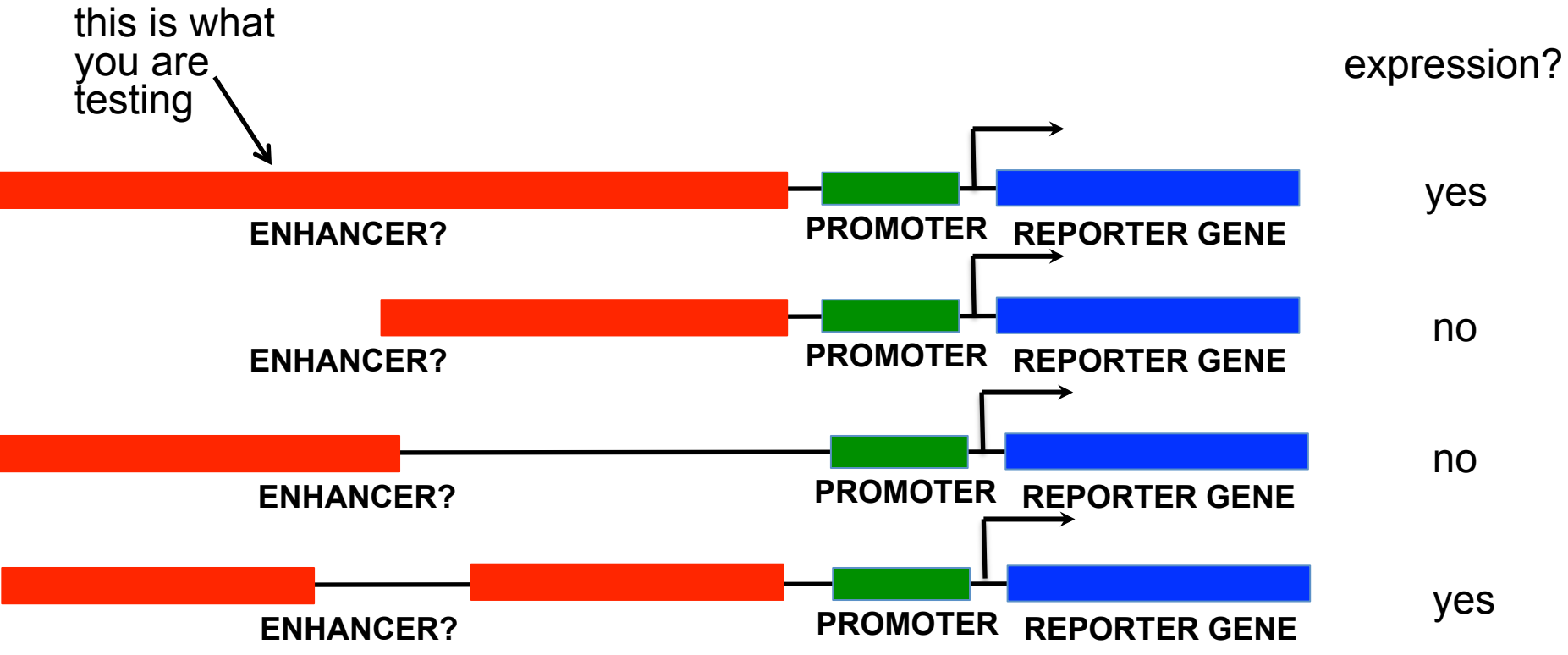
- make a DNA construct (fusion gene) with a functional promoter upstream of a reporter gene
- this reporter gene will make a protein that you can see (color or fluorescence)
- put a DNA sequence that you think might be an enhancer upstream of the promoter + reporter gene
- does this DNA sequence drive transcription? Do you see the product of the reporter gene?

which sequences here drive expression of Gene X?



make reporter constructs with different pieces of non-coding DNA from nearby Gene X

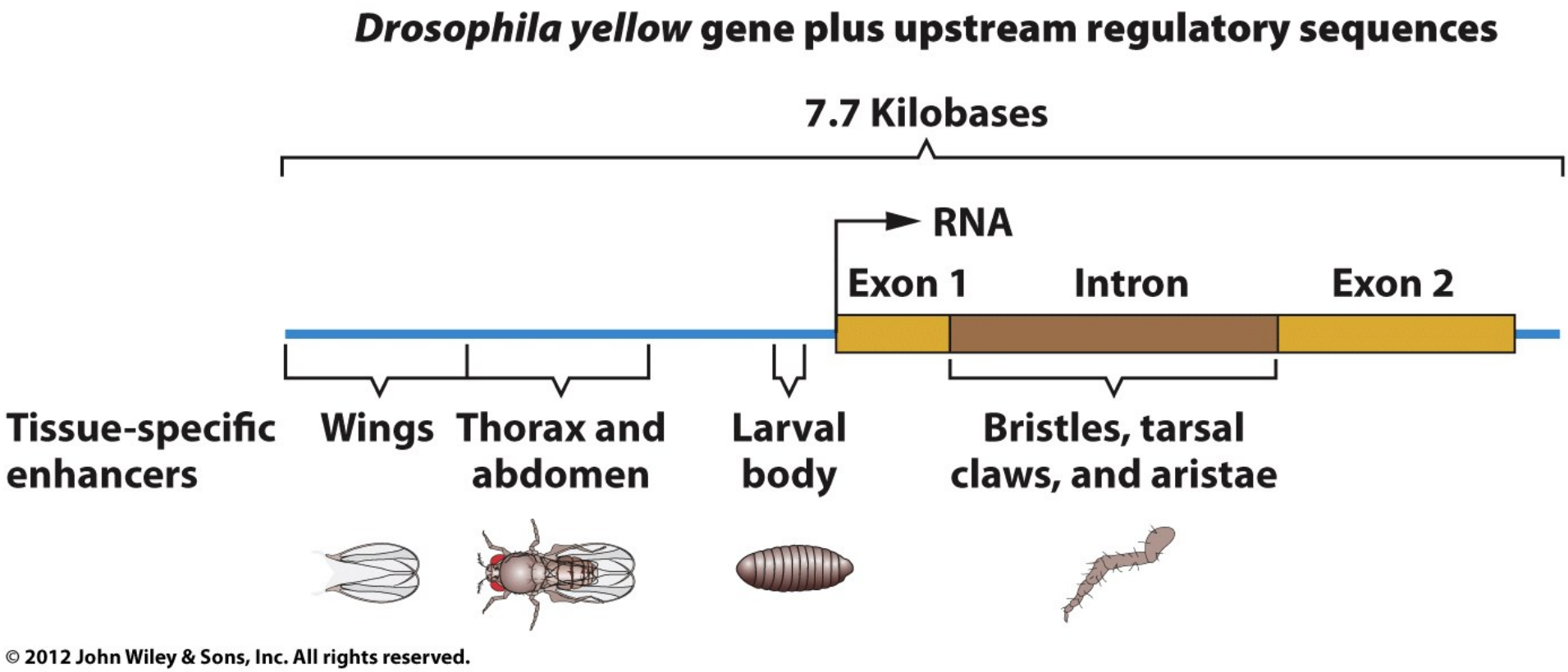
do they drive expression?



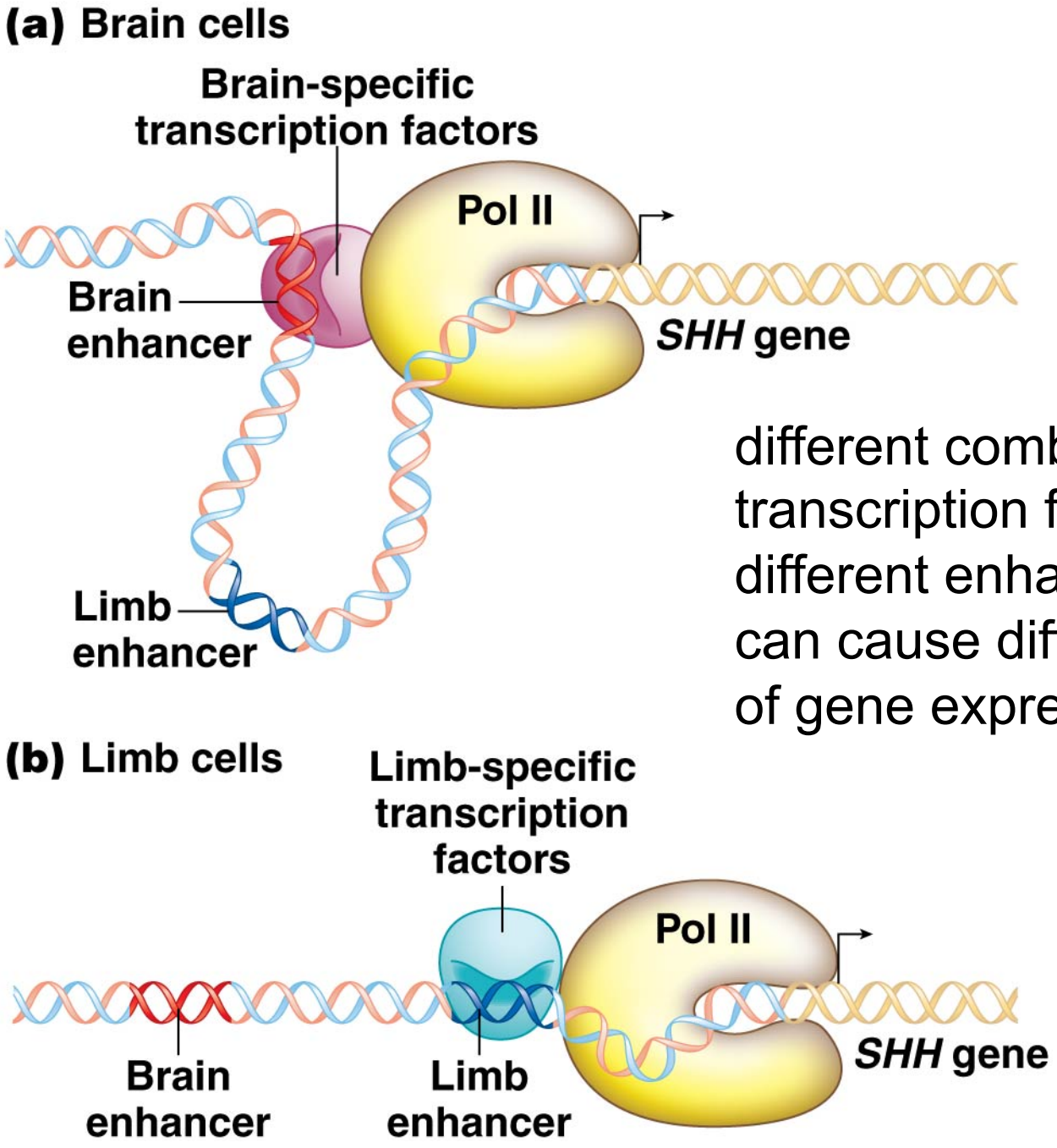
enhancer-TF interactions can be differentially regulated

- different enhancers are used in different tissues, or at different times in development, or under different conditions
- different transcription factors are present in different tissues/times/conditions
- this produces differential gene expression!

Modular tissue-specific enhancers of the *Drosophila yellow* gene



© 2012 John Wiley & Sons, Inc. All rights reserved.



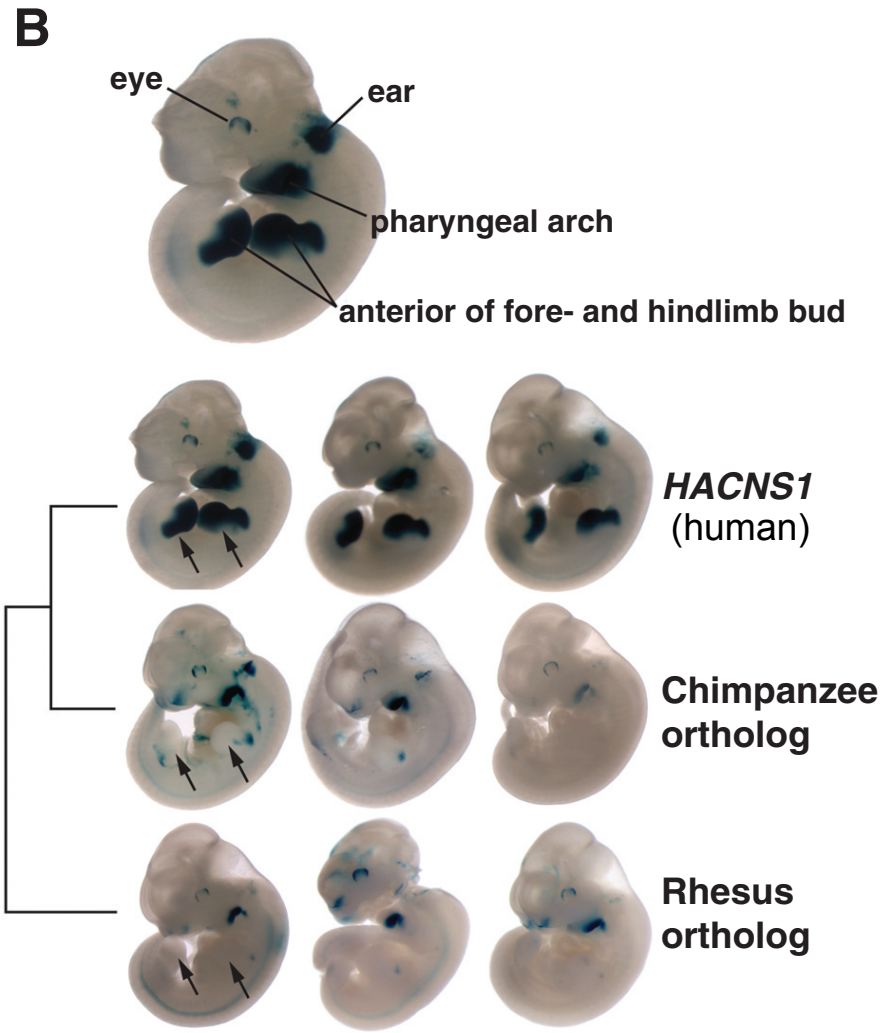
different combinations of transcription factors binding to different enhancers/silencers can cause differential regulation of gene expression

Human-Specific Gain of Function in a Developmental Enhancer

Shyam Prabhakar,[1]* Axel Visel,[1] Jennifer A. Akiyama,[1] Malak Shoukry,[1] Keith D. Lewis,[1]† Amy Holt,[1] Ingrid Plajzer-Frick,[1] Harris Morrison,[2] David R. FitzPatrick,[2] Veena Afzal,[1] Len A. Pennacchio,[1,3] Edward M. Rubin,[1,3]‡ James P. Noonan[1]‡§

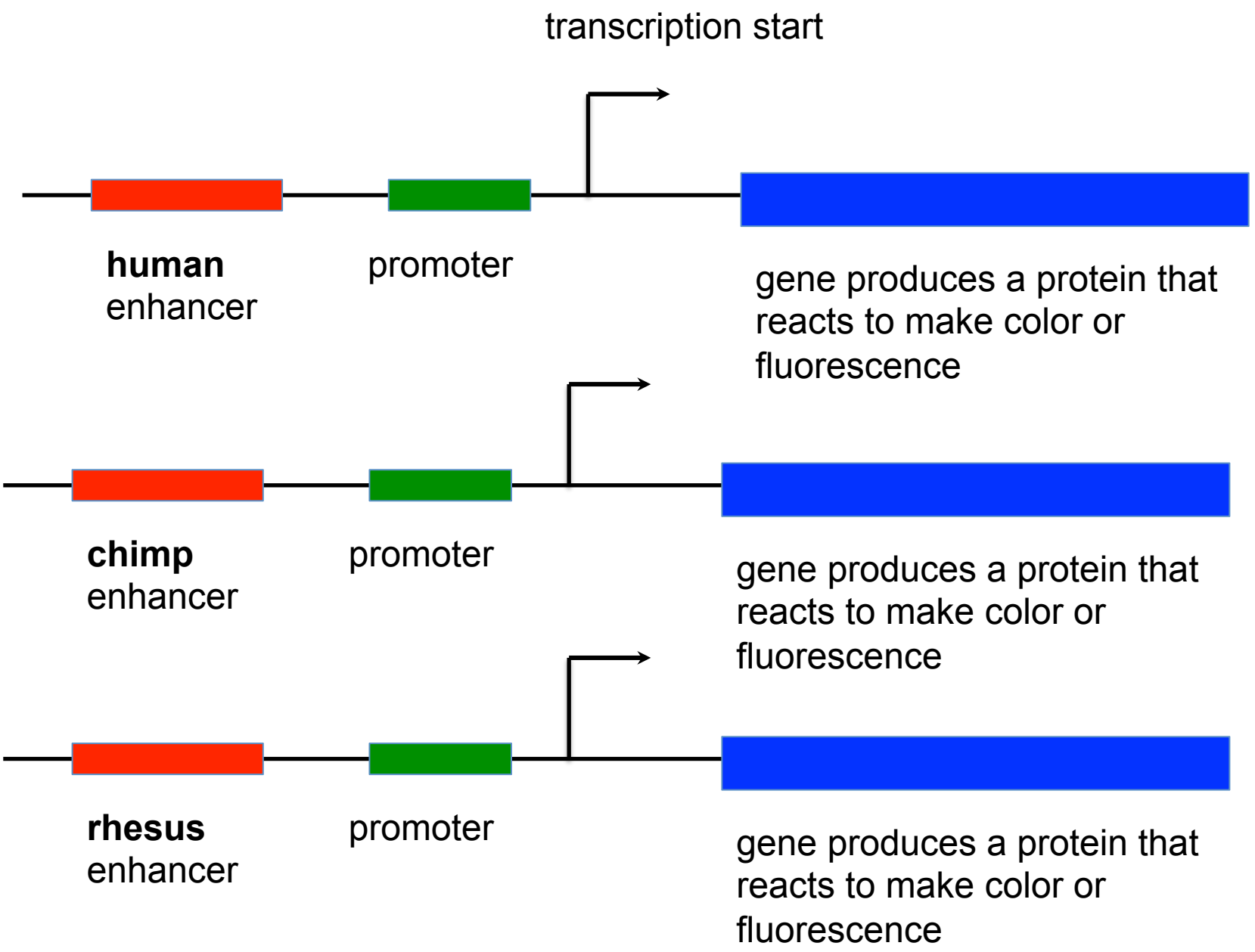
5 SEPTEMBER 2008 VOL 321 SCIENCE

another use of reporter genes: in what tissues are enhancers driving gene expression?



find the fastest evolving non-coding region between human and chimpanzee – is it an enhancer?

place this piece of DNA from this region in human, chimp, and rhesus macaque in reporter genes

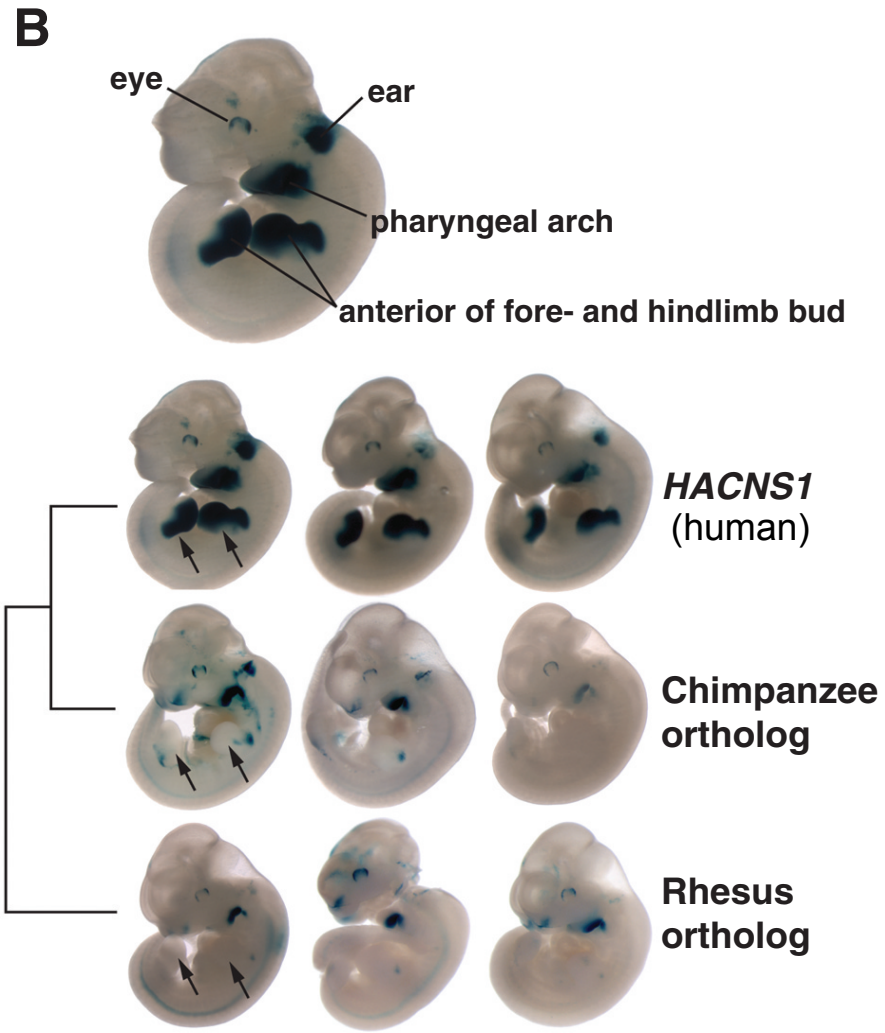


Where in the body are these reporter genes expressed?

Human-Specific Gain of Function in a Developmental Enhancer

Shyam Prabhakar,[1]* Axel Visel,[1] Jennifer A. Akiyama,[1] Malak Shoukry,[1] Keith D. Lewis,[1]†
Amy Holt,[1] Ingrid Plajzer-Frick,[1] Harris Morrison,[2] David R. FitzPatrick,[2] Veena Afzal,[1]
Len A. Pennacchio,[1,3] Edward M. Rubin,[1,3]‡ James P. Noonan[1]‡§

5 SEPTEMBER 2008 VOL 321 **SCIENCE**



find the fastest evolving non-coding region between human and chimpanzee

place the piece of DNA from this region in human, chimp, and rhesus macaque in reporter genes

make transgenic mice expressing those reporter genes

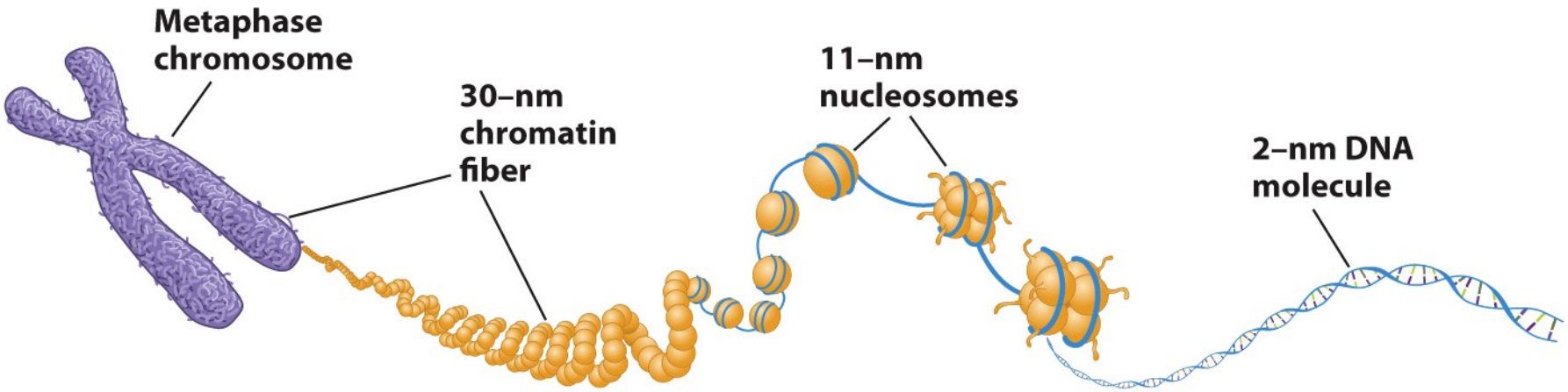
determine where the gene affected by this enhancer is expressed

only the human enhancer causes expression in the developing thumb

regulation of gene expression in eukaryotes

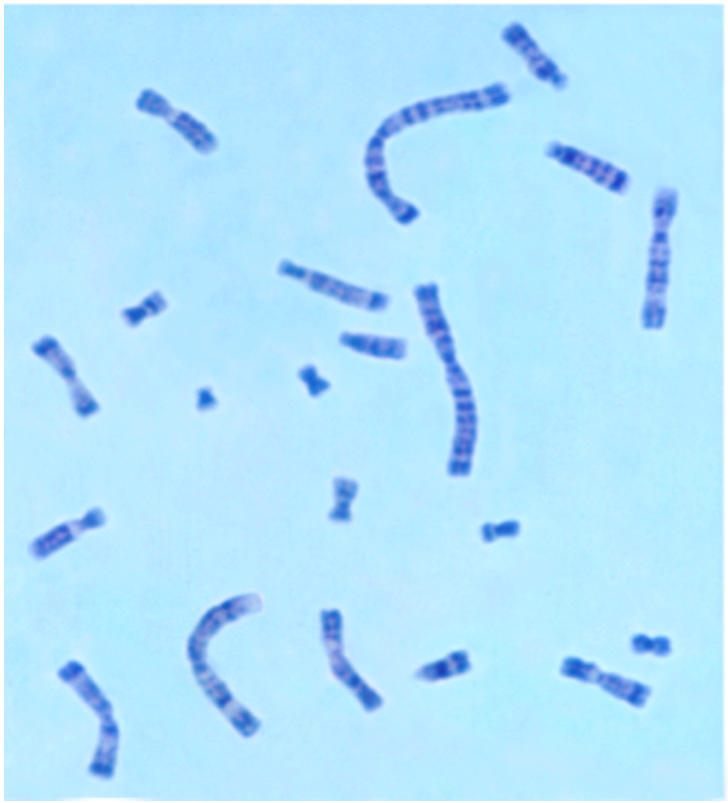
- overview
- transcriptional regulation
  - *cis* regulatory sequences and *trans* acting factors
  - chromatin organization
- post-transcriptional regulation
  - factors that affect mRNA stability
  - RNA interference
- regulation of whole chromosomes

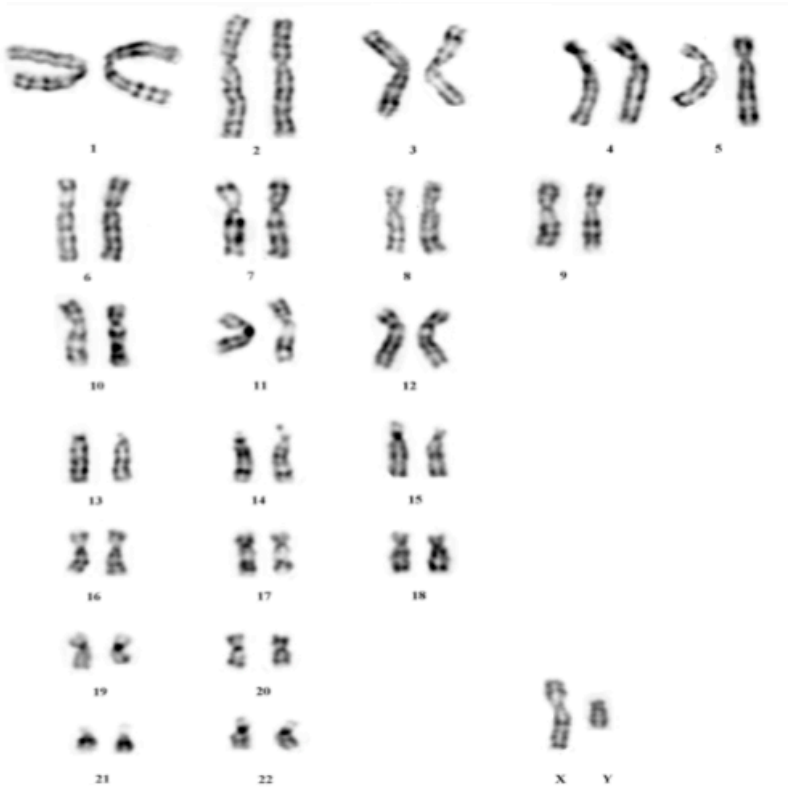
organization of eukaryotic DNA



After Figure 1 in The ENCODE Project Consortium. *Science* 306:636–640, Oct. 22, 2004

chromatin state varies across the chromosome



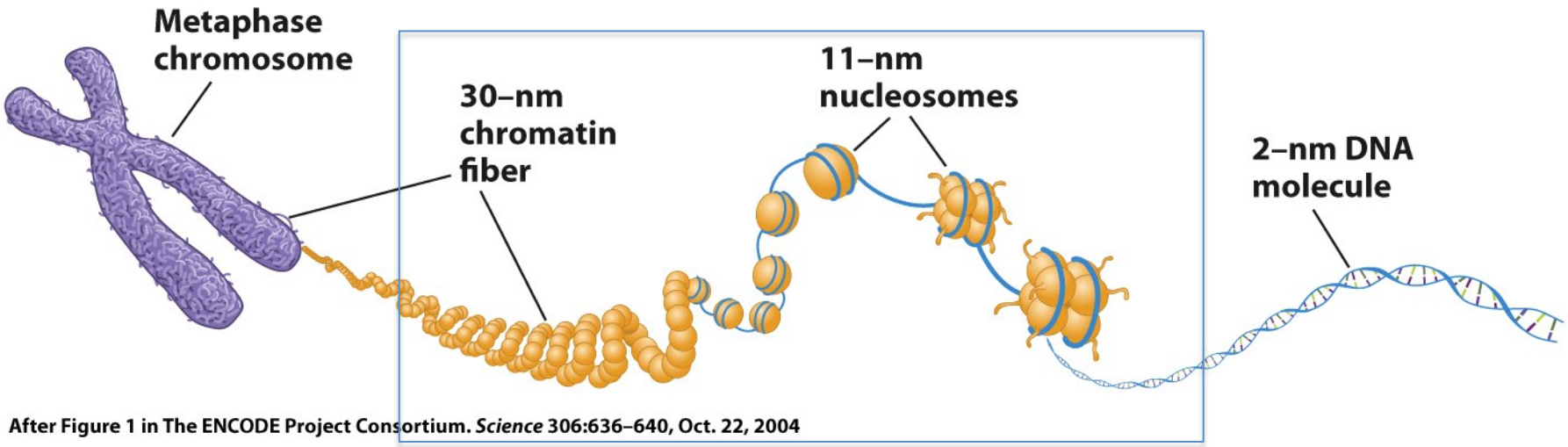


- G-banded CHO9 metaphase (left) and a male human peripheral lymphocyte karyotype exhibiting G-bands (right).

  - dark bands correspond to heterochromatin
  - light bands correspond to euchromatin

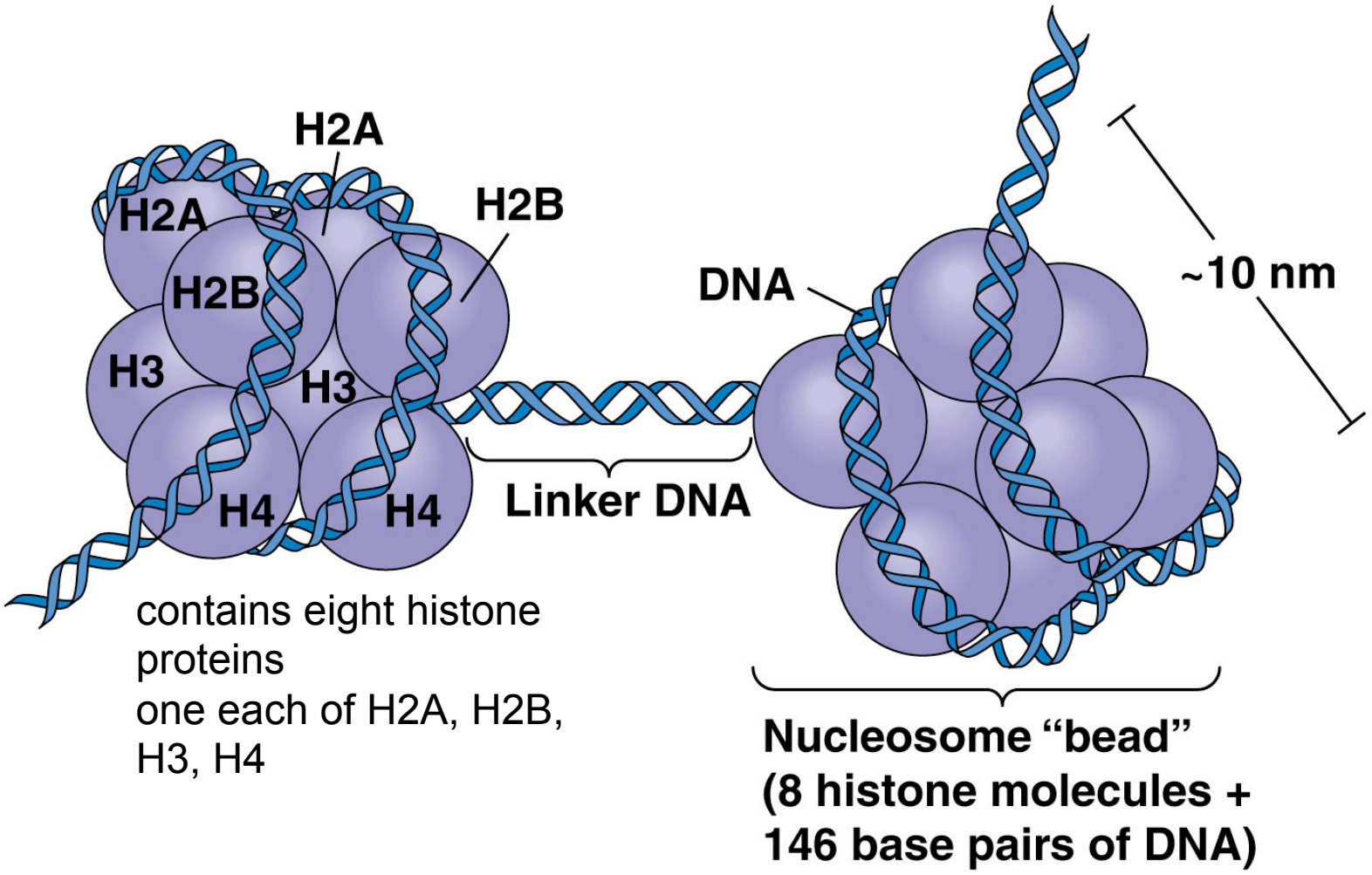
Di Tomaso, et al, 2013

organization of eukaryotic DNA



After Figure 1 in The ENCODE Project Consortium. *Science* 306:636–640, Oct. 22, 2004

nucleosomes are histone proteins + DNA

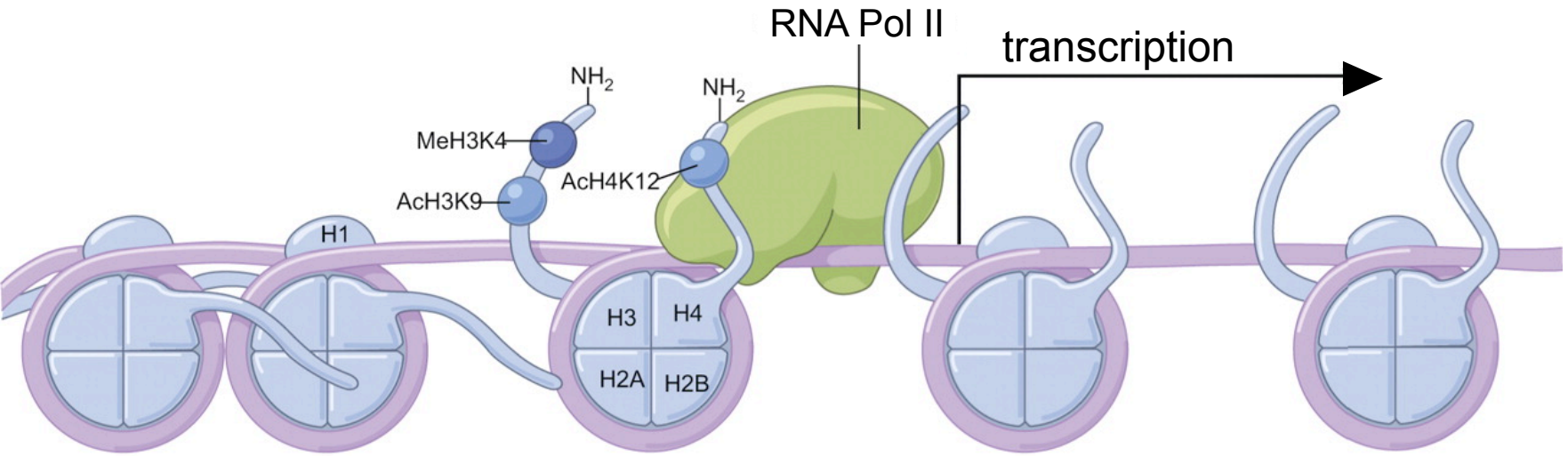


contains eight histone proteins
one each of H2A, H2B, H3, H4

© 2012 Pearson Education, Inc.

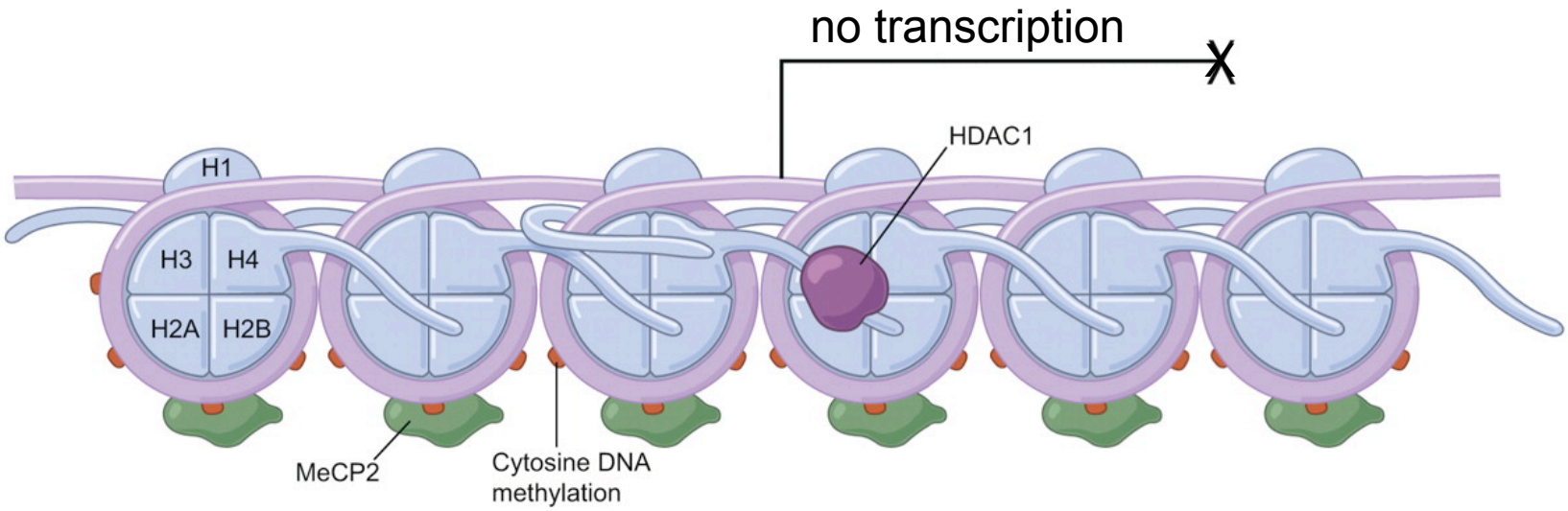
Open chromatin = transcription can occur

A eNOS promoter in vascular endothelial cells



Closed chromatin = transcription unable to occur

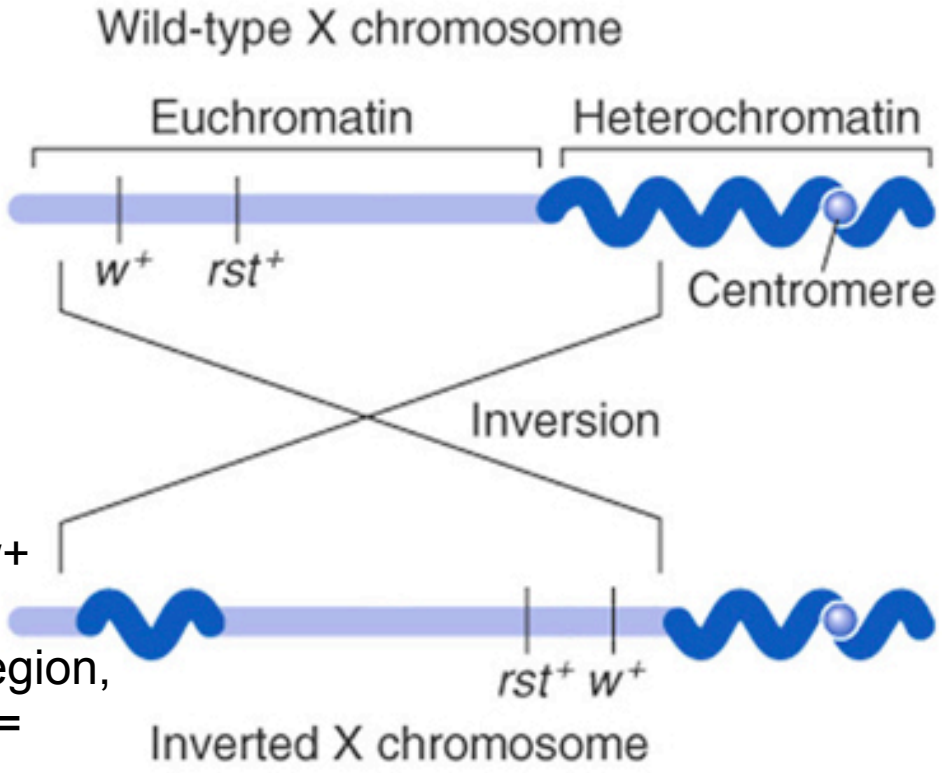
B eNOS promoter in vascular smooth muscle cells

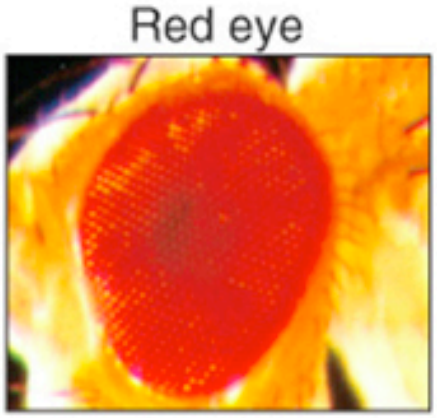


nucleosome density confirms chromatin state

Chromatin state effects gene expression

w+ gene = red eyes

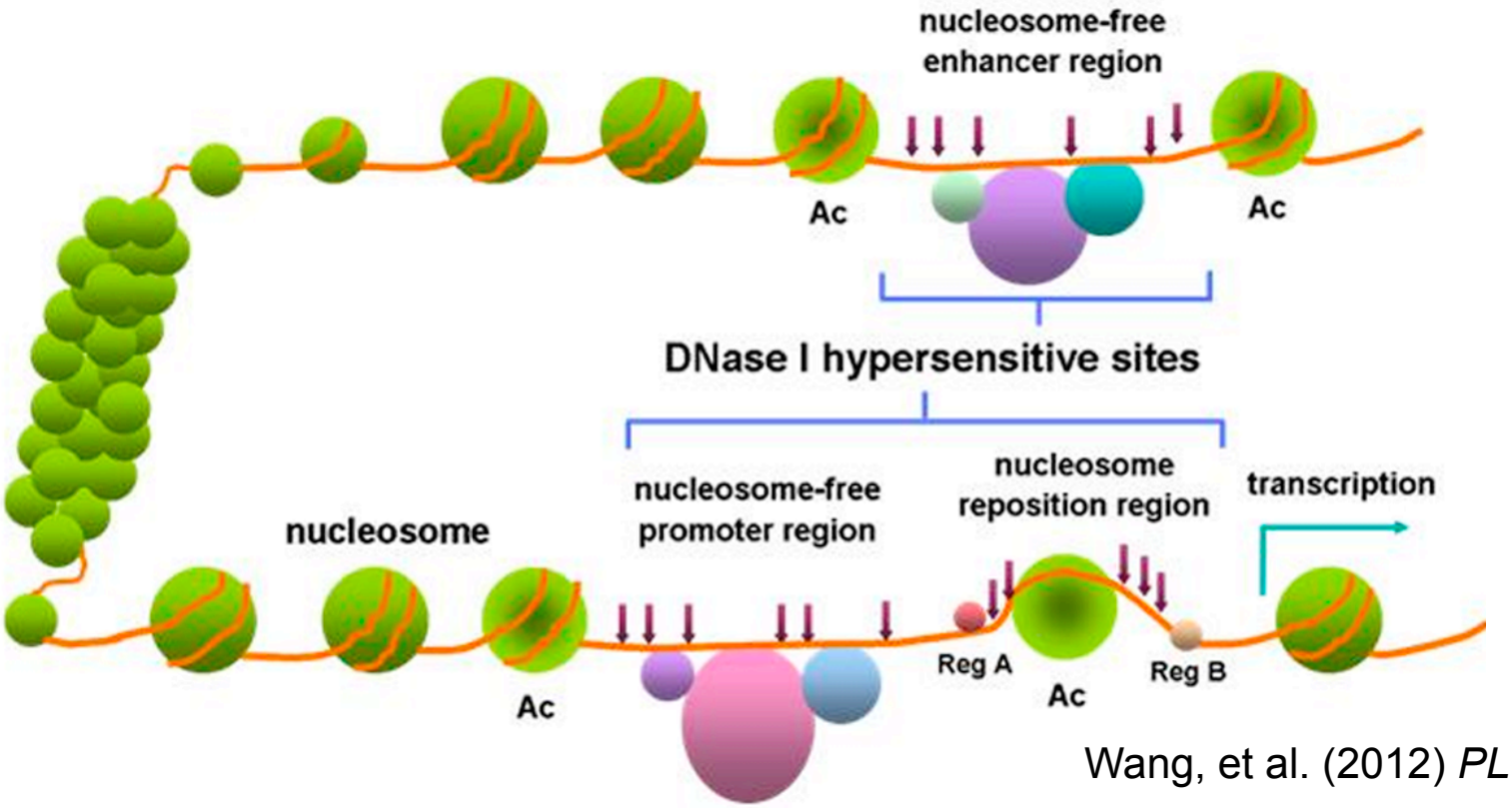




inversion moves w+ gene close to heterochromatic region, which can spread = variegated (red & white) eye

nearby chromatin state affects gene expression!

DNase I sensitivity assay



Wang, et al. (2012) *PLOS One*

open chromatin is cut by DNase I: DNA is open and can be cut by DNase I

closed chromatin is not cut by DNase I: DNA bound tightly by nucleosomes is protected from cutting by DNase I

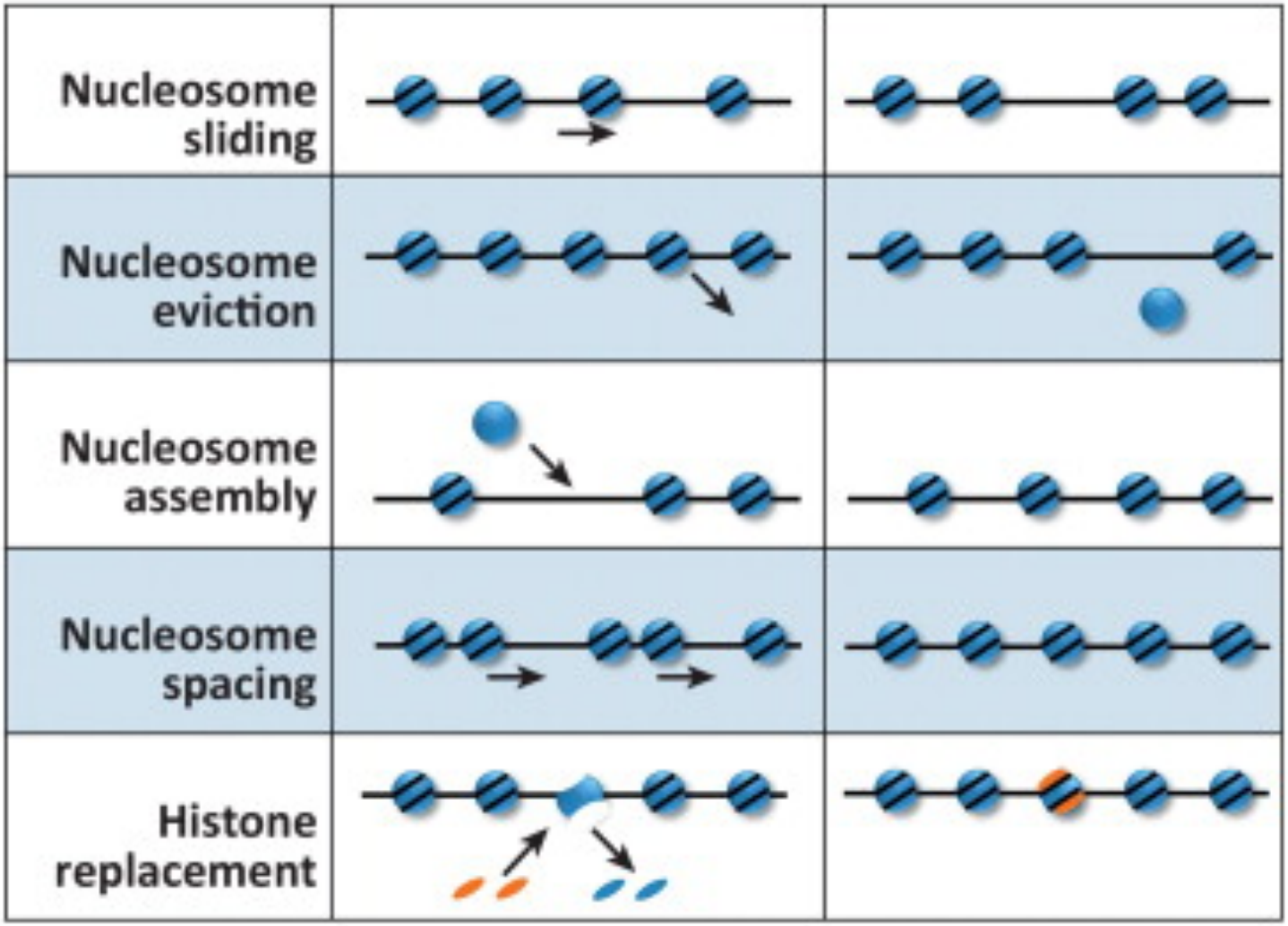
chromatin organization affects transcription through these mechanisms:

- control of nucleosome density
- histone modifications
- DNA methylation

control of nucleosome density

- constitutively active genes have promoters that are generally “open”
- inducible genes have nucleosomes that need to be moved
- chromatin remodeling complexes control positioning of nucleosomes to allow or disallow access to promoters

functions of chromatin remodelers



replace histones with easier to move variants of those histones

Petty & Pillus (2013) *TRENDS in Genetics*

examples of types of chromatin remodeling complexes

- SWI/SNF- sliding or ejection of nucleosomes opens chromatin
- SWR1- replaces a stable histone protein for one more easily displaced, opening chromatin
- ISWI- controls position of histones to keep chromatin closed

chromatin organization affects transcription through these mechanisms:

- control of nucleosome density
- histone modifications
- DNA methylation

chemical modifications of histone proteins

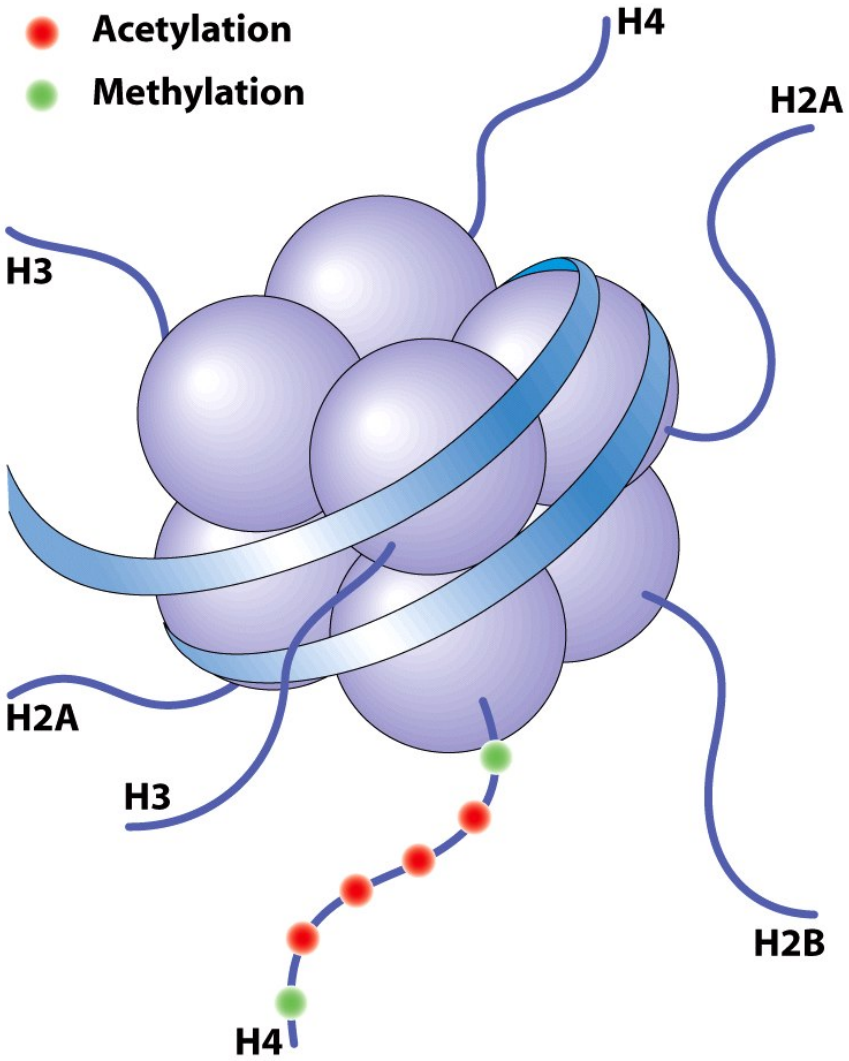
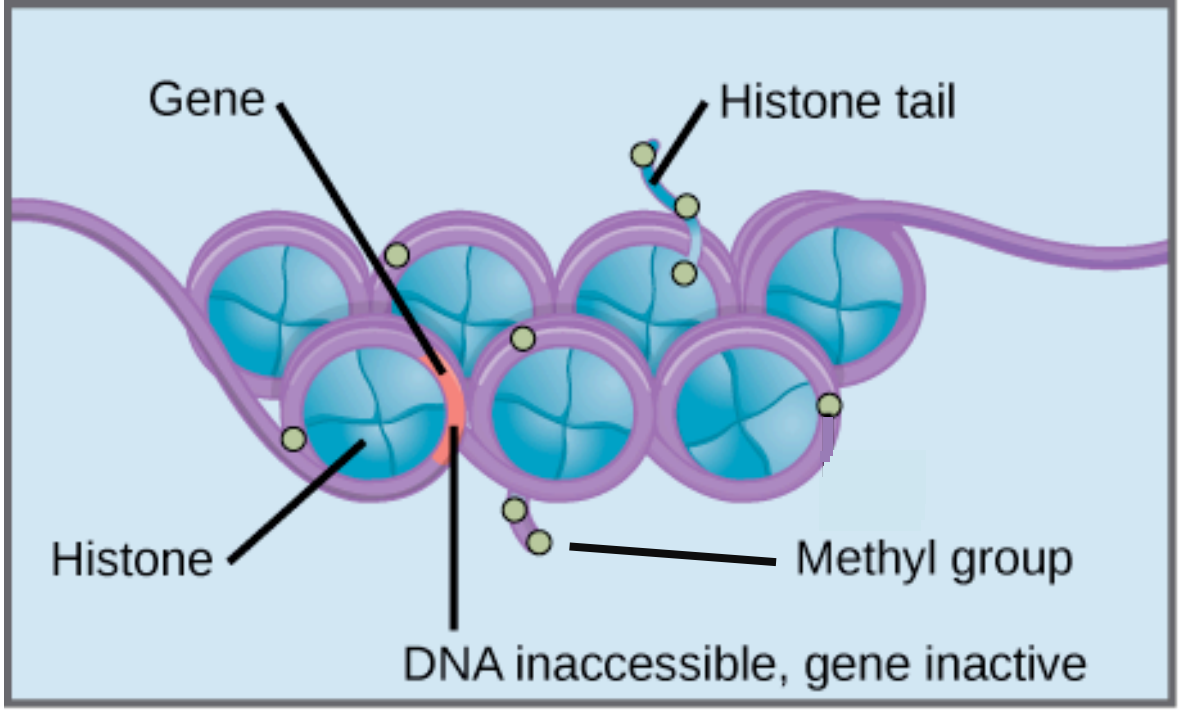


Figure 11-12
Introduction to Genetic Analysis, Ninth Edition
© 2008 W. H. Freeman and Company

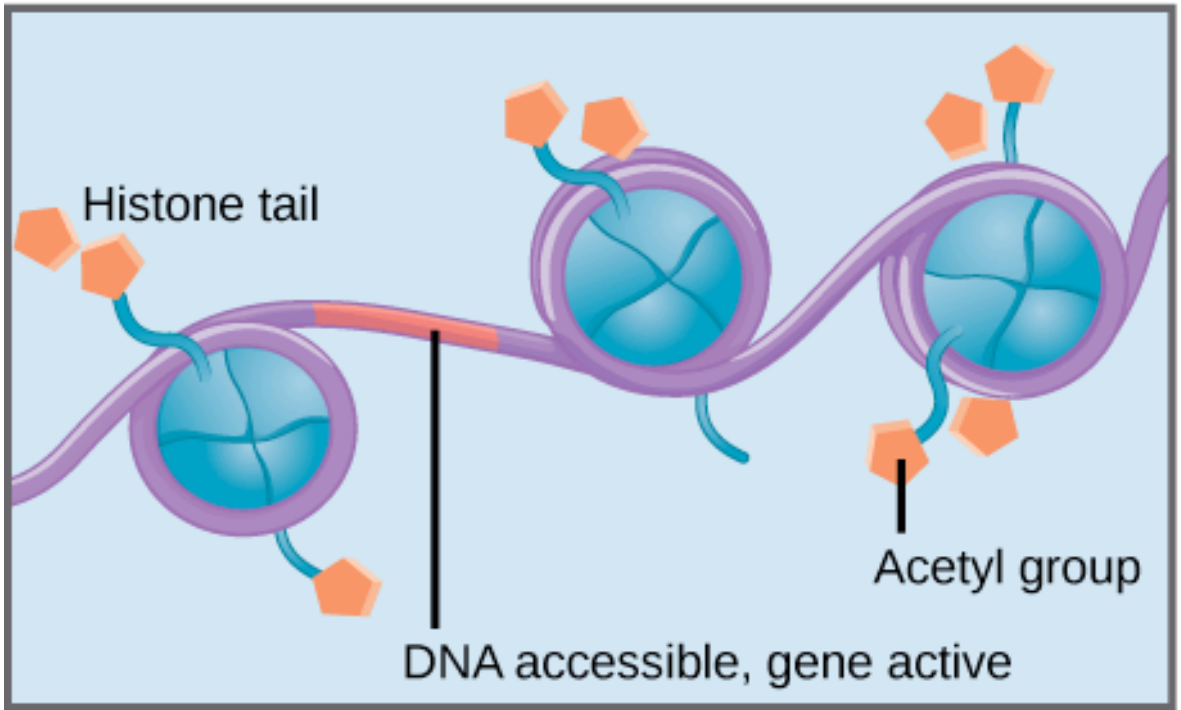
Histone tail modifications control chromatin stability and condensation

chemical groups are added to particular amino acids on particular histone proteins

this is a type of post-translational modification of the histones



Methylation of histones causes nucleosomes to pack tightly together. Transcription factors cannot bind the DNA, and genes are not expressed.



Histone acetylation results in loose packing of nucleosomes. Transcription factors can bind the DNA and genes are expressed.

Mahr, J., 2014

Histone Acetylation

- The most common chemical modification associated with opening chromatin structure is addition of acetyl ($COCH_3$) groups by **histone acetyltransferases** (**HATs**)

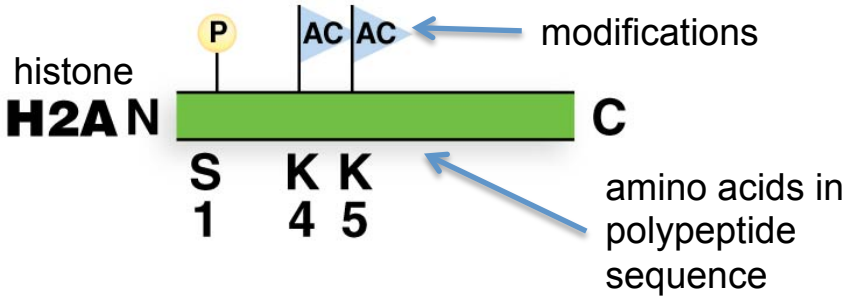
- These add acetyl groups to positively charged residues in the N-terminal histone tails; the acetylation neutralizes the positive charge and relaxes the histone/DNA interaction

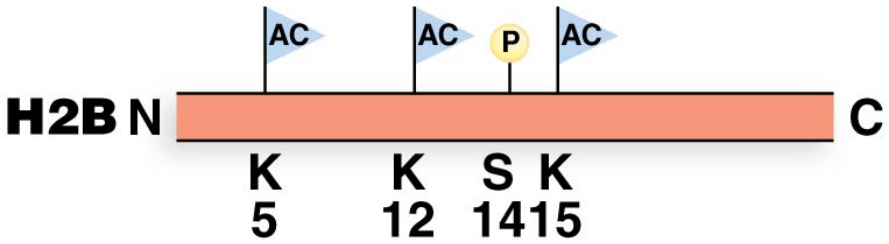
- Acetyl groups are removed by **histone deacetylases** (**HDACs**)

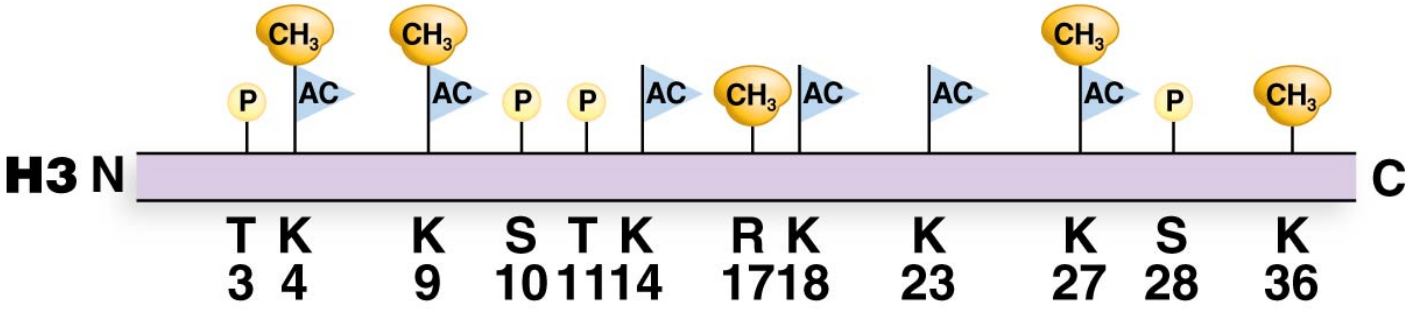
Histone Methylation

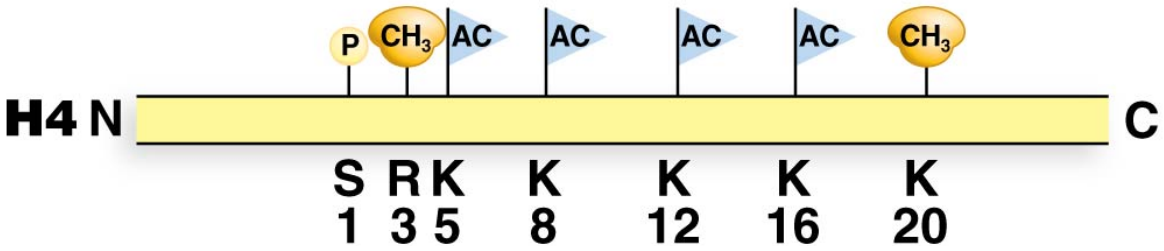
- Methyl (CH3) groups are added to N-terminal histone tails by **histone methyltransferases** (**HMTs**)
- Lysine and arginine are both targeted for methylation
- Methylation plays a role in converting open to closed chromatin
- Demethylation is carried out by **histone demethylases** (**HDMTs**)

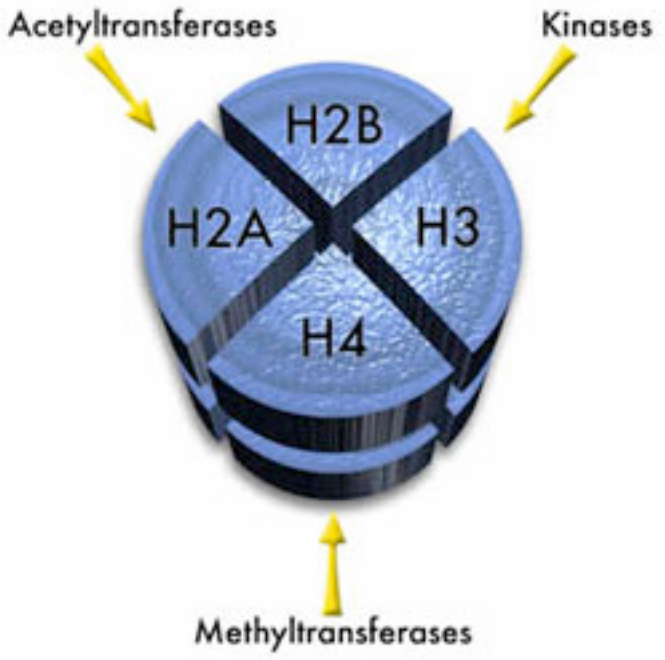
Different amino acids on the histone tails of different histone components can be targeted for modification





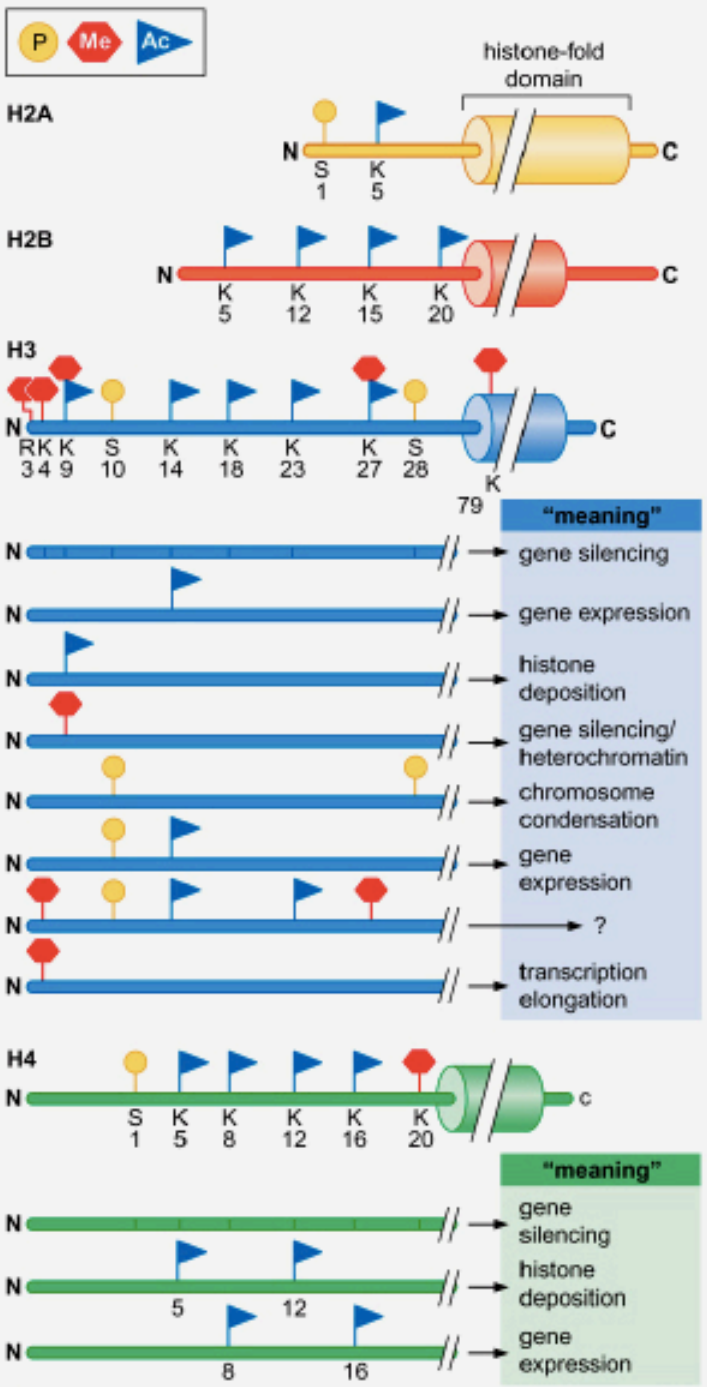






The "histone code" hypothesis

Strahl & Allis (2000) *Nature*



© 2004 Pearson Education, Inc., publishing as Benjamin Cummings

Different combinations of modifications have different "meanings" for chromatin structure and transcriptional regulation

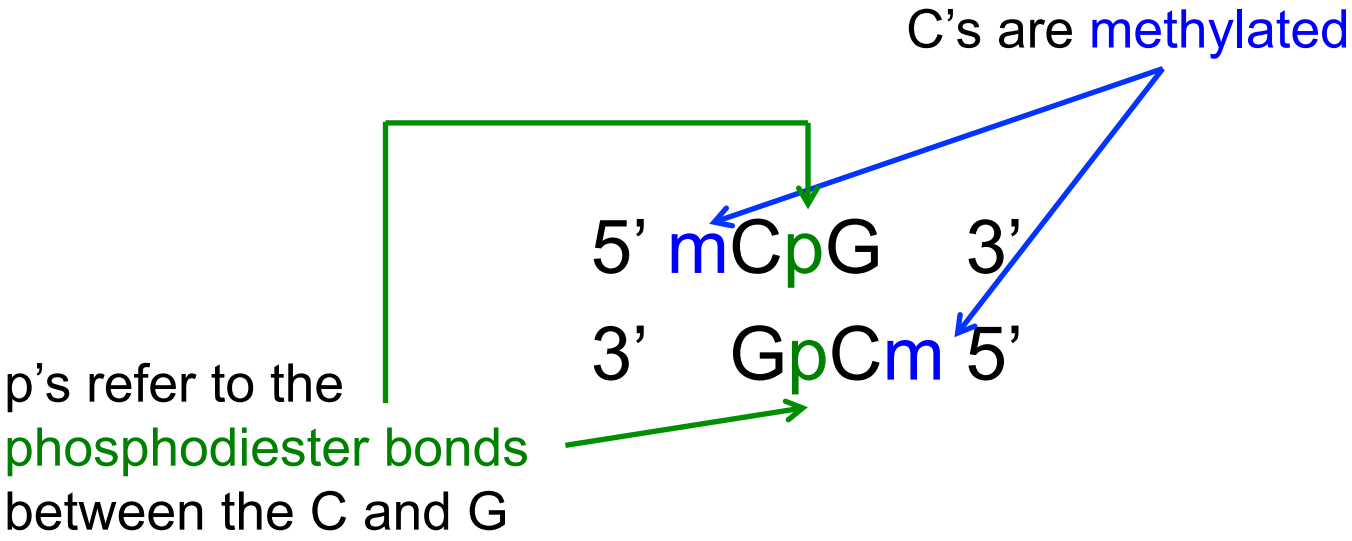
no real success "solving" the histone code

just too complicated? For example, histone H3 has 10+ lysines (K) that can have 4 methylation states (un-, mono-, di-, and tri-), so each H3 K alone has 4^{10} (= about 1 million) possible states relative to methylation alone (K can also be acetylated!)

chromatin organization affects transcription through these mechanisms:

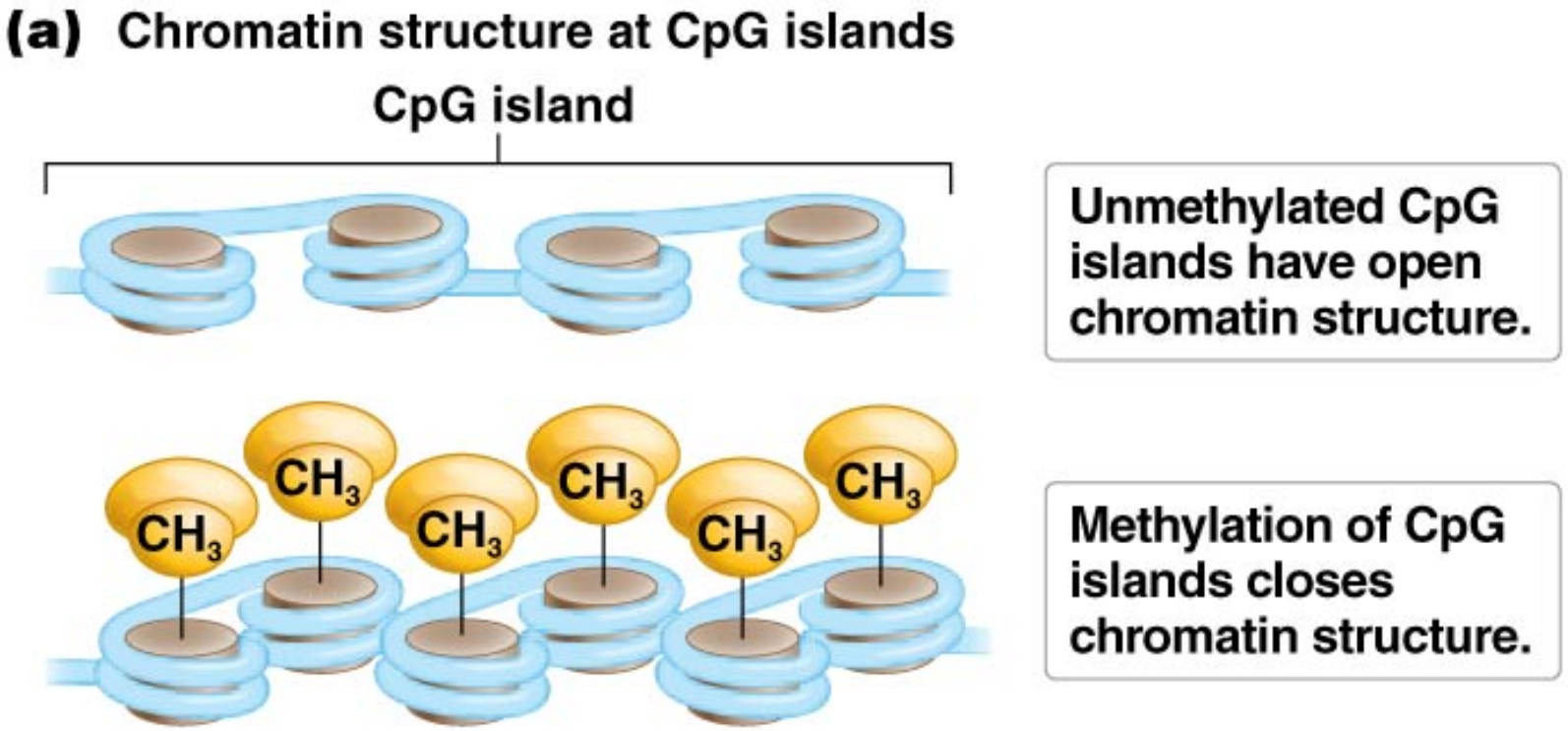
- control of nucleosome density
- histone modifications
- DNA methylation

DNA methylation is important for regulation of genes, especially mammals



- methyltransferases add methyl groups to cytosines located in **CpG dinucleotides**
- Sequences rich in CpG, **CpG islands**, are targeted for methylation; these islands are clustered at promoters in mammals
- **Unmethylated** CpG islands = chromatin structure is **open**, active transcription can take place
- **Methylated** CpG islands = promoter regions are **closed**, transcription is repressed

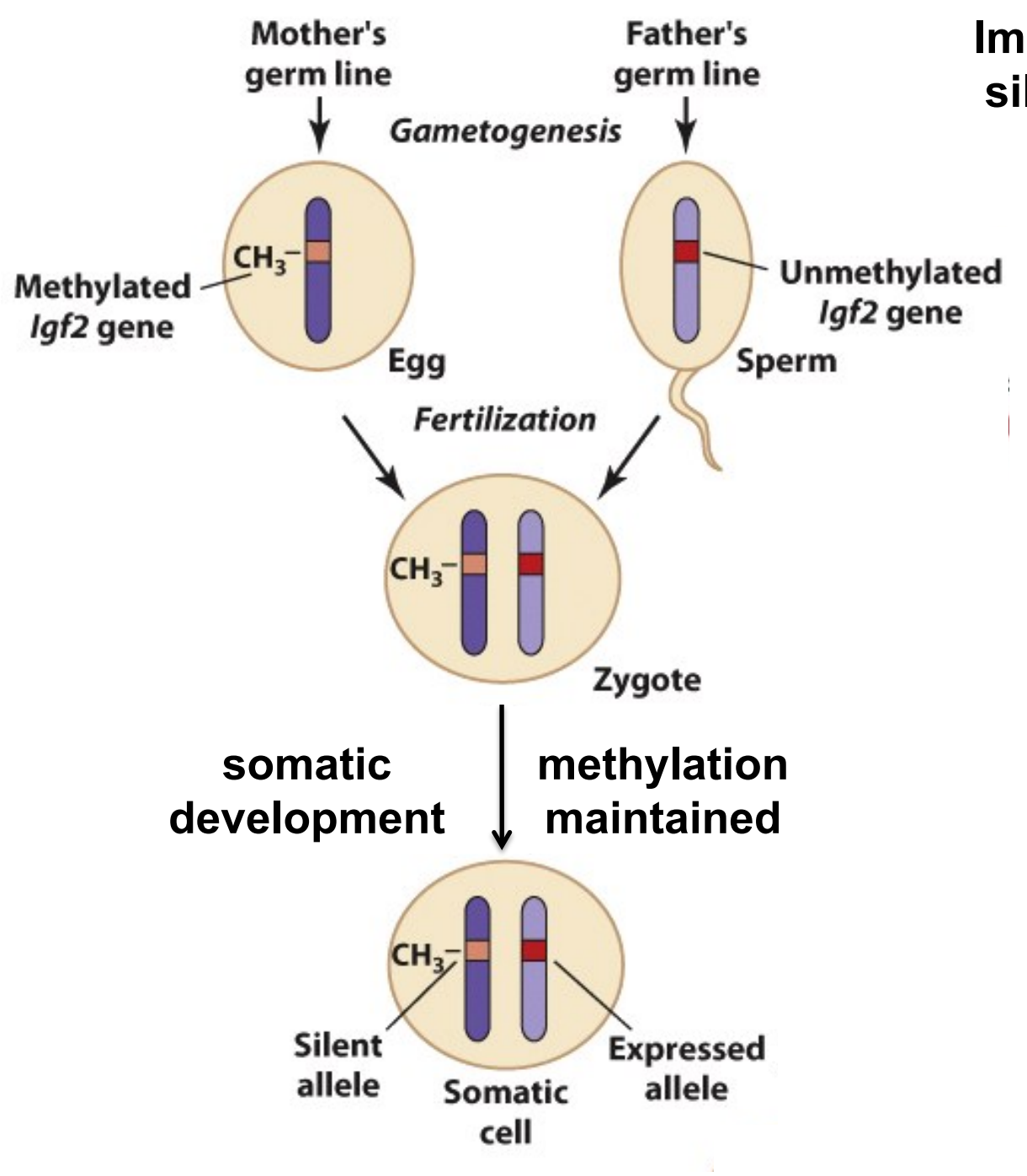
nucleotide methylation



DNA methylation is mechanism for imprinting

- with imprinting, genes are expressed differently depending upon the parent of origin
- For imprinted genes, *expression occurs from only one parent's allele* - the other parent's allele is silenced
- silencing of one parent's allele usually occurs through DNA methylation
- this is a required part of normal development, primarily in mammals (also in some plants, fungi)

Imprinting of Igf2 in mouse silences maternal allele in somatic cells



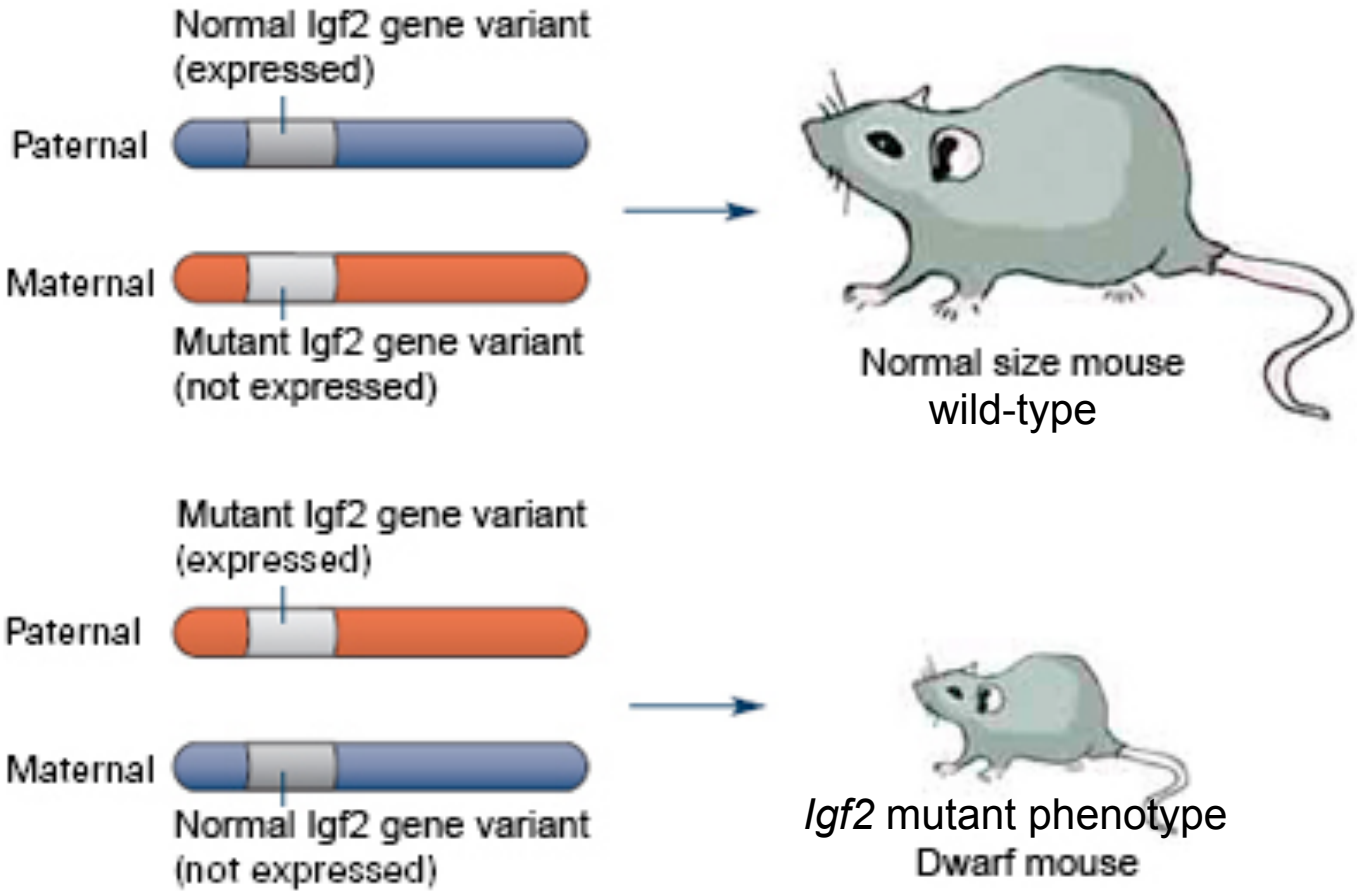
Alleles of Igf2 imprinted in parental germ lines, mom's copy is methylated, dad's copy isn't

Imprinted alleles in *Igf2* are combined in zygote at fertilization

In somatic cells, the methylation of maternal allele is maintained and thus this allele is not expressed

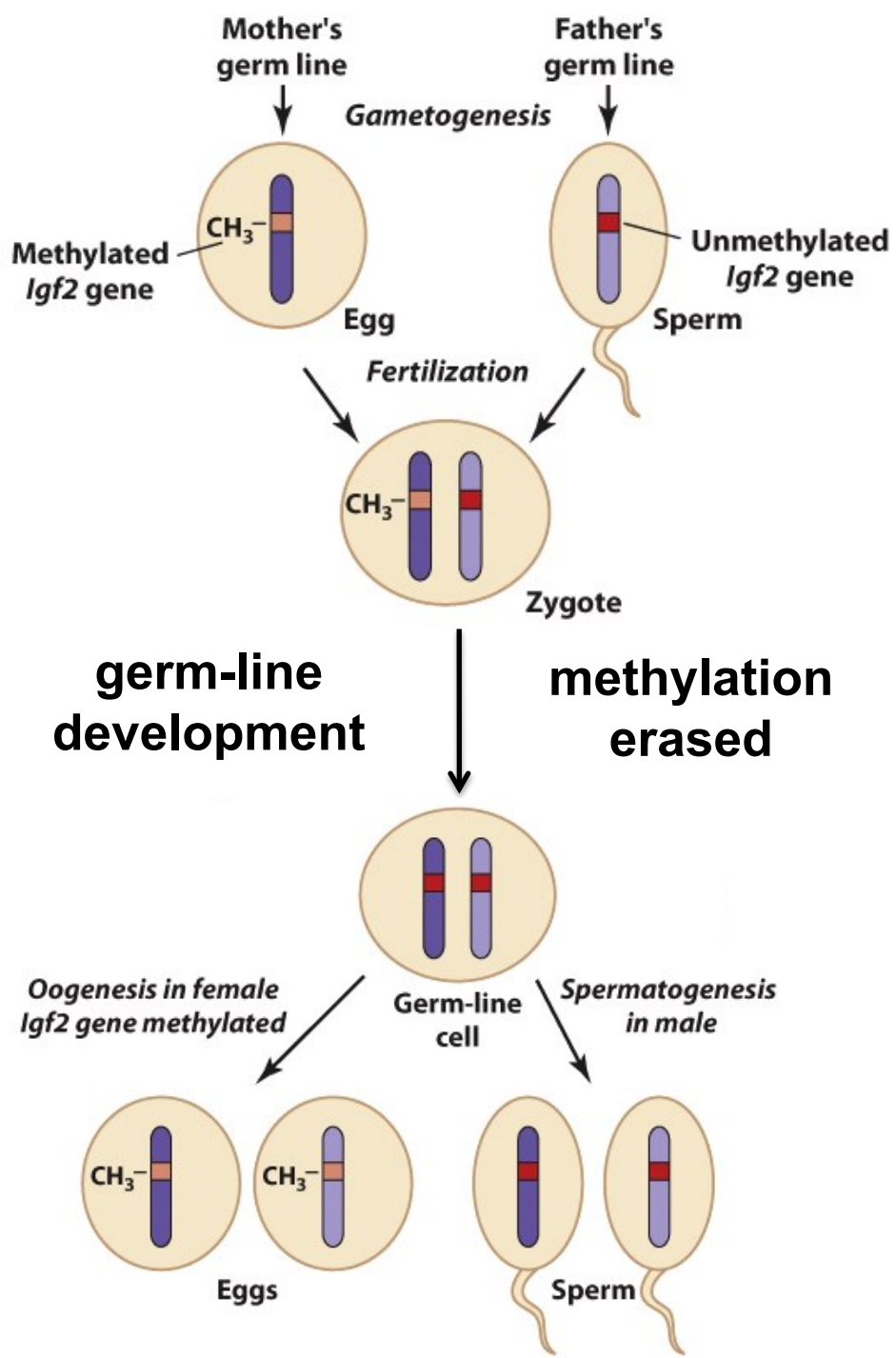
the paternal allele maintains unmethylated state, so is expressed

differential effects of mutation of maternal or paternal allele



http://publications.nigms.nih.gov/thenewgenetics/chapter2.html

Igf2 = Insulin growth factor 2, regulates growth during development



Imprinting of *Igf2* in mouse is reset in germline

Alleles of Igf2 imprinted in parental germ lines, mom's copy is methylated, dad's copy isn't

Imprinted alleles in *Igf2* are combined in zygote at fertilization

during development of the germ-line, the imprint is erased

methylation is reestablished in oogenesis but not spermatogenesis- if mouse is female, all her oocytes will have methylated *Igf2*, all males will produce sperm with unmethylated *Igf-2*

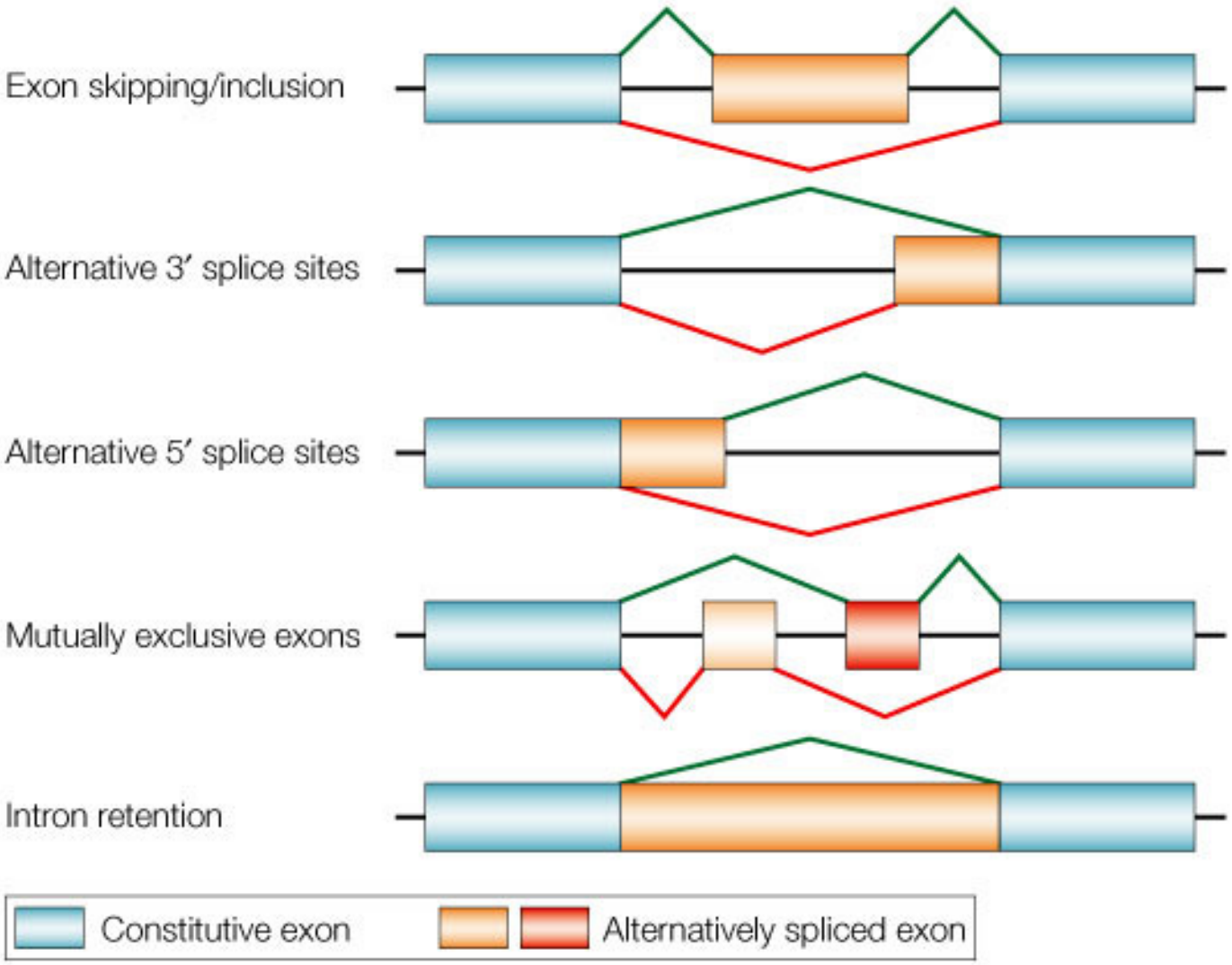
a short note on epigenetics

- epigenetics are functionally relevant changes that do not alter DNA sequence
- histone marks, DNA methylation are examples of these kinds of changes
- most epigenetic marks are involved in establishing and maintaining cell identity
- most epigenetic marks are erased in the germline, so are not passed to offspring
- evidence of some epigenetic changes transmitted over short numbers of generations *in plants*

regulation of gene expression in eukaryotes

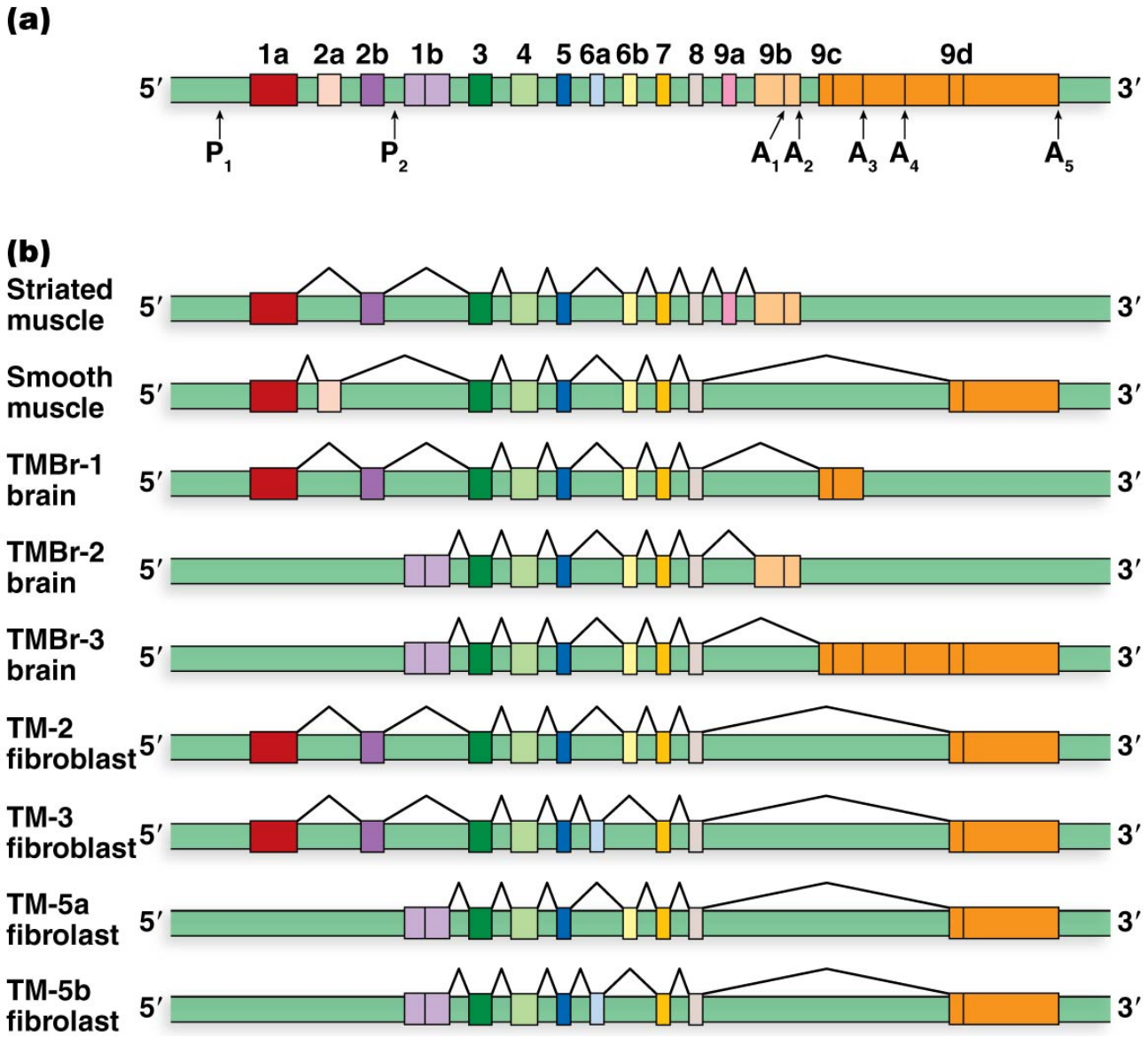
- overview
- transcriptional regulation
- post-transcriptional regulation
  - alternative splicing
  - factors that affect mRNA stability
  - RNA interference
- regulation of whole chromosomes

Alternative splicing



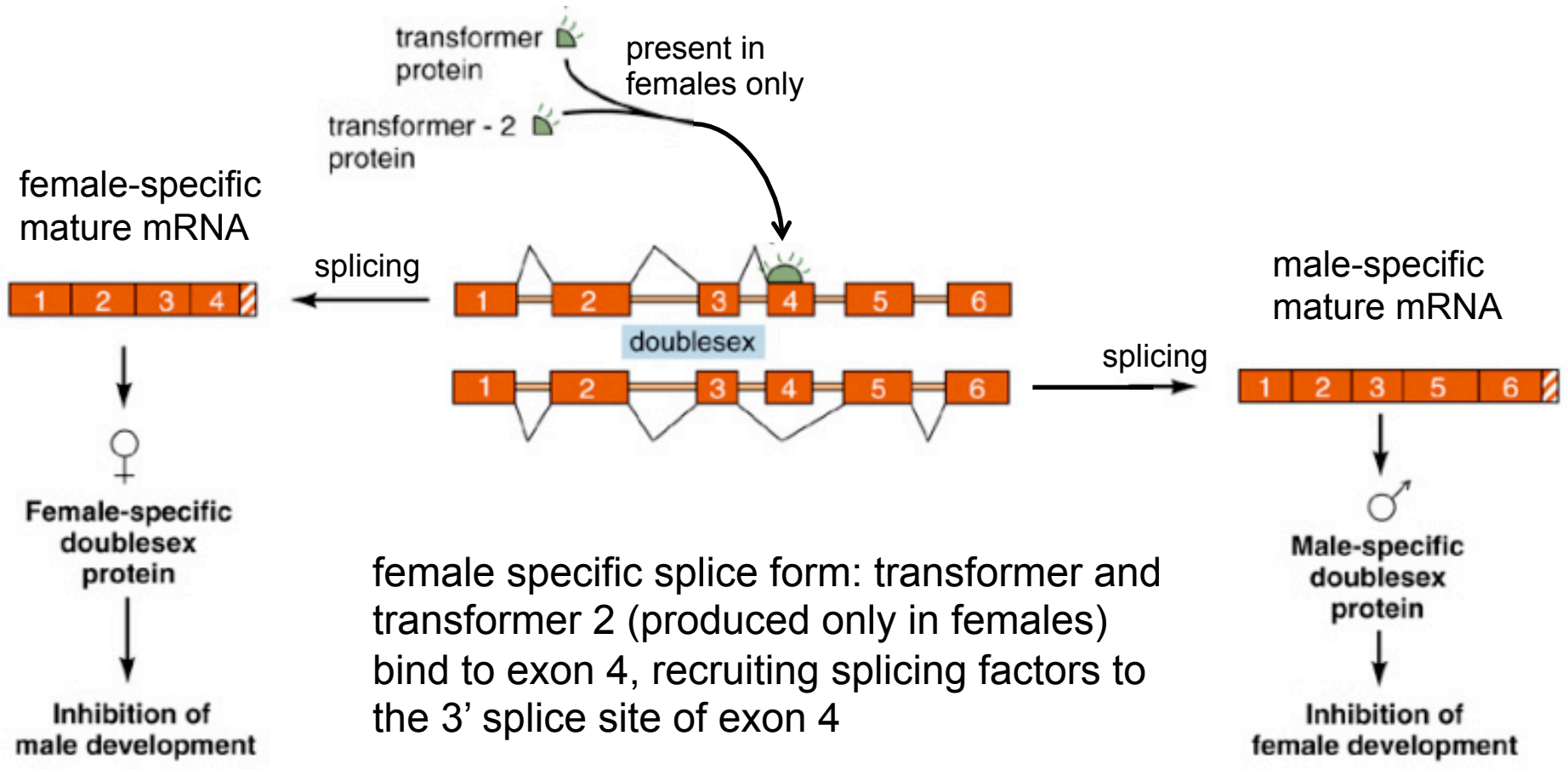
Cartegni, et al., (2002) *Nature Reviews Genetics*

Alternative pre-mRNA processing of the rat α-tropomyosin gene.



Nine distinct mature mRNAs produced by different types of muscle, brain, and fibroblast cells each produce a different tropomyosin protein.

regulation of sex-specific development via alternative splice forms in *Drosophila*



female specific splice form: transformer and transformer 2 (produced only in females) bind to exon 4, recruiting splicing factors to the 3' splice site of exon 4

male specific splice form: without these proteins, 3' splice site on exon 5 is used instead

http://www.bio.miami.edu/dana/pix/sxl_splicing.jpg

regulation of gene expression in eukaryotes

- overview
- transcriptional regulation
- post-transcriptional regulation
  - alternative splicing
  - factors that affect mRNA stability
  - RNA interference
- regulation of whole chromosomes

control of mRNA stability

- mRNA stability is influenced by several factors
  - The poly(A) tail
  - The sequence of the 3’ UTR
  - Chemical factors (e.g., hormones)
  - Small interfering RNAs (siRNAs) or microRNAs (miRNAs)

- stability of mRNA will affect the level of that mRNA in the cell
  - less stable mRNA can lead to fewer mRNAs and thus fewer proteins produced

regulation of gene expression in eukaryotes

- overview
- transcriptional regulation
- post-transcriptional regulation
  - alternative splicing
  - factors that affect mRNA stability
  - RNA interference
- regulation of whole chromosomes

first noted as a phenomenon termed "co-suppression"

- In the 1990's, Jorgensen and colleagues added an additional copy of a pigment gene to a petunia

- instead of having more pigment, the flowers were variegated with regions of little or no pigment

- they called this "co-suppression" as these pigment-less regions reflected suppression of *both* the added copy and the original native copy of the gene

Matzke & Matzke, 2004

Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*

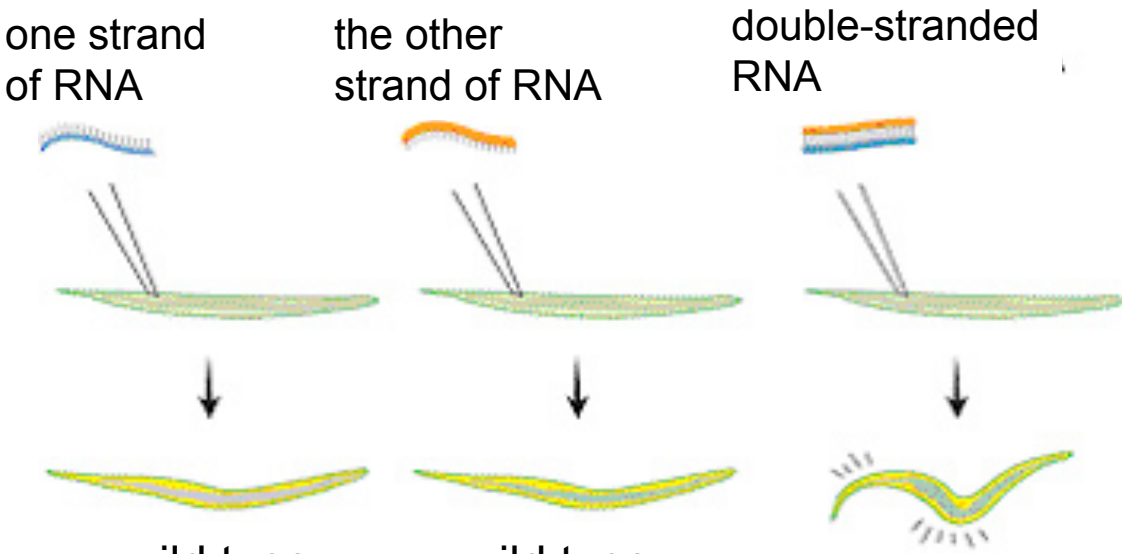
Andrew Fire*, SiQun Xu*, Mary K. Montgomery*, Steven A. Kostas*†, Samuel E. Driver‡ & Craig C. Mello‡

** Carnegie Institution of Washington, Department of Embryology, 115 West University Parkway, Baltimore, Maryland 21210, USA*
† Biology Graduate Program, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA
‡ Program in Molecular Medicine, Department of Cell Biology, University of Massachusetts Cancer Center, Two Biotech Suite 213, 373 Plantation Street, Worcester, Massachusetts 01605, USA

NATURE | VOL 391 | 19 FEBRUARY 1998

Fire & Mello, Nobel 2006

- injected worms with single and double stranded RNA of *unc-22* muscle gene
- those injected with only double stranded *unc-22* RNA had a *unc-22* mutant phenotype (twitchy worms)
- the native *unc-22 m*RNAs were being degraded when double-stranded *unc-22* RNAs were injected



RNA Interference

- Double stranded RNA that has one strand complementary to a particular target mRNA is supplied
- This causes destruction of the target mRNA or can block its translation
- this results in reduction or elimination of the protein that the mRNA codes for
- RNAi has been documented in a number of organisms, including *C. elegans, Drosophila, Arabidopsis*, and in mammals, including humans.

why can cells do RNAi?

- surely not just to make the lives of scientists easier
- small RNAs have regulatory functions in normal cellular processes
- small RNAs play a big role in immune response in plants and some animals
- small RNAs have been demonstrated to be critical to normal development

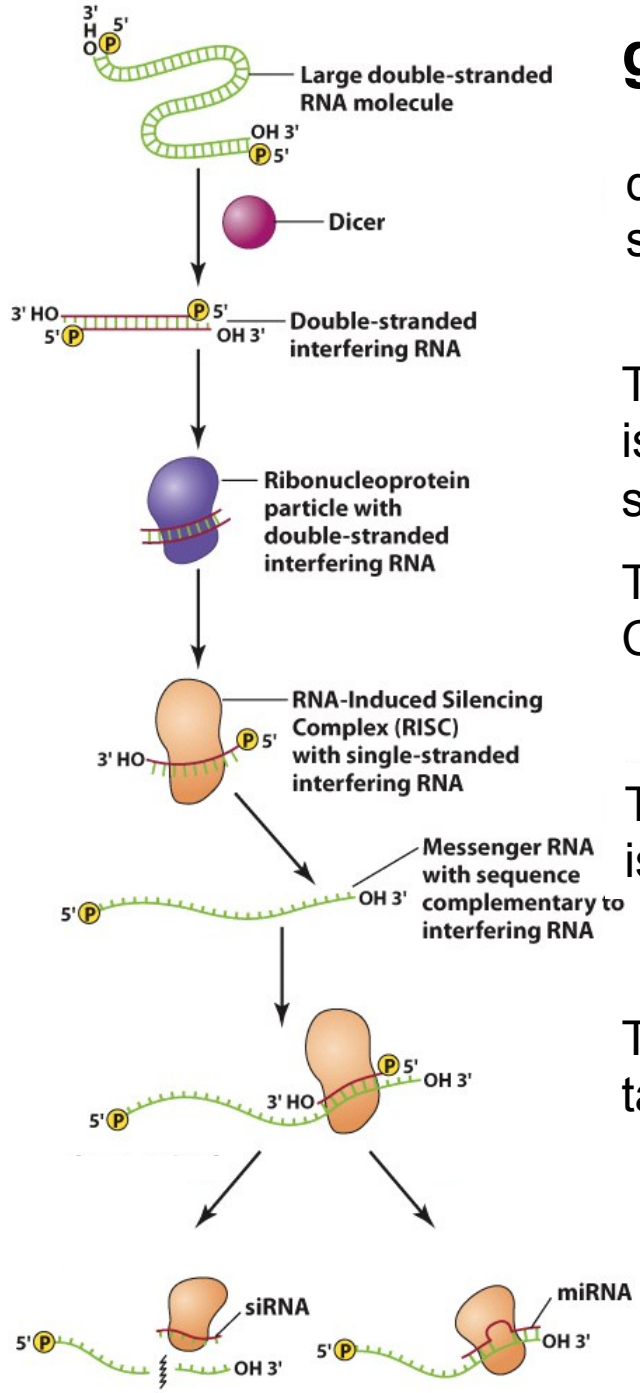
RNAs capable of RNAi are numerous

Name	Organism	Length (nt)	Proteins	Source of trigger	Function	Refs
miRNA	Plants, algae, animals, viruses, protists	20–25	Drosha (animals only) and Dicer	Pol II transcription (pri-miRNAs)	Regulation of mRNA stability, translation	93–95, 200–202,226
casiRNA	Plants	24	DCL3	Transposons, repeats	Chromatin modification	38,44,51, 52,61–63
tasiRNA	Plants	21	DCL4	miRNA-cleaved RNAs from the *TAS* loci	Post-transcriptional regulation	64–68
natsiRNA	Plants	22	DCL1	Bidirectional transcripts induced by stress	Regulation of stress-response genes	71,72
		24	DCL2			
		21	DCL1 and DCL2			
Exo-siRNA	Animals, fungi, protists	~21	Dicer	Transgenic, viral or other exogenous dsRNA	Post-transcriptional regulation, antiviral defense	4,5,8,227
	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	75–79,82, 83,86,87, 200,201, 228
piRNA	Metazoans excluding *Trichoplax adhaerens*	24–30	Dicer-independent	Long, primary transcripts?	Transposon regulation, unknown functions	157, 163–169, 177,202
piRNA-like (soma)	*Drosophila melanogaster*	24–30	Dicer-independent	In *ago2* mutants in *Drosophila*	Unknown	76
21U-RNA piRNAs	*Caenorhabditis elegans*	21	Dicer-independent	Individual transcription of each piRNA?	Transposon regulation, unknown functions	114, 173–175
26G RNA	*Caenorhabditis elegans*	26	RdRP?	Enriched in sperm	Unknown	114

ago2, Argonaute2; casiRNA, *cis*-acting siRNA; DCL, Dicer-like; endo-siRNA, endogenous small interfering RNA; exo-siRNA, exogenous small interfering RNA; miRNA, microRNA; natsiRNA, natural antisense transcript-derived siRNA; piRNA, Piwi-interacting RNA; Pol II, RNA polymerase II; pri-miRNA, primary microRNA; RdRP, RNA-dependant RNA polymerase; tasiRNA, *trans*-acting siRNA.

Ghildiyal & Zamore, 2009, Nat. Rev. Gen

general features of RNAi pathways



© 2012 John Wiley & Sons, Inc. All rights reserved.

double-stranded RNA (dsRNA) is cut by Dicer into small RNA 21-28bp long

The small dsRNA assemble with proteins, and is unwound or cleaved to produce a single stranded RNA (ssRNA)

This produces an RNA-Induced Silencing Complex (RISC)

The RISC targets a sequence in a mRNA that is complementary to the ssRNA in the RISC

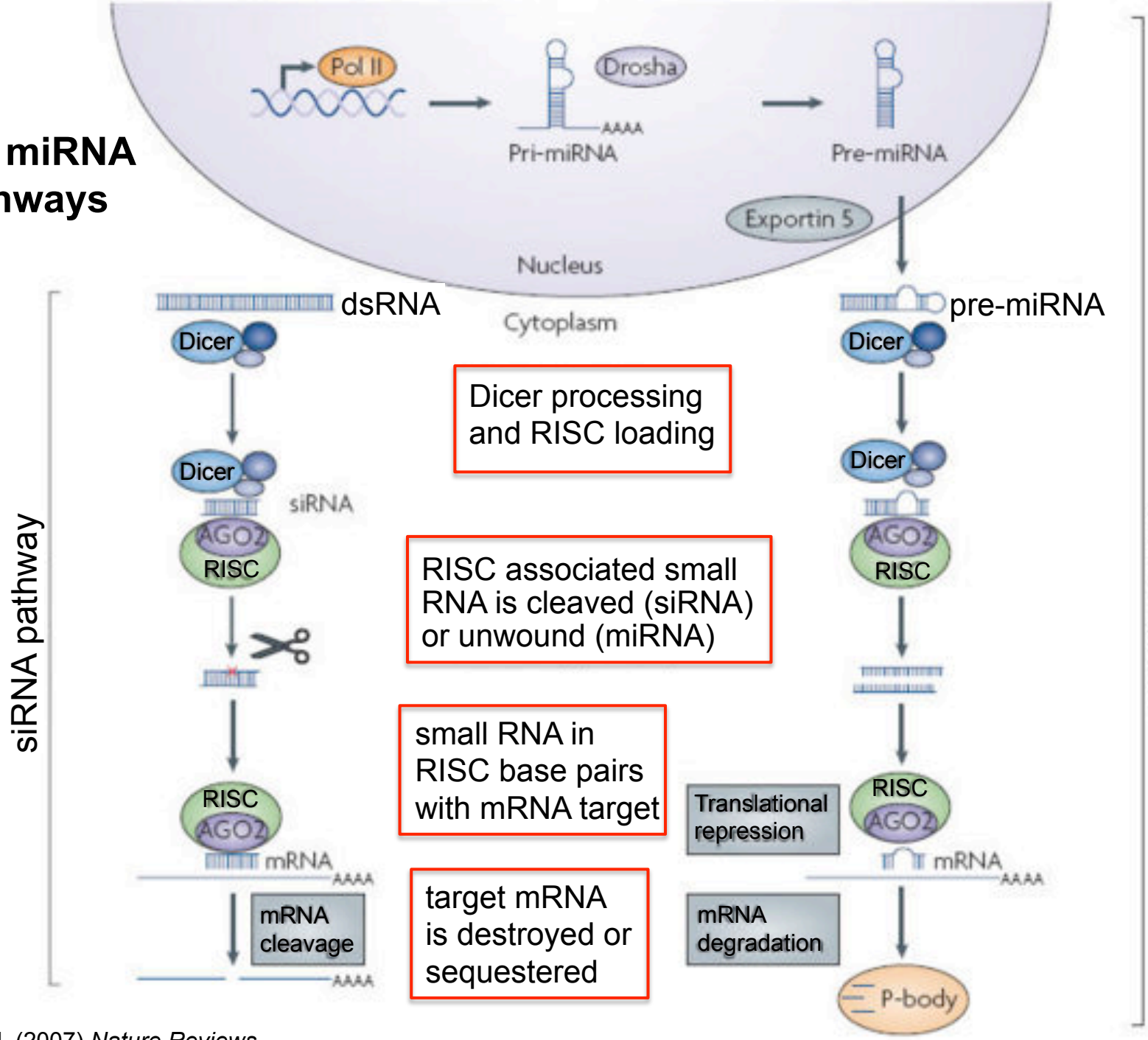
The ssRNA in the RISC base-pairs with its target mRNA

The target mRNA is either cleaved and degraded, its translation is repressed, or it is sequestered

common features of RNAi

- **Dicer** activity cleaves double stranded RNA (dsRNA) molecules to 21-28 bp dsRNA fragments, these are the active form of small RNA
- RNA-Induced Silencing Complex (**RISC**) binds ds small RNA, cleaves or unwinds dsRNA into a single-stranded “guide strand”
- RISC-guide strand binds target mRNA by complementary base pairing
- target mRNAs are degraded, sequestered, or otherwise prevented from being translated

siRNA and miRNA RNAi pathways



de Fougerolles et al. (2007) *Nature Reviews*

where do double-stranded RNAs come from?

- microRNAs (miRNAs) are present in the genome
  - transcribed by RNA Pol II
  - produce transcript that fold back up on themselves to produce dsRNA
  - imperfect base-pairing with target mRNA
  - Target mRNA is prevented from being translated, stored, or degraded
- small interfering RNAs (siRNAs)
  - start as dsRNA
  - can be exogenous or endogenous
  - perfect base-pairing with target
  - Target mRNA is degraded

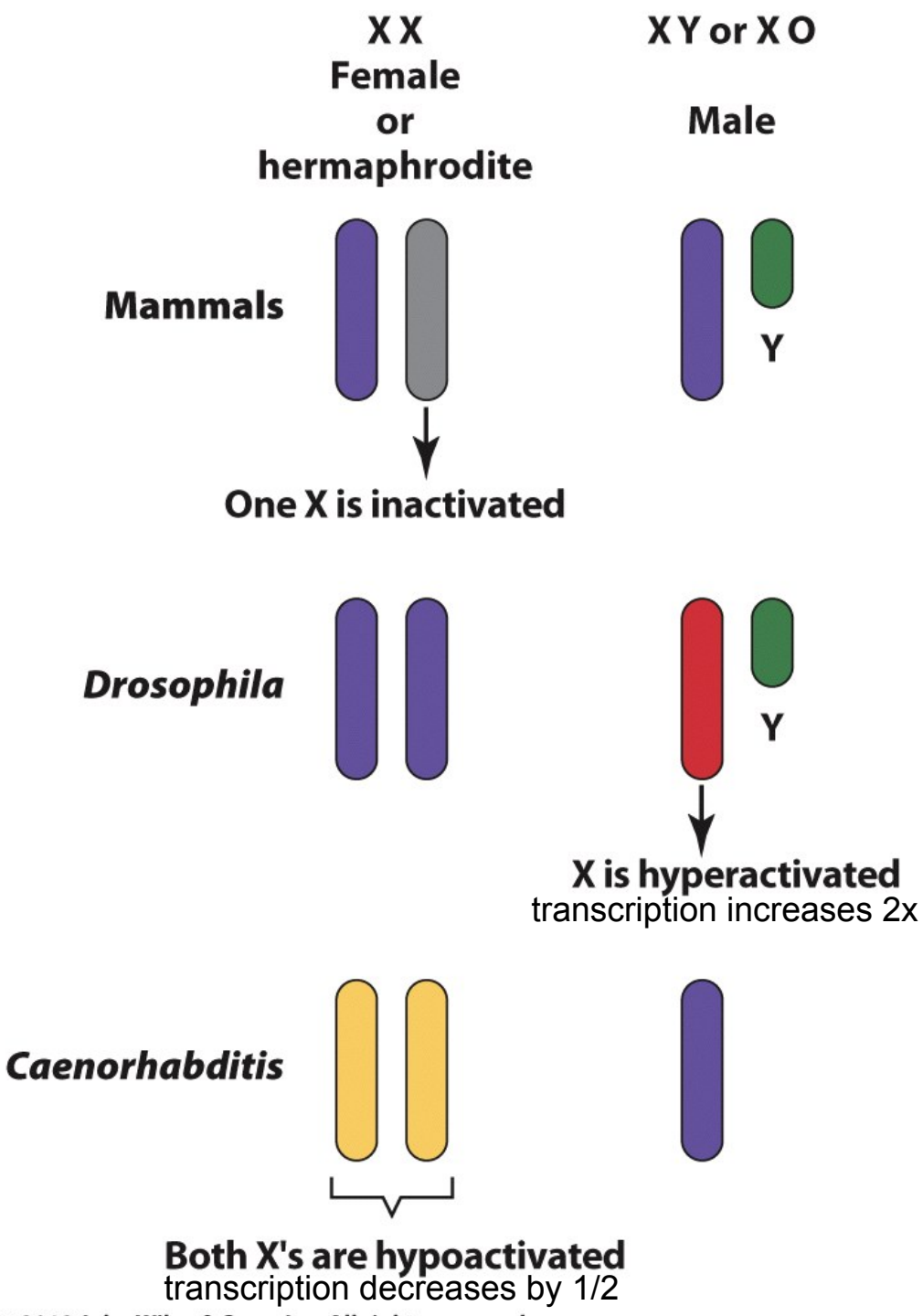
regulation of gene expression in eukaryotes

- overview
- transcriptional regulation
- post-transcriptional regulation
- regulation of whole chromosomes

activation and inactivation of whole chromosomes

- Transcription from entire chromosomes can be activated or inactivated
- This is most often observed with sex chromosomes
- This is necessary to equalize transcription between females and males
  - in XY systems, females have two X's and males have one (as they are XY)
  - in ZW systems, males have two Zs and females have one (as they are ZW)
- Different organisms have evolved distinct ways of compensating for different dosages of X/Z chromosomes in males and females

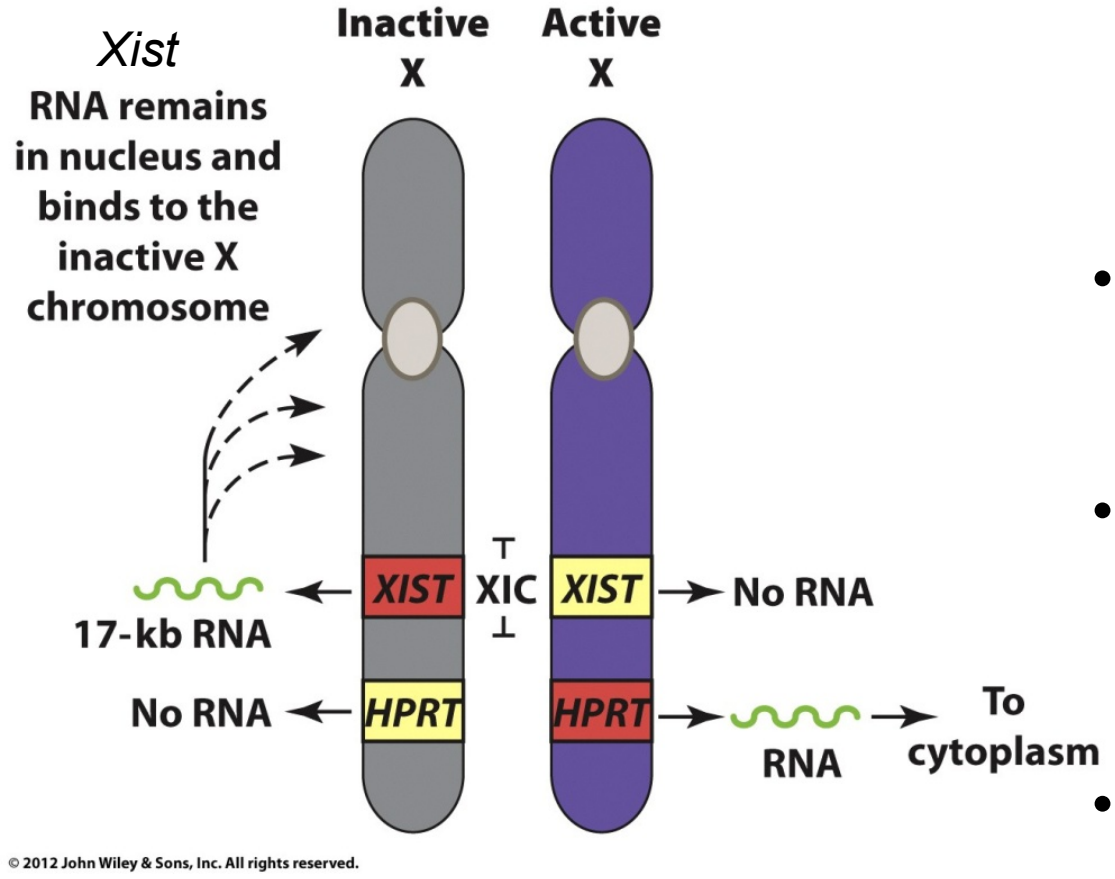
X chromosome dosage compensation



entire chromosomes are activated or inactivated

to equalize transcription between females and males

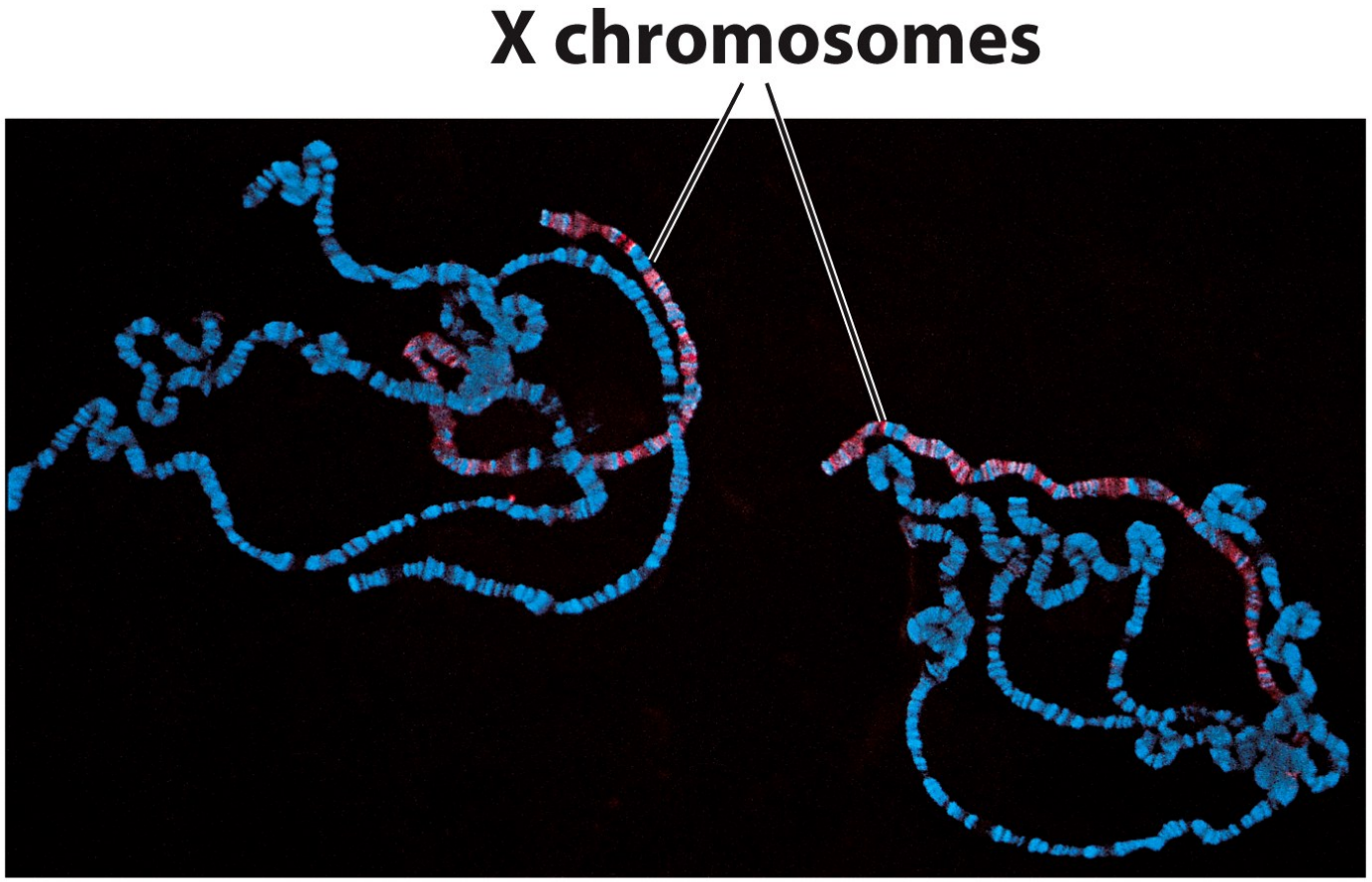
X-inactivation in mammals



© 2012 John Wiley & Sons, Inc. All rights reserved.

HPRT = example gene for comparison

- X inactivation begins at the X inactivation center (XIC) and spreads toward the ends of the chromosomes
- The X inactive specific transcript (*XIST)* gene remains active
- *XIST* encodes a functional RNA that coats the inactive X chromosome
- inactive X has establishes very compact chromatin structure, forms Barr body



In *Drosophila*, a protein complex made up of different proteins (including the MSL or male-specific lethal proteins) and non-coding RNAs (roX RNAs) bind the single male X

This protein complex permits hypertranscription from the single male X by opening up chromatin on this chromosome, through histone modifications

regulation of gene expression in eukaryotes

- overview
- transcriptional regulation
  - *cis* regulatory sequences and *trans* acting factors
  - chromatin organization
- post-transcriptional regulation
  - alternative splicing
  - factors that affect mRNA stability
  - RNA interference
- regulation of whole chromosomes