MEDICAL GENOMICS

Василий Евгеньевич Раменский
ramensky@gmail.com

Анастасия Александровна Жарикова
azharikova89@gmail.com

НМИЦ Профилактической медицины
Факультет биоинженерии и биоинформатики МГУ

MEDICAL GENOMICS

- Big genomic data enable inference without intervention (aka reverse genetics)

- Data are just around the corner

MEDICAL GENOMICS

Part I. Кто виноват?

1. Mutations: origins and rates
2. Mutations: transmisson
3. Mutations in time: some basics of population genetics
4. Mutations in space: genes and consequences
5. Mutations in individuals and populations

Part II. Что делать?

6. Some basics of genetic epidemiology
7. Mendelian diseases: gene discovery and diagnostics

7½. Invited lecture?

8. Complex diseases: gene discovery and allelic architecture
9. Oncogenomics
10. Pharmacogenomics

Remarks

- English
- Molecular genetics + population genetics + medical genetics + statistical genetics + genetic epidemiology + bioinformatics $\Longrightarrow$ no textbook
- Many topics not covered: immunology, pathogens, microbiome

- Definitions
- Slides for individual work at home
- Summary, concepts, further reading

Remarks

- Окончательный список слушателей курса к 2й лекции

- Mid-term test: письменный, короткие вопросы и задачи.
  Без переписывания
- Журнальный клуб: ~6 статей, 15 минут на статью
- Зачет автоматом по результатам мидтерма и журклуба
- Итоговый зачет: письменный + собеседование

Textbooks

1. T.Strachan, A.Read. Human Molecular Genetics. 2011. ISBN 0815345895.
2. J. Gillespie. Population genetics. A concise guide 1998 ISBN 0-8018-5764-6
3. S. Szalai, et al. Medical genetics and genomics. 2016. https://www.researchgate.net/publication/303309837_Medical_genetics_and_genomics_2016
4. A.Griffiths et al. An Introduction to Genetic Analysis. Freeman/Worth, 11 ed. 2015 ISBN 1464109486
5. Бочков Н.П., Пузырев В.П., Смирнихина С.А. Клиническая генетика. Учебник. Под ред. Н.П. Бочкова. ГЭОТАР-Медиа, 4-е издание, 2018. ISBN 978-5-9704-4628-7

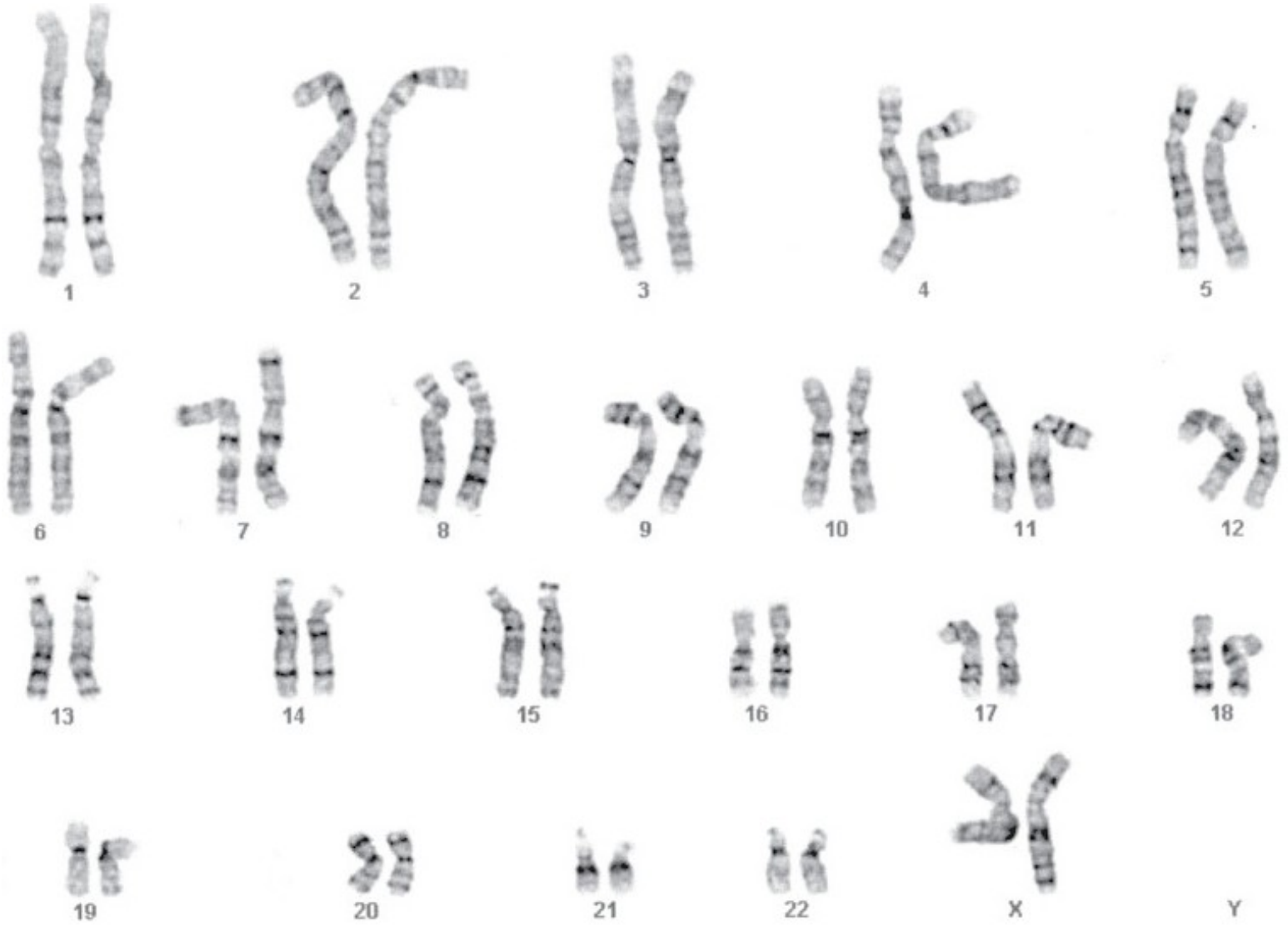
MUTATIONS:

ORIGINS AND RATES

Lecture plan

- Human karyotype
- Mitosis and DNA replication
- Replication fidelity and mutation rate
- Exogenous and endogenous DNA damage
- DNA repair mechanisms
- Single nucleotide variants
- Structural variants and CNVs
- Aneuploidy

Human karyotype

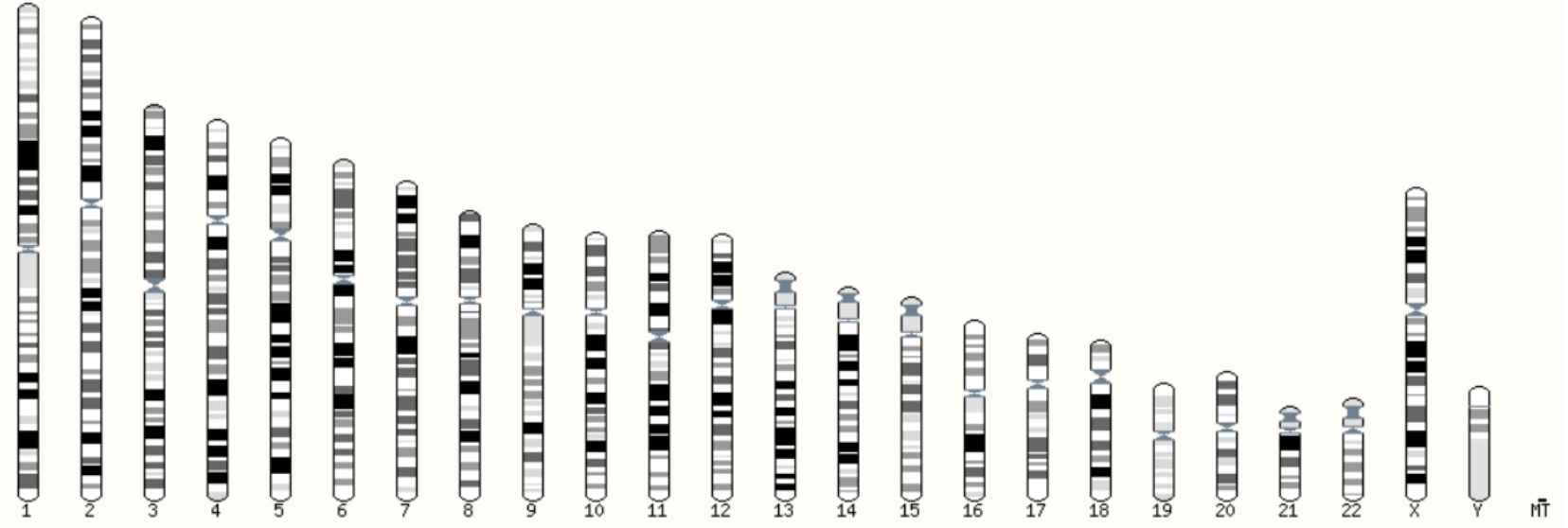


22 pairs of **autosomal** chromosomes + 2 **sex** chromosomes

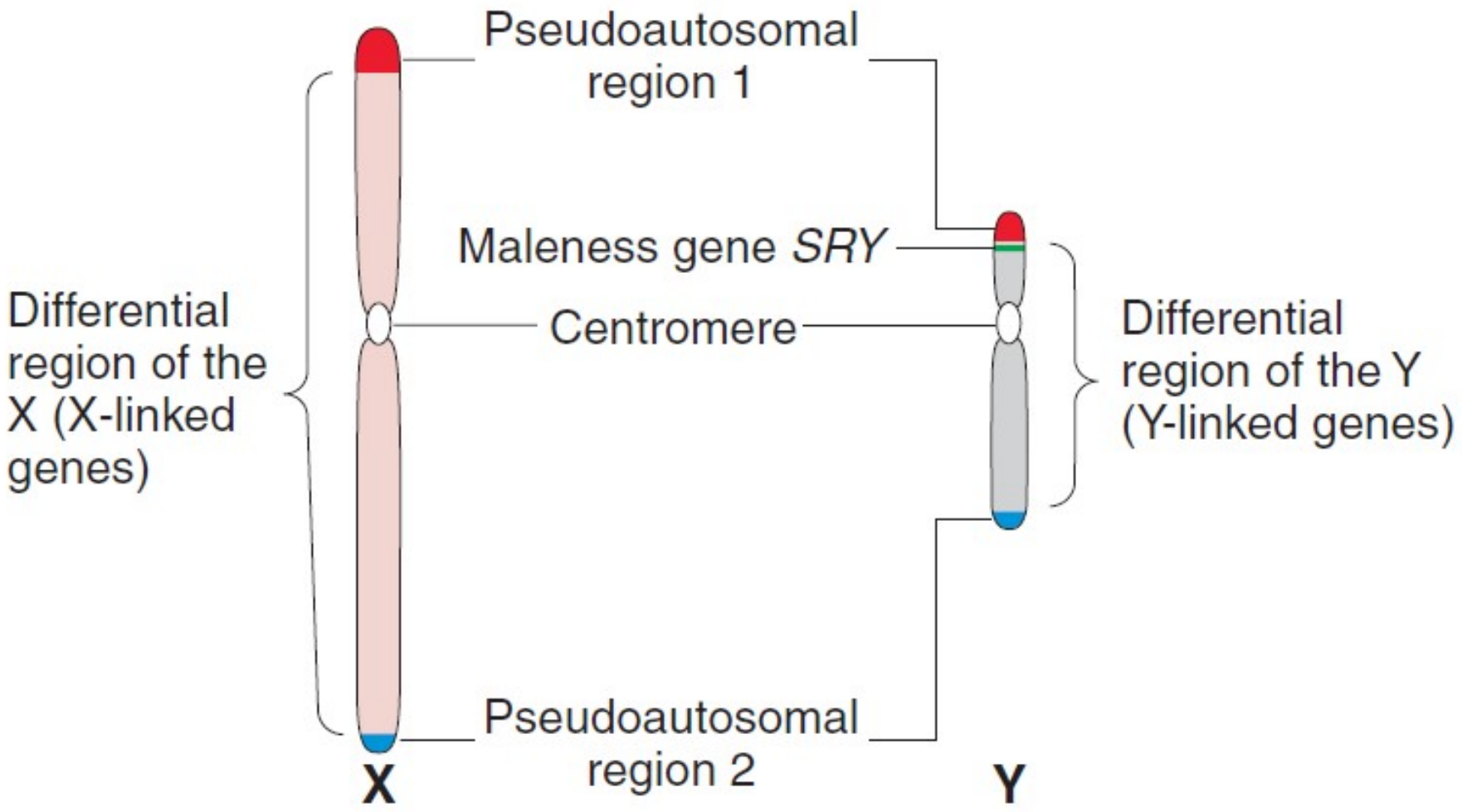
Human karyotype

- **Euchromatin** (2.9 Gbp): the gene-rich, transcriptionally active regions of the nuclear genome
- **Heterochromatin** (0.2 Gbp): tightly packed (condensed), transcriptionally inactive, highly repetitive DNA. Location: centromeres, telomeres.

- **Metacentric chromosomes** have the centromere in the center, such that both arms are of nearly equal length.
- **Acrocentric chromosomes** (13, 15, 21, 22) have unequal arms.



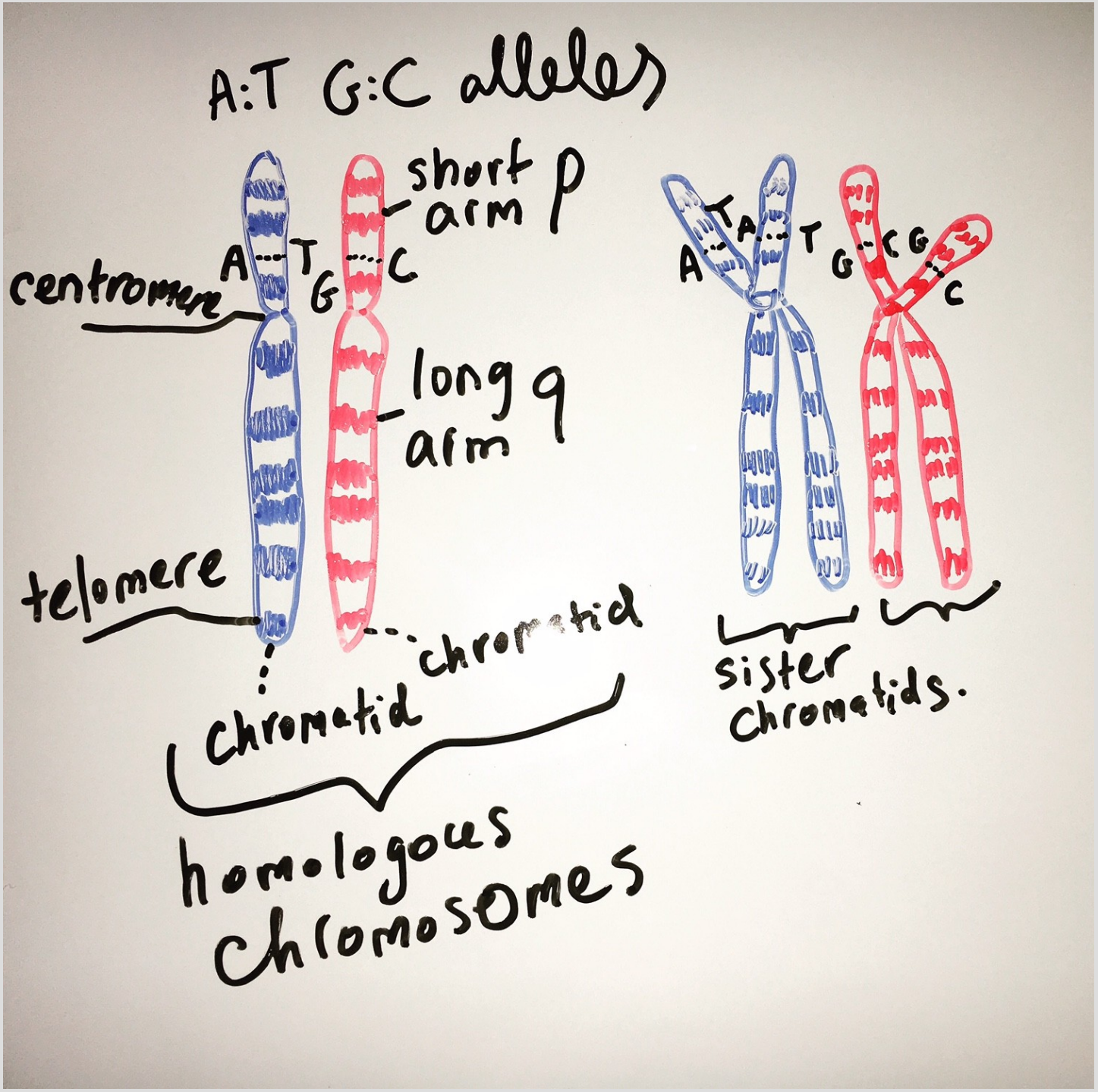
Sex chromosomes



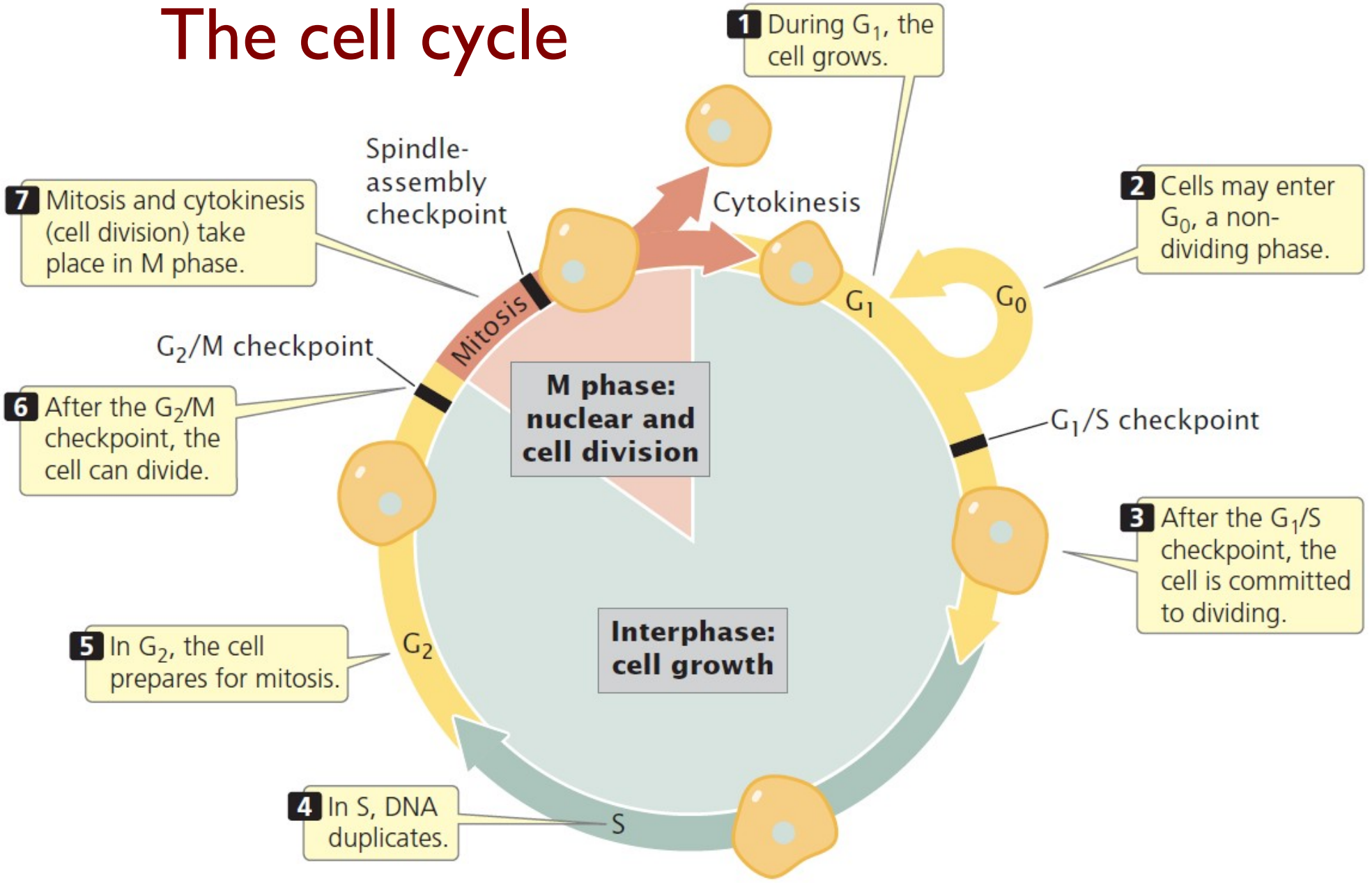
Women: XX, men: XY

Q: transmission of Y chromosome

Griffiths -- *Introduction to Genetic Analysis*

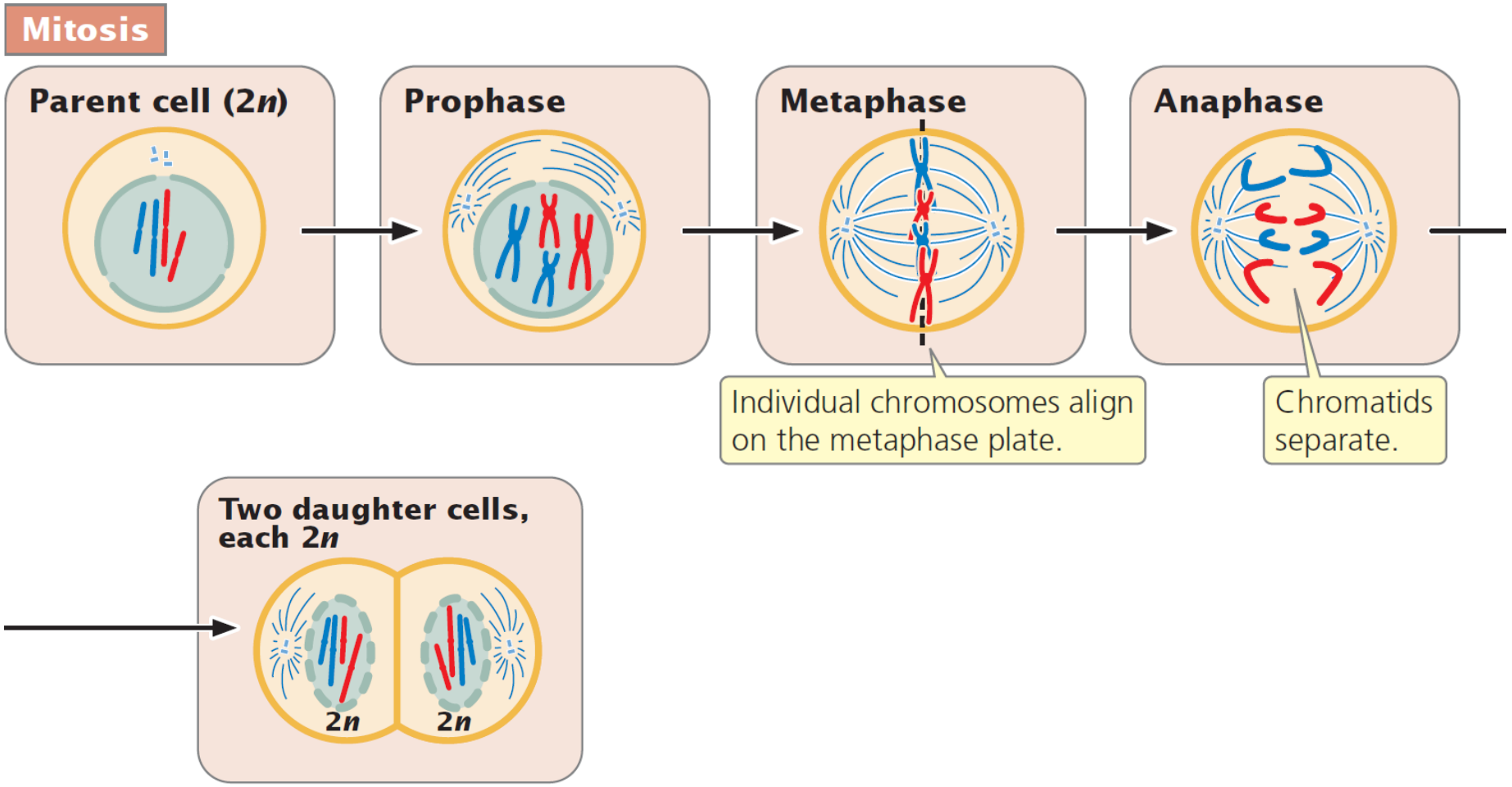

A:T G:C alleles
short arm p
centromere
A
T
G
C
long q arm
telomere
chromatid
chromatid
homologous chromosomes
A
T
T
A
G
C
C
G
sister chromatids.

The cell cycle



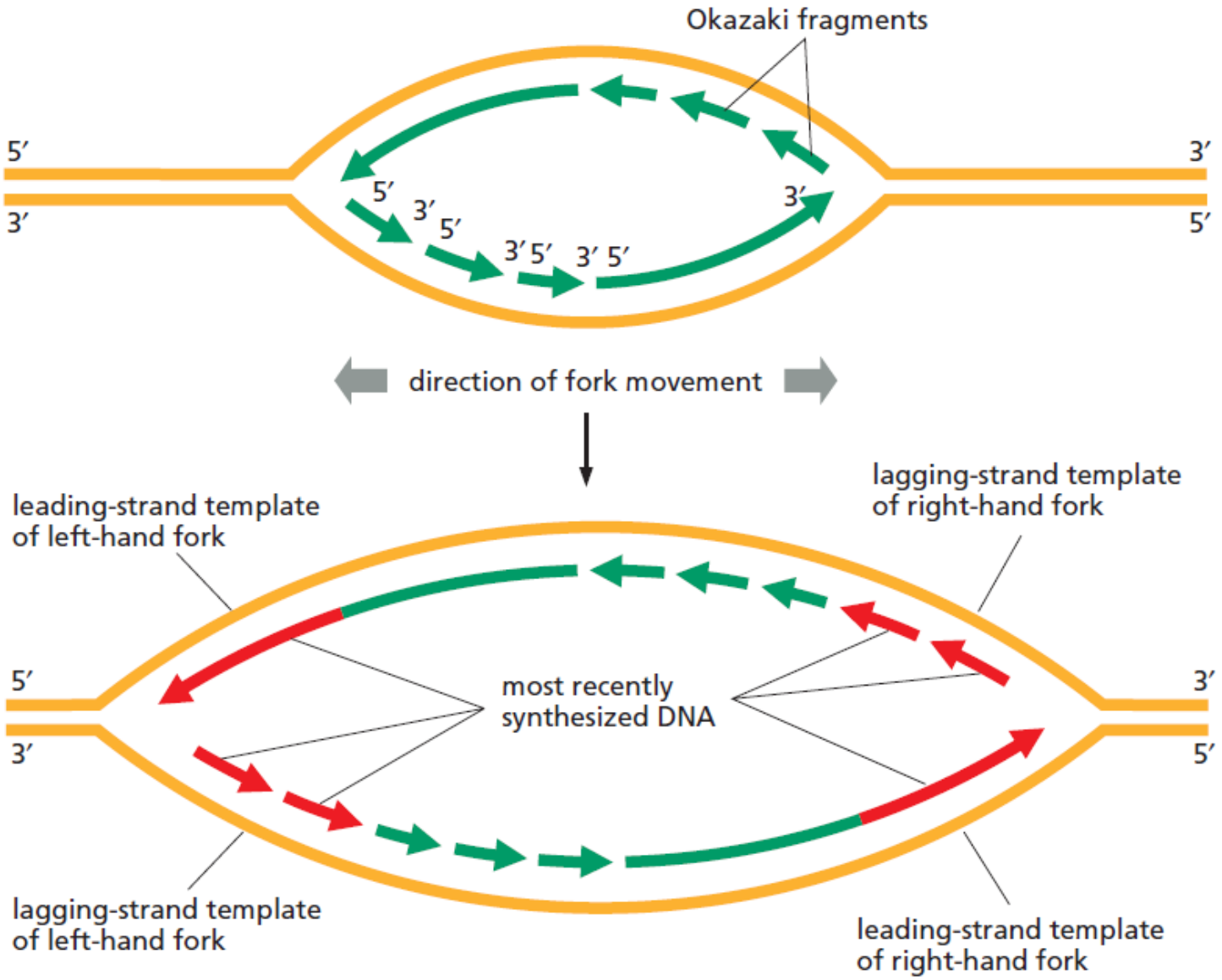
2.7 The cell cycle consists of interphase and M phase.

Mitosis



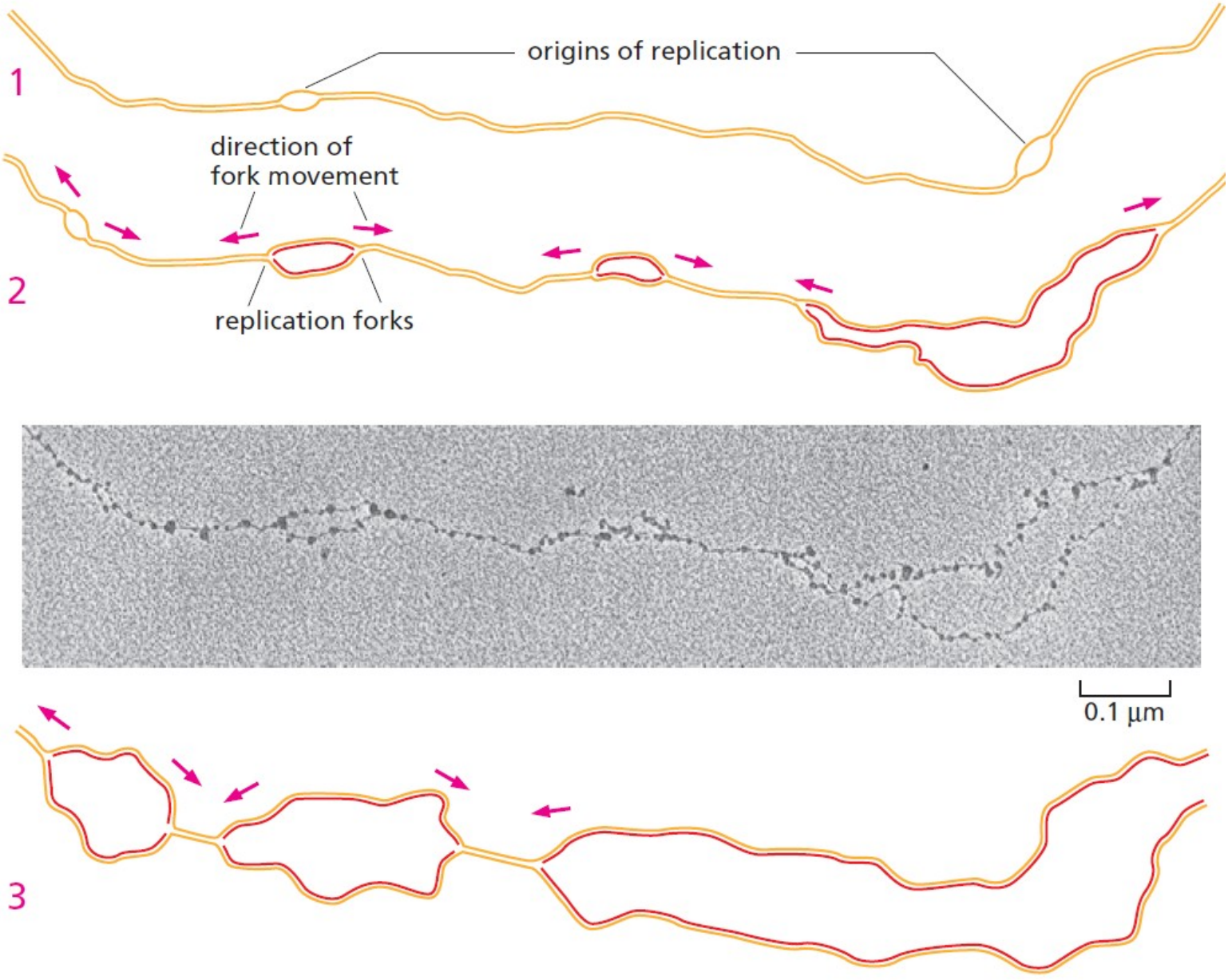
Mitosis: a type of cell division that results in two daughter cells with the set of chromosomes as the parent nucleus, typical of ordinary tissue growth

DNA replication forks



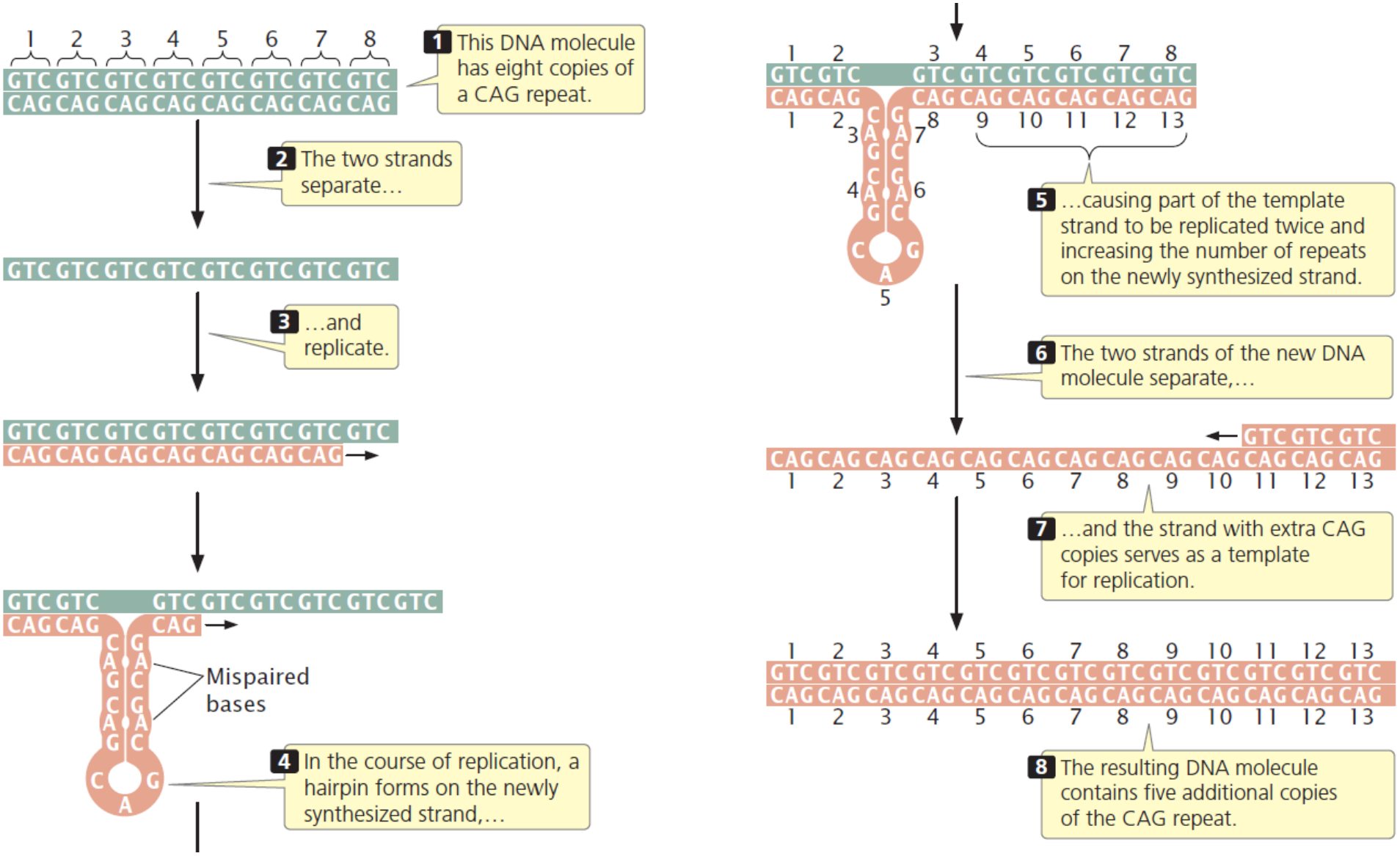
Alberts -- *Essential Cell Biology*

DNA replication forks

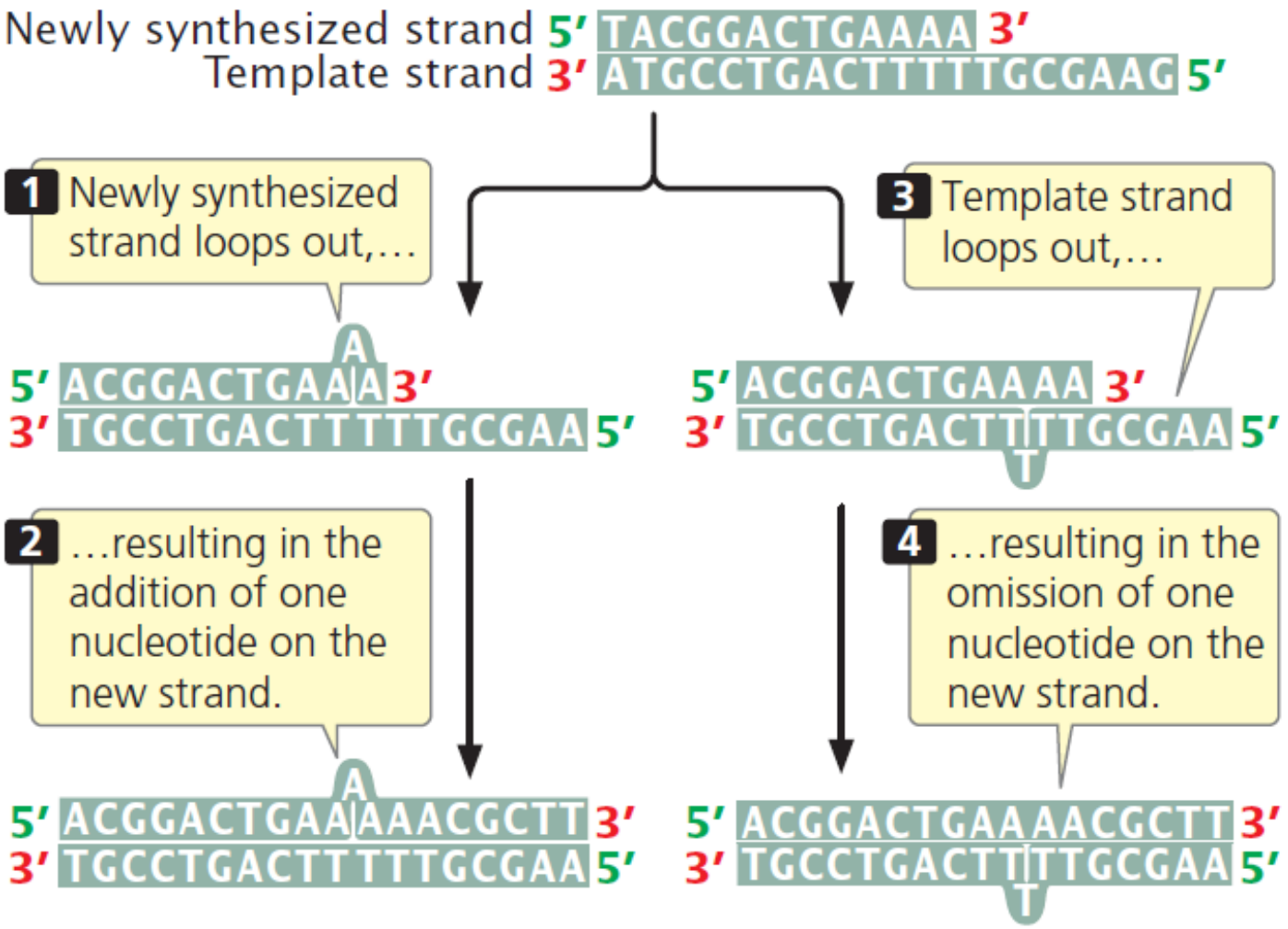


Alberts -- *Essential Cell Biology*

Repeat expansion during replication



Repeat expansion during replication



13.13 Insertions and deletions may result from strand slippage.

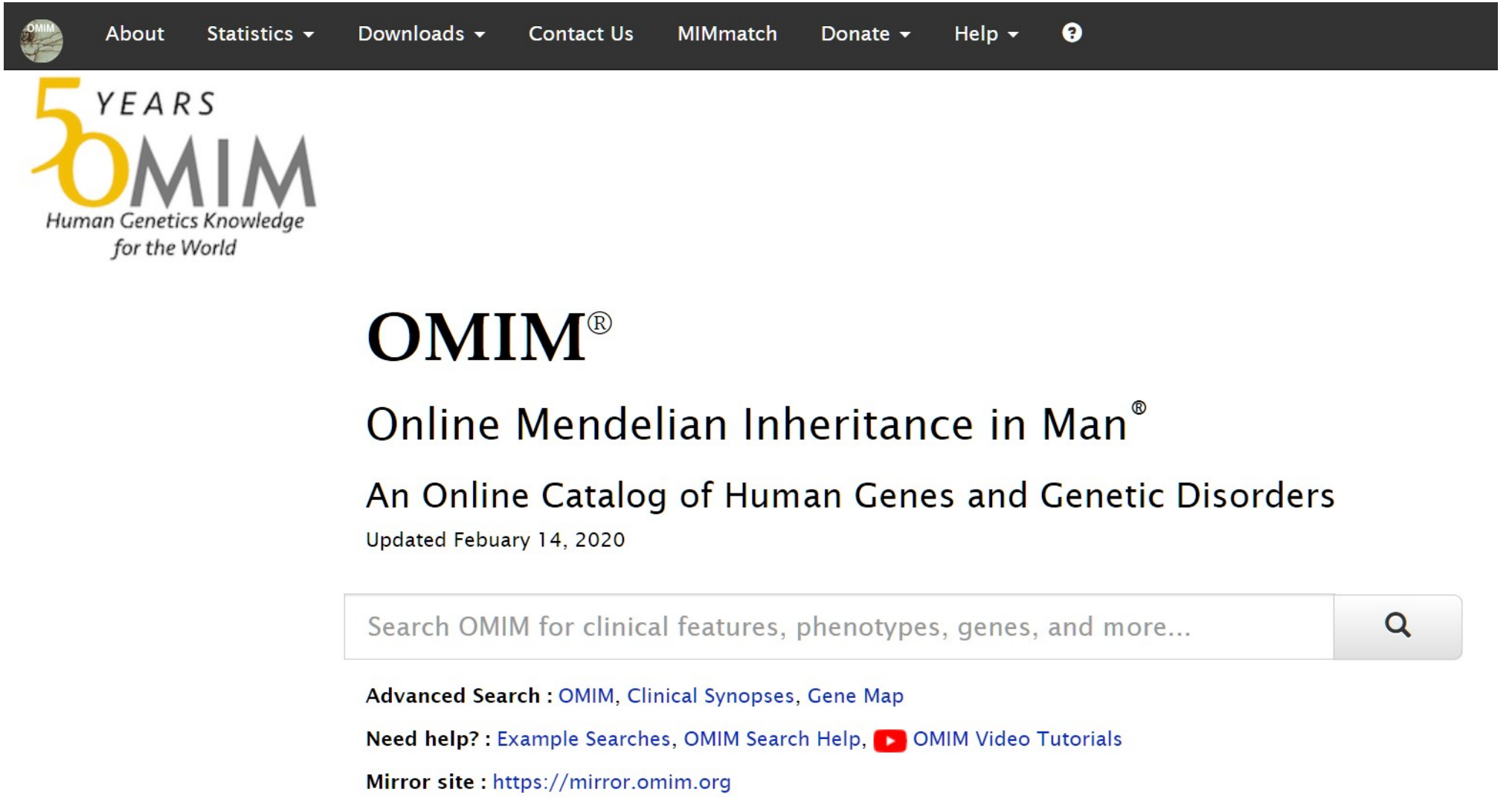
Repeat expansion and disease

Table 13.1 Examples of genetic diseases caused by expanding trinucleotide repeats

Disease	Repeated Sequence	Number of Copies of Repeat: Normal Range	Number of Copies of Repeat: Disease Range
Spinal and bulbar muscular atrophy	CAG	11–33	40–62
Fragile-X syndrome	CGG	6–54	50–1500
Jacobsen syndrome	CGG	11	100–1000
Spinocerebellar ataxia (several types)	CAG	4–44	21–130
Autosomal dominant cerebellar ataxia	CAG	7–19	37–220
Myotonic dystrophy	CTG	5–37	44–3000
Huntington disease	CAG	9–37	37–121
Friedreich ataxia	GAA	6–29	200–900
Dentatorubral-pallidoluysian atrophy	CAG	7–25	49–75
Myoclonus epilepsy of the Unverricht–Lundborg type*	CCCCGCCCCGCG	2–3	12–13

Exercise: find related genes in OMIM database

What is MIM/OMIM ?



Mutations

Mutations are random changes in DNA sequences

Mutations are the cause of all genetic variation and genetic disease.

Variants: mutations (recent changes) OR polymorphisms (segregating in a population) OR engineered (non-random) changes

Mechanisms of mutation:

- Spontaneous replication errors
- Endogenous (spontaneous) DNA damage: deamination, depurination
- Exogenous (induced) DNA damage: chemical agents, radiation

Mutations

Single nucleotide variant: change of the base of a single DNA nucleotide (90%)

- Transition (G>A, C>T)
- Transversion (C>G, etc)

Short deletion: removal of few (<50bp?) nucleotides (6%)

- Deletion of a unique sequence
- Contraction of a short repeat

Short insertion: addition of few (<50bp?) nucleotides (2%),

- Insertion of a unique sequence
- Expansion of a short repeat

Structural variant (2%): sequence change ~1 kbp and larger in size

- Balanced

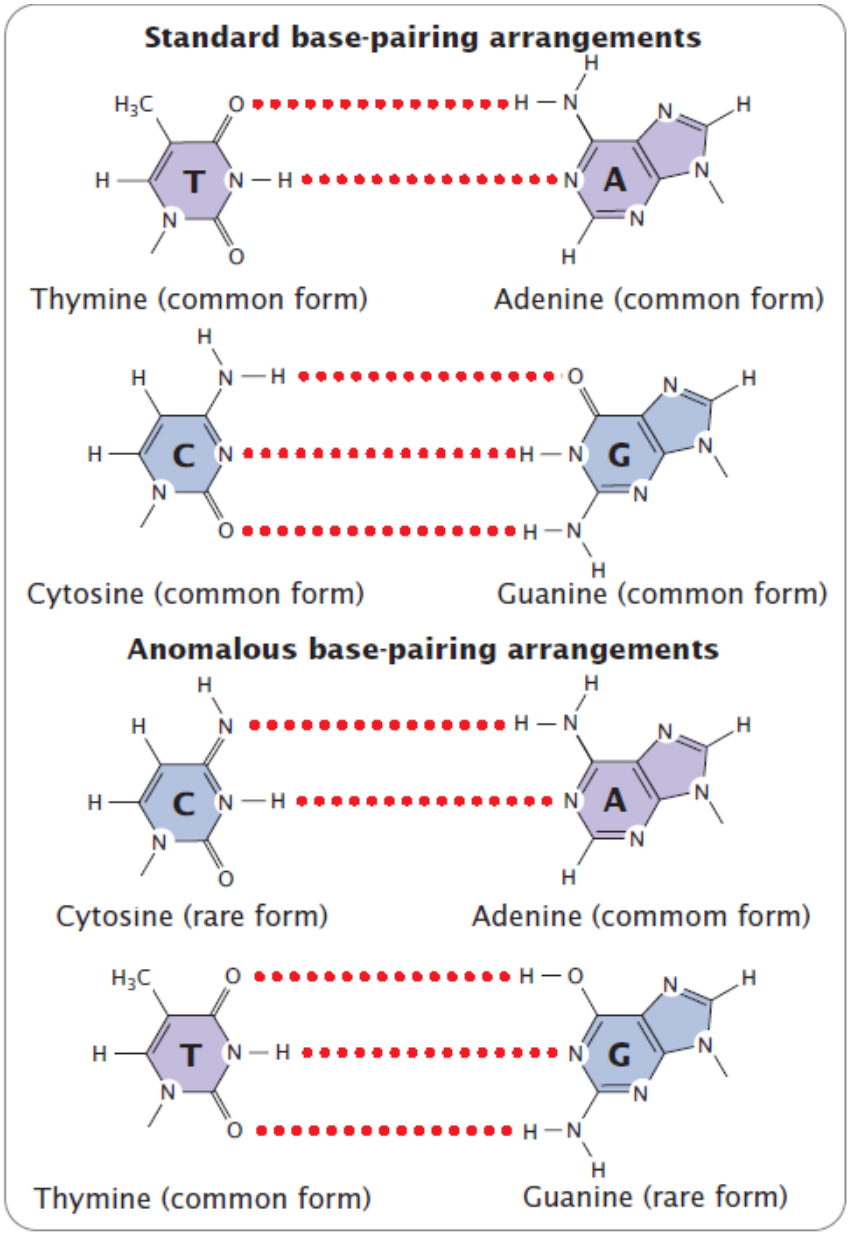
  Inversion or translocation

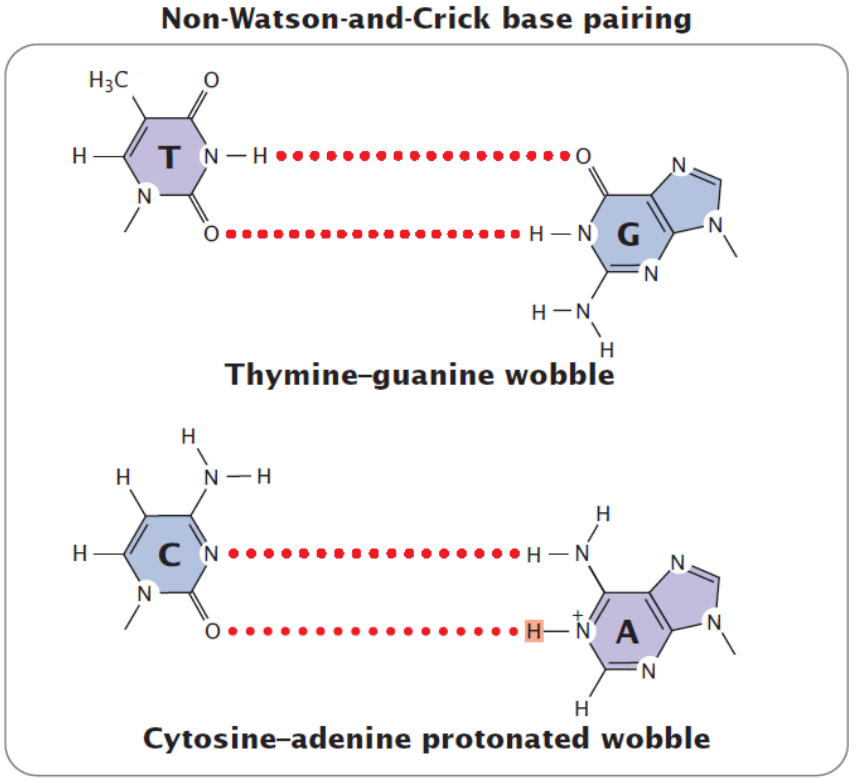
- Unbalanced (aka **CNV, copy number variant**)

  Tandem or dispersed duplication, deletion, insertion

Aneuploidy: wrong number of whole chromosomes: nullisomy, monosomy, trisomy

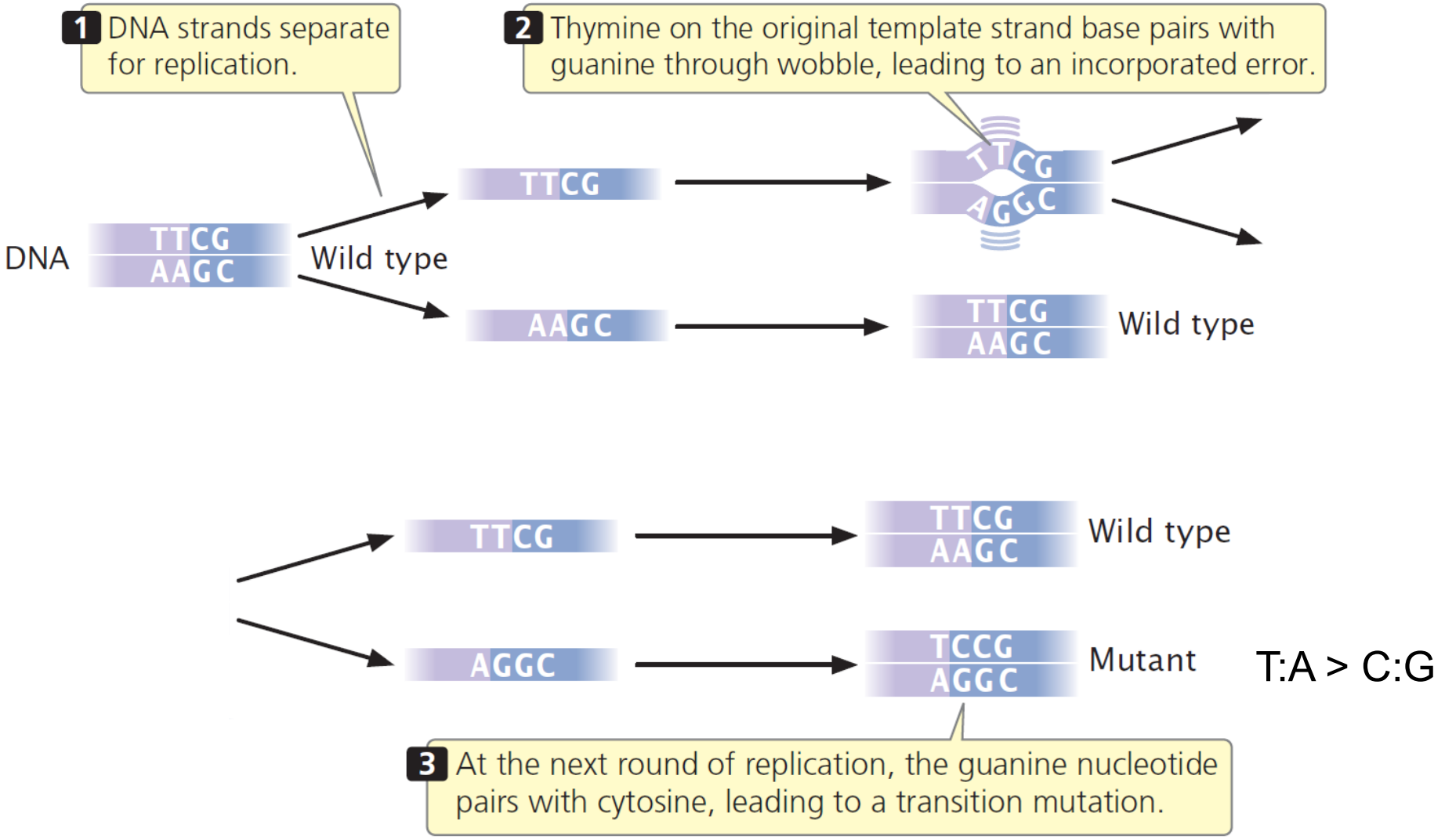
Standard and non-standard base pairing





Pierce -- *Genetics Essentials. Concepts and Connections*

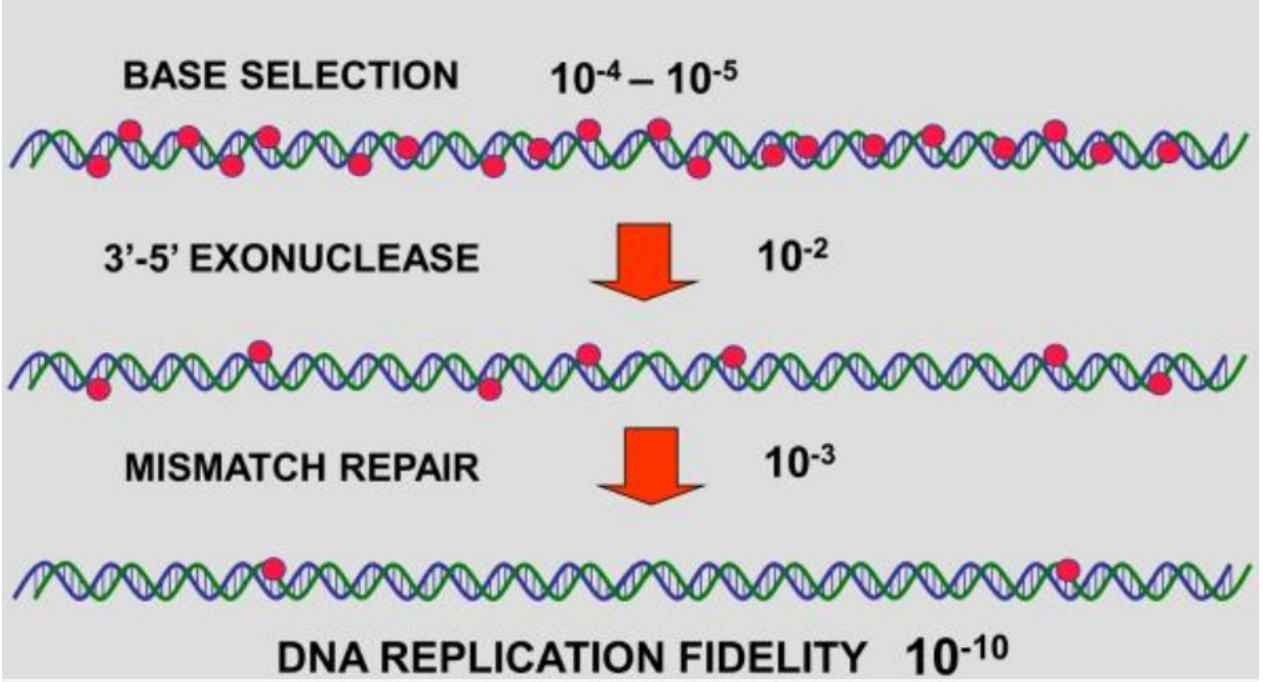
Replication errors become mutations



Mechanisms of replication fidelity

Overall mutation rate: 10^{-10} per nucleotide per replication

1. DNA polymerase: ~10^{-5} error rate
2. Proofreading 3'>5' exonuclease removes 99% mispairings: ~10^{-2}
3. Mismatch repair (MMR) machinery removes and restores DNA fragment around the mismatch: ~10^{-3}



Fijalkowska (2012) *FEMS Microbiology Rev*

Mechanisms of replication fidelity

Overall mutation rate: 10^{-10} per nucleotide per replication

TABLE 6–1 ERROR RATES	
US Postal Service on-time delivery of local first-class mail	13 late deliveries per 100 parcels
Airline luggage system	1 lost bag per 200
A professional typist typing at 120 words per minute	1 mistake per 250 characters
Driving a car in the United States	1 death per 10^4 people per year
DNA replication (without mismatch repair)	1 mistake per 10^7 nucleotides copied
DNA replication (including mismatch repair)	1 mistake per 10^9 nucleotides copied

Mutation rate and its consequences

S: mutation rate per nucleotide per cell division
K: the average number of germline cell divisions per generation, from zygote to zygote (~30 in females, ~60–500 in males)
N: genome size
Mutation rate per genome: $S \times K \times N$

~10^{-10} per nucleotide per cell division (or ~10^{-8} per generation, because there are ~100 cell divisions and rounds of DNA replication per human generation
$\Longrightarrow$ ~100 *de novo* mutations in a newborn

1) ~1% of all newborns being affected by a serious disease due to a de novo mutation. If the mutation rate were 100 times higher, 10^{-8} per cell division, we would immediately **go extinct**.
2) 10^{14} cells in human body $\Longrightarrow$ total number of somatic mutations in each person ?

Mutation rate and its consequences

Genes Genet. Syst. (2019) **94**, p. 13–22

Spontaneous *de novo* germline mutations in humans and mice: rates, spectra, causes and consequences

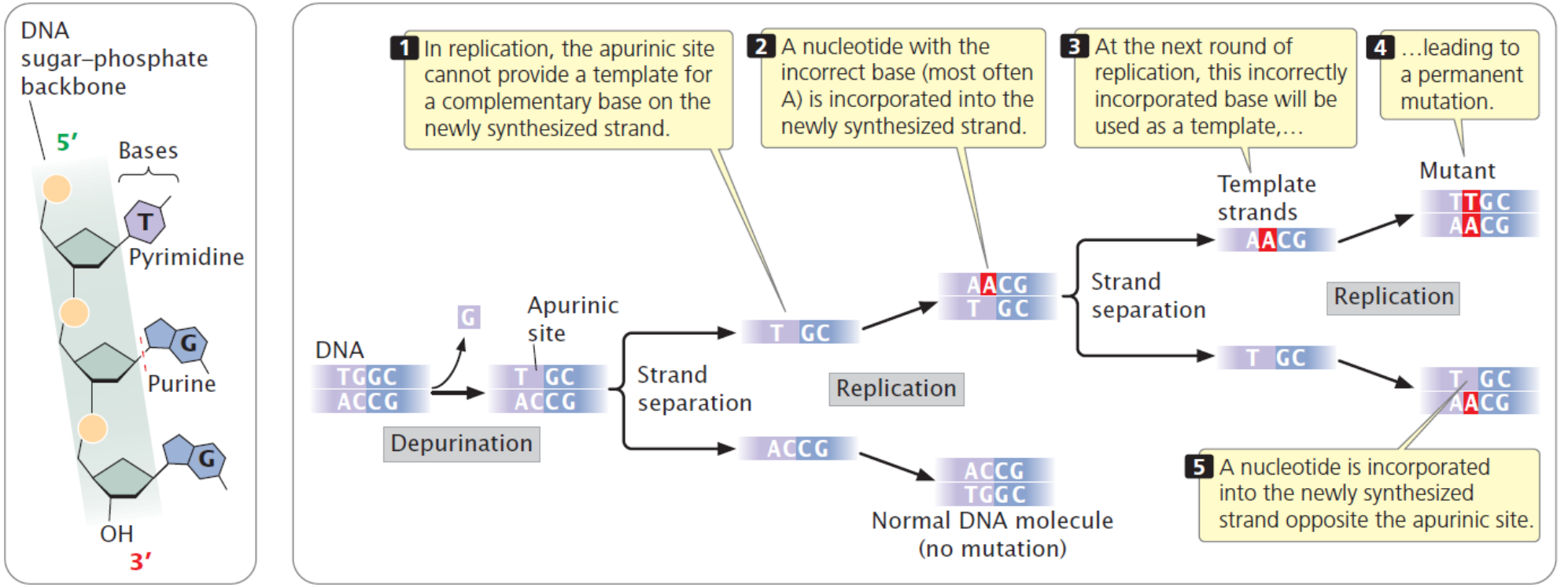
Mizuki Ohno*

The human body consists of approximately 10^{14} cells and undergoes approximately 10^{16} cell divisions in a lifetime, resulting in **over 10^{15} cumulative mutations per individual** (Frank, 2014).

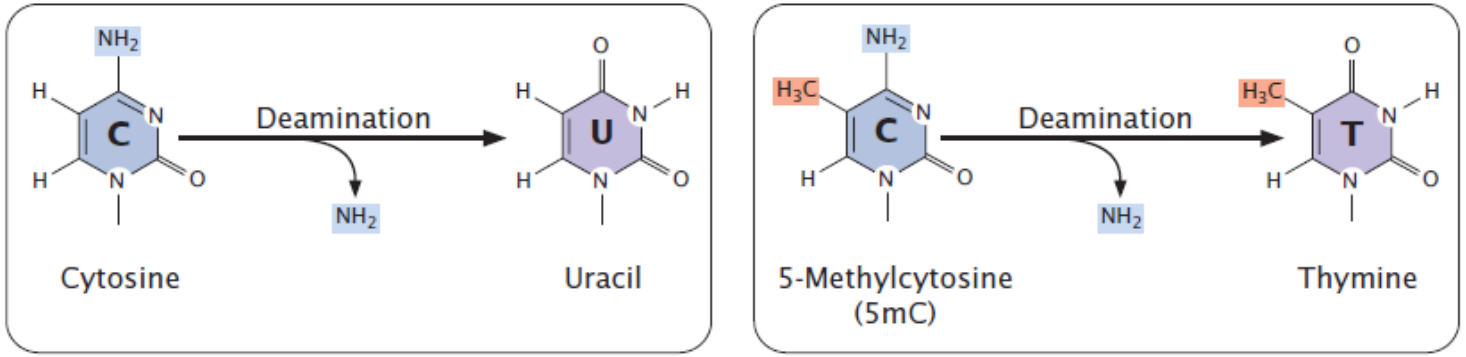
If 10^6 stem cells in intestinal tissue generate transient daughter cells once a week with a mutation rate of approximately 10^{-9} per nucleotide per cell division, the intestinal epithelium of a 60-year-old human would have accumulated more than 10^9 independent mutations. Thus, **nearly every genomic site is likely to be mutated in at least one cell in this organ** (Lynch, 2010a, 2010b).

Endogenous DNA damage

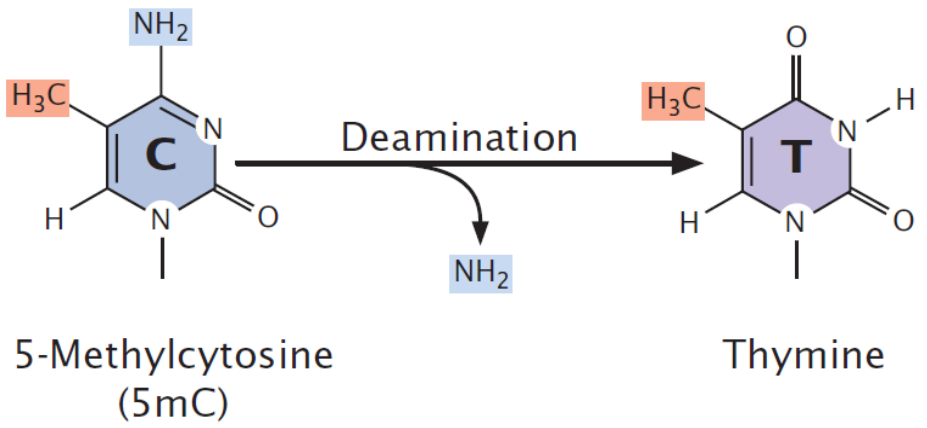
Depurination G:C → A:T



Deamination C:G → U:A → T:A C:G → 5mC:G → T:G → T:A



Deamination of 5'-methylcytosine



CpG → TpG

Complementary:
CpG → CpA

- Cannot be detected by DNA repair system, because it produces a normal base
- Most mutations occur in male germ cells (M/F = 7:1), because of heavy methylation of sperm DNA and high number of cell divisions
- Example: 46% of point mutations in coagulation factor VIII (*F8*) in unrelated hemophilia A patients
- 23% of all mutations in Human Gene Mutation Database (1998)

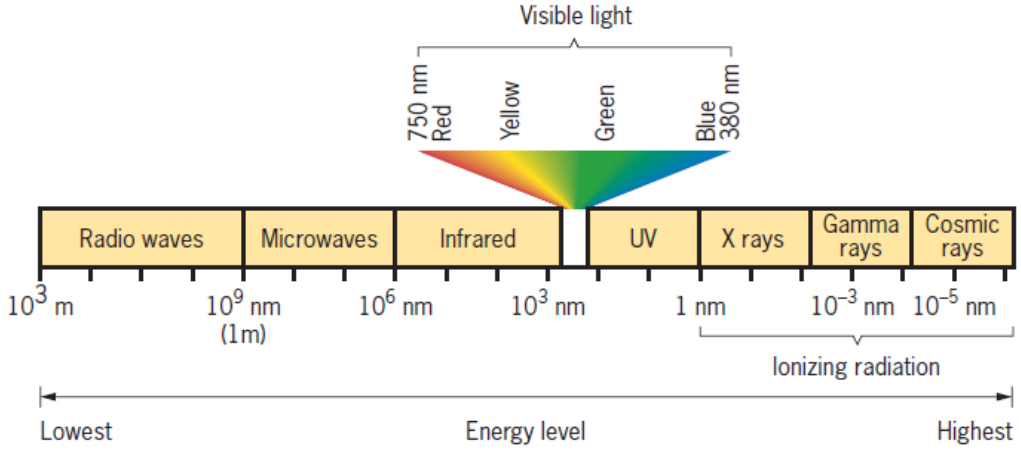
Exogenous DNA damage

Chemical mutagens

- **Base analogs**: 5-bromouracil, 2-aminopurine
- **Alkylating agents**: methyl (–CH_3) and ethyl (–CH_3–CH_2) groups added to nucleotide bases
- **Deamination**: nitrous acid deaminates cytosine, creating uracil
- **Hydroxylamine**: adds a hydroxyl group (–OH) to cytosine
- **Intercalating agents**: proflavin, acridine orange, ethidium bromide, dioxin

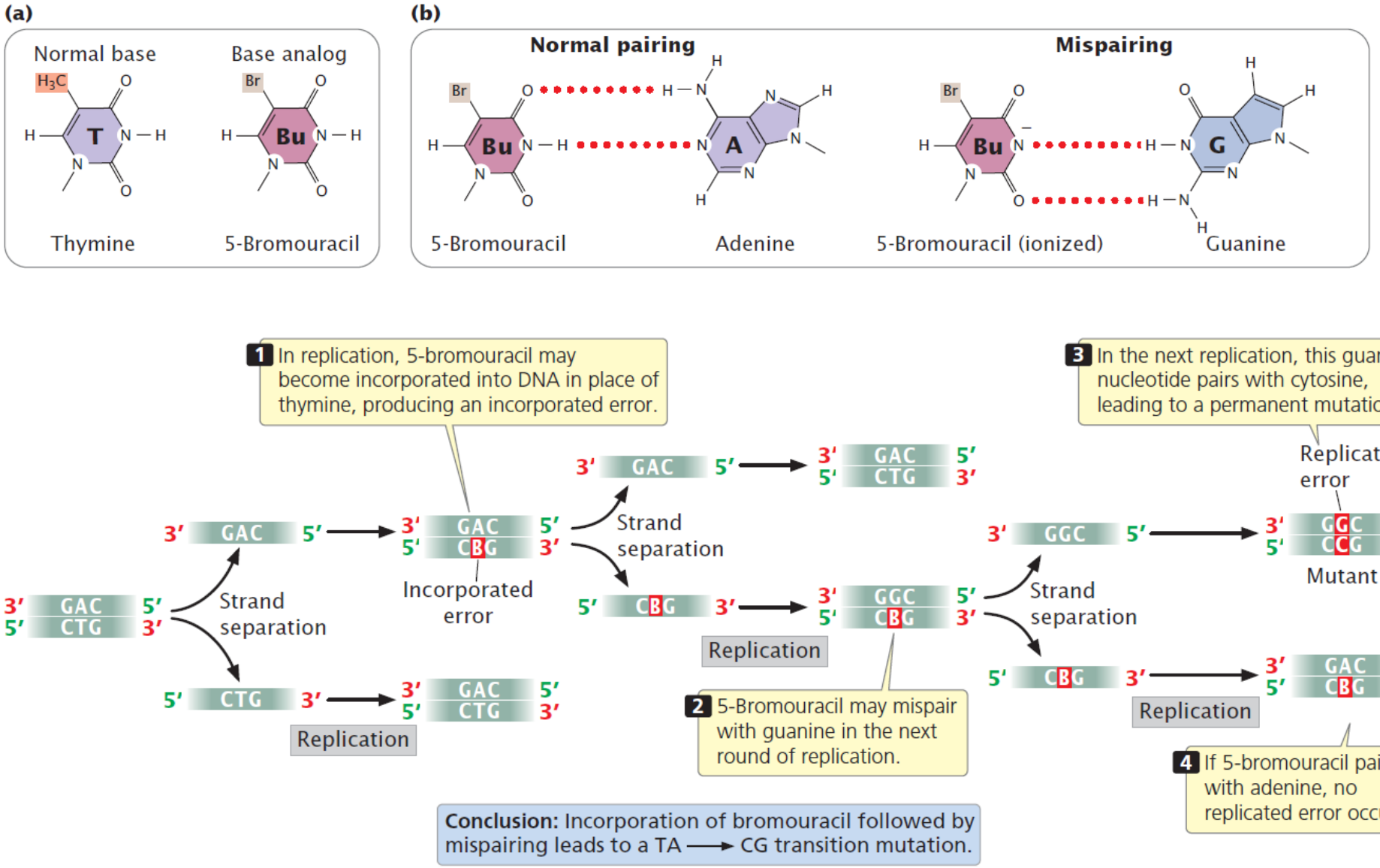
Radiation

- **Ionizing:** ~10^{-5} – 1 nm
- **Ultra-violet:** ~1 – 380 nm



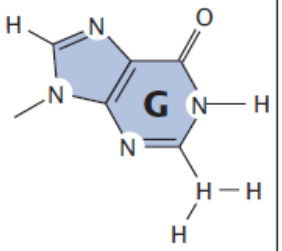
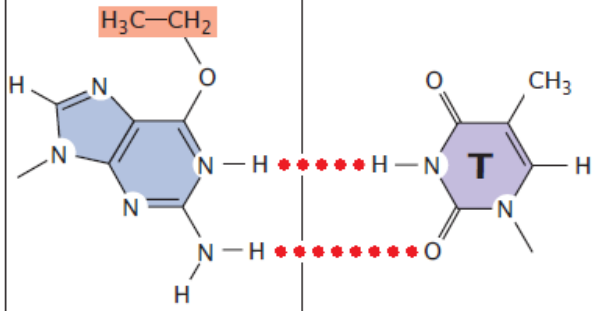
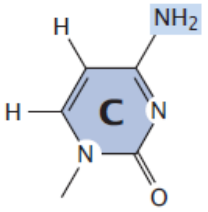
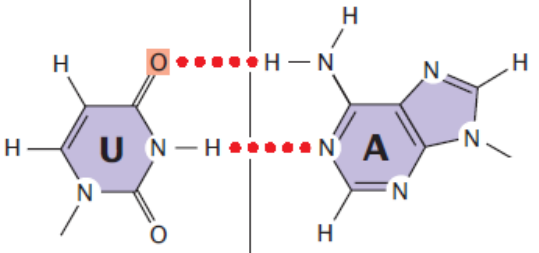
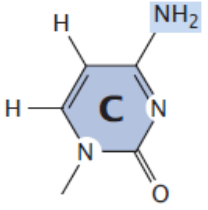
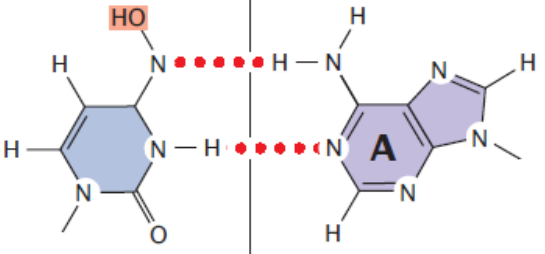
Exogenous DNA damage

Chemical mutagens: 5-bromouracil

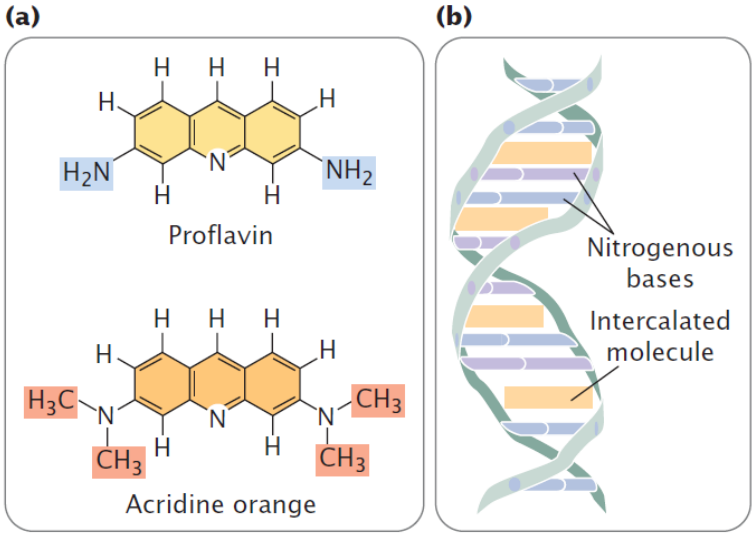


Exogenous DNA damage

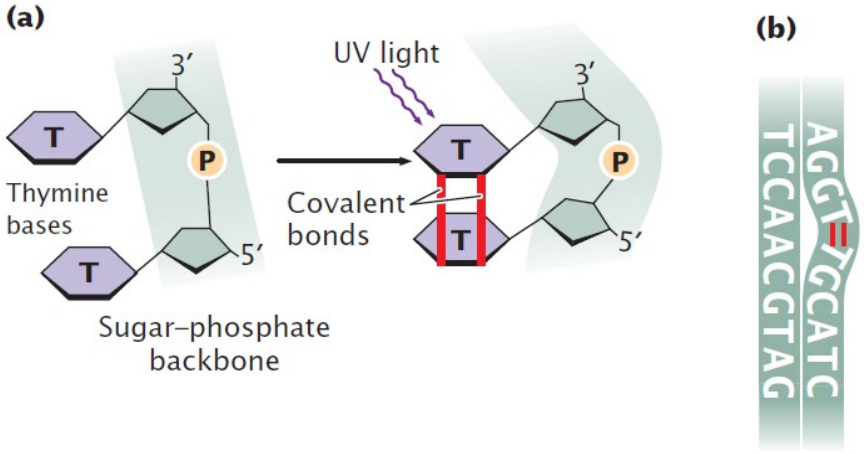
Chemical mutagens

	Original base	Mutagen	Modified base	Pairing partner	Type of mutation
(a)	 Guanine	EMS → Alkylation	 O^6-Ethylguanine	Thymine	CG → TA
(b)	 Cytosine	Nitrous acid (HNO_2) → Deamination	 Uracil	Adenine	CG → TA
(c)	 Cytosine	Hydroxylamine (NH_2OH) → Hydroxylation	 Hydroxylamino-cytosine	Adenine	CG → TA

Exogenous DNA damage



13.20 **Intercalating agents such as proflavin and acridine orange insert themselves between adjacent bases in DNA, distorting the three-dimensional structure of the helix and causing single-nucleotide insertions and deletions in replication.**



13.21 **Pyrimidine dimers result from ultraviolet light.**
(a) Formation of thymine dimer. (b) Distorted DNA.

Intercalating agents: distorted DNA ⟹ insertions and deletions

Ionizing radiation:

- Free radicals, reactive ions ⟹ altered bases
- Double-strand breaks

UV light: Pyrimidune dimers (TpT, CpC, CpT) ⟹ distorted DNA ⟹ replication blocked ⟹ apoptosis or continued error-prone replication

Endogenous DNA damage

Depurination: about 5000 adenine or guanine bases are lost every day from each nucleated human cell by spontaneous hydrolysis of the base-sugar link

Deamination: at least 100 cytosines each day in each nucleated human cell are spontaneously deaminated to produce uracil.

Attack by reactive oxygen species: highly reactive superoxide anions and related molecules are generated as a by-product of oxidative metabolism in mitochondria. They can also be produced by the impact of ionizing radiation on cellular constituents. These reactive oxygen species attack purine and pyrimidine rings.

Nonenzymatic methylation: accidental nonenzymatic DNA methylation by S-adenosyl methionine produces about 300 molecules per cell per day of the cytotoxic base 3-methyl adenine, plus a quantity of the less harmful 7-methyl guanine.

Strachan, Read. *Human Molecular Genetics*, Chapter 13

Exogenous DNA damage

Ionizing radiation: gamma- and X-rays can cause single-strand or double-strand breaks in the sugar-phosphate backbone.

Ultraviolet radiation: UV-C rays (with a wavelength of about 260 nm) are especially damaging, but the major source of UV damage in humans is from the UV-B rays (260-315 nm) in sunlight that can penetrate the ozone layer. UV radiation causes cross-linking between adjacent pyrimidines on a DNA strand to form cyclobutane pyrimidine dimers and other abnormal photoproducts.

Environmental chemicals: these include hydrocarbons (for example, in cigaretre smoke), some plant and microbial products such as the aflatoxins found on moldy peanuts, and chemicals used in cancer chemotherapy. Alkylating agents can transfer a methyl or other alkyl group onto DNA bases and can cause cross-linking between bases within a strand or between different DNA strands.

Strachan, Read. *Human Molecular Genetics*, Chapter 13

DNA repair mechanisms

One strand affected:

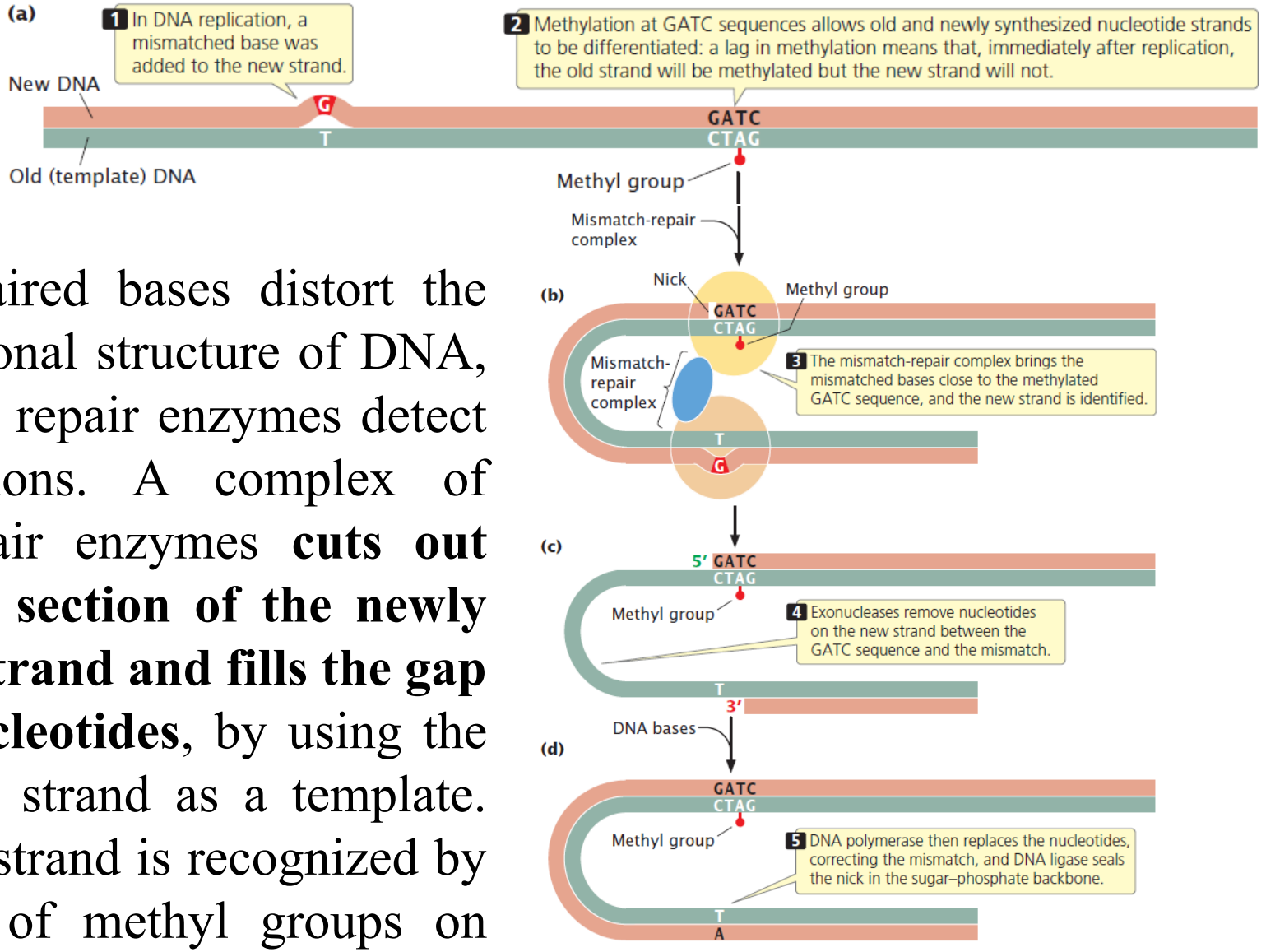
- **Mismatch repair (MMR)** during replication
- **Direct reversal**
- **Base excision repair (BER)** before replication
- **Nucleotide excision repair (NER)** before replication

Both strands affected:

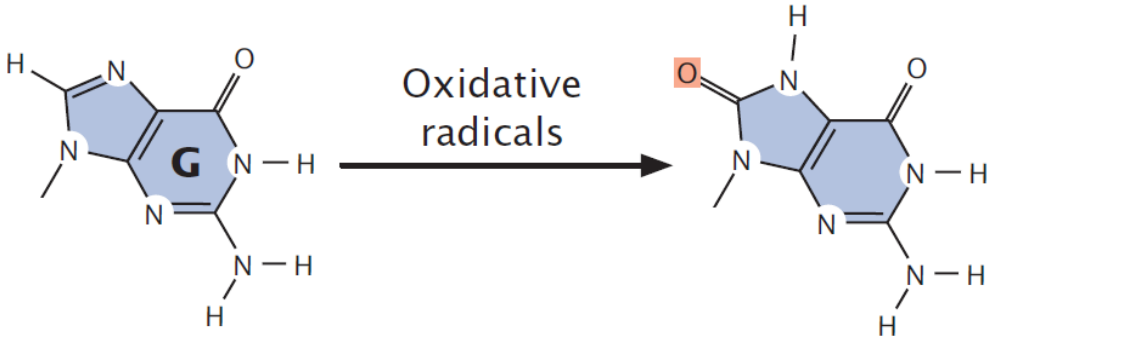
- **Non-homologous end joining (NHEJ)**: ionizing radiation; errors at the replication fork; strong oxidizing agents; metabolites produced in the cell
- **Homologous recombination (HR)**: when a double-strand break occurs shortly after a stretch of DNA has been replicated; at that time, the duplicated helices are still in close proximity to one another

DNA mismatch repair mechanism (MMR) (1)

Incorrectly paired bases distort the three-dimensional structure of DNA, and mismatch repair enzymes detect these distortions. A complex of mismatch-repair enzymes **cuts out the distorted section of the newly synthesized strand and fills the gap with new nucleotides**, by using the original DNA strand as a template. The template strand is recognized by the presence of methyl groups on special sequences of the old strand.



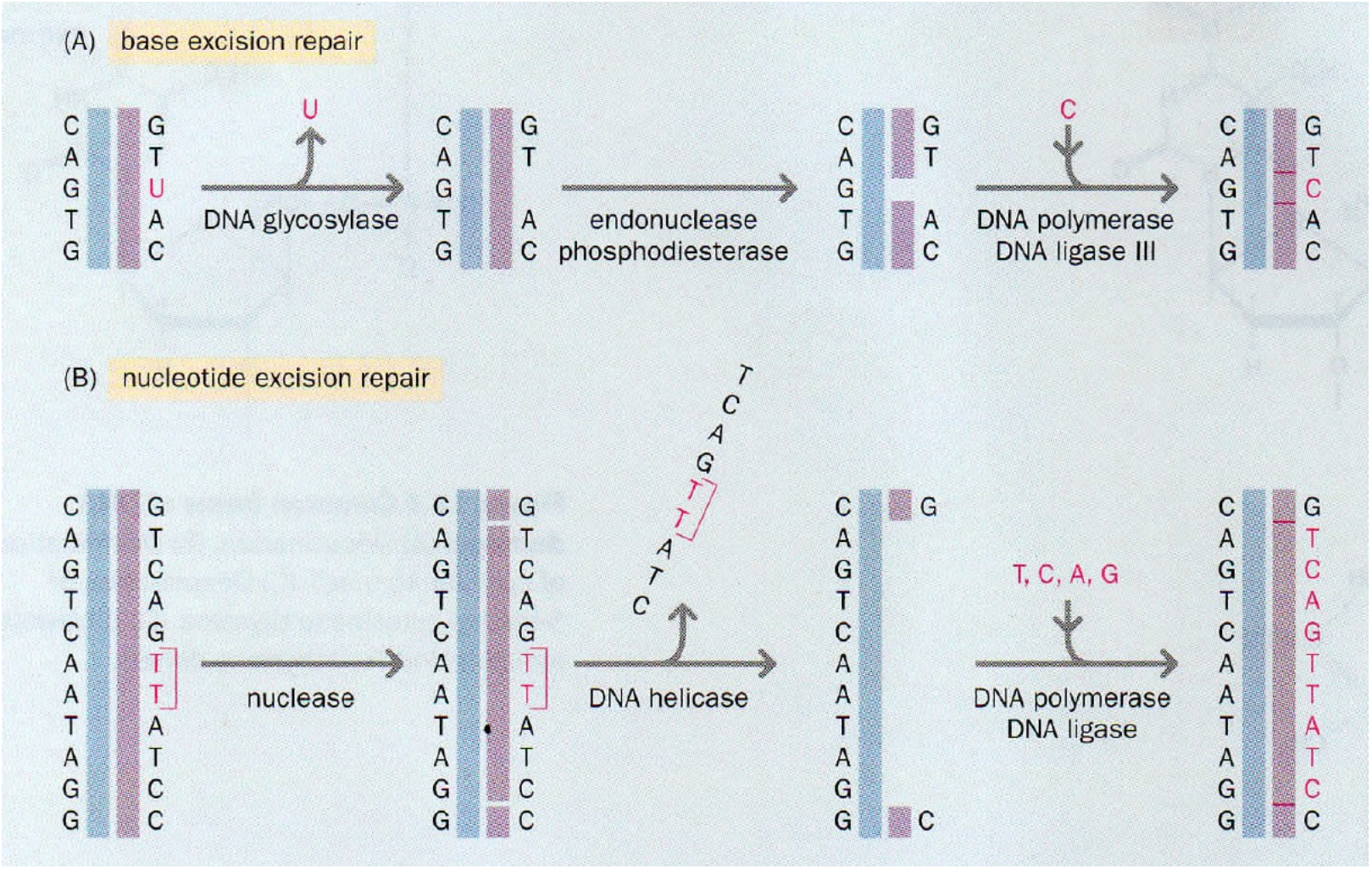
Direct repair (2)



13.28 **Direct repair changes nucleotides back into their original structures.**

Direct repair does not replace altered nucleotides but, instead, changes them back into their original structures. For example, direct repair corrects O^6-methylguanine, an alkylation product of guanine that pairs with adenine, producing G:C→T:A transversions. An enzyme called O^6-methylguanine-DNA methyltransferase removes the methyl group from O^6-methylguanine, restoring the base to guanine.

Base and nucleotide excision repair (3, 4)

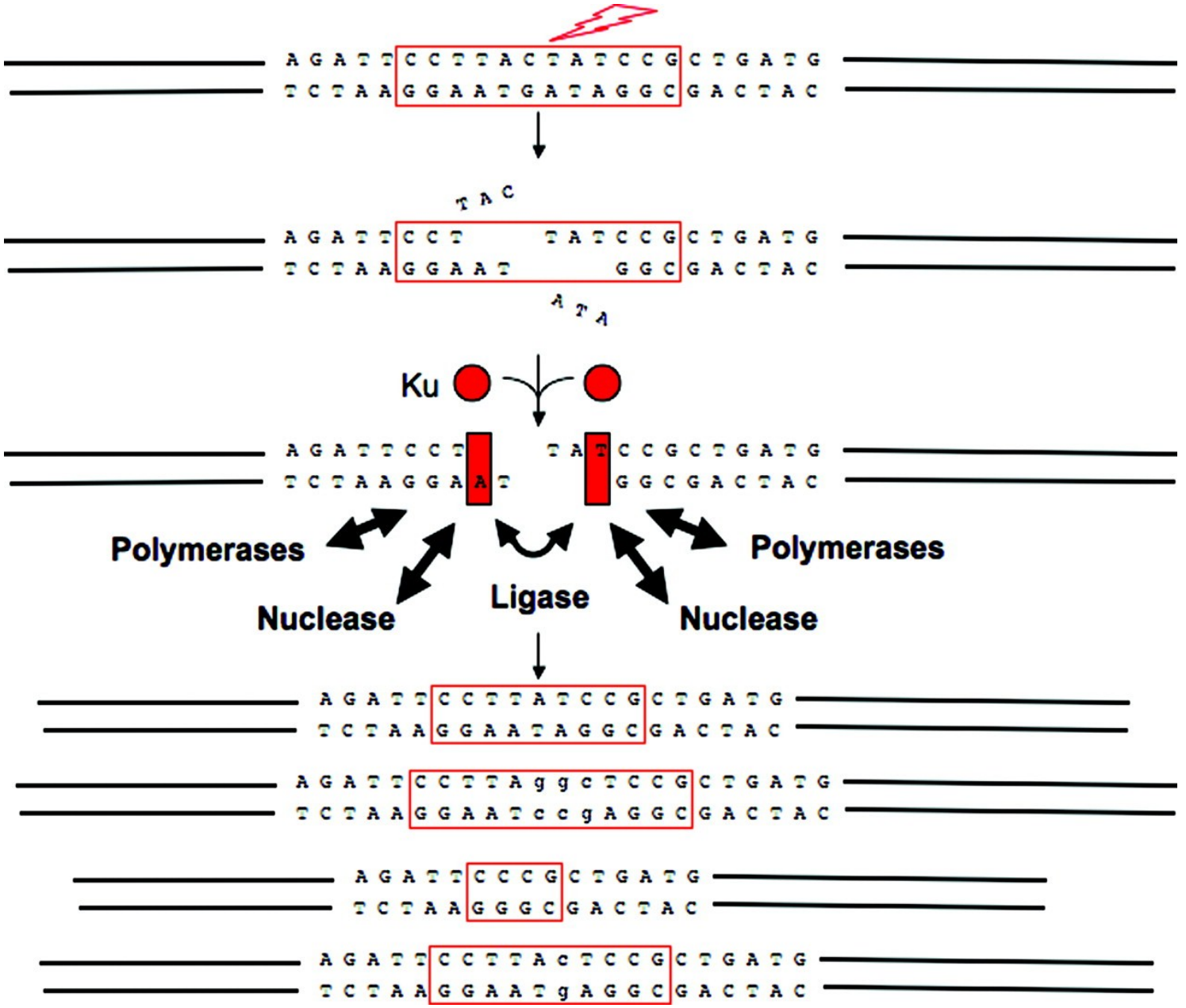


(A) **Base excision repair (BER)** corrects most common DNA damages: ~20,000 bases in each cell per day

(B) **Nucleotide excision repair (NER)** remove thymide dimers and large chemical alterations

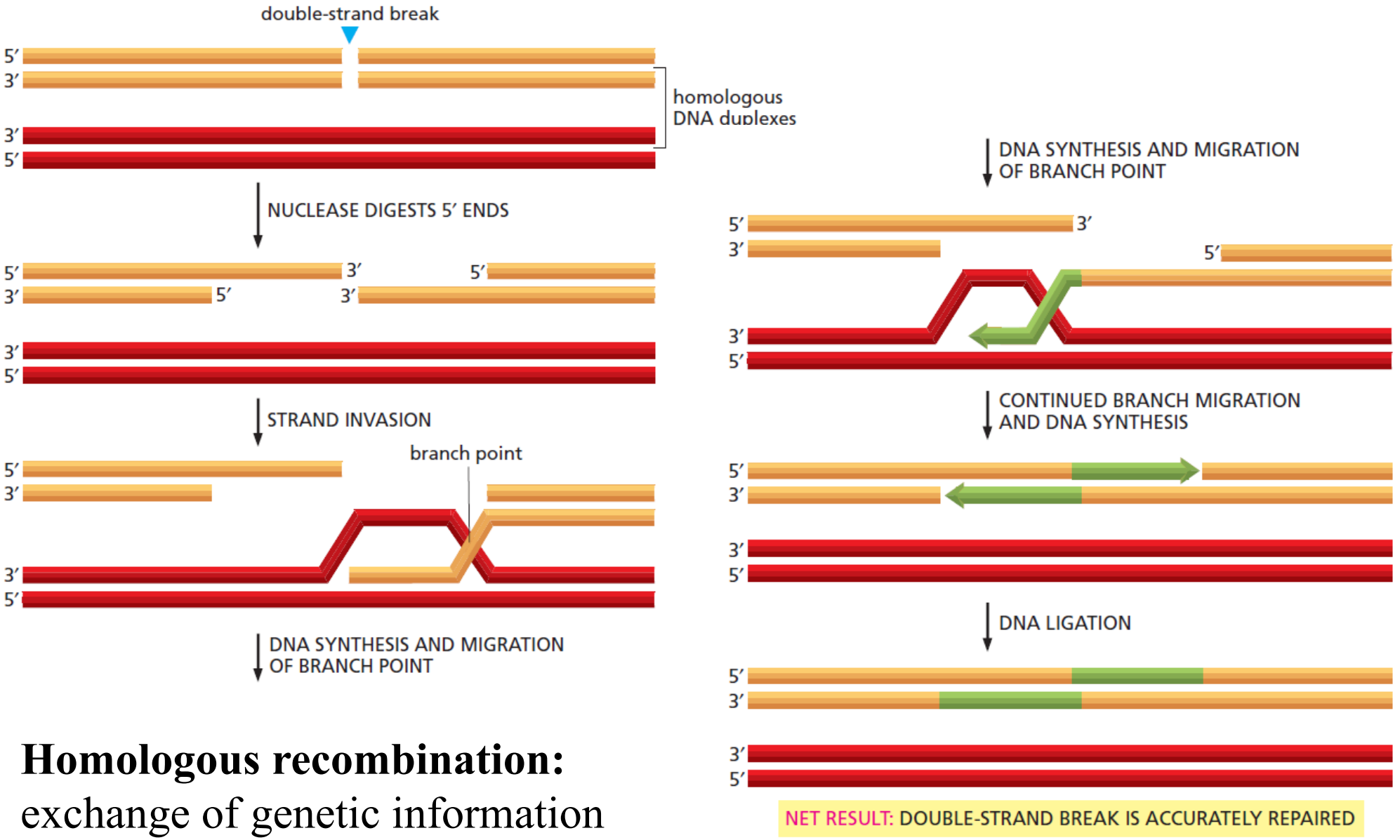
Strachan, Read. *Human Molecular Genetics*

Non-homologous end joining (NHEJ)



Michael R. Lieber (2008) *J. Biol. Chem*.

Repair by homologous recombination (HR)



Homologous recombination:
exchange of genetic information
between homologous DNA molecules

Alberts – *Essential Cell Biology*

G

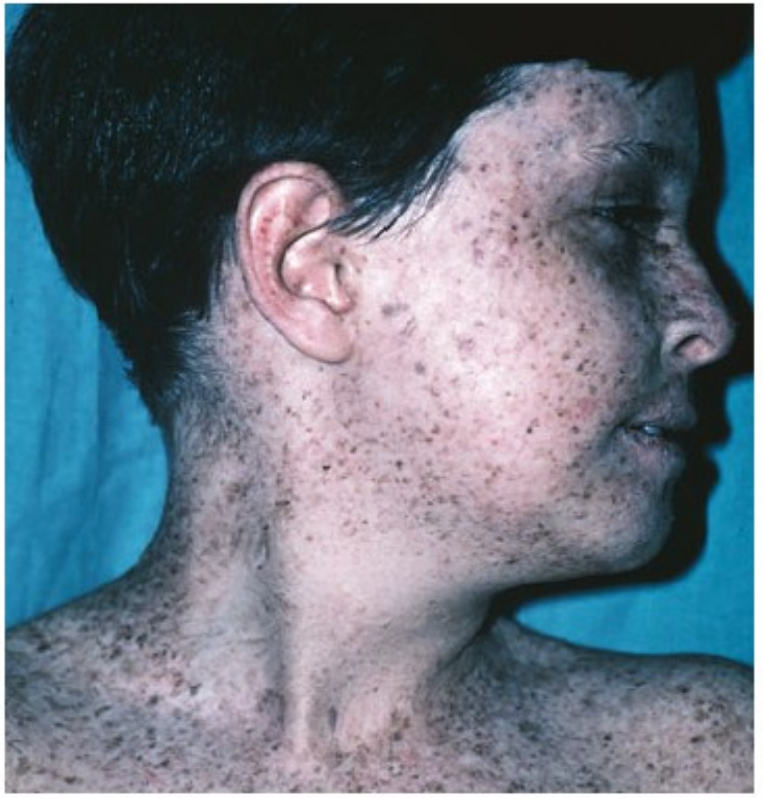
defects in DNA-repair systems

Disease	Symptoms	Genetic Defect
Xeroderma pigmentosum	Frecklelike spots on skin, sensitivity to sunlight, predisposition to skin cancer	Defects in nucleotide-excision repair
Cockayne syndrome	Dwarfism, sensitivity to sunlight, premature aging, deafness, mental retardation	Defects in nucleotide-excision repair
Trichothiodystrophy	Brittle hair, skin abnormalities, short stature, immature sexual development, characteristic facial features	Defects in nucleotide-excision repair
Hereditary nonpolyposis colon cancer	Predisposition to colon cancer	Defects in mismatch repair
Fanconi anemia	Increased skin pigmentation, abnormalities of skeleton, heart, and kidneys, predisposition to leukemia	Possibly defects in the repair of interstrand cross-links
Ataxia telangiectasia	Defective muscle coordination, dilation of blood vessels in skin and eyes, immune deficiencies, sensitivity to ionizing radiation, predisposition to cancer	Defects in DNA-damage detection and response
Li–Fraumeni syndrome	Predisposition to cancer in many different tissues	Defects in DNA-damage response

G

defects in DNA-repair systems

Xeroderma pigmentosum, a rare autosomal recessive condition that includes **abnormal skin pigmentation and acute sensitivity to sunlight**. Persons who have this disease also have a **strong predisposition to skin cancer**, with an incidence ranging from 1000 to 2000 times that found in unaffected people.



The cells of most people with xeroderma pigmentosum are defective in nucleotide excision repair, and many of their pyrimidine dimers (UV from sunlight) remain uncorrected and may lead to cancer.

De novo mutations

De novo mutations (DNM) detected in a genome (exome), for example, by sequencing a mother-father-child trio

DNM rate variation: across the genome; in families; mutational clusters (within an individiual) and mutational hotspots (across individuals)

Factors contributing to DNM rate variation:

- sequence composition and functional context
- replication timing: early / late
- transcriptional activity and chromatin state
- the number of mitoses a cell has undergone (parental age)
- exposure to damaging agents
- the efficiency of the DNA repair
- the amount of time between mitoses

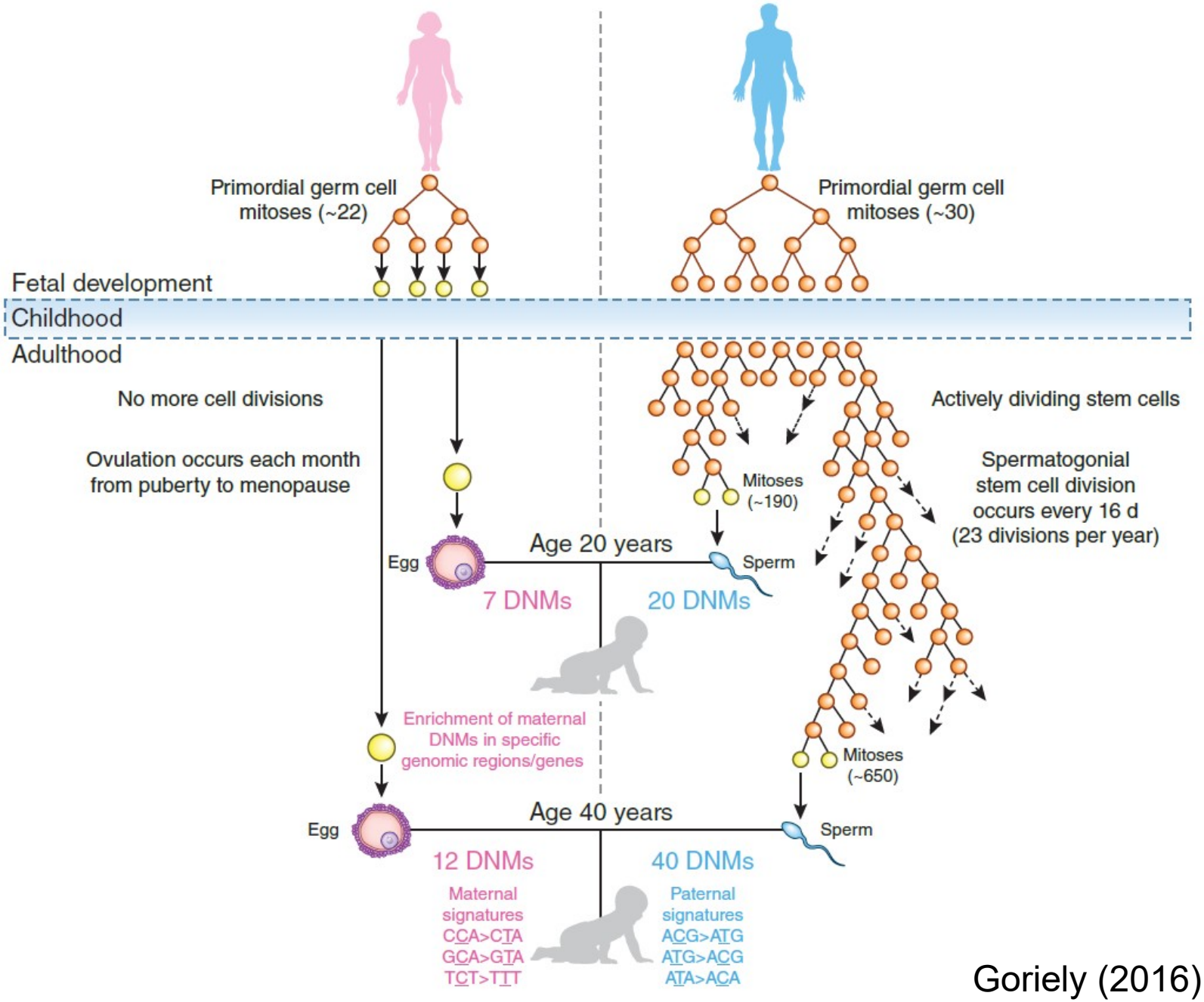
Mutation rates in disease-causing genes

Disorder	MIM	Mutations per Million Gametes	Signs and Symptoms (Phenotype)
X-linked			
Duchenne muscular dystrophy	310200	40–105	Muscle atrophy
Hemophilia A	306700	30–60	Severe impairment of blood clotting
Hemophilia B	306900	0.5–10	Mild impairment of blood clotting
Autosomal Dominant			
Achondroplasia	100800	10	Very short stature
Aniridia	106200	2.6	Absence of iris
Huntington disease	143100	<1	Uncontrollable movements, personality changes
Marfan syndrome	154700	4–6	Long limbs, weakened blood vessel walls
Neurofibromatosis type 1	162200	40–100	Brown skin spots, benign tumors under skin
Osteogenesis imperfecta	166200	10	Easily broken bones
Polycystic kidney disease	600666	60–120	Benign growths in kidneys
Retinoblastoma	180200	5–12	Malignant tumor of retina

Exercise: find related genes in OMIM database

Lewis – Human genetics. Concepts and applications 2009

De novo mutations



Goriely (2016) *Nat Genet*

De novo mutations

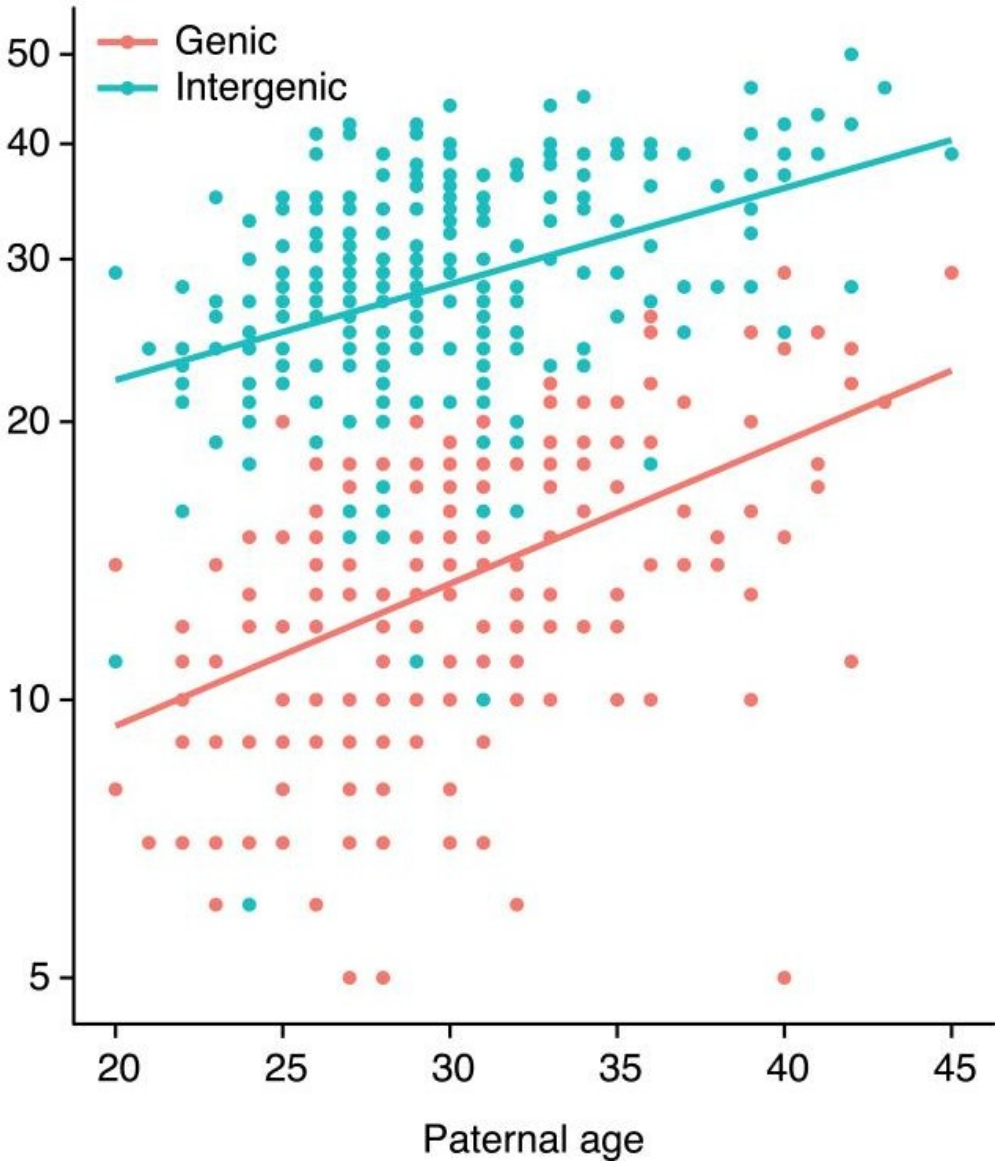
Paternal and maternal DNMs

+1-3 DNMs for each year of paternal age at conception.

+0.24 DNMs for each year of maternal age at conception (non-replicative DNA damage)

~80% of all DNMs are paternal

This effect varies considerably between families



Francioli (2015) *Nat Genet*

De novo mutation spectra

Transitions

- C:G>T:A, deamination of 5-mC and C: 40%
- T:A>C:G, cause unknown: 25%

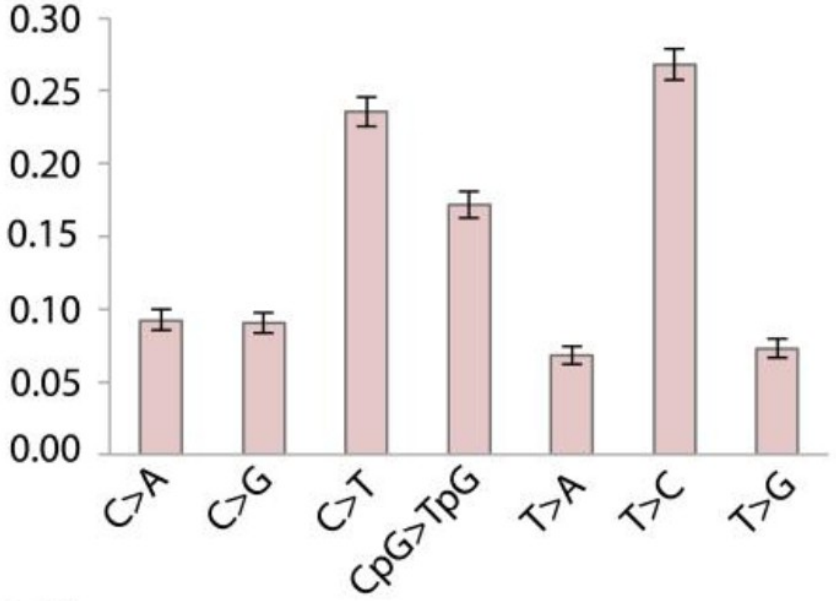
Note: CpG are only ~1% of the genome, so also at non-CpG; but transitions at CpG are ~18x more frequent than non-CpG

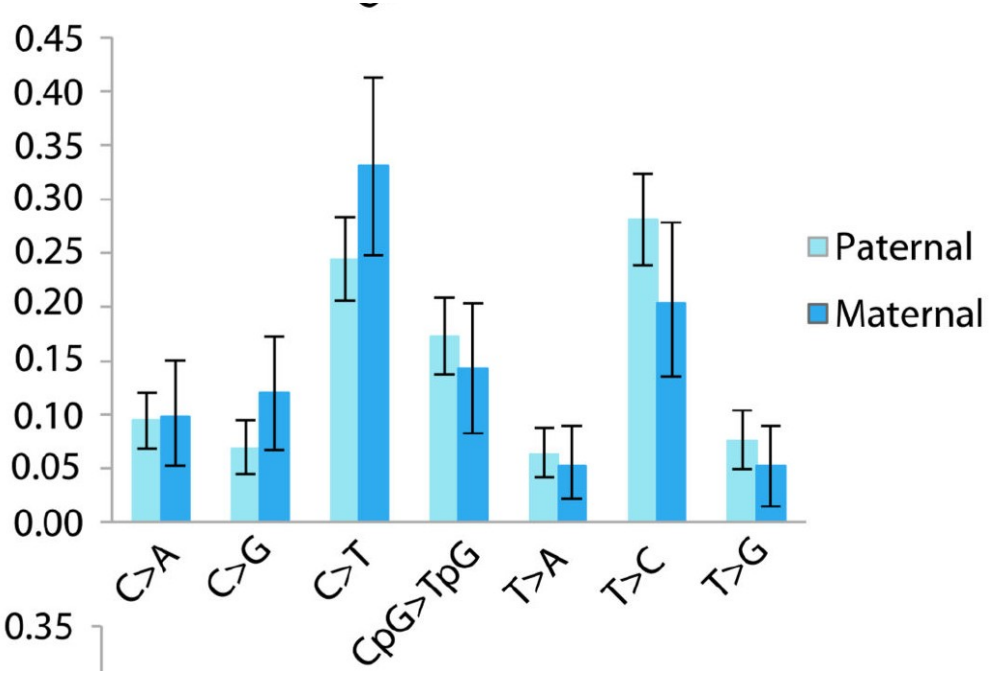
Transversions // occur ~2.5x more frequently at CpG sites

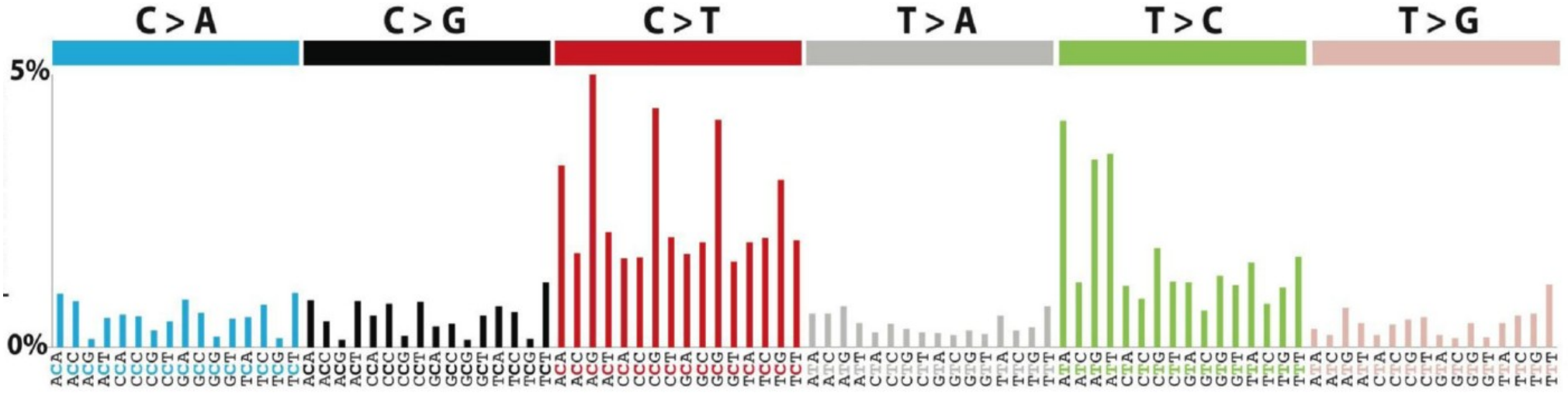
- G:C>T:A: 10%
- G:C>C:G: 10%
- A:T>C:G: <8%
- A:T>T:A: <8%

Ohno (2019) *Genes & Genet Systems*

De novo mutation spectra







6,570 high confidence DNMs from 109 trios

Rahbari et al. (2016) *Nat Genet*

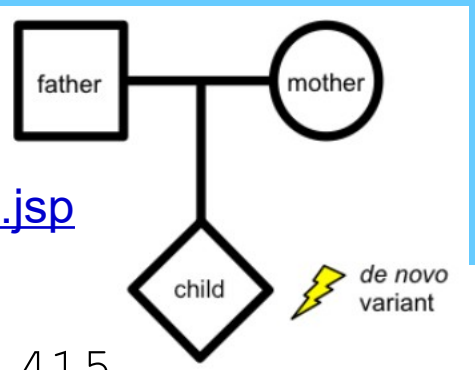
De novo mutation spectra

Validated mutational signatures found in human cancer
Each signature is displayed according to the 96 substitution classification defined by the substitution class and sequence context immediately 3' and 5' to the mutated base.

Alexandrov (2013) *Nature*

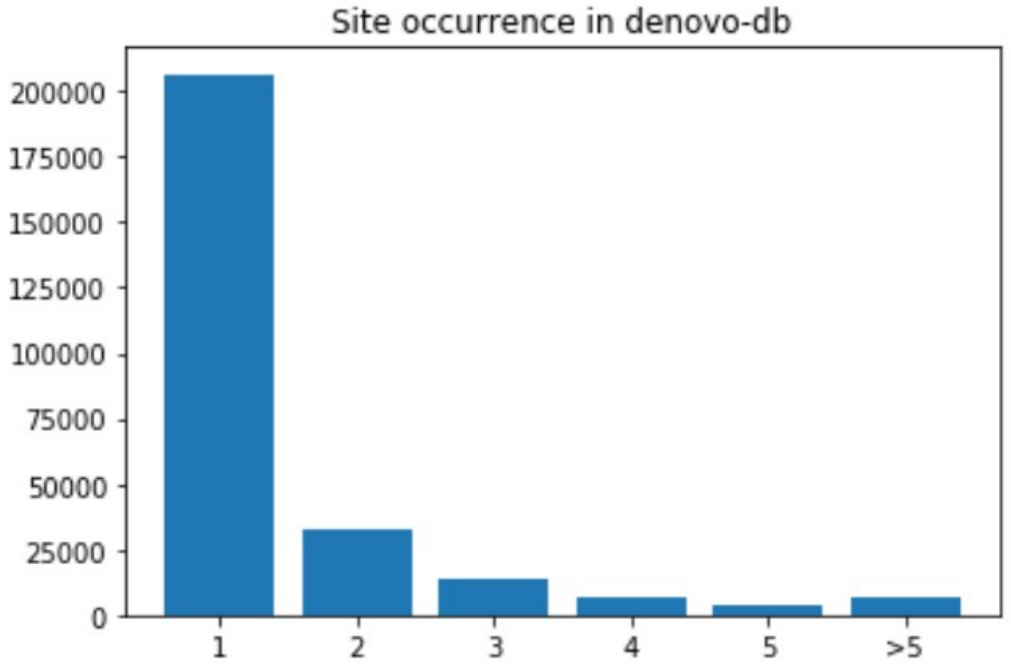
Denovo-db v.1.6.1

http://denovo-db.gs.washington.edu/denovo-db/index.jsp

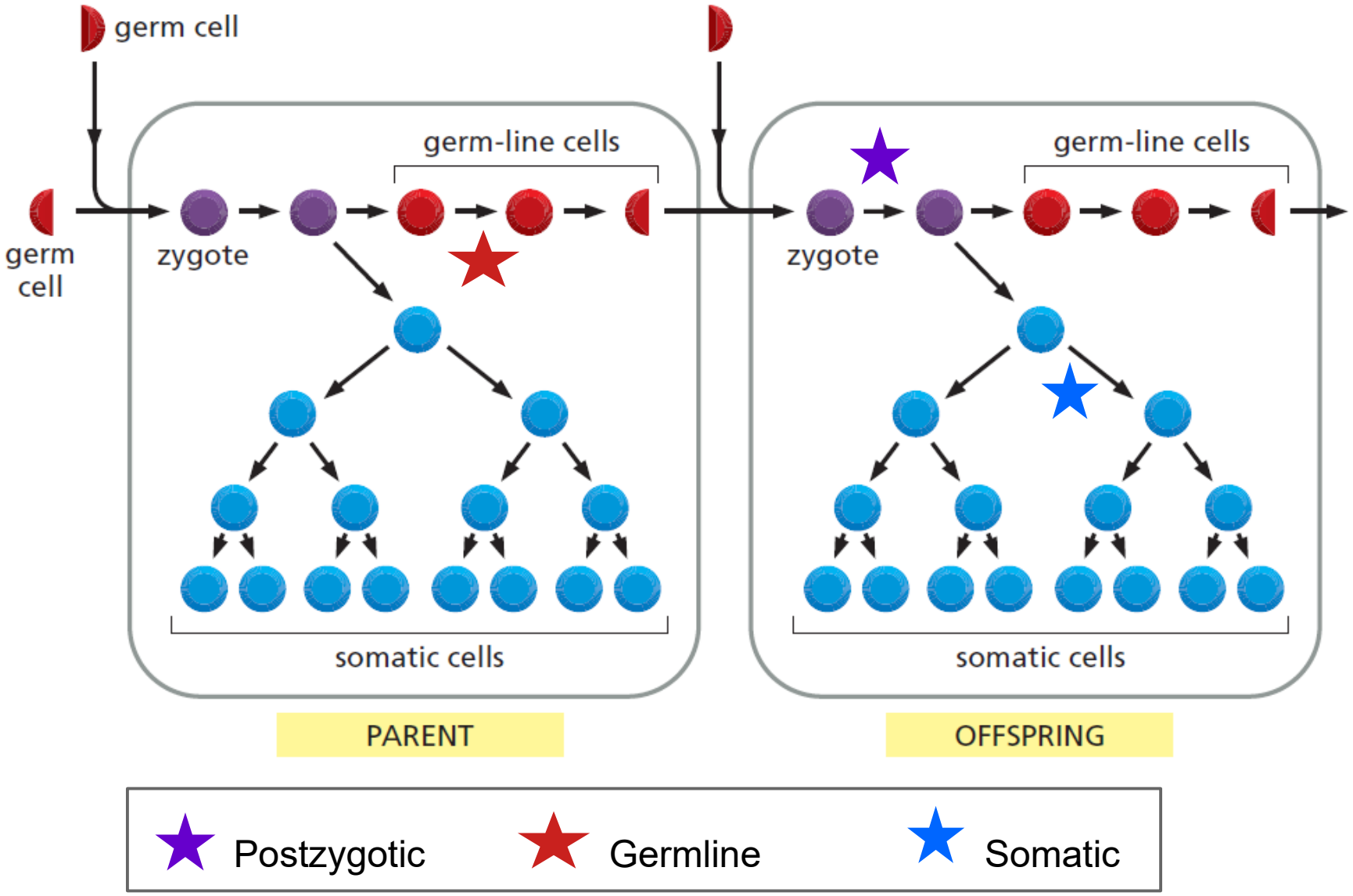


Variant	Counts	%
C>T	79,110	19.0
G>A	79,016	19.0
A>G	47,666	11.5
T>C	47,584	11.5
C>G	17,431	4.2
G>C	17,154	4.1
C>A	15,719	3.8
G>T	15,234	3.7
A>C	12,744	3.1
T>G	12,464	3.0
T>A	11,637	2.8
A>T	11,482	2.8
T>TA	739	0.2
CT>C	737	0.2
...		
Total	*415,515*	

Variants:	415,415
Samples:	11,518
Genes:	17,431
Sites:	270,506
Missense:	20,815
Pred. Damaging:	~75%



Mutation timing and mosaicism



Alberts - *Essential Cell Biology*, Fig 9-3