MEDICAL GENOMICS

Василий Евгеньевич Раменский
ramensky@gmail.com

Анастасия Александровна Жарикова
azharikova89@gmail.com

НМИЦ Терапии и 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. Mendelian diseases: gene discovery and diagnostics
7. Some basics of quantitative genetics
8. Complex diseases: gene discovery and allelic architecture

Remarks

- Important info: kodomo
- English
- Molecular genetics + population genetics + medical genetics + statistical genetics + genetic epidemiology + bioinformatics $\Longrightarrow$ no single textbook
- Many topics not covered: immunology, pathogens, microbiome, therapy, genome editing

- Definitions
- Questions and exercises
- Homework slides 
- Summary, concepts, further reading

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

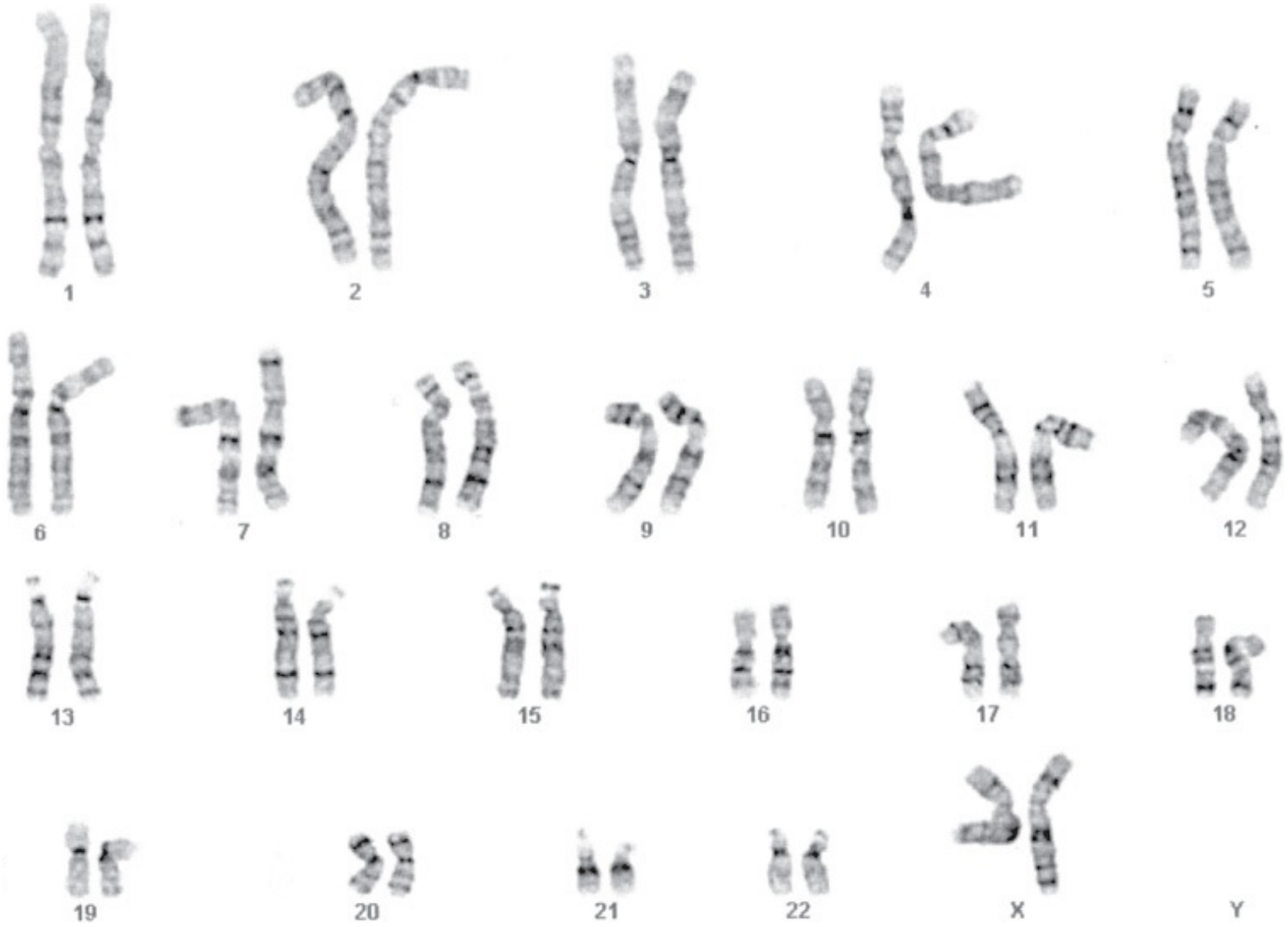
MUTATIONS:

ORIGINS AND RATES

Lecture plan

- Human karyotype
- Mitosis and DNA replication
- Replication fidelity and mutation rate
- Exogenous and endogenous DNA damage
- DNA repair mechanisms
- De novo mutations: single nucleotide variants
- Structural variants and CNVs
- Aneuploidy

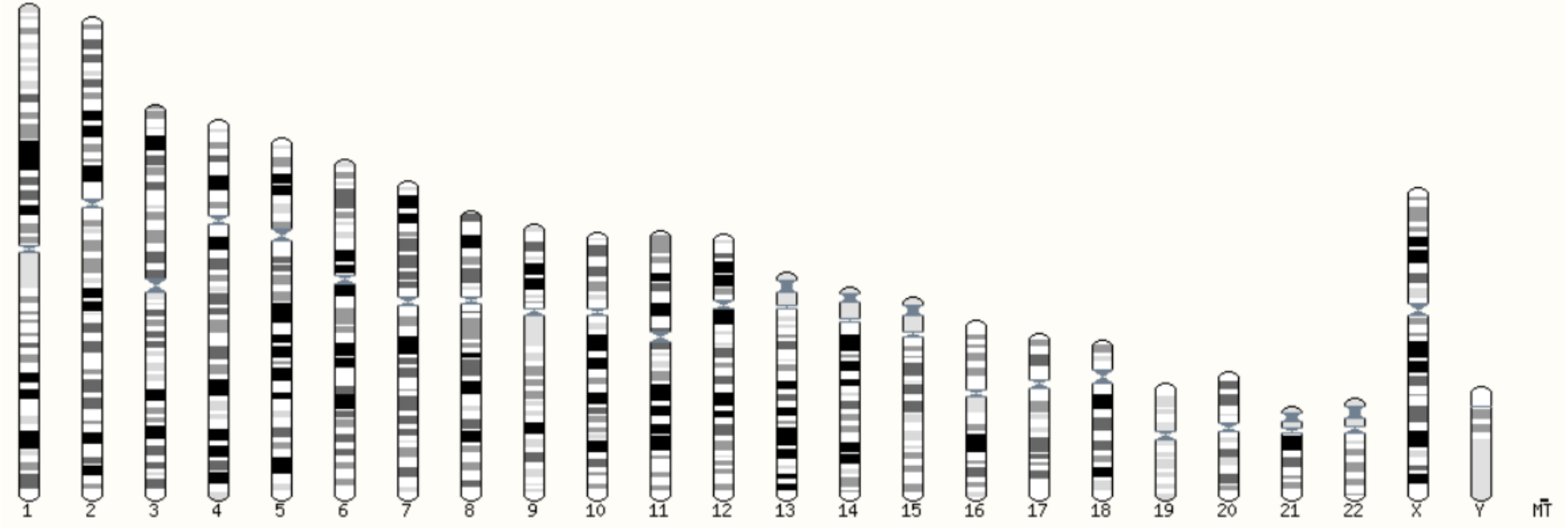
Human karyotype

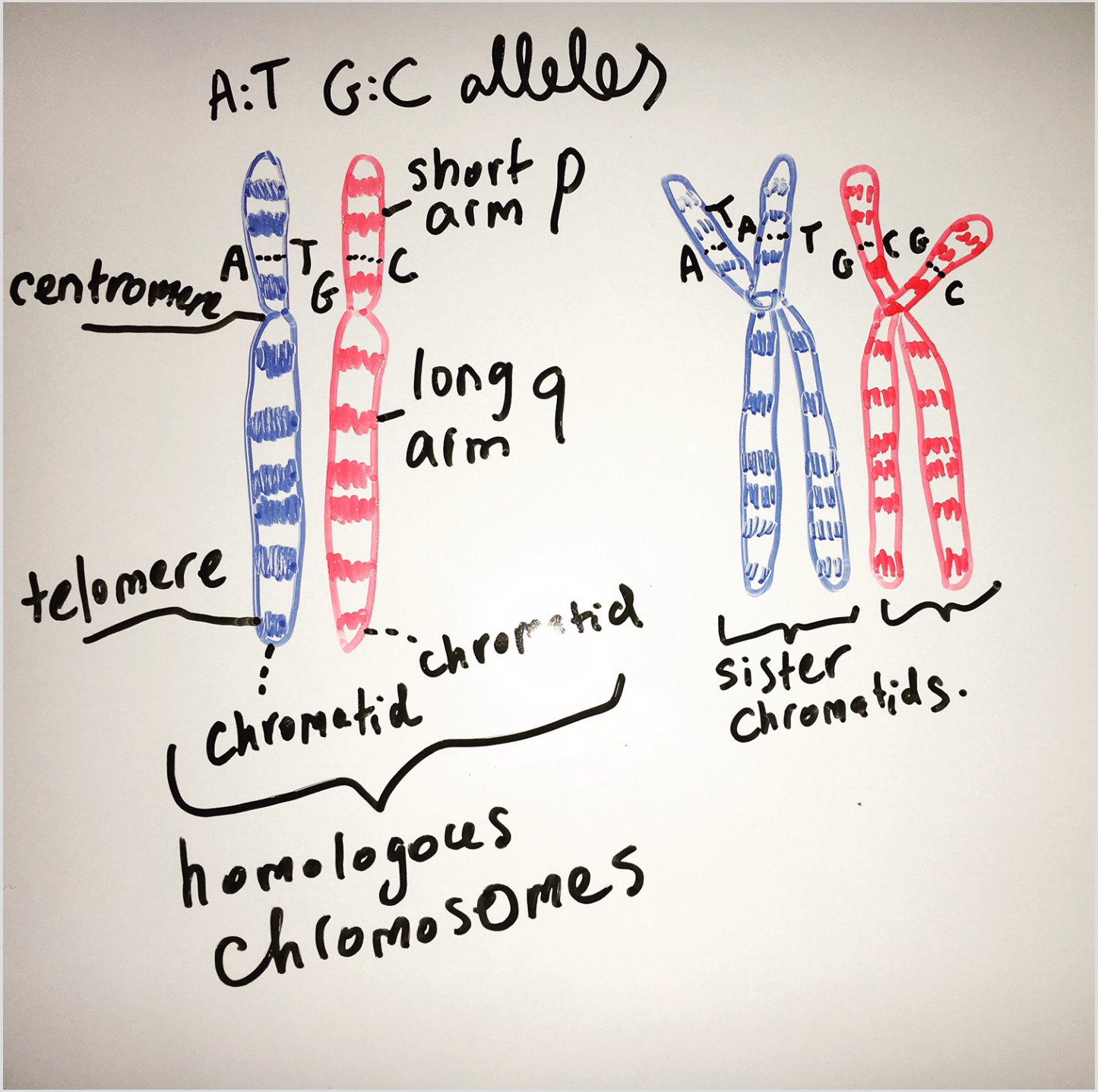


22 pairs of **autosomal** chromosomes + 2 **sex** chromosomes

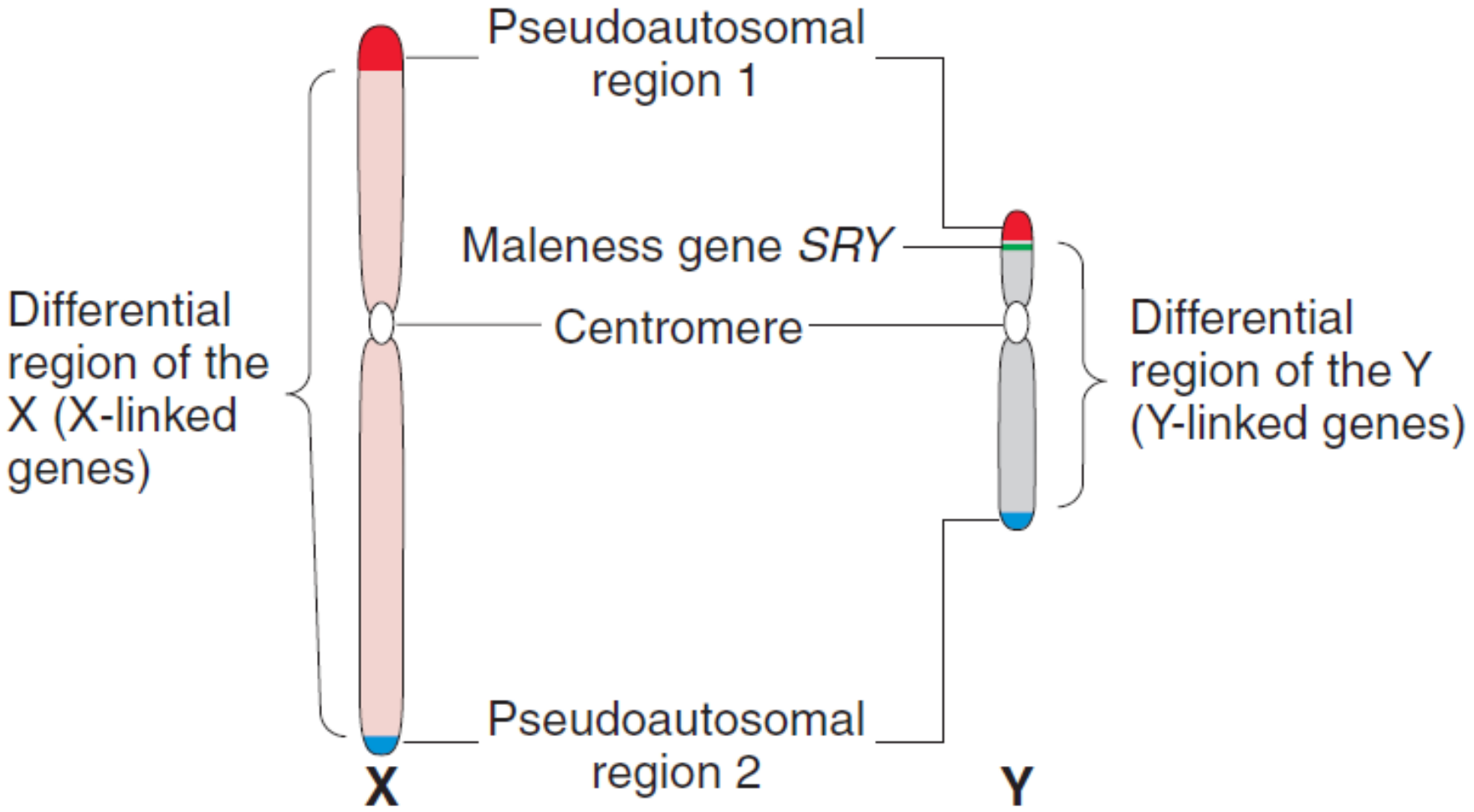
Human karyotype

- **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.




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.

Sex chromosomes



Women: XX, men: XY

Q: transmission of Y chromosome

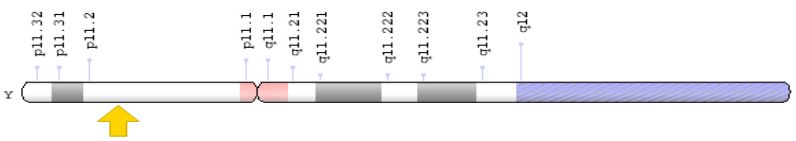
Griffiths -- *Introduction to Genetic Analysis*

How to check sex of an NGS sample?

How to check sex of an NGS sample?

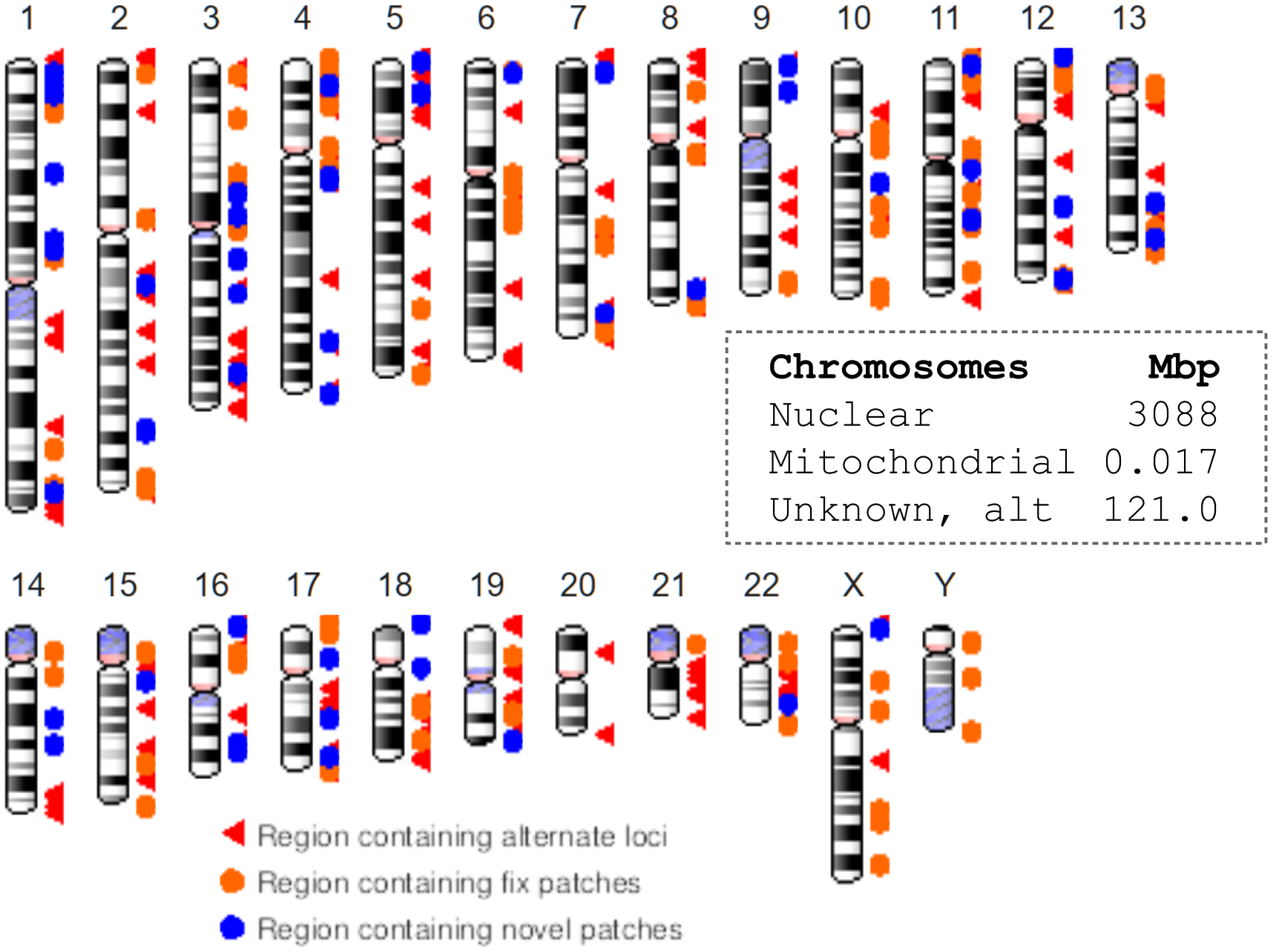
• **Heterozygous / Homozygous ratio on chromosome X**
Het/Hom<0.8 => M, Het/Hom>0.8 => F; suspicious : 0.5-1.0

• ***SRY* gene** (Sex-determining Region Y): intronless sex-TF protein, responsible for the initiation of male sex determination in mammals



• **The human amelogenin genes: *AMELX* and *AMELY***
Short arms of X and Y sex chromosomes, share 84% sequence identity. A 6 bp insertion/deletion difference in the first intron of the *AMELY* and *AMELX* genes is typically targeted in forensic sex identification (Tzvetkov 2010 *Pharmacogenomics*)

Genome Reference Consortium GRCh38.p13



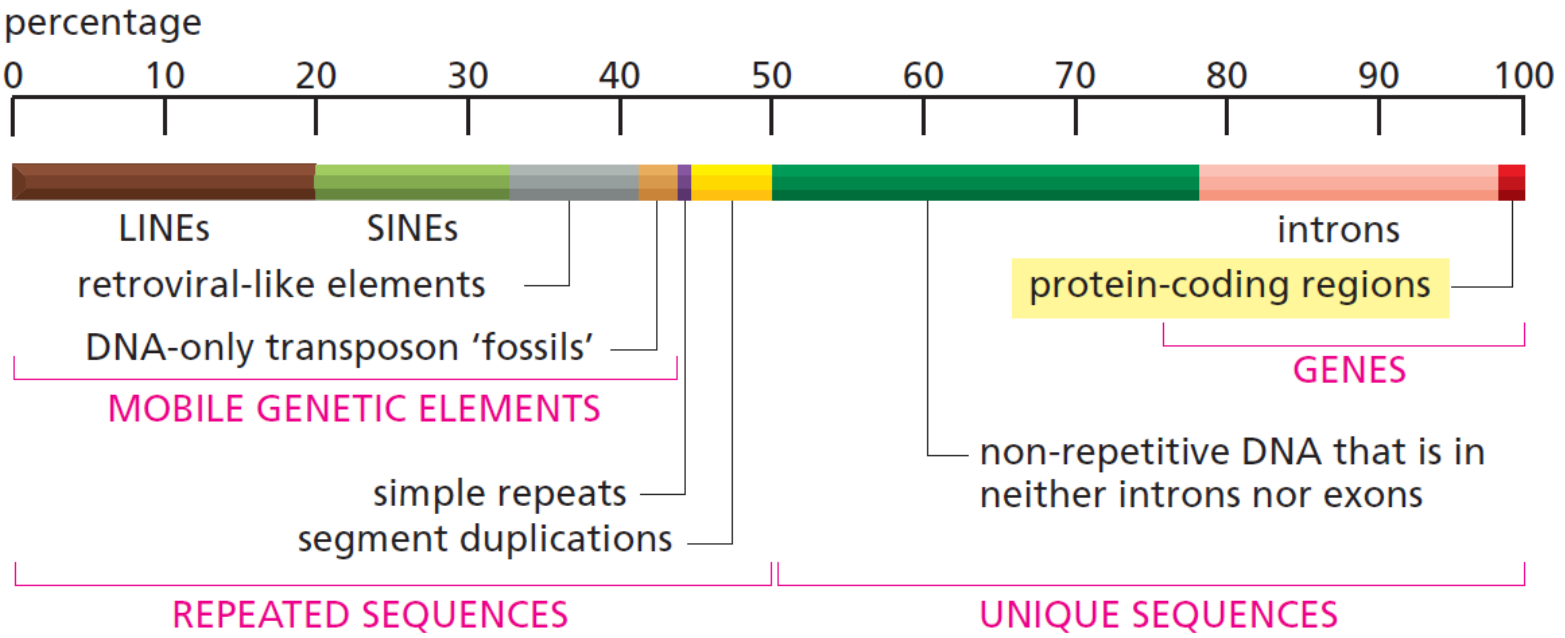
Chromosomes	Mbp
Nuclear	3088
Mitochondrial	0.017
Unknown, alt	121.0

https://www.ncbi.nlm.nih.gov/grc/human

Human genome contents

Regions	Length, Mbp	%	Description
Genes	1,200	37.5	Genomic locus where transcription occurs
Exons	48	1.5	Transcribed genomic region that remain in the RNA after splicing
Other (introns, UTRs)	1,152	36.0	Regions of a coding cDNA which are not translated
Unique and regulatory sequences	510	15.9	
Interspersed repeats	1,400	43.8	
LINEs	640	20.0	~850,000 Long Interspersed Elements (~7,000 bp). Retrotransposed elements containing open reading frames encoding (often inactive) reverse transcription machinery
SINEs, Alu repeats	420	13.1	~1,500,000 Short Interspersed Elements. Retrotransposed elements <500 bp that contain tRNA, snRNA and rRNA, which require other mobile elements to be transposed.
LTR retrotransposons	250	7.8	Transposable elements characterized by the presence of Long Terminal Repeats (LTRs) directly flanking an internal coding region
DNA transposons	90	2.8	Class II transposable elements that move through a DNA intermediate
Microsatellites	90	2.8	A region in the genomic sequence containing short tandem repeats of 2-10bp
Total	3,200	100.0	

Human genome contents



Alberts - *Essential Cell Biology*

Element	Transposition	Structure	Length	Copy number	Fraction of genome
LINEs	Autonomous	ORF1 ORF2 *(pol)* AAA	1–5 kb	20,000–40,000	21%
SINEs	Nonautonomous	AAA	100–300 bp	1,500,000	13%
DNA transposons	Autonomous	transposase	2–3 kb	300,000	3%
	Nonautonomous		80–3000 bp		

Griffiths -- *Introduction to Genetic Analysis*

ENSEMBL gene annotation GRCh38 v.99

Gene biotype	Genes (Transcripts)	%	Description
Proten coding	19,968 (153,197)	32.9	Genes that contain an open reading frame (ORF)
Pseudogenes	15,263	25.2	Genes that have homology to known protein-coding genes but contain a frameshift and/or stop codon(s) which disrupts the ORF
To be confirmed	1,060	1.7	Require experimental validation
T-cell receptors, immunoglobulins	408	0.7	Undergo somatic recombination before transcription
RNA genes	23,977	39.5	
lncRNA	16,880		A non-coding gene >200bp in length
snRNA	1,910		Processing of pre-messenger RNA
miRNA	1,879		A small RNA (~22bp) that silences the expression of target mRNA
snoRNA	942		Post-transcriptional modification of other RNAs
Other	2,366		rRNA, sRNA, scRNA, scaRNA, miscRNA
Total	60,676 (227,818)	100	

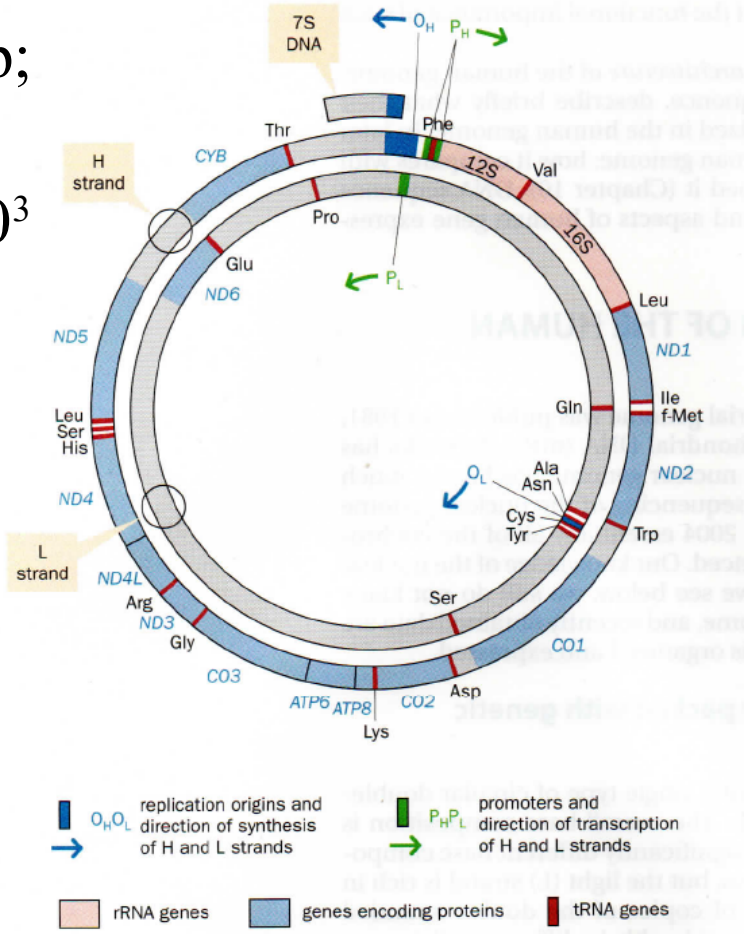
ENSEMBL gene annotation GRCh38 v.99

Chromosome	Approximate length (bp)	Protein-coding genes	Non-protein coding genes	Pseudogenes
1	248956422	2047	1964	1233
2	242193529	1303	1605	1033
3	198295559	1075	1160	768
4	190214555	753	984	732
5	181538259	881	1200	710
6	170805979	1041	989	803
7	159345973	989	977	893
8	145138636	670	1041	629
9	138394717	778	786	678
10	133797422	728	880	568
11	135086622	1312	1053	815
12	133275309	1036	1197	627
13	114364328	321	586	378
14	107043718	820	857	519
15	101991189	613	986	513
16	90338345	867	1033	467
17	83257441	1185	1198	531
18	80373285	269	608	246
19	58617616	1474	895	514
20	64444167	543	594	250
21	46709983	231	403	183
22	50818468	492	513	332
X	156040895	843	640	872
Y	57227415	63	108	392
Mitochondrial	16569	13	24	

Jackson (2018) *Essays Biochem*

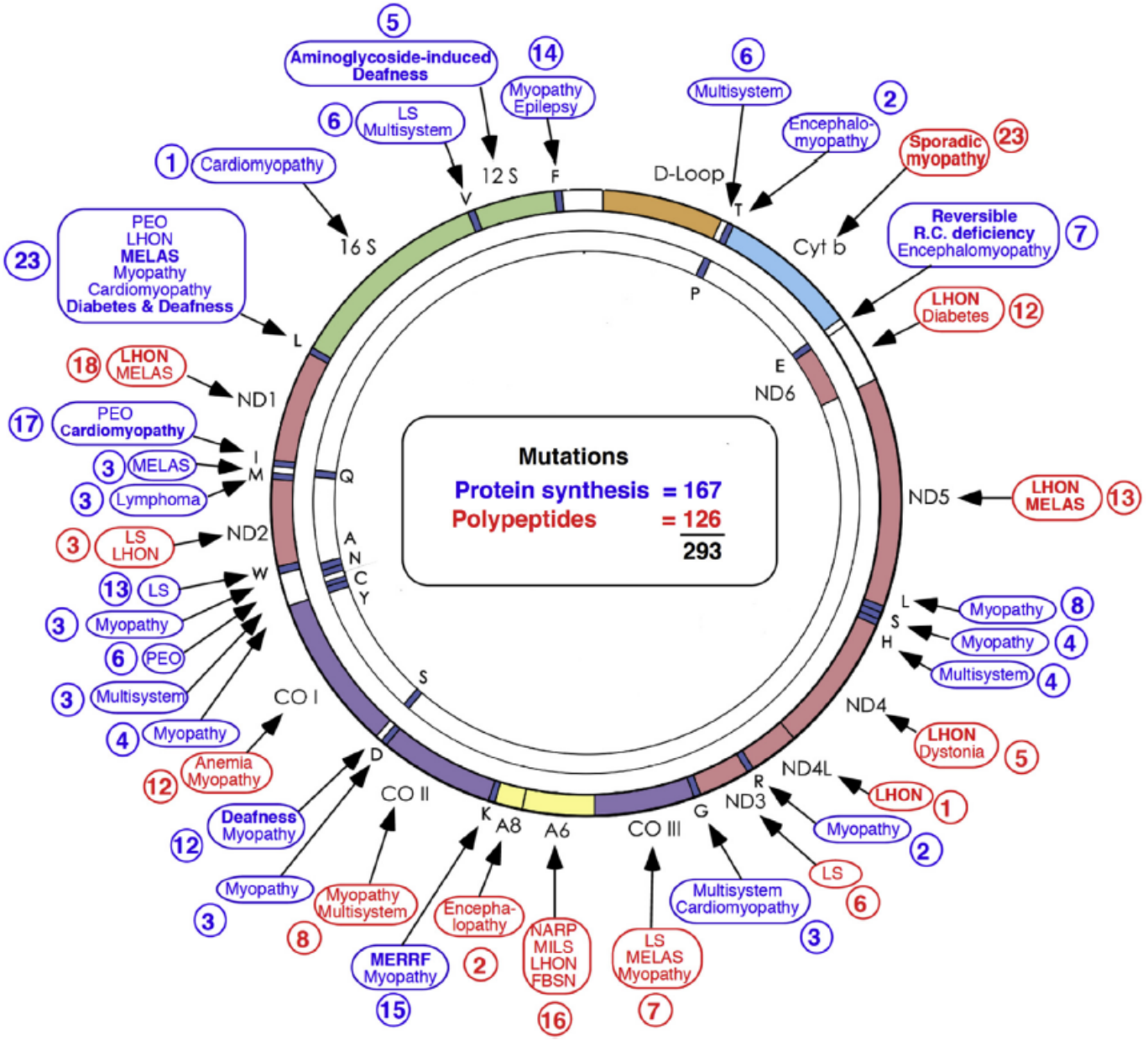
Mitochondrial genome

- **mtDNA**: circular, double-stranded, 16,569 bp; H and L chains; similar to bacteria
- Egg only, maternally inherited; each cell: ~10^3 copies; highly heterogeneous
- 37 genes: 22 tRNA + 2 rRNA + 13 coding
- 13 polypeptides are part of **mitochondrial respiratory complex** (Sugars → ATP), together with multiple nuclear genes
- mtDNA is to some extent autonomous, with its own genetic code
- Stop codons: TAA, TAG, AGA, AGG



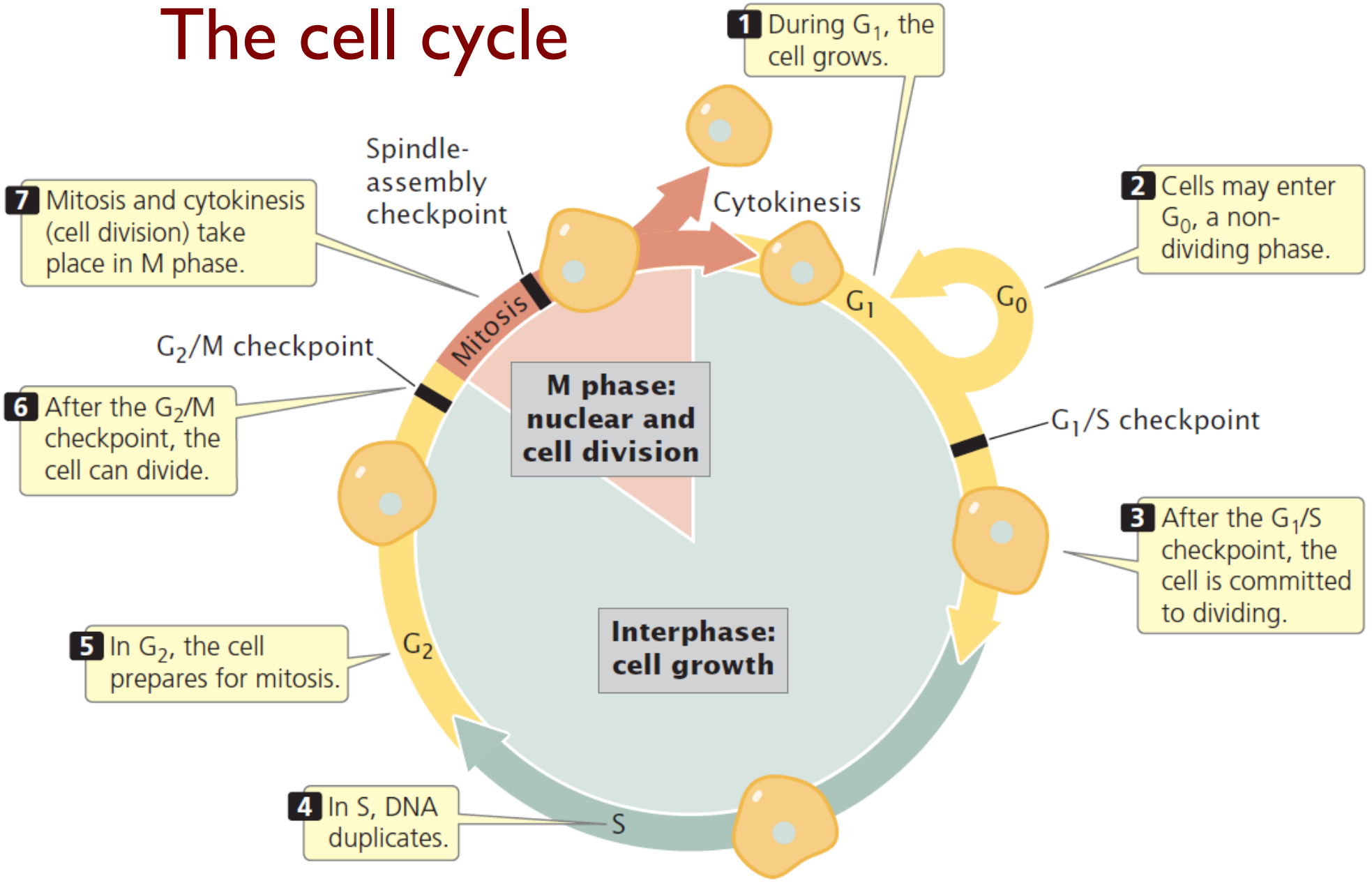
- **Mitochondrial diseases**: a heterogeneous group of inherited anomalies in oxidative phosphorylation due to mutations in the mitochondrial (70%) or nuclear DNA (30%)
- ~300 disease-causing point mutations known in mtDNA

Pathogenic mutations in mtDNA



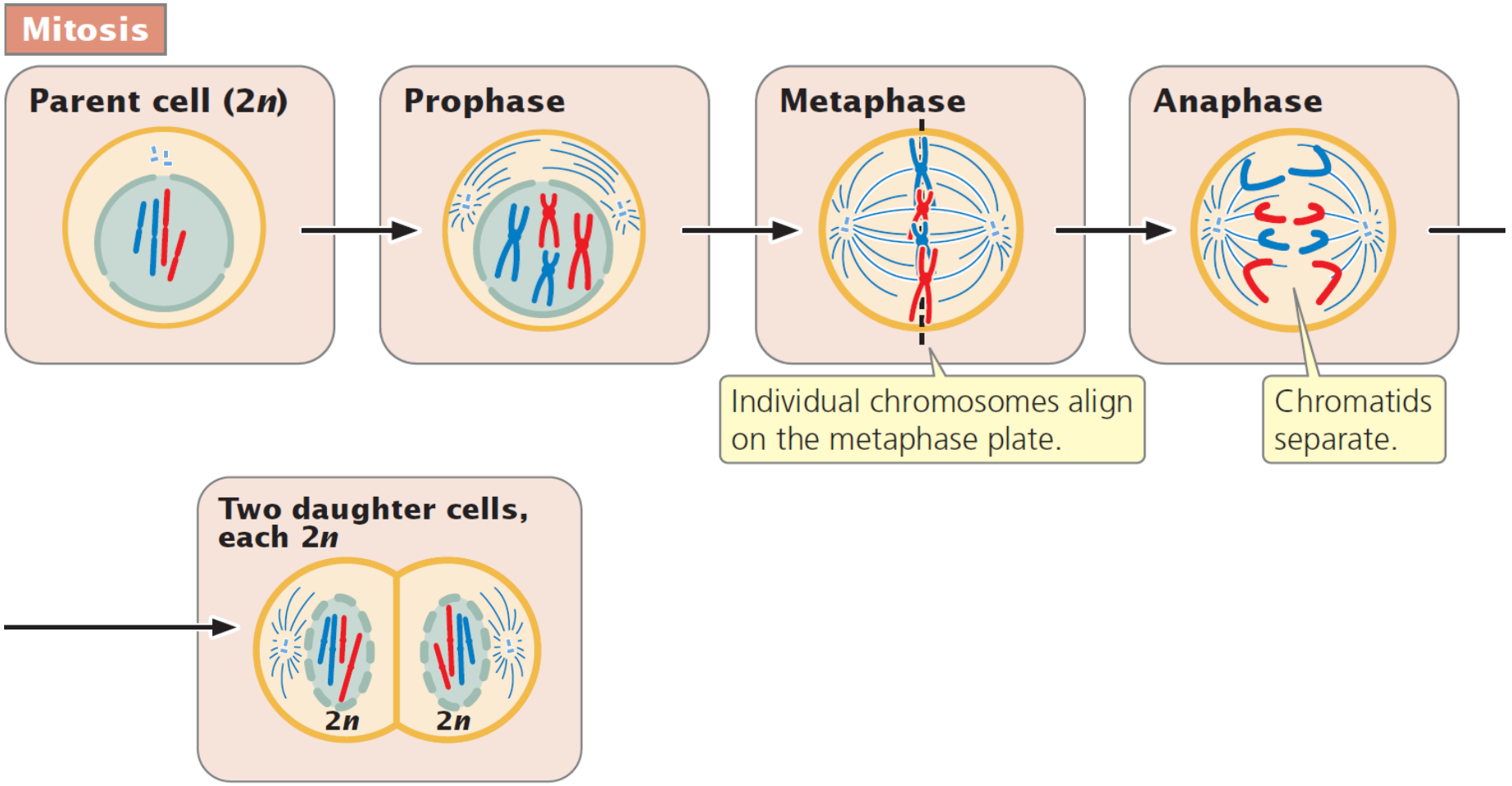
DiMauro (2017) Mitochondrial Encephalomyopathies

The cell cycle



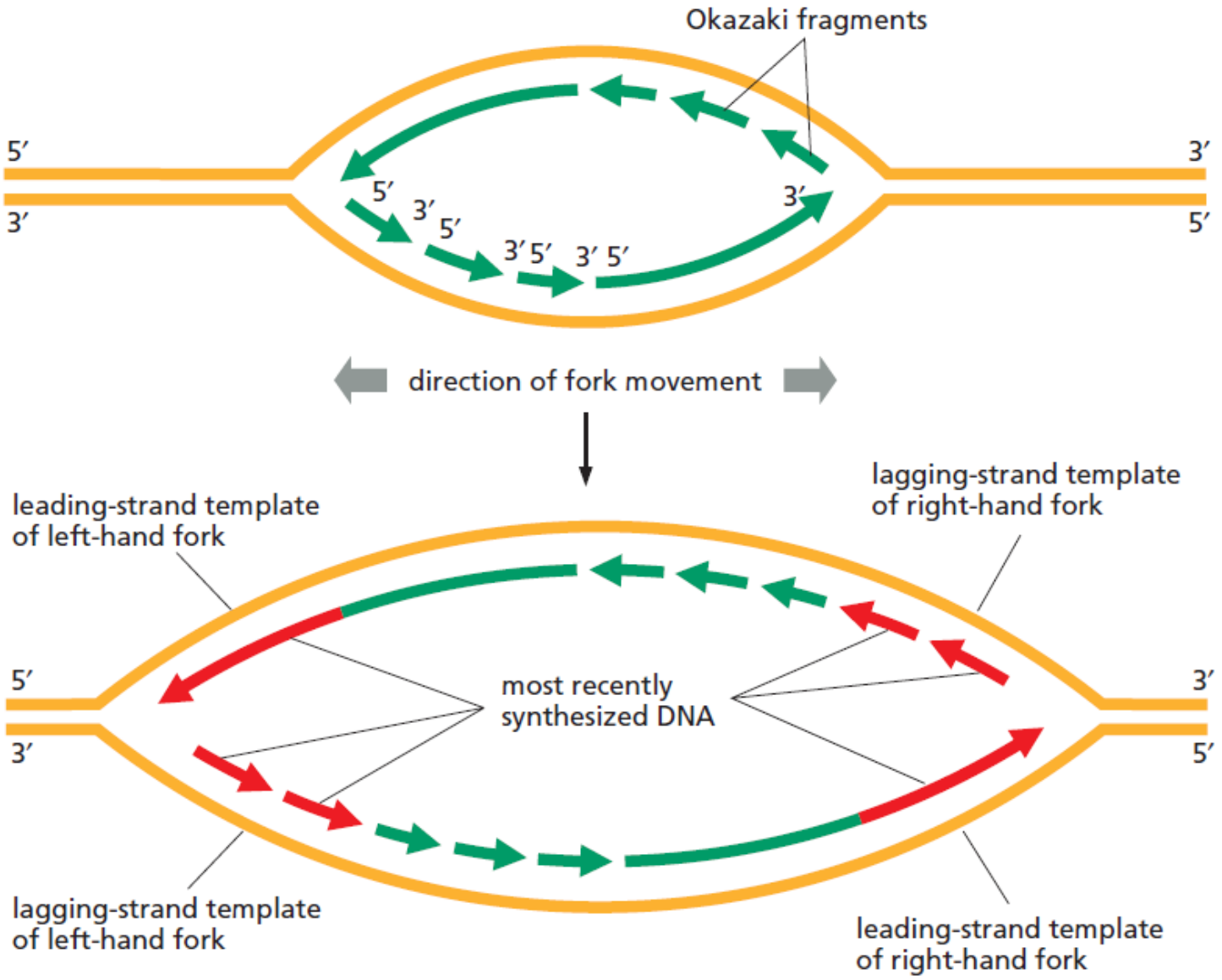
2.7 The cell cycle consists of interphase and M phase.

Mitosis



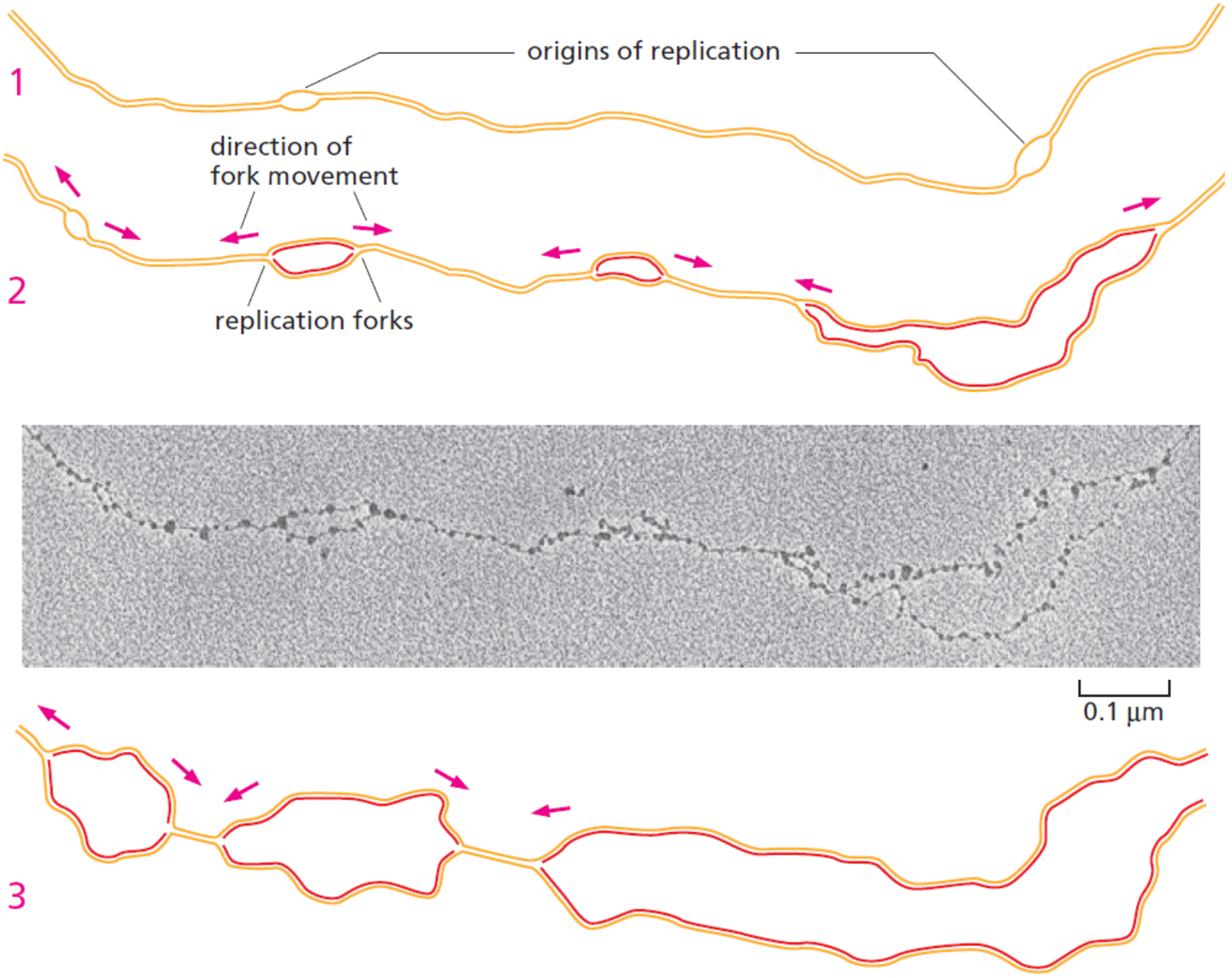
Mitosis: a type of cell division that results in two daughter cells with the set of chromosomes as the parent nucleus, typical of ordinary tissue growth

DNA replication forks



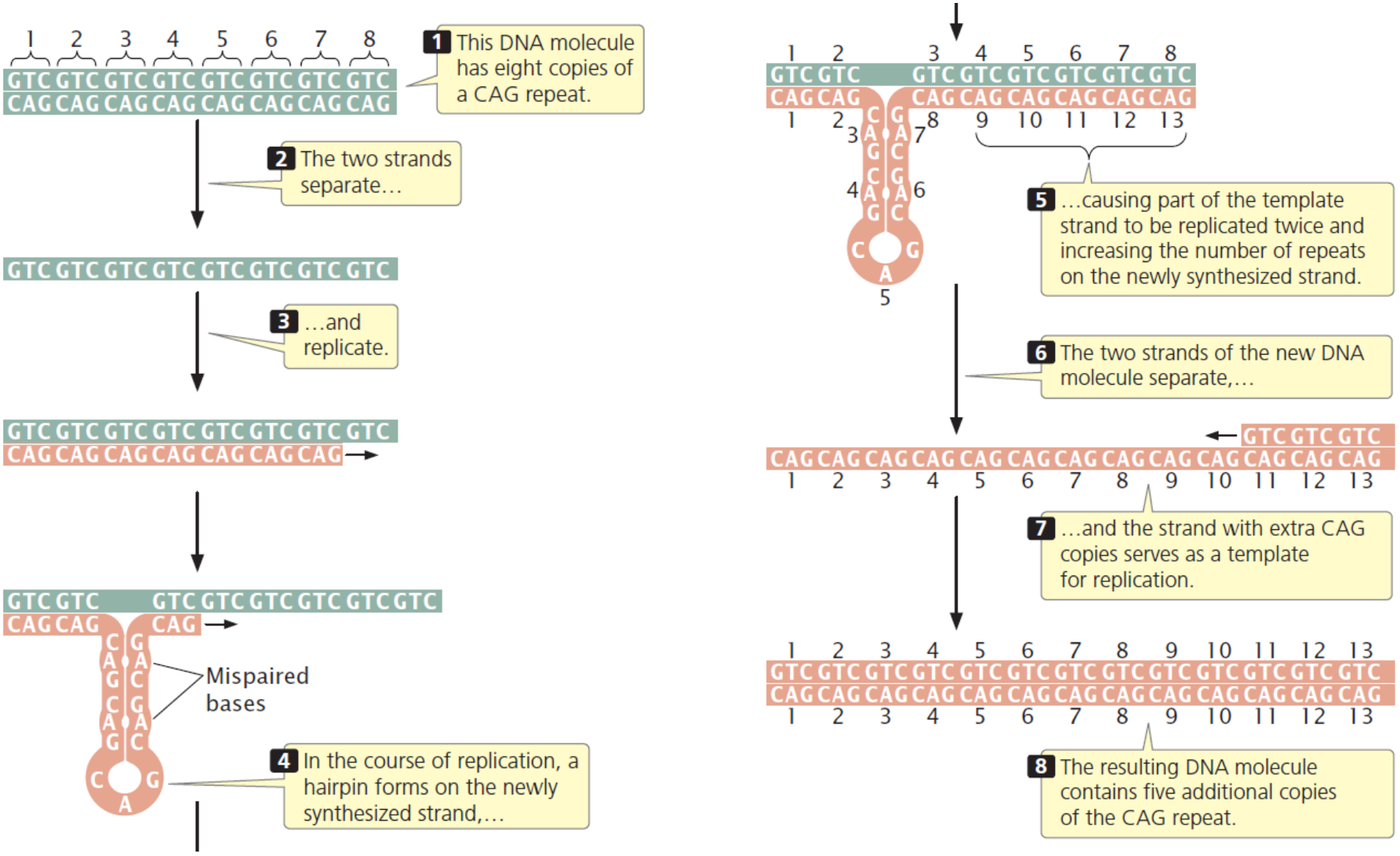
Alberts -- *Essential Cell Biology*

DNA replication forks

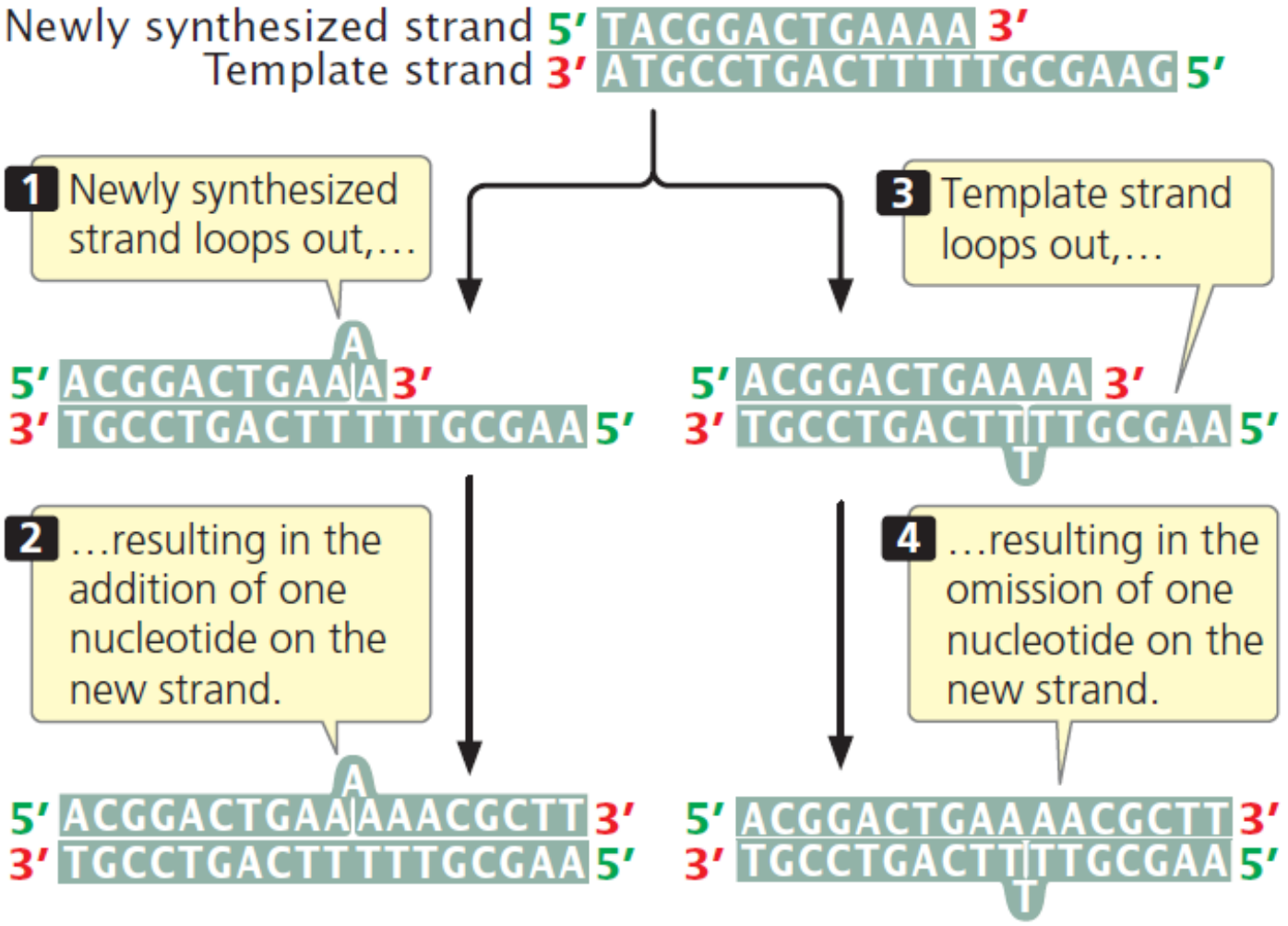


Alberts -- *Essential Cell Biology*

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

OMIM®

Online Mendelian Inheritance in Man®

An Online Catalog of Human Genes and Genetic Disorders

Updated February 12, 2021

Search OMIM for clinical features, phenotypes, genes, and m

Dissected OMIM Morbid Map Scorecard (Updated February 12th, 2021) :

Class of phenotype	Phenotype	Gene *
Single gene disorders and traits	5,740	4,006
Susceptibility to complex disease or infection	694	499
"Nondiseases"	151	119
Somatic cell genetic disease	231	130

*Some genes may be counted more than once because mutations in a gene may cause more than one phenotype and the phenotypes may be of different classes (e.g., activating somatic BRAF mutation underlying cancer, 164757.0001. and germline BRAF mutation in Noonan syndrome, 164757.0022.)

Mutations

Mutations are random changes in DNA sequences

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

Mechanisms of mutation:

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

Variants = mutations (recent changes), polymorphisms (segregating in a population), engineered (non-random) changes

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

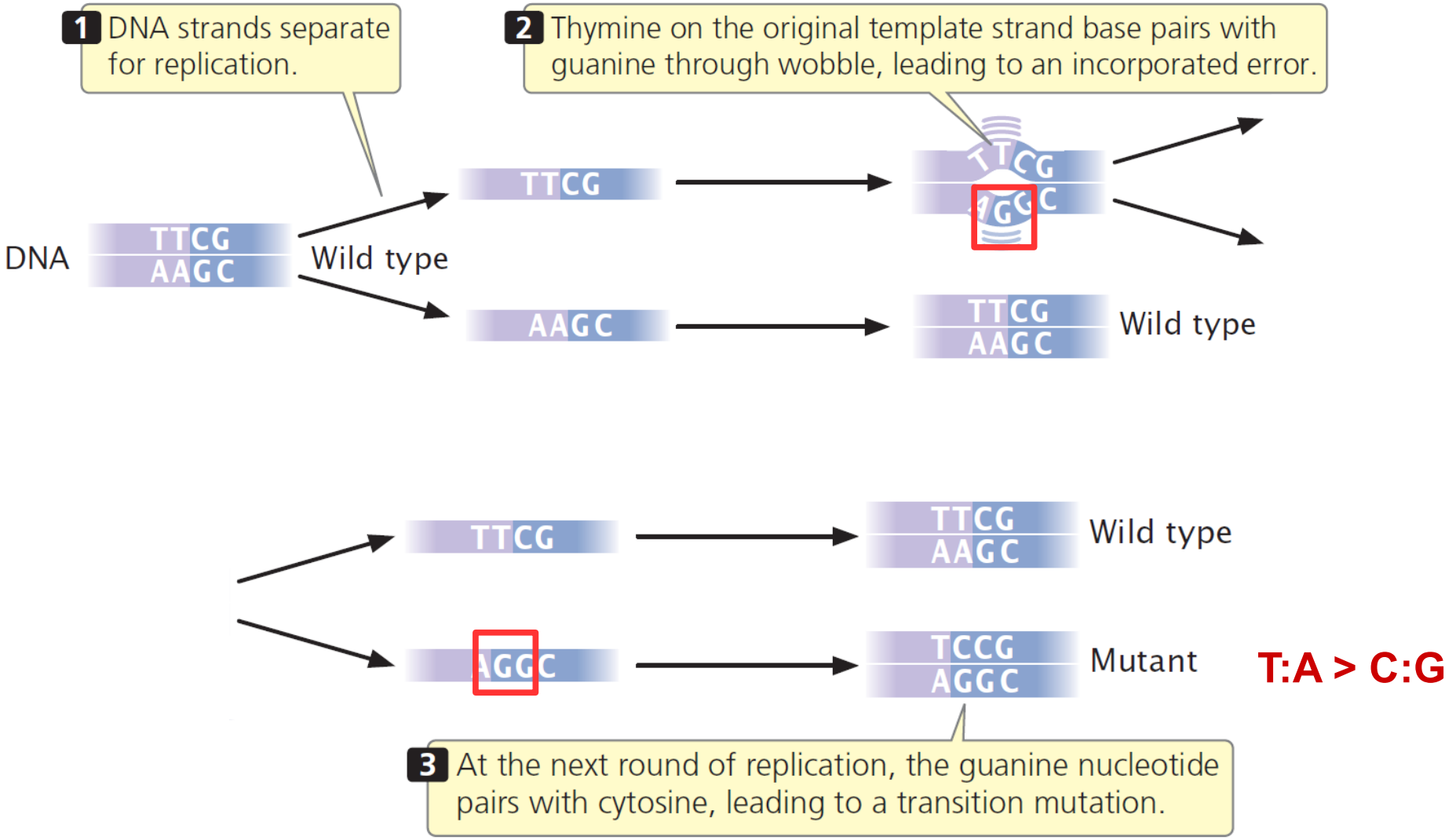
  Inversion or translocation

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

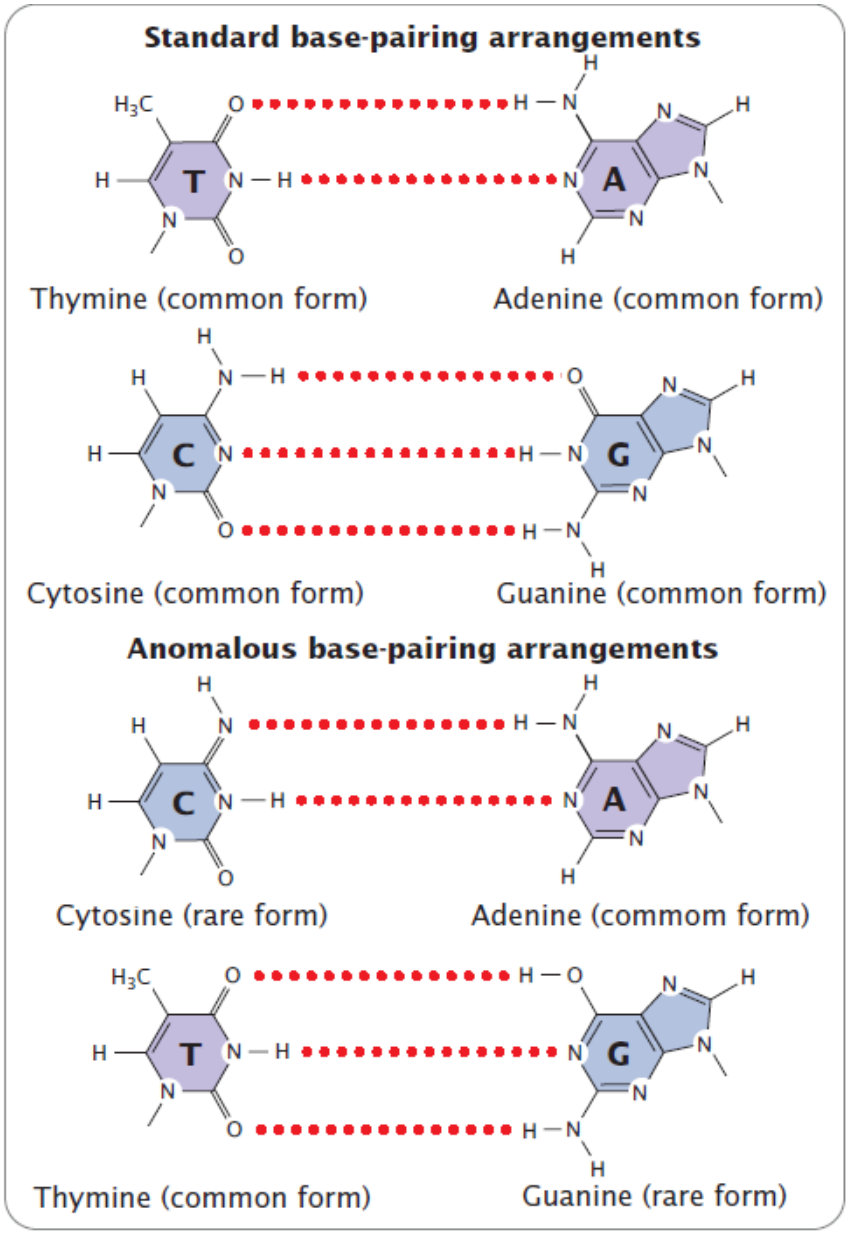
  Tandem or dispersed duplication, deletion, insertion

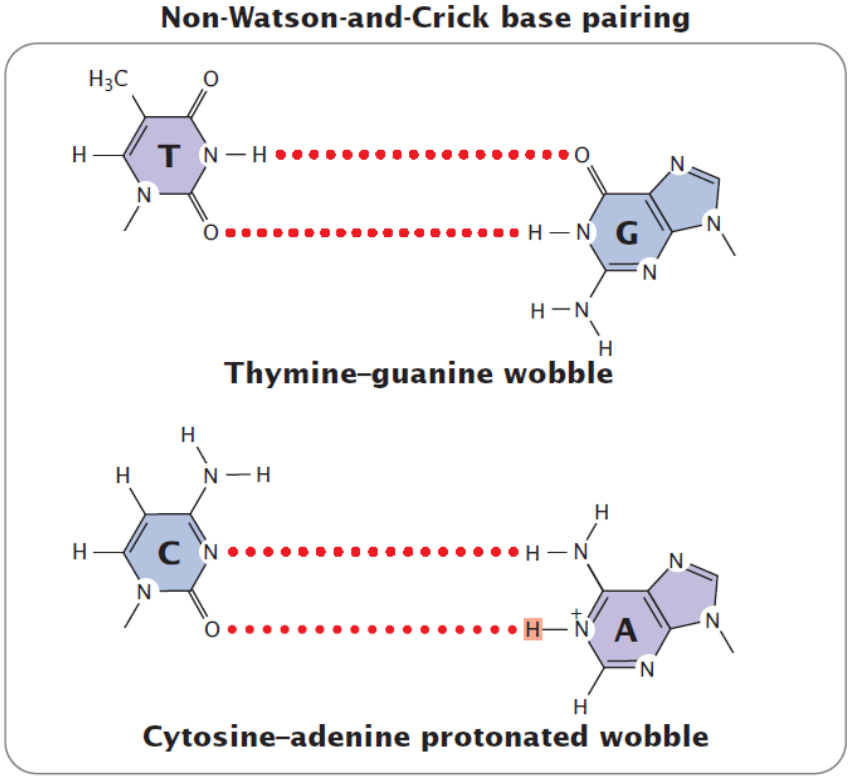
Aneuploidy: wrong number of whole chromosomes: nullisomy, monosomy, trisomy

Replication errors become mutations



Standard and non-standard base pairing



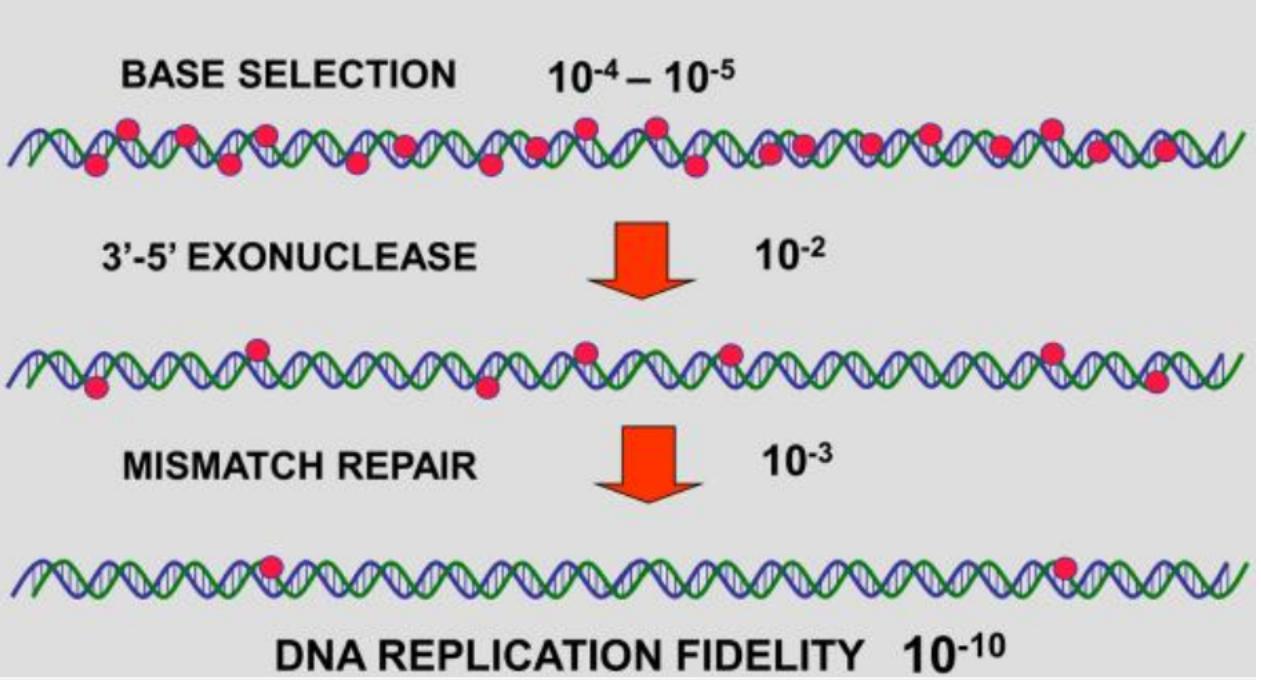


Pierce -- *Genetics Essentials. Concepts and Connections*

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 ?

Crumbling Genome: The Impact of Deleterious Mutations on Humans, First Edition. Alexey S. Kondrashov.
© 2017 John Wiley & Sons, Inc. Published 2017 by John Wiley & Sons, Inc.

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

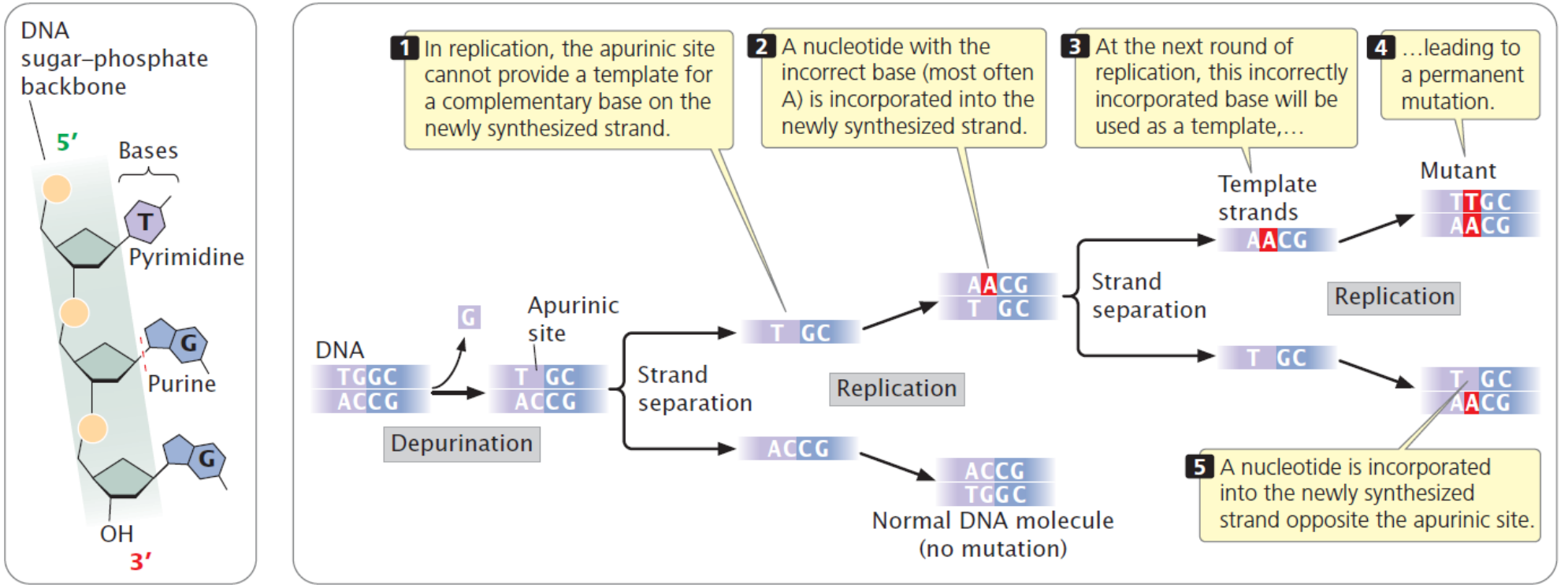
Mizuki Ohno*

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).

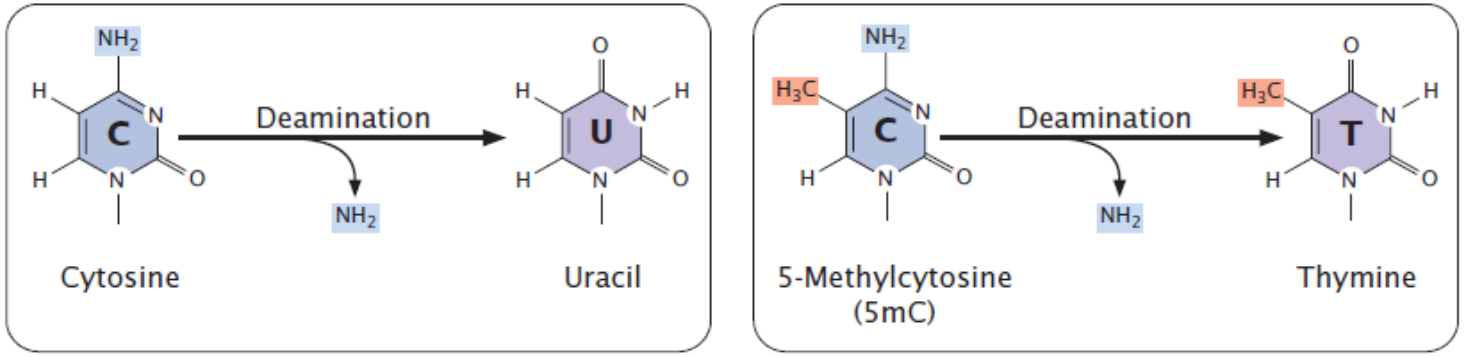
If 10^6 stem cells in intestinal tissue generate transient daughter cells once a week with a mutation rate of approximately 10^{-9} per nucleotide per cell division, the intestinal epithelium of a 60-year-old human would have accumulated more than 10^9 independent mutations. Thus, **nearly every genomic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

Endogenous DNA damage

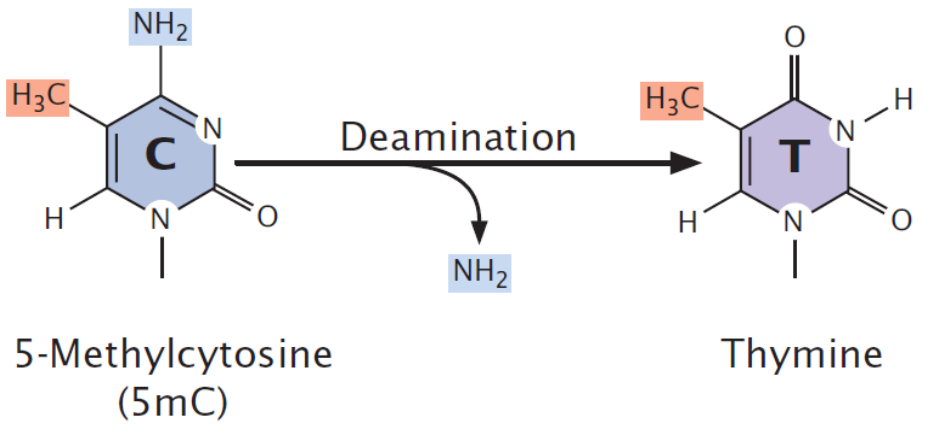
Depurination G:C → A:T



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)

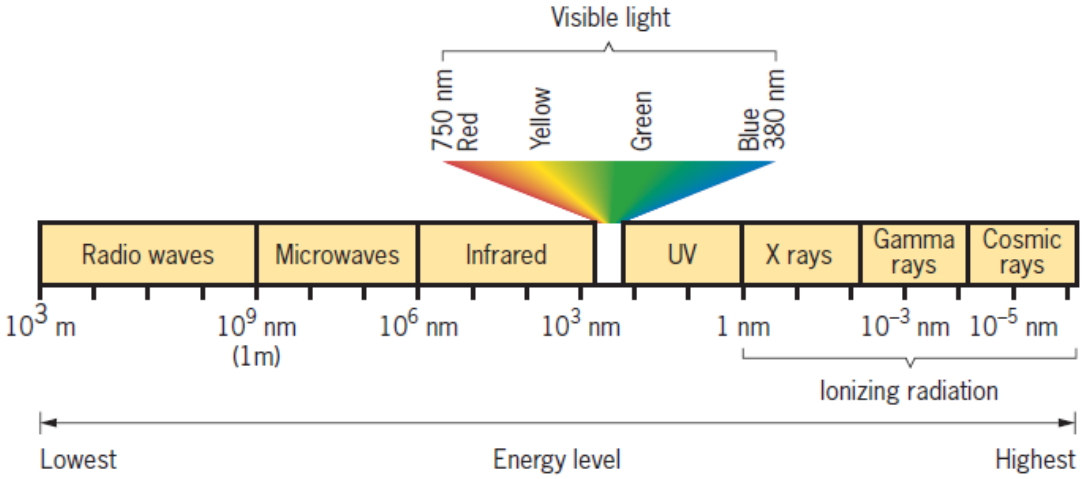
Exogenous DNA damage

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

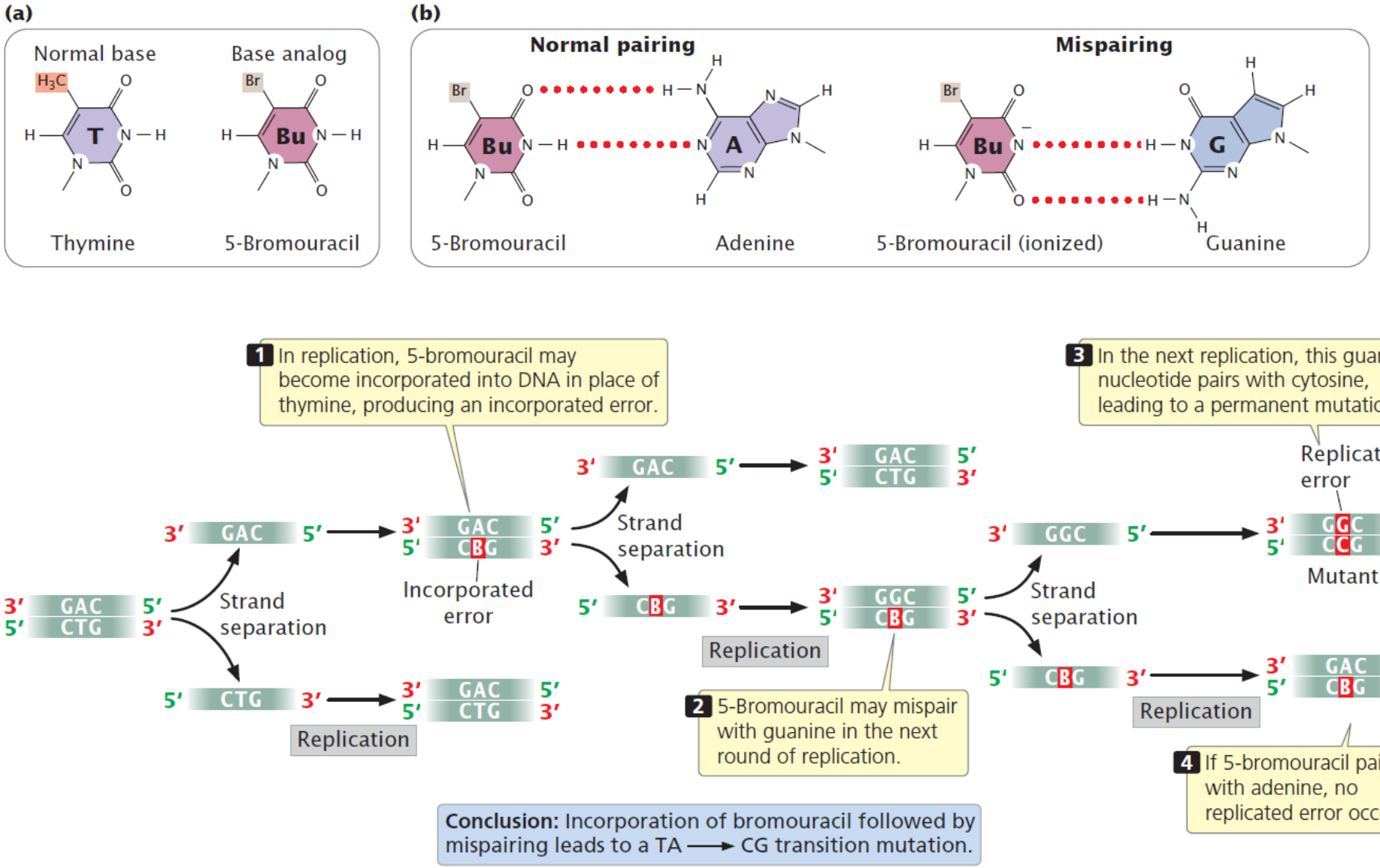
Radiation

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



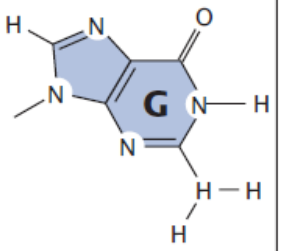
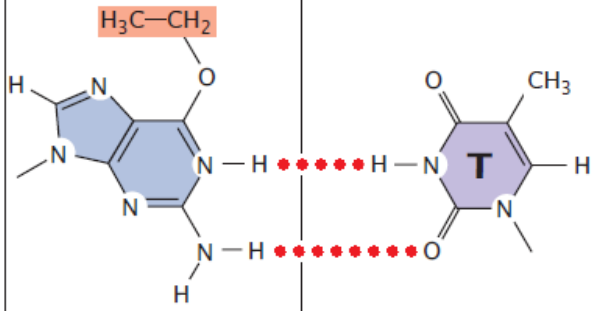
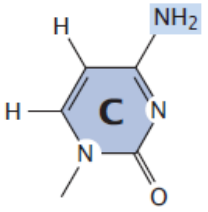
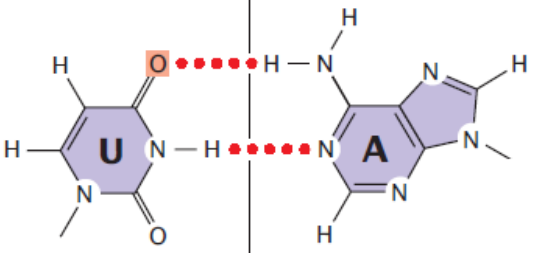
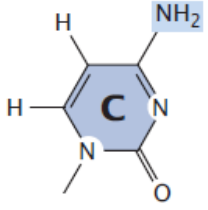
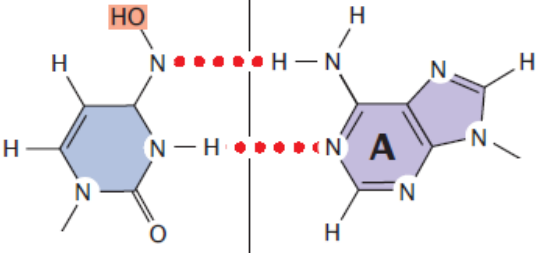
Exogenous DNA damage

Chemical mutagens: 5-bromouracil

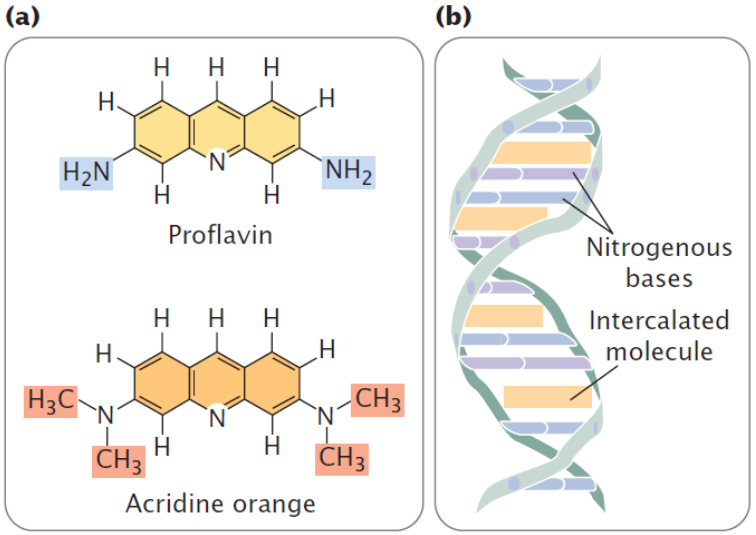


Exogenous DNA damage

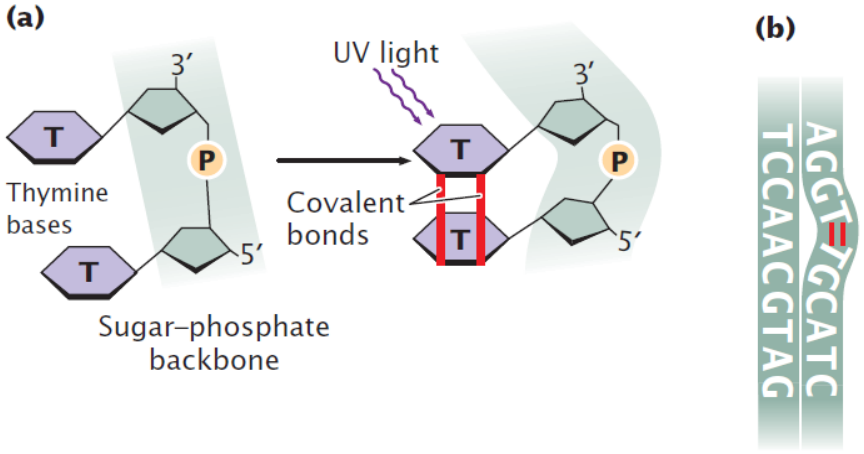
Chemical mutagens

	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

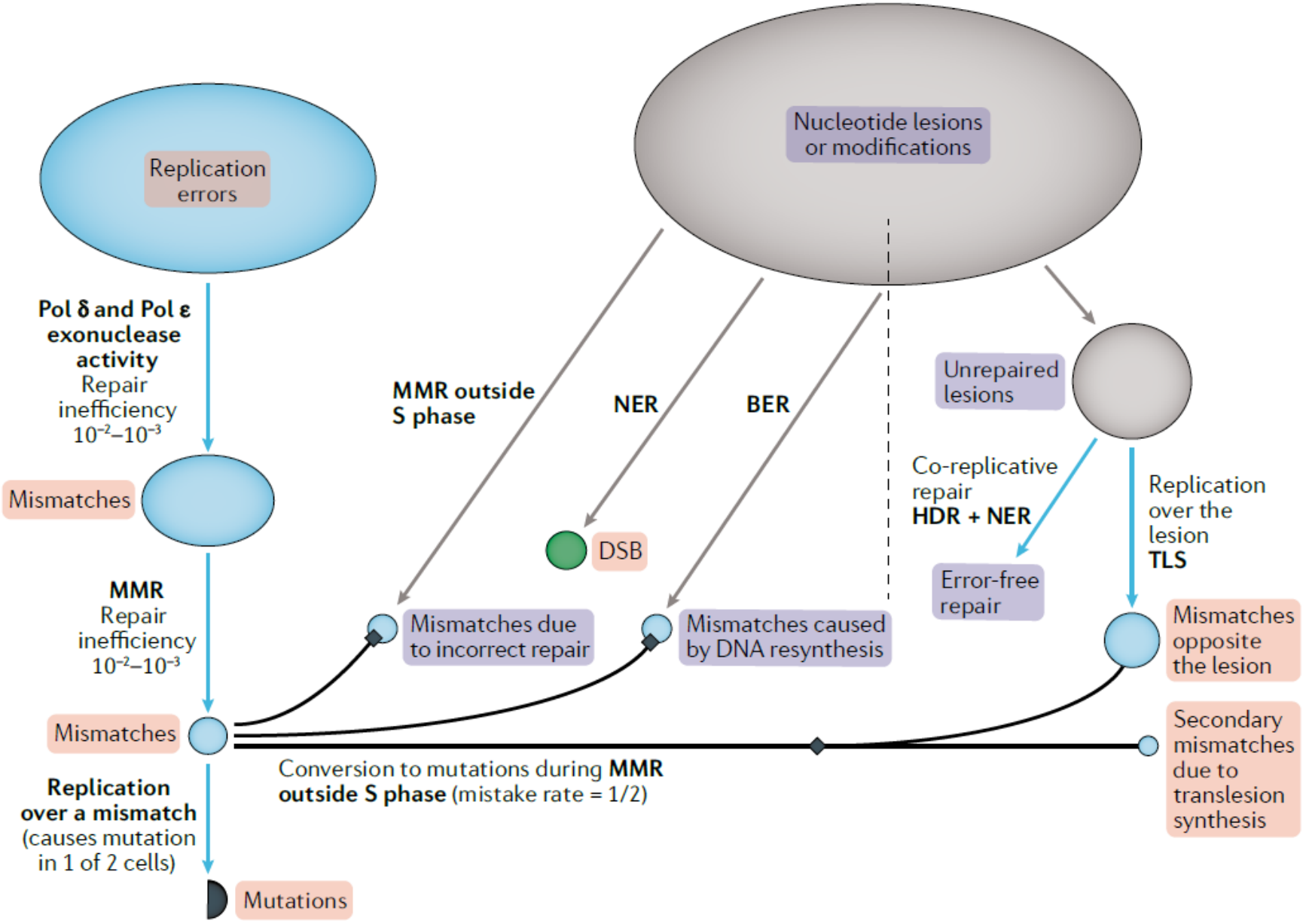
Ionizing radiation: gamma- and X-rays can cause single-strand or double-strand breaks in the sugar-phosphate backbone.

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

Sources of point mutations



Seplyarskiy and Sunyaev (2021) *Nat Rev Genet*

DNA repair mechanisms

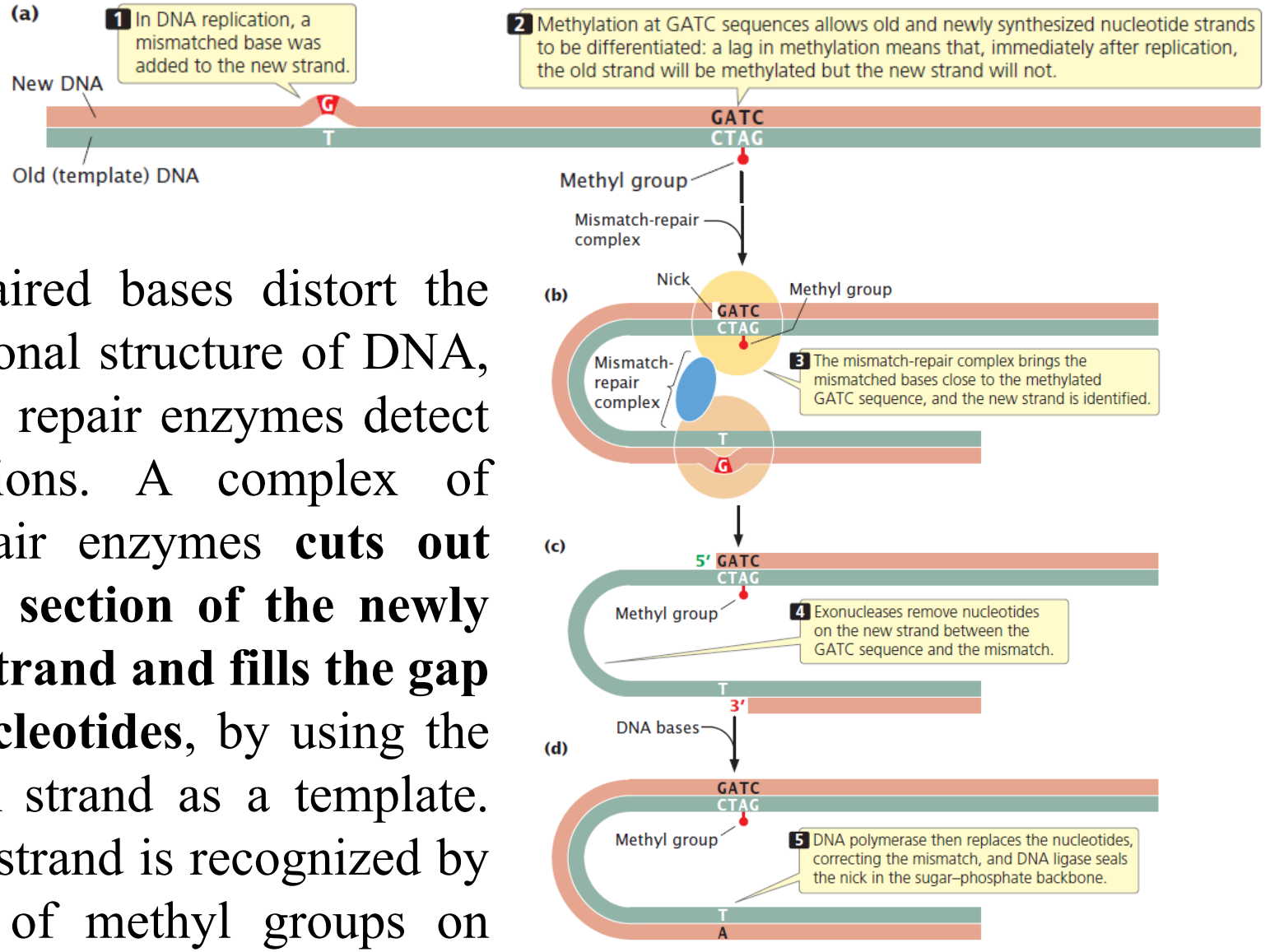
One strand affected:

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

Both strands affected:

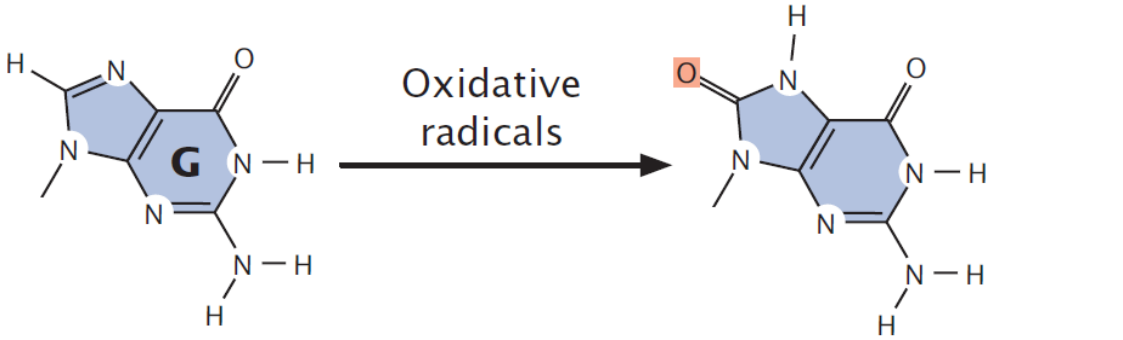
- **Non-homologous end joining (NHEJ)**: ionizing radiation; errors at the replication fork; strong oxidizing agents; metabolites produced in the cell
- **Homologous recombination (HR)**: when a double-strand break occurs shortly after a stretch of DNA has been replicated; at that time, the duplicated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

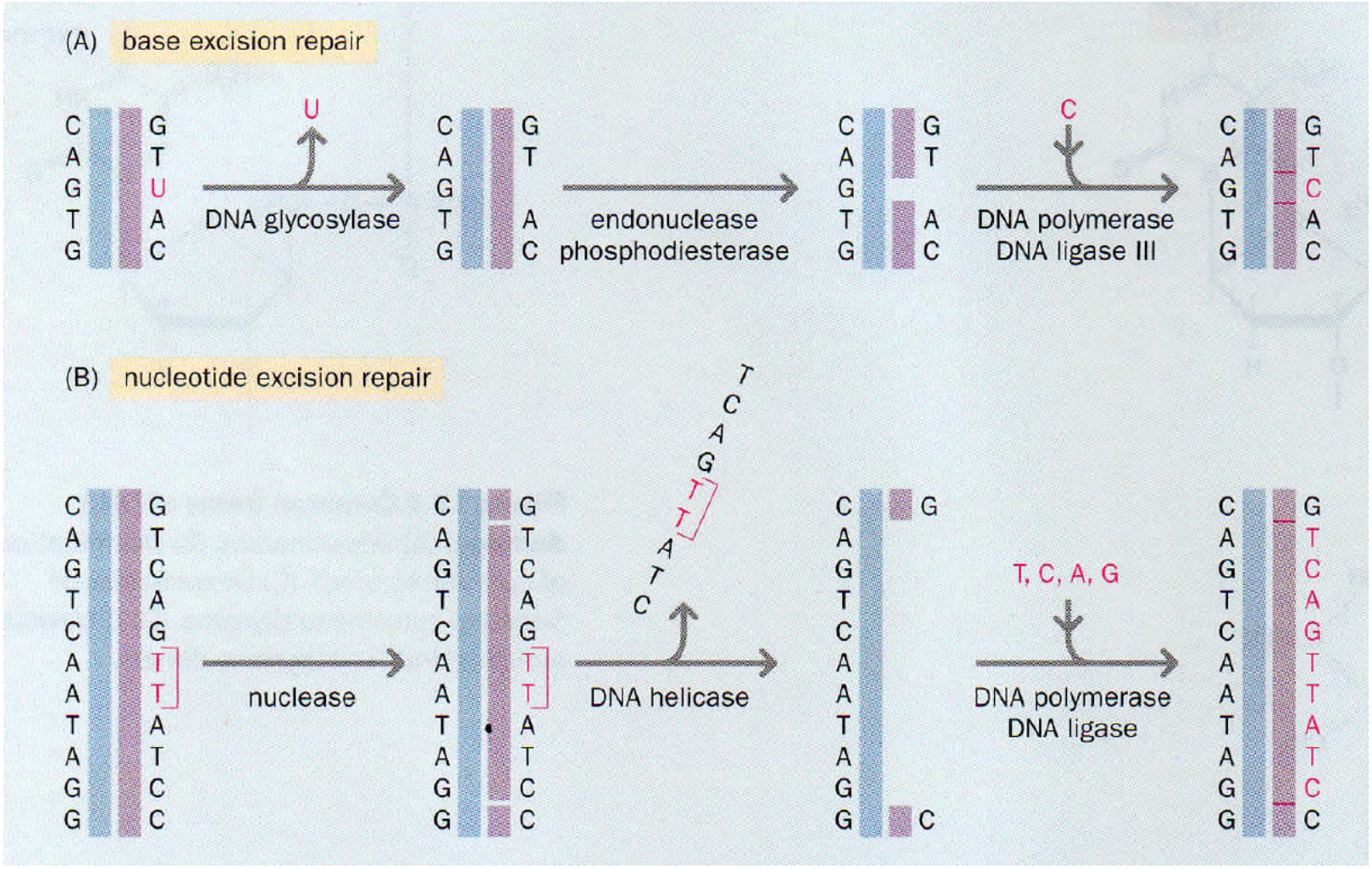
Repair by direct reversal (2)



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)

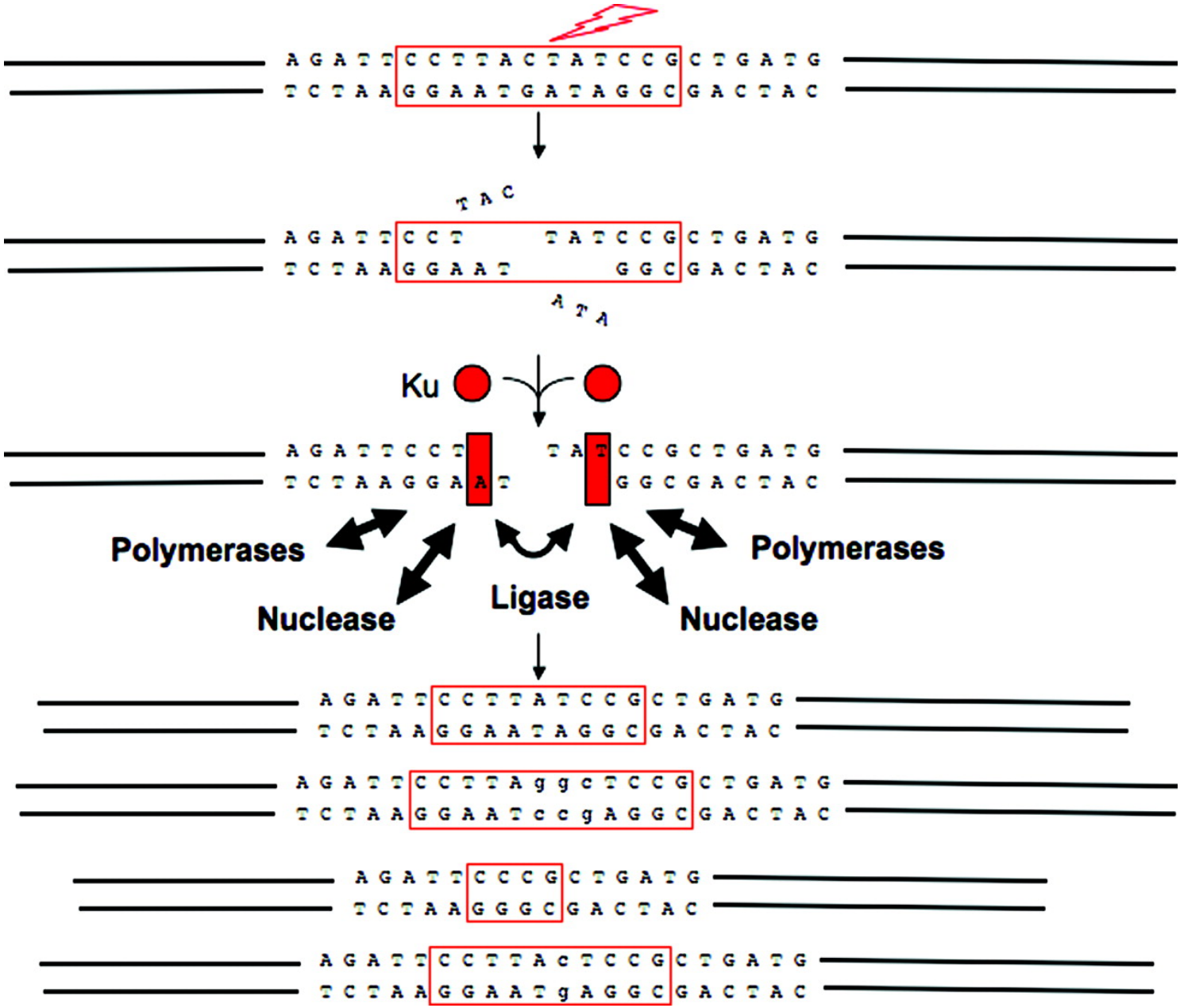


(A) **Base excision repair (BER)** corrects most common DNA damages: ~20,000 bases in each cell per day

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

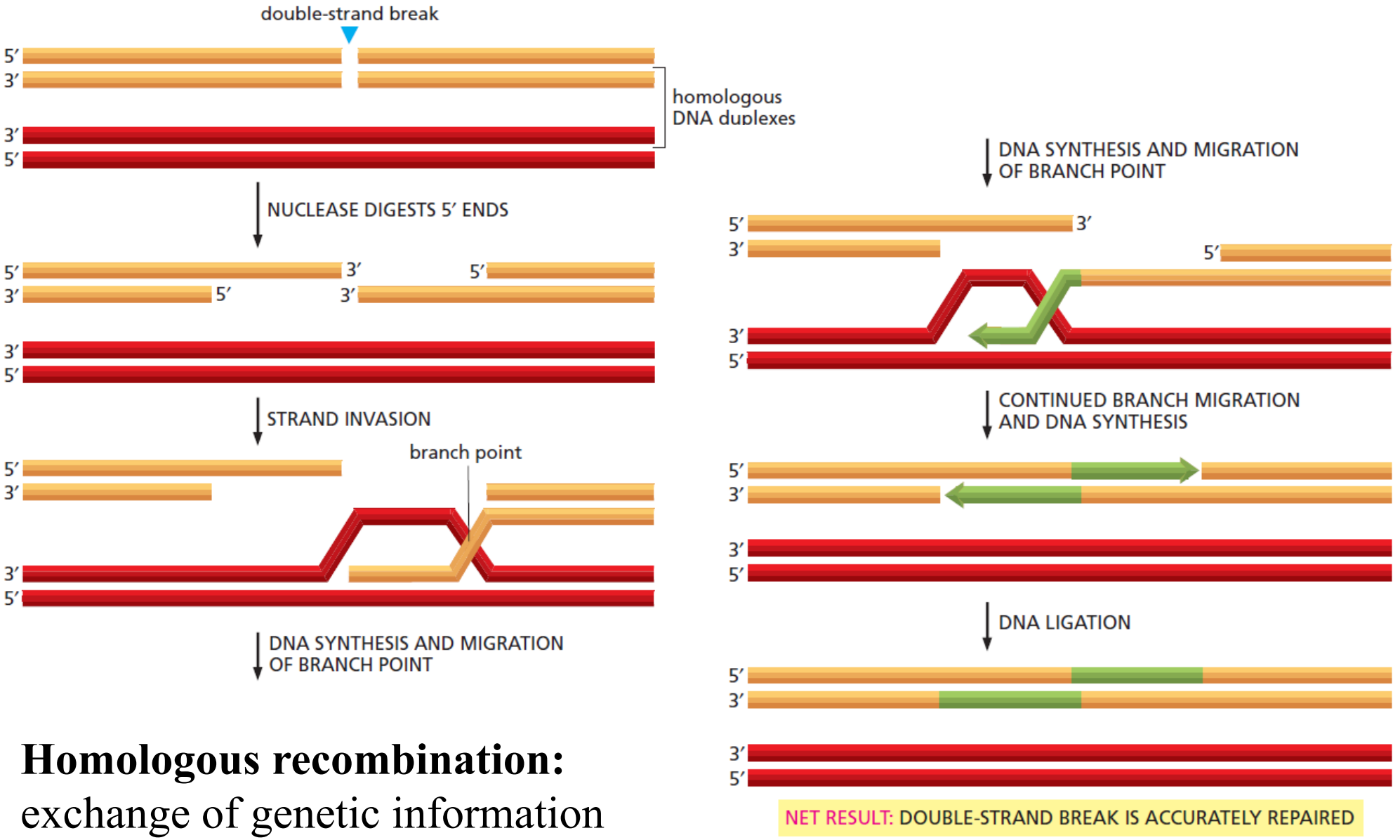
Strachan, Read. *Human Molecular Genetics*

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*

Genetic diseases associated with 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

Genetic diseases associated with defects in DNA-repair systems

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key	Gene/Locus	Gene/Locus MIM number
17p13.1	Li-Fraumeni syndrome	151623	AD	3	TP53	191170

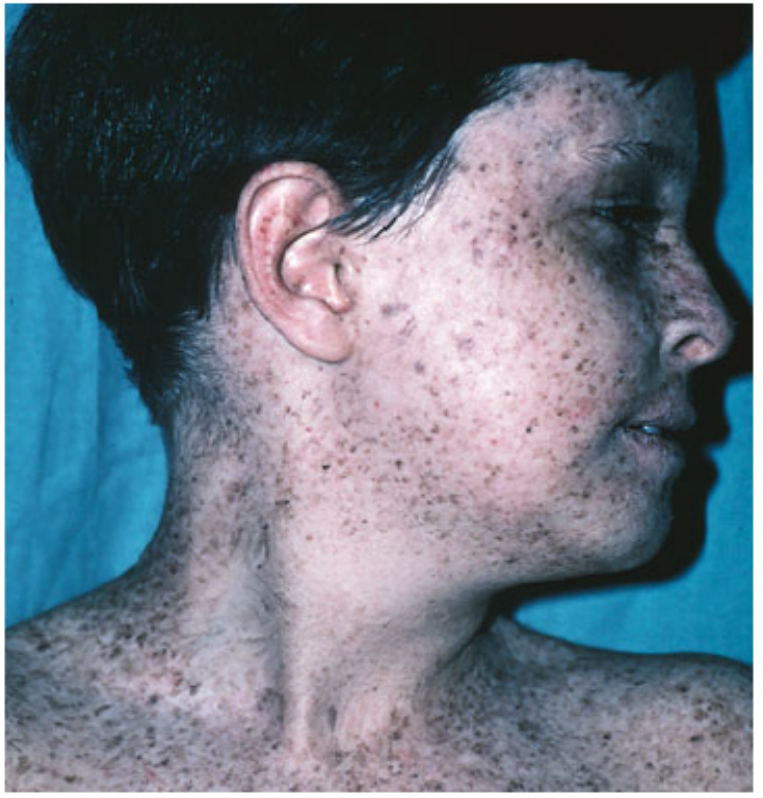
- **Li–Fraumeni syndrome** is a rare, autosomal dominant, hereditary disorder that predisposes carriers to cancer development.

- The risk of developing any invasive cancer (excluding skin cancer) is about 50% by age 30 (1% in the general population) and is 90% by age 70.

- The syndrome is linked to germline mutations of the ***TP53*** **tumor suppressor gene**, which encodes a transcription factor P53 that normally assists in the control of cell division and growth. TP53 typically becomes expressed due to cellular stressors, such as DNA damage, and can halt the cell cycle to assist with either the repair of repairable DNA damage, or can induce apoptosis of a cell with irreparable damage.

Genetic diseases associated with 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

Overall dnSNV rate: 40-80 in a newborn

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

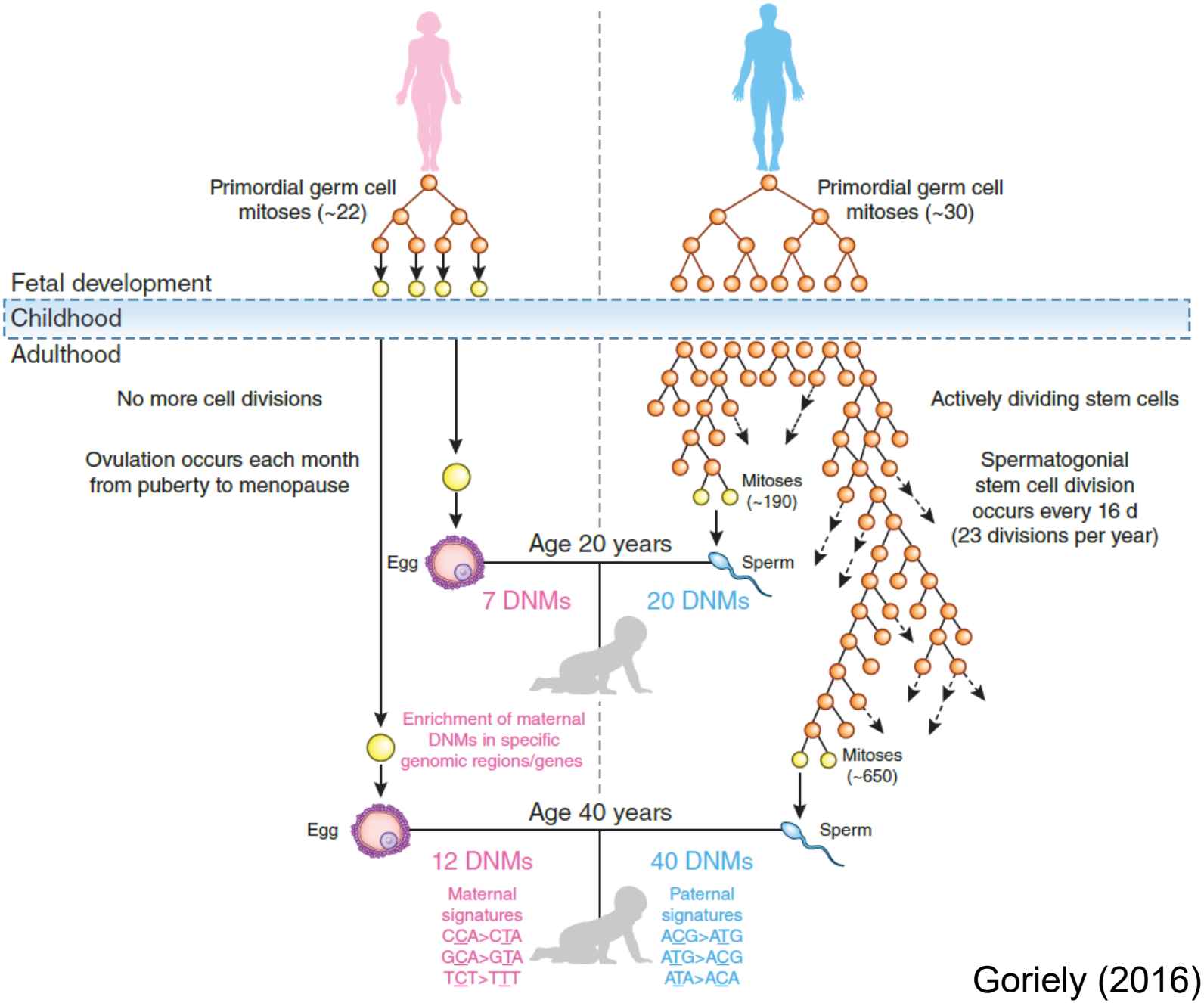
Mutation rates in disease-causing genes

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

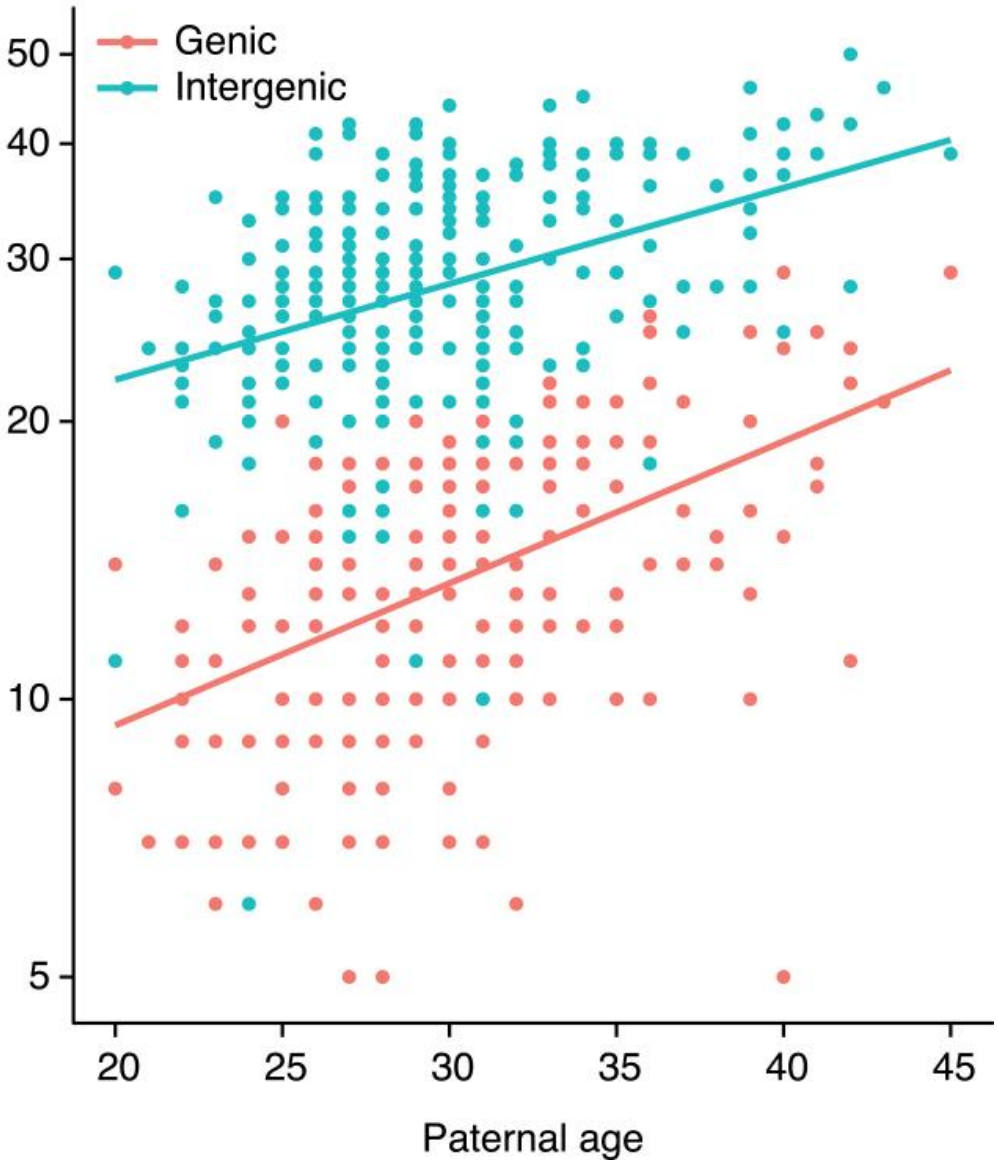
Paternal and maternal DNMs

+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-methyl-C and C: 40%
- T:A>C:G, cause unknown: 25%

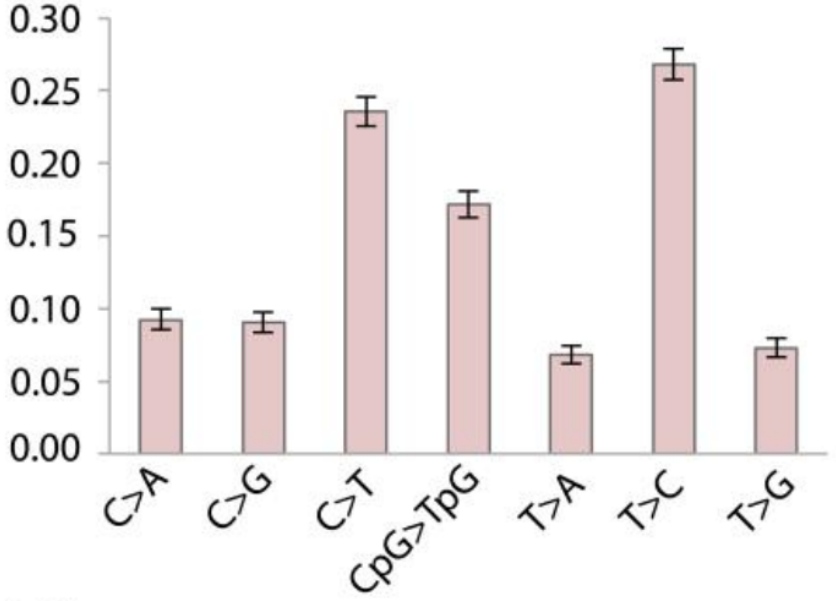
Note: CpG are only ~1% of the genome, so also at non-CpG; but transitions at CpG are ~18x more frequent than non-CpG

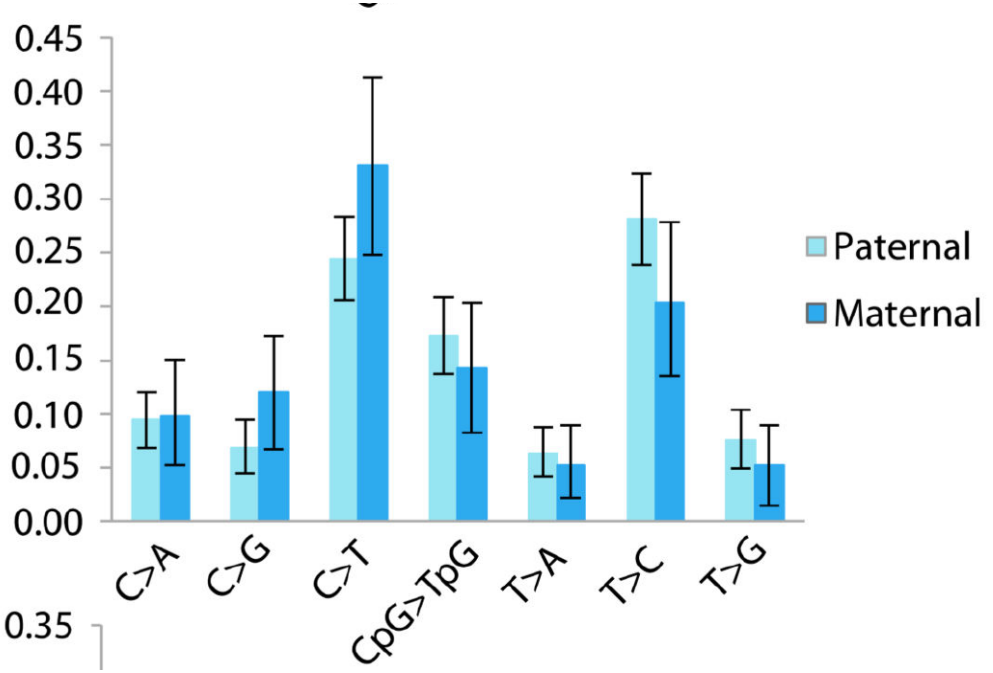
Transversions // occur ~2.5x more frequently at CpG sites

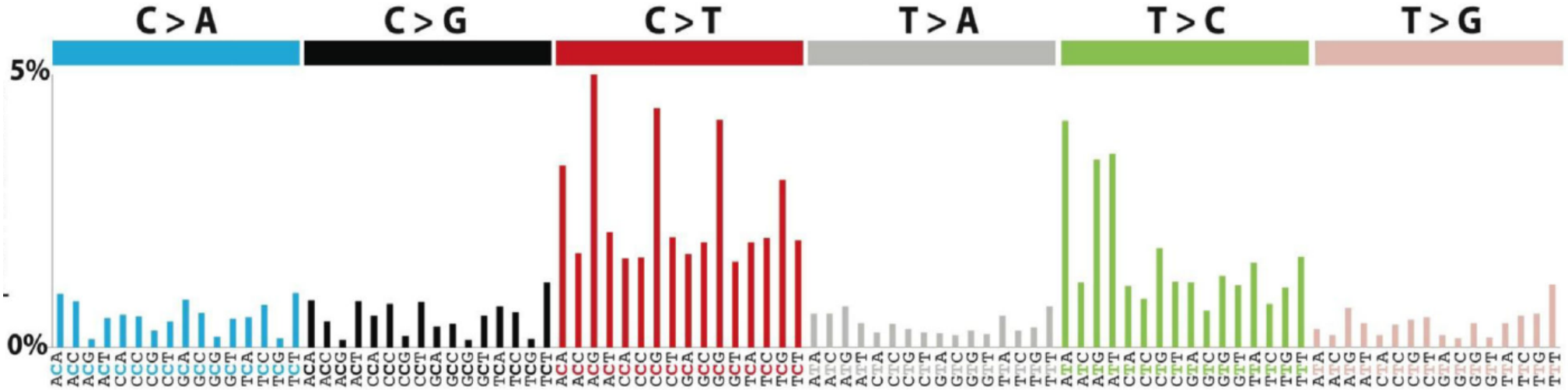
- G:C>T:A: 10%
- G:C>C:G: 10%
- A:T>C:G: <8%
- A:T>T:A: <8%

Ohno (2019) *Genes & Genet Systems*

De novo mutation spectra







6,570 high confidence DNMs from 109 trios

Rahbari et al. (2016) *Nat Genet*

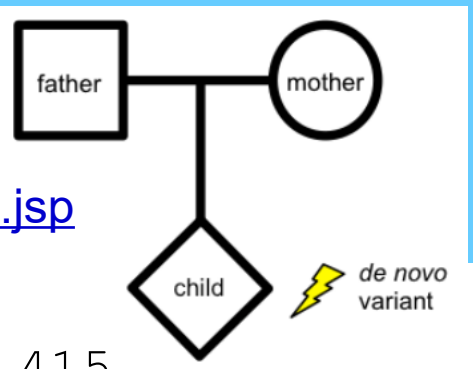
De novo mutation spectra

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*

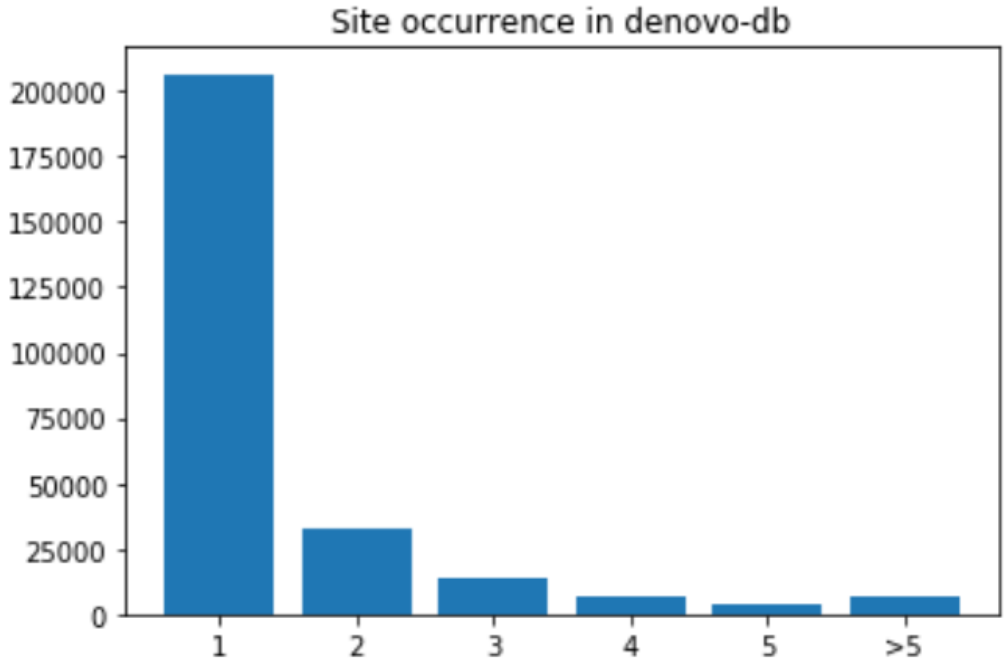
Denovo-db v.1.6.1

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

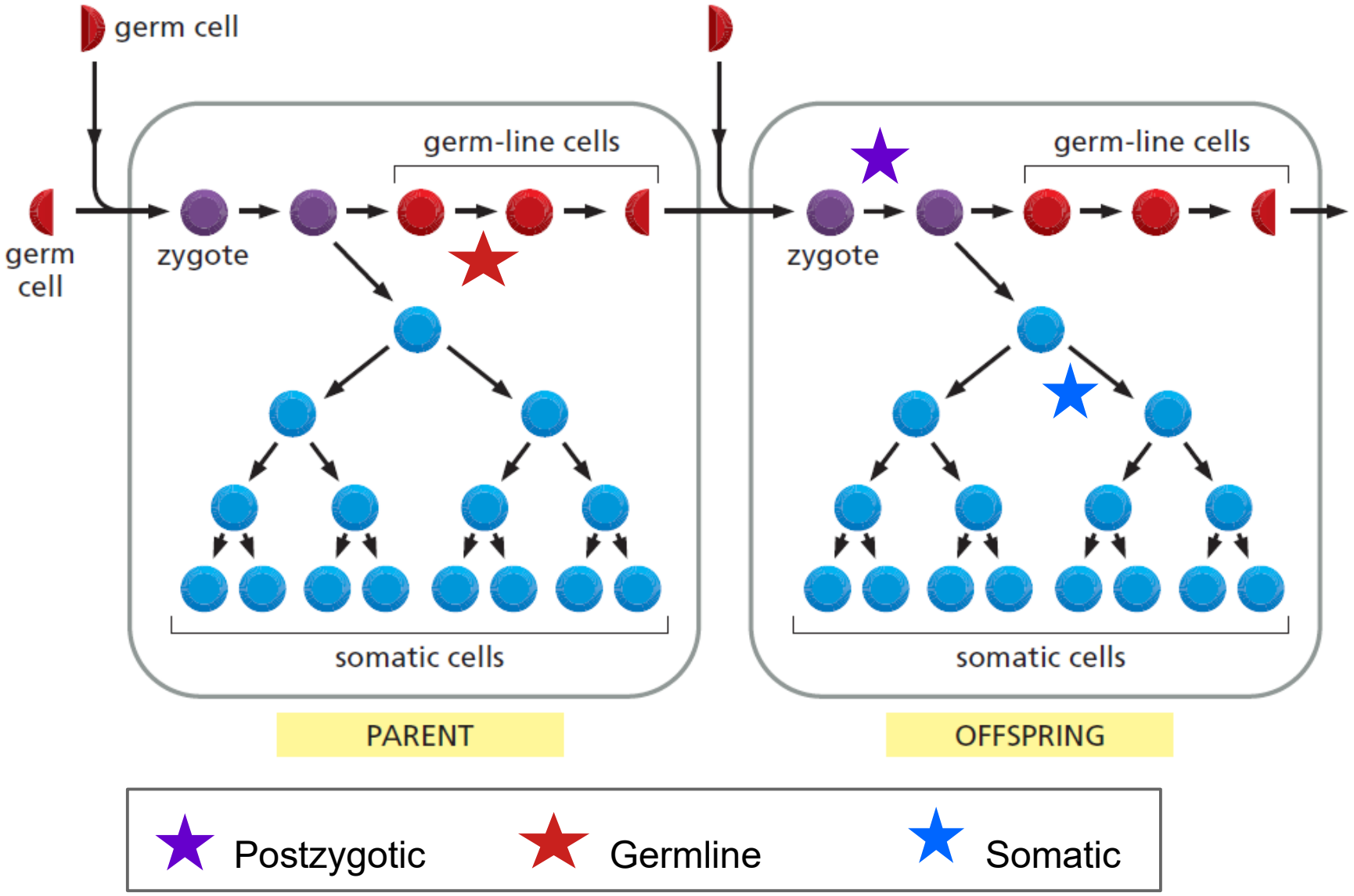


Variant	Counts	%
C>T	79,110	19.0
G>A	79,016	19.0
A>G	47,666	11.5
T>C	47,584	11.5
C>G	17,431	4.2
G>C	17,154	4.1
C>A	15,719	3.8
G>T	15,234	3.7
A>C	12,744	3.1
T>G	12,464	3.0
T>A	11,637	2.8
A>T	11,482	2.8
T>TA	739	0.2
CT>C	737	0.2
...		
Total	*415,515*	

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

De novo mutations in human disease

- Ultra-rare individually, but significant collectively: 60-75% of all sporadic disease cases are DNMs
- More damaging than inherited; effect depends on timing
- **Severe pediatric disorders in outbred populations**: sporadic malformation syndromes (Schinzel–Giedion, Kabuki, Bohring–Opitz), neurodevelopmental (severe intellectual disability, ID), congenital heart disease (CHD)
- **Late-onset neurological and psychiatric disorders**: Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), schizophrenia (SCZ), epilepsy, Tourette syndrome (TS), autistic-spectrum disorder (ASD), and bipolar disorder (BP)
  Example: 10% SCZ cases have DNM CNV vs 1.26% controls
- **Inherited cancers**: Li-Fraumeni syndrome (TP53), familial adenomatous polyposis (APC), ~7% of non-somatic mutations are DNMs

Structural variants and CNVs

Structural variant (aka **chromosomal rearrangement)**: sequence change >1 kbp in size

- Balanced
  Inversion or translocation
- Unbalanced (aka **CNV, copy number variant**)
  Tandem or dispersed duplication, deletion, insertion

Mechanisms

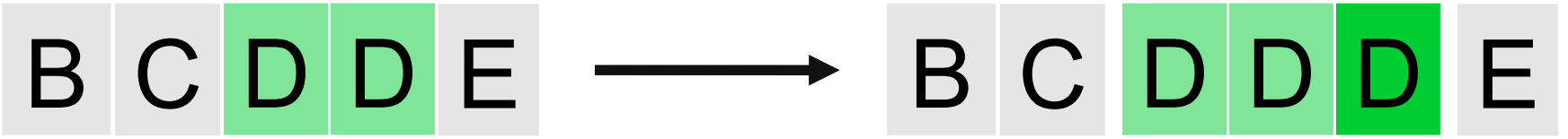
- Recombination: non-allelic homologous recombination (NAHR)
- Nonreplicative: Nonhomologous end joining (NHEJ) repair
- Replication-based:
  Fork stalling and template switching (FoSTeS)
  Microhomology-mediated break-induced replication (MMBIR)
- Retrotransposition (LINE1, Alu repeat)

Structural variants and CNVs

1. Unbalanced structural variants (CNVs)

Tandem duplication

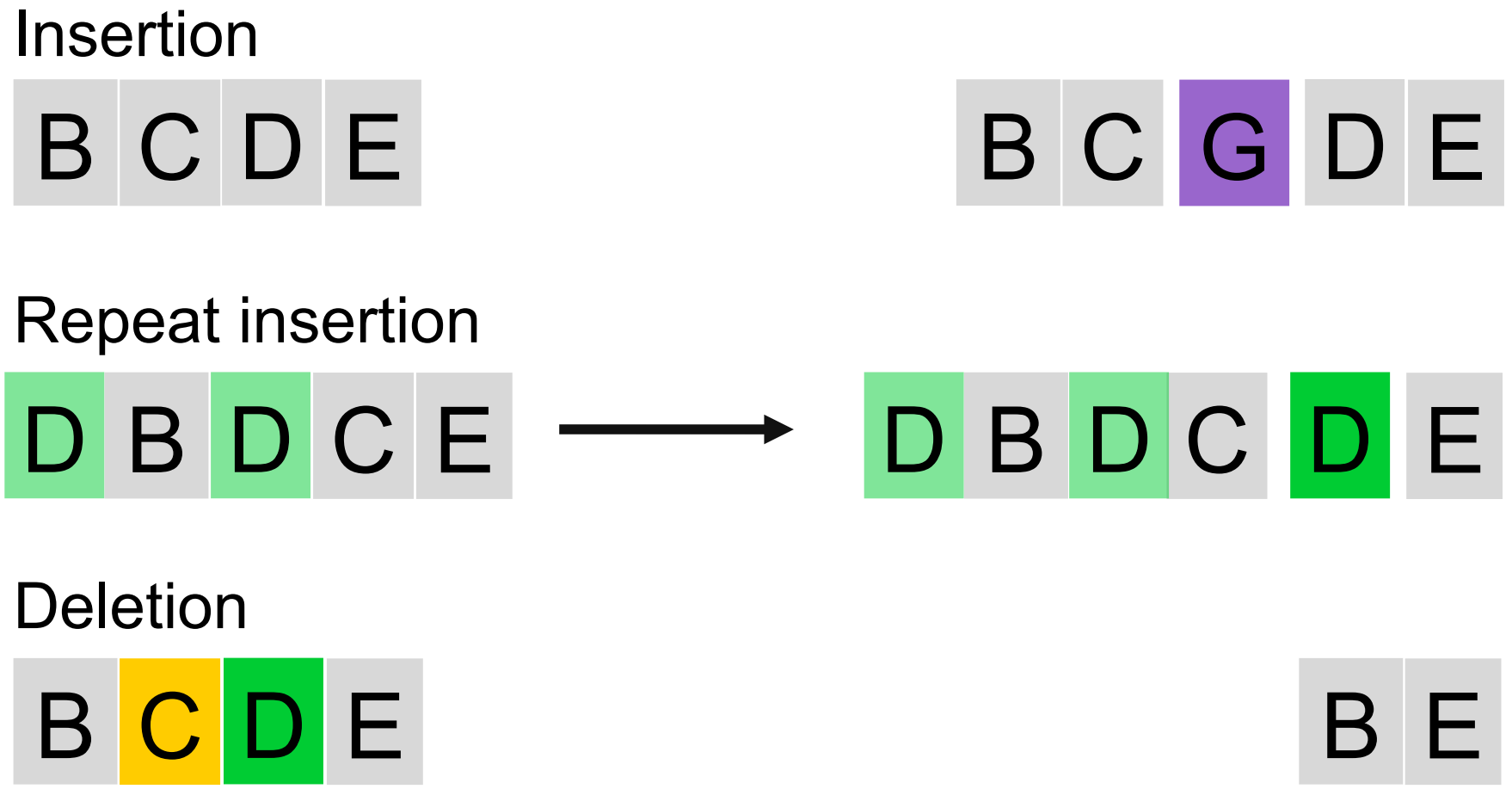
Variable Number of Tandem Repeats



Dispersed duplication

Structural variants and CNVs

1. Unbalanced structural variants (CNVs)

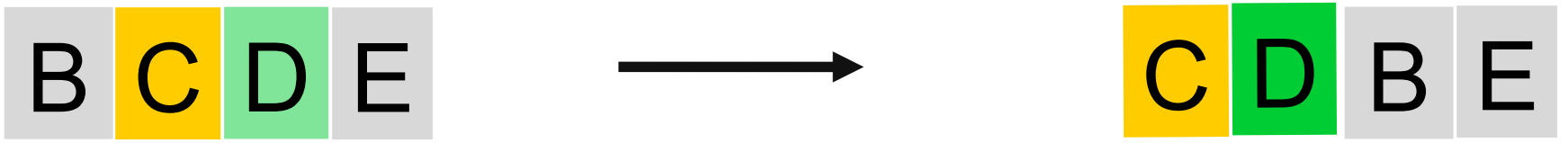


Structural variants and CNVs

2. Balanced structural variants

Inversion

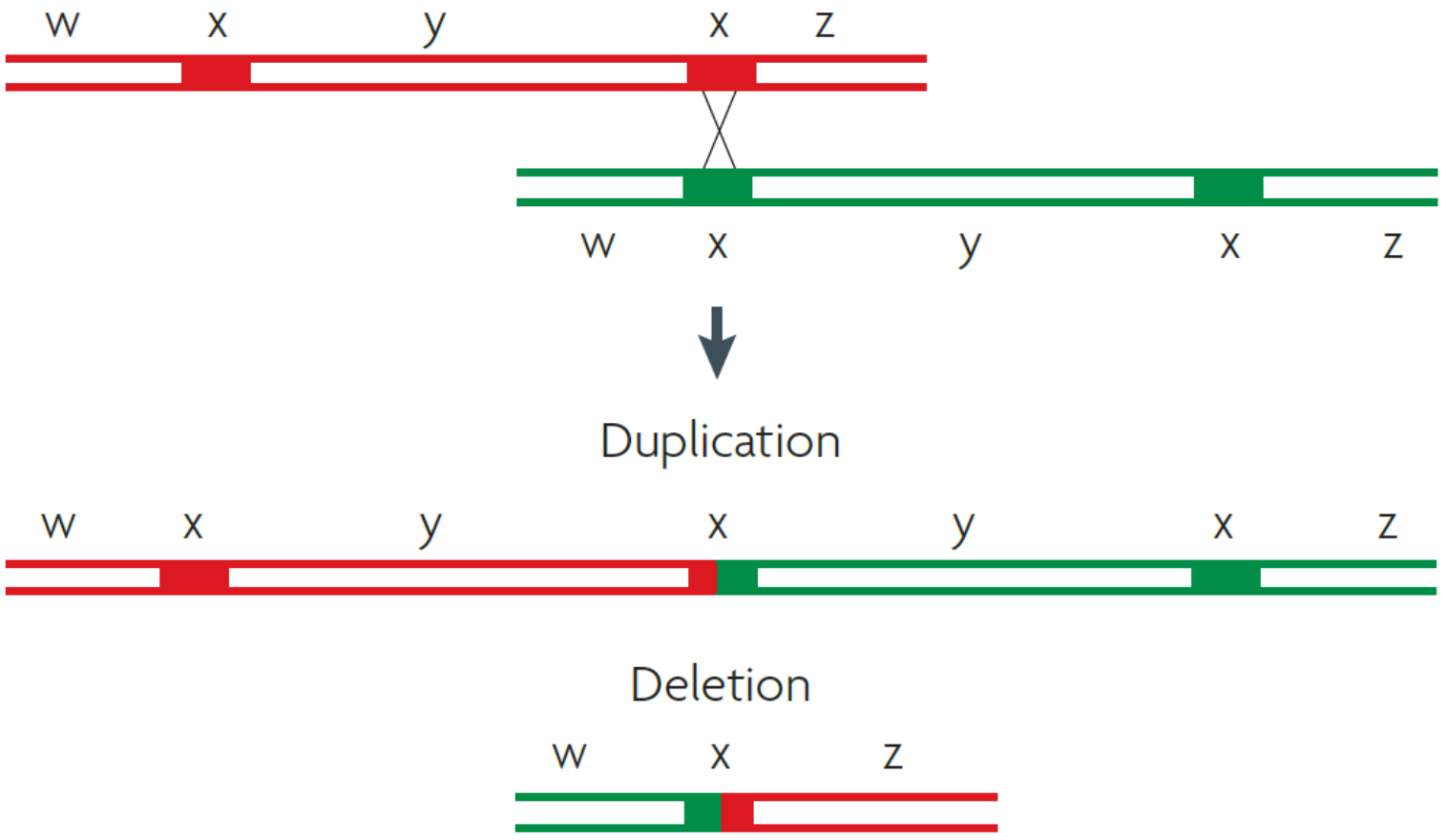
Intra-chromosomal translocation (ITX)



Inter-chromosomal translocation (CTX)

Mechanisms of chromosomal rearrangements

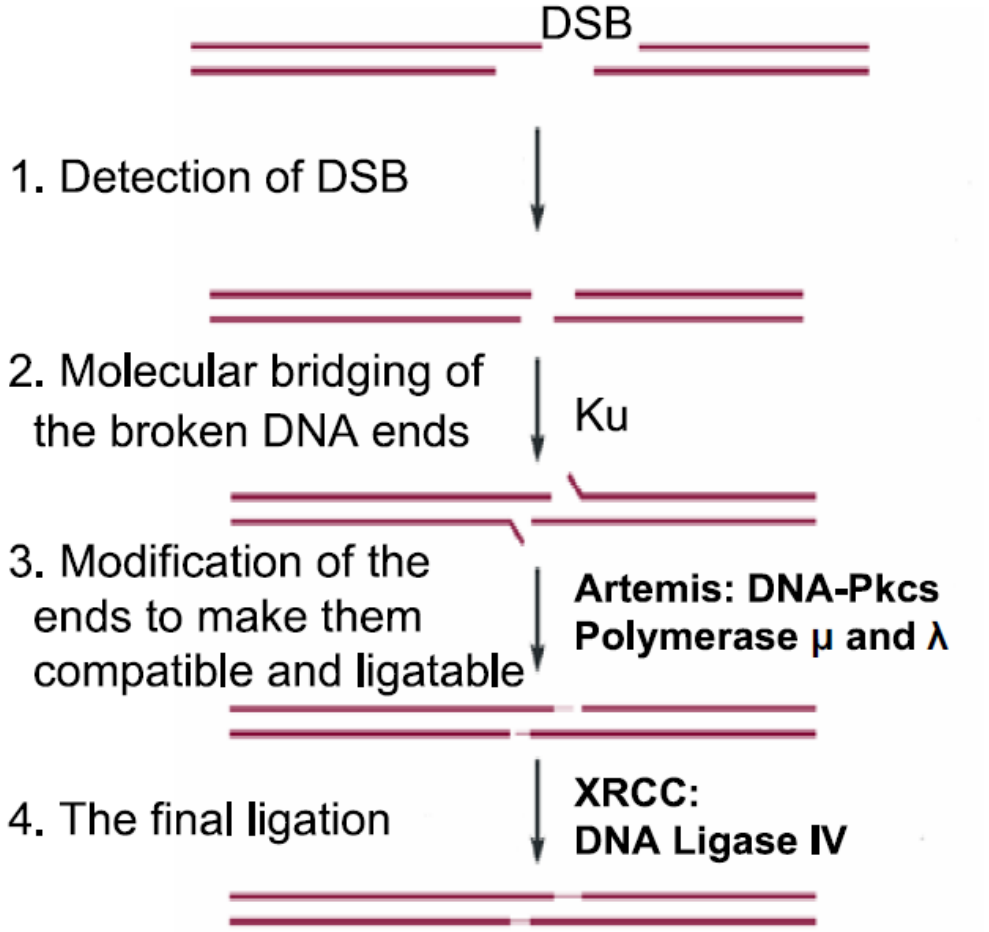
Non-allelic homologous recombination (NAHR)



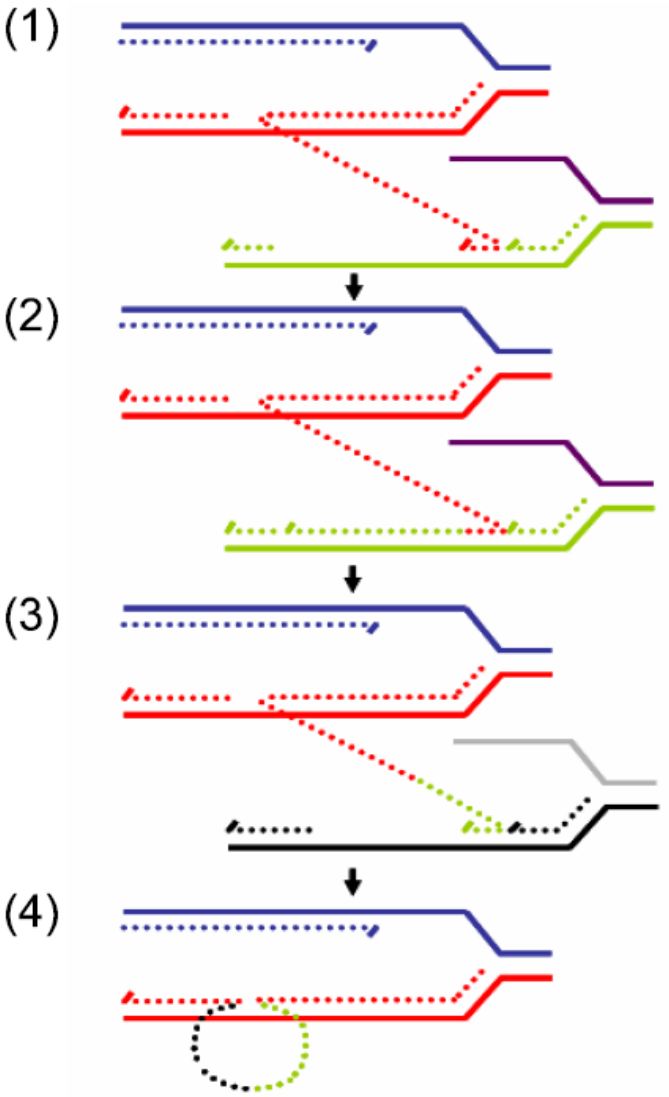
Hastings (2009) *Nat Rev Genet*

Mechanisms of chromosomal rearrangements

Nonhomologous end joining (NHEJ)

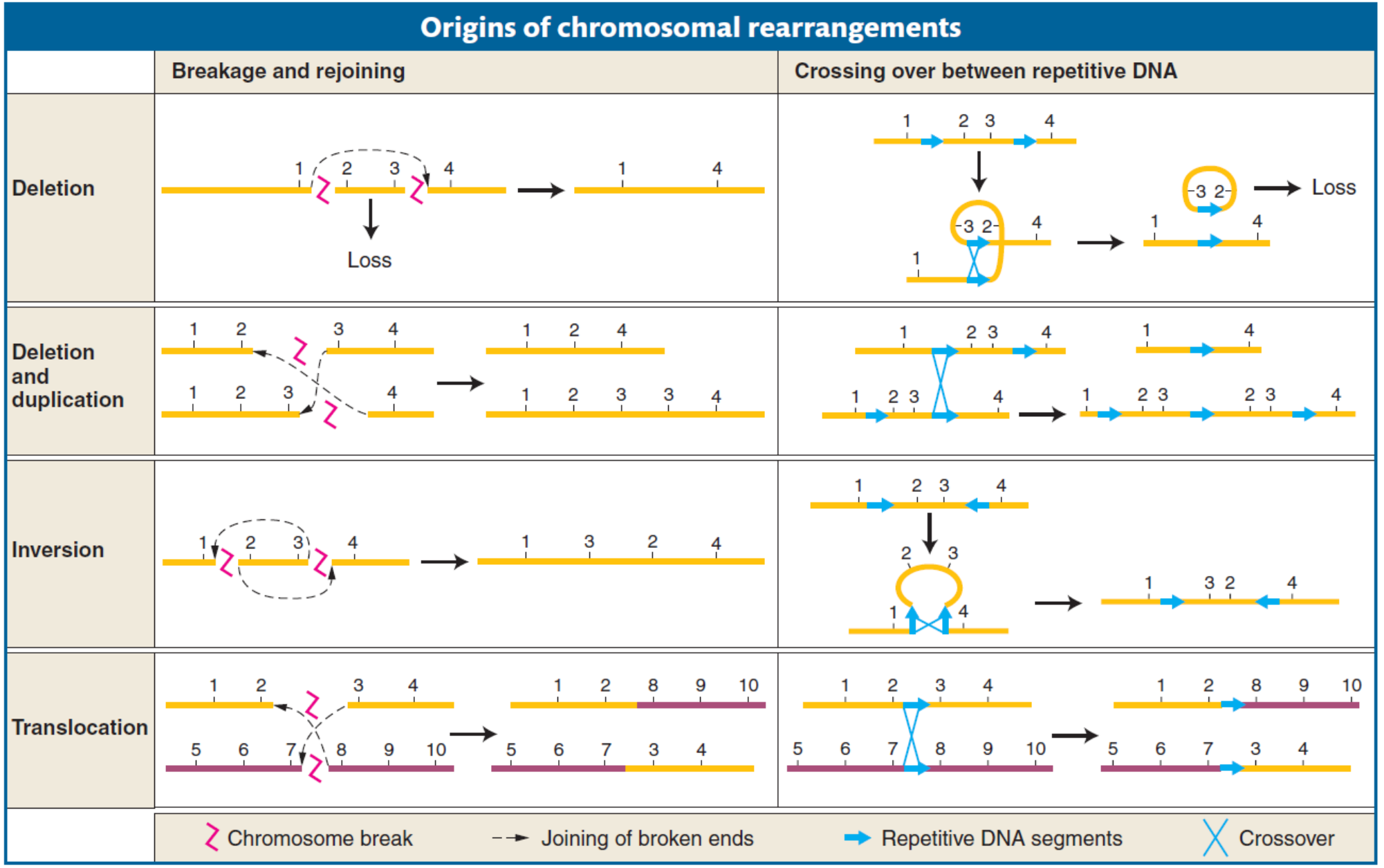


Fork stalling and template switching (FoSTeS)



Gu(2008) *Pathogenetics*

Mechanisms of chromosomal rearrangements



Griffiths -- *Introduction to Genetic Analysis*

Am J Hum Genet. 2021 Apr 1;108(4):597-607. doi: 10.1016/j.ajhg.2021.02.012. Epub 2021 Mar 5.

De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families

Jonathan R Belyeu [1], Harrison Brand [2], Harold Wang [2], Xuefang Zhao [2], Brent S Pedersen [1], Julie Feusier [3], Meenal Gupta [1], Thomas J Nicholas [1], Joseph Brown [1], Lisa Baird [1], Bernie Devlin [4], Stephan J Sanders [5], Lynn B Jorde [6], Michael E Talkowski [7], Aaron R Quinlan [8]

Abstract

Each human genome includes *de novo* mutations that arose during gametogenesis. While these germline mutations represent a fundamental source of new genetic diversity, they can also create deleterious alleles that impact fitness. Whereas the rate and patterns of point mutations in the human germline are now well understood, far less is known about the frequency and features that impact de novo structural variants (dnSVs).

Am J Hum Genet. 2021 Apr 1;108(4):597-607. doi: 10.1016/j.ajhg.2021.02.012. Epub 2021 Mar 5.

De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families

Jonathan R Belyeu [1], Harrison Brand [2], Harold Wang [2], Xuefang Zhao [2], Brent S Pedersen [1], Julie Feusier [3], Meenal Gupta [1], Thomas J Nicholas [1], Joseph Brown [1], Lisa Baird [1], Bernie Devlin [4], Stephan J Sanders [5], Lynn B Jorde [6], Michael E Talkowski [7], Aaron R Quinlan [8]

Introduction

Several mechanisms, including replication infidelity, genomic damage, non-allelic recombination, and double-strand break repair, are known to create de novo mutations (DNMs) in the human germline. These mutations contribute to genomic diversity and often are primary targets in the analysis of rare, dominant genetic disorders. There is therefore a long-standing interest in understanding the frequency at which DNMs occur and the patterns that affect these rates. Numerous studies have measured the rate of germline de novo single-nucleotide variants (dnSNVs) and small insertion-deletion mutations (indels) at approximately 70 events per individual, and it has been established that the majority of these small point mutations arise on the paternal gamete. The frequency of single-nucleotide and insertion-deletion DNMs increases with parental age, especially paternal age.

Am J Hum Genet. 2021 Apr 1;108(4):597-607. doi: 10.1016/j.ajhg.2021.02.012. Epub 2021 Mar 5.

De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families

Jonathan R Belyeu [1], Harrison Brand [2], Harold Wang [2], Xuefang Zhao [2], Brent S Pedersen [1], Julie Feusier [3], Meenal Gupta [1], Thomas J Nicholas [1], Joseph Brown [1], Lisa Baird [1], Bernie Devlin [4], Stephan J Sanders [5], Lynn B Jorde [6], Michael E Talkowski [7], Aaron R Quinlan [8]

In contrast, precise estimates of germline mutations affecting the structure of the human genome (structural variants [SVs]) have been far more difficult to discern.

De novo SVs (dnSVs) largely arise from mutational mechanisms that are distinct from those responsible for point mutations. The larger size of SVs, defined here and in many other studies as variants affecting at least 50 base pairs, increases the likelihood that any given SV will impact protein-coding genes or other critical genomic regions. Understanding the selective constraints on dnSV-specific mechanisms is essential because a broad spectrum of balanced, unbalanced, and complex structural mutations are known to underlie many developmental disorders. However, dnSVs are predicted to occur several hundred-fold less frequently than point mutations, requiring a much larger sample size to achieve accurate estimates of dnSV rates.

Am J Hum Genet. 2021 Apr 1;108(4):597-607. doi: 10.1016/j.ajhg.2021.02.012. Epub 2021 Mar 5.

De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families

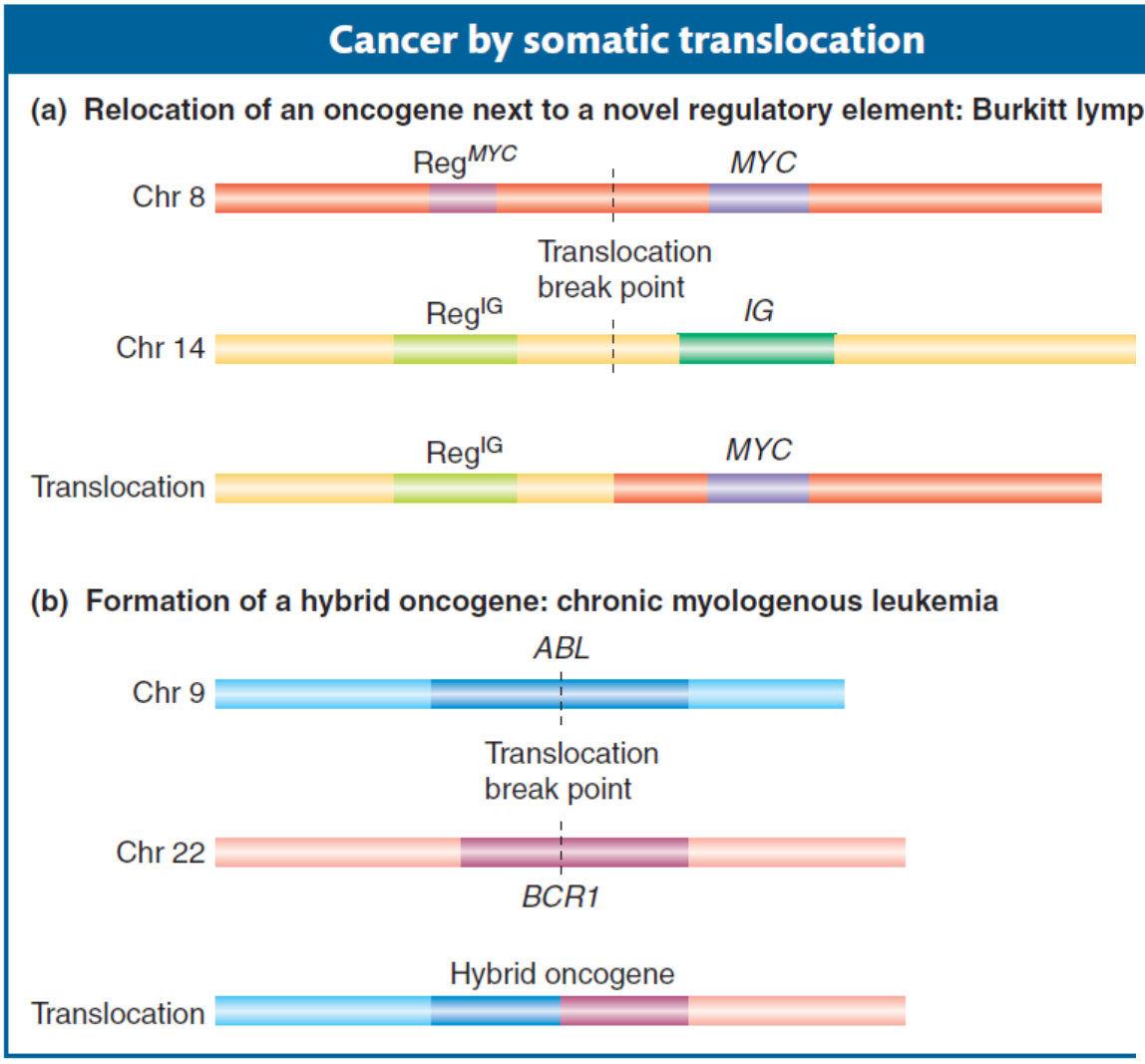
Jonathan R Belyeu [1], Harrison Brand [2], Harold Wang [2], Xuefang Zhao [2], Brent S Pedersen [1], Julie Feusier [3], Meenal Gupta [1], Thomas J Nicholas [1], Joseph Brown [1], Lisa Baird [1], Bernie Devlin [4], Stephan J Sanders [5], Lynn B Jorde [6], Michael E Talkowski [7], Aaron R Quinlan [8]

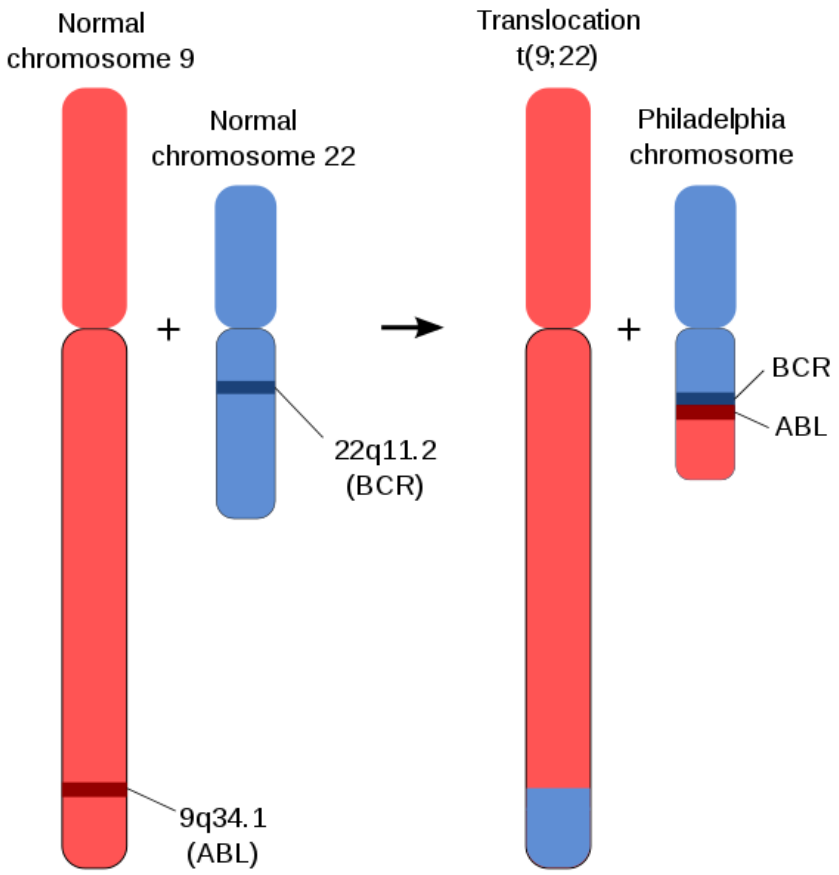
- Family-based study of germline mutations among 9,599 human genomes from 33 multigenerational CEPH-Utah families and 2,384 families from the Simons Foundation Autism Research Initiative; short-read WGS

- dnSV rate: 0.160 events per genome in unaffected individuals, 0.206 per genome) in ASD-affected individuals.
- In both probands and unaffected samples, ~73% of dnSVs arose in paternal gametes
- Most de novo structural mutations to be caused by mutational mechanisms that do not require sequence homology.
- No statistically significant correlation between parental age and dnSV in offspring.

Conclusion: dnSVs have different mechanisms than dnSNVs

Chromosomal rearrangements and disease





The *MYC* proto-oncogene is a transcription factor that plays a role in cell cycle progression, apoptosis and cellular transformation. The *ABL* proto-oncogene encodes a protein kinase in a cell proliferation signaling pathway. The Bcr1-Abl fusion protein has a permanent kinase activity, regardless of the initiating signal.

Griffiths -- *Introduction to Genetic Analysis*

Chromosomal rearrangements and disease

Syndrome	Chromosomal location and key genes (if identified)	Typical size of deletion/duplication	Estimated incidence among live-births	Typical phenotypic features (not exhaustive, and not all these features are seen in all cases)
Di George syndrome/22q11 deletion syndrome	**22q11.2** *TBX1, COMT*	3 Mb deletion (90% of cases)	1/4000	Congenital heart defects, cleft palate, developmental delay, learning difficulty, increased risk of mental illness, recurrent infections
Williams syndrome/Williams–Beuren syndrome	**7q11.3** *CLIP2, ELN, GTF2I, GTF2IRD1, LIMK1*	1.5–1.8 Mb deletion	1/7500 to 1/10000	Supravalvular aortic stenosis, joint problems and loose skin, mild to moderate intellectual disability, characteristic 'elfin' facial appearance
Smith–Magenis syndrome	**17p11.2** *RAI1*	Approximately 3.6 Mb deletion	1/15000 to 1/25000	Mild to moderate intellectual disability, disturbed sleep patterns, behaviour problems including aggression and self-harm
Cri-du-chat syndrome	**5p15.2** *CTNND2*	Approximately 5–40 Mb deletion	1/15000 to 1/50000	Cat-like cry, microcephaly, severe psychomotor problems and severe intellectual disability
Wolf–Hirschhorn syndrome	**4p16.3** *NSD2, LETM1, MSX1*	Approximately 5–18 Mb deletion	1/50000	Characteristic 'Greek warrior helmet' facial appearance, delayed growth and development, mild to severe intellectual disability
Potocki–Lupski syndrome	**17p11.2** *RAI1*	Approximately 3.6 Mb duplication	1/25000	Developmental delay, mild to moderate learning disability, behavioural problems
Cat eye syndrome/Schmid–Fraccaro syndrome	**22q11** *ADA2, CECR2*	2–5 Mb duplication or triplication	1/50000 to 1/150000	Preauricular skin tags or pits, ocular coloboma, anal atresia with fistula, heart and renal malformations

Jackson (2018) *Essays in Biochemistry*

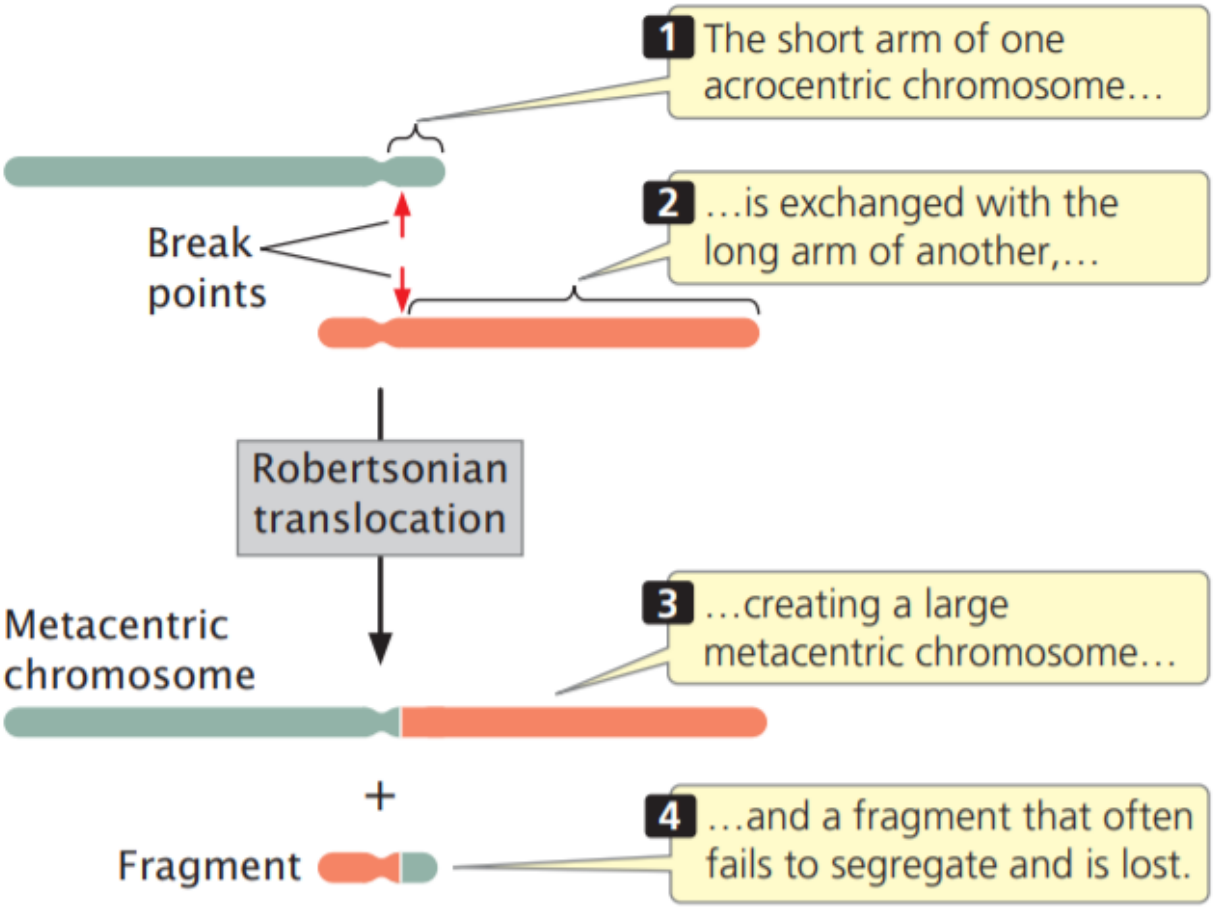
Aneuploidy

Aneuploidy: wrong number of complete chromosomes: nullisomy, monosomy, trisomy. Results from aberration in mitosis or meiosis

- Major cause of spontaneous abortions (~30% of all conceptions)
- Detected in ~0.3-0.6% live human births

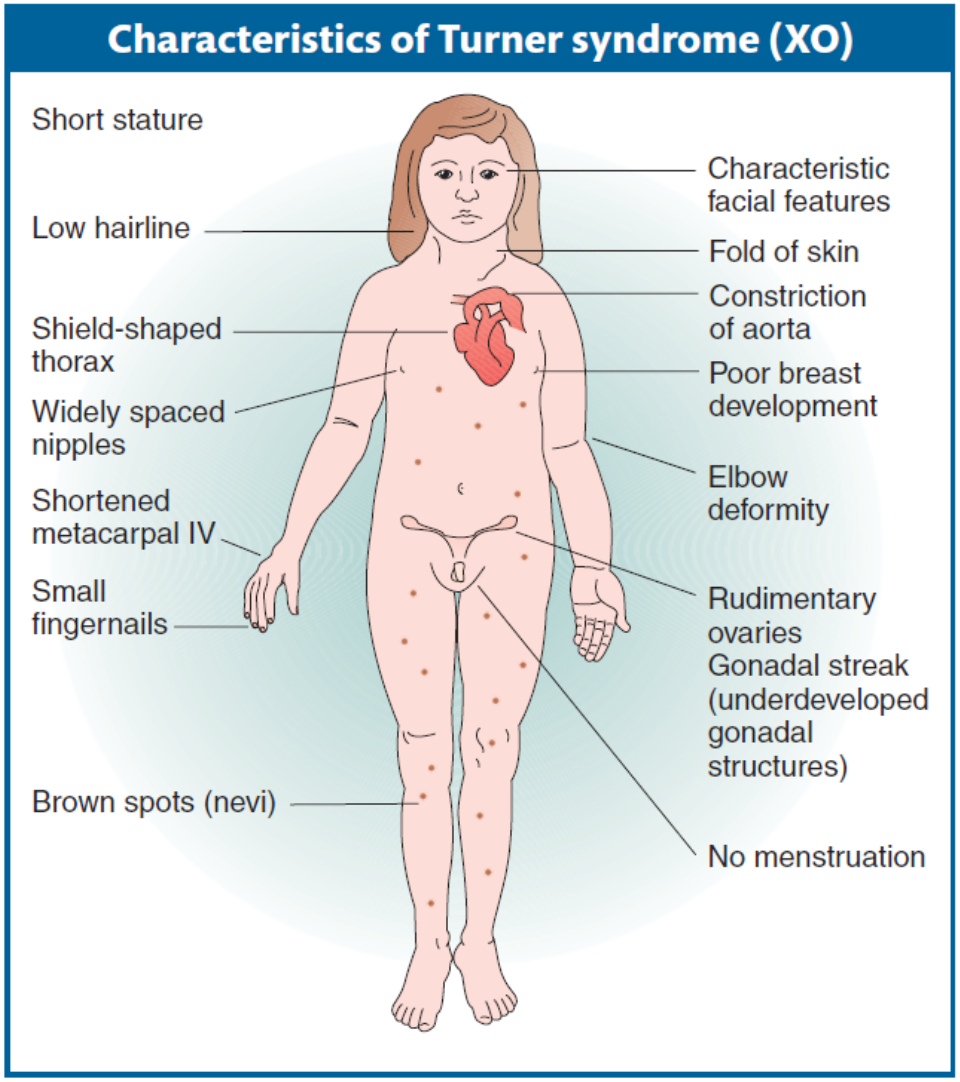
Name	Karyotype	Frequency
Turner syndrome	XO (Females, X monosomy)	1:2000- 1:2500
Klinefelter syndrome	XXY (XXXY, XXXXY, XXYY)	1:1000
Poly-X females	XXX	1:1000
Down syndrome	Trisomy 21	1:1100
Edwards syndrome	Trisomy 18	1:6000
Patau syndrome	Trisomy 13	1:7000-1:14000
Trisomy 8	Trisomy 8	1:25000 – 1:50000

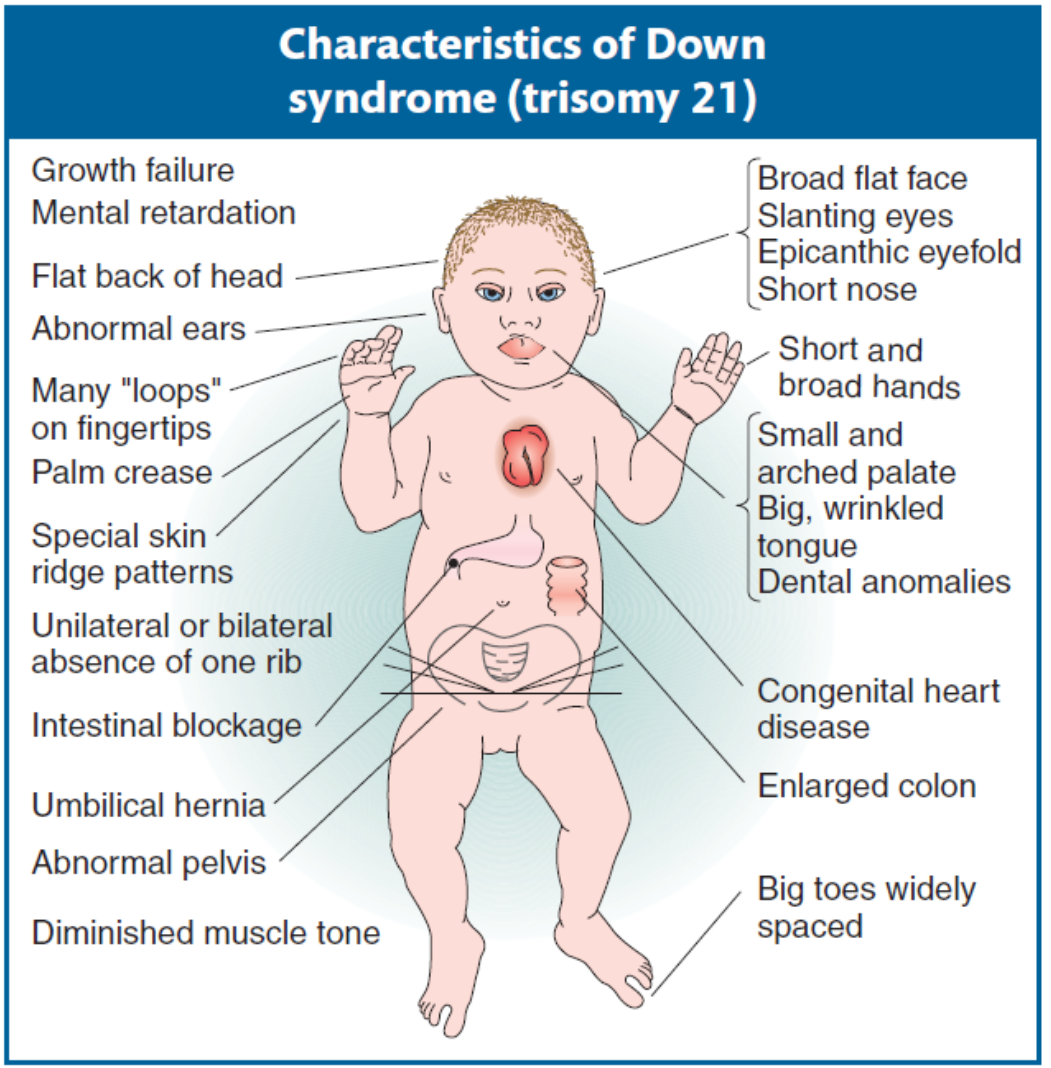
Robertsonian translocation



Robertsonian translocation: the long arms of two acrocentric chromosomes (13,14,15,21) become joined to a common centromere, resulting in a chromosome with two long arms and usually another chromosome with two short arms. **Affects ~1/1000 newborns.**

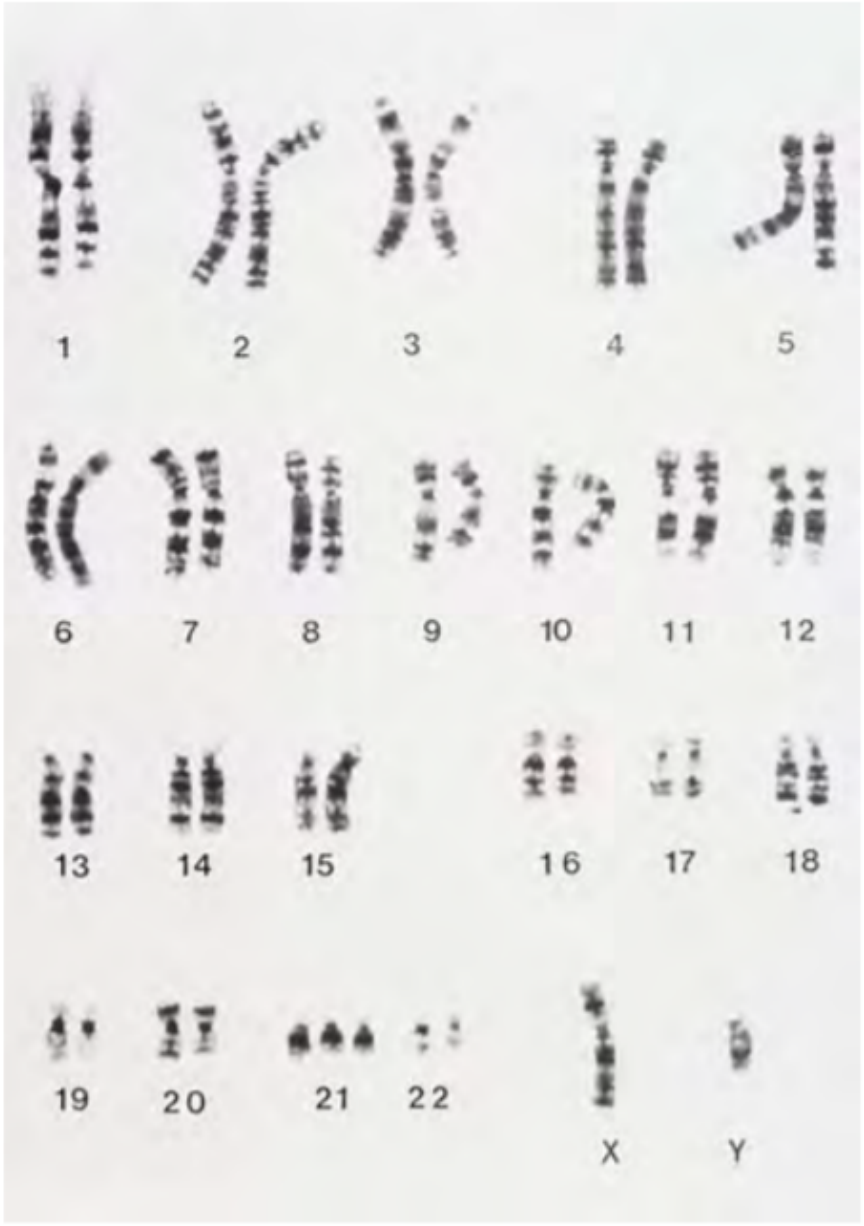
Aneuploidy

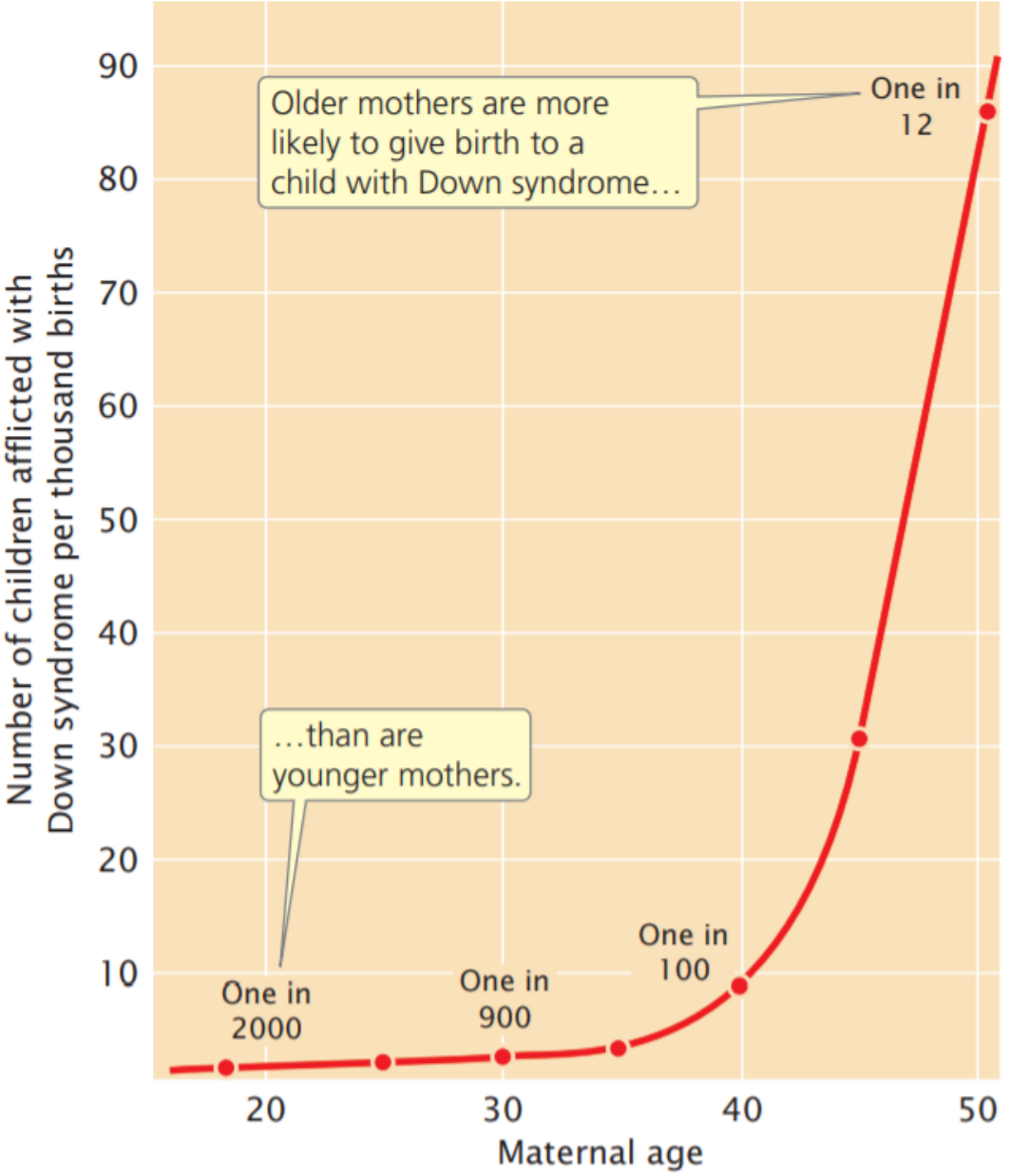




Griffiths -- *Introduction to Genetic Analysis*

Aneuploidy





Down syndrome

Pierce -- *Genetics Essentials. Concepts and Connections*

Aneuploidy

Aneuploidy	Common name	Estimated incidence among life-births	Symptoms can include
Trisomy 13	Patau syndrome	Approximately 1:16000	Severe intellectual disability, heart defects, brain or spinal cord abnormalities, small or poorly developed eyes, extra fingers or toes, cleft lip and palate, weak muscle tone
Trisomy 18	Edwards syndrome	Approximately 1:5000	Intrauterine growth retardation, low birth weight, heart defects and abnormalities of other organs, small, abnormally shaped head, small jaw and mouth, clenched fists, severe intellectual disability
Trisomy 21	Down syndrome	Approximately 1:800	Mild to moderate intellectual disability, characteristic facial appearance, weak muscle tone, heart defects, digestive abnormalities, hypothyroidism, increased risk of hearing and vision problems, leukaemia, Alzheimer's disease
Trisomy X	Triple X syndrome	Approximately 1:1000	Increased height, increased risk of learning disabilities, delayed development of speech, language and motor skills, weak muscle tone, behavioural and emotional difficulties, seizures, kidney abnormalities
47,XYY		Approximately 1:1000	Increased height, increased risk of learning disabilities, delayed development of speech, language, and motor skills, weak muscle tone, hand tremors, seizures, asthma, scoliosis, behavioural and emotional difficulties
47,XXY	Klinefelter syndrome	1:500 to 1:1000	Small testes, low testosterone levels, delayed and incomplete puberty, breast enlargement, reduced facial and body hair, infertility, increased height, increased risk of breast cancer, learning disabilities, delayed speech and language development
48,XXXY		Approximately 1:18000 to 1:40000	Small testes, low testosterone levels, delayed and incomplete puberty, breast enlargement, reduced facial and body hair, infertility, increased height, tremors, dental problems, peripheral vascular disease, deep vein thrombosis, asthma, type 2 diabetes, seizures, heart defects, delayed speech and language development, learning disabilities
45,X	Turner syndrome	Approximately 1:2500	Short stature, early loss of ovarian function, infertility, absence of puberty, webbing of the neck, skeletal abnormalities, kidney problems, heart defects

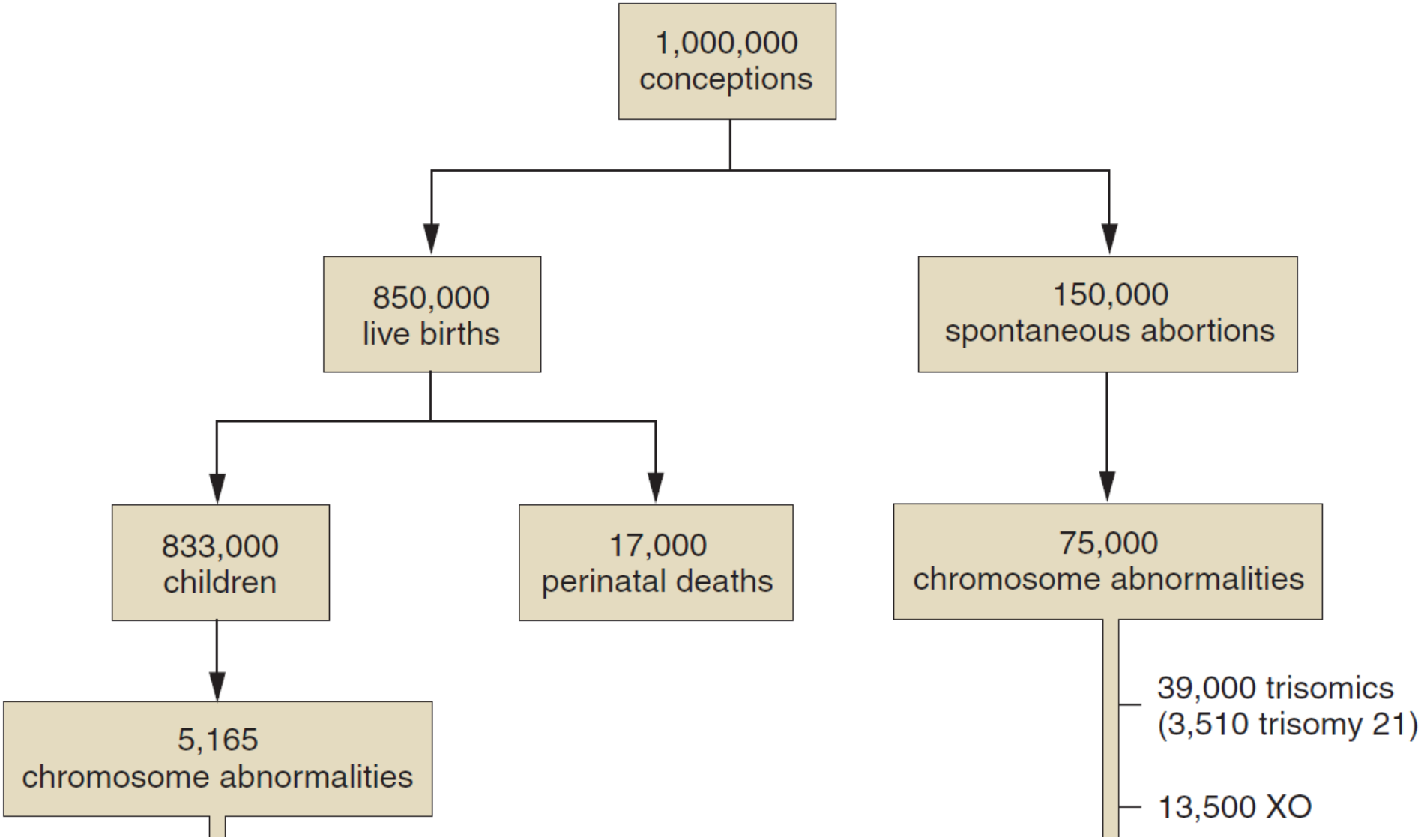
Jackson (2018) *Essays in Biochemistry*

X-inactivation

X-inactivation: in every cell in the female embryo, one of the two X chromosomes becomes inactivated and condensed.

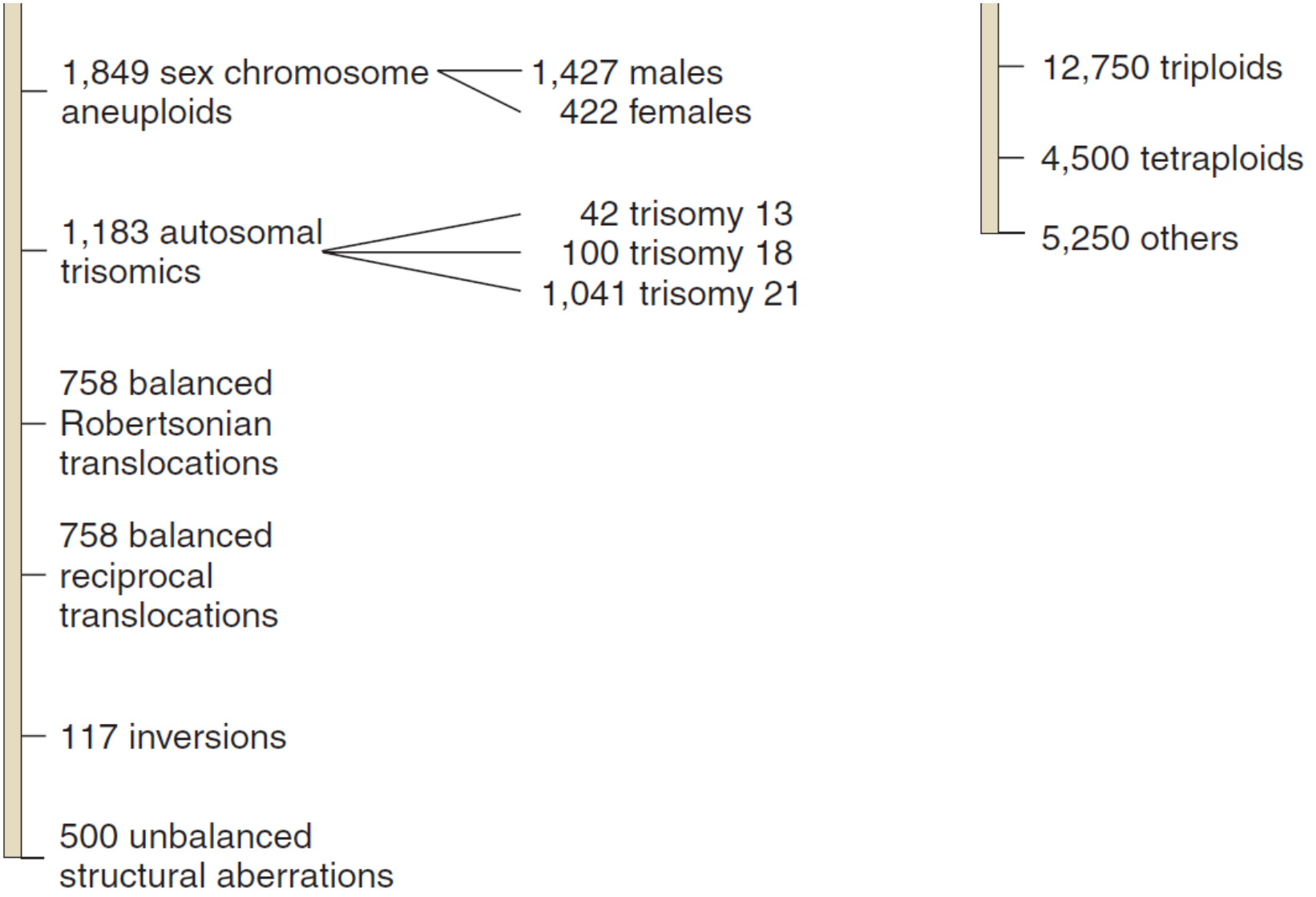
- Early in development
- Random in different cells
- Persists through subsequent cell divisions, but not generations
- Female tissues are a patchwork, with 50/50 inactivated paternal and maternal chromosomes
- X-inactivation provides **dosage compensation**: comparable levels of expresssion for ~1,500 X-chromosome genes in males and females

The fates of a 1 mln implanted human zygotes



Griffiths -- *Introduction to Genetic Analysis*

The fates of a 1 mln implanted human zygotes



Griffiths -- *Introduction to Genetic Analysis*

De novo variants rates and counts

DNM type	*Rate per generation*	*Total in an individuum*
Single nucleotide variants (SNVs)	**$1.20 \cdot 10^{-8}$** per bp **$1.66 \cdot 10^{-8}$** ‡	44..82
Dinucleotide repeats	$2.73 \cdot 10^{-4}$ per locus	N/A
Coding SNVs	N/A	1-2
Small indels (<50bp)	**$0.53\text{-}1.5 \cdot 10^{-9}$** per bp $1.26 \cdot 10^{-9}$ ‡	3-9
Large indels	0.16	0.16
Copy number variants (CNVs)	$10^{-6} – 10^{-4}$ per locus per generation	0.0154

‡ Ref: Palamara (2015) *AJHG*

De novo variants rates and counts

DNM type	*Rate per generation*	*Total in an individuum*
Single nucleotide variants (SNVs)	**$1.20 \cdot 10^{-8}$** per bp **$1.66 \cdot 10^{-8}$** ‡	44..82
Dinucleotide repeats	$2.73 \cdot 10^{-4}$ per locus	N/A
Coding SNVs	N/A	1-2
Small indels (<50bp)	**$0.53\text{-}1.5 \cdot 10^{-9}$** per bp $1.26 \cdot 10^{-9}$ ‡	3-9
Large indels	0.16	0.16
Copy number variants (CNVs)	$10^{-6} – 10^{-4}$ per locus per generation	0.0154

‡ Ref: Palamara (2015) *AJHG*

Summary

- Several mechanisms ensure the high rate of accuracy in DNA replication, including precise nucleotide selection, proofreading, and mismatch repair
- However, mutations are inevitable due to spontaneous replication errors and endogenous and exogenous DNA damage
- Human mutation rate is a trade-off between extinction and need for evolutionary change
- There is a wide spectrum of de novo mutations with varying rates and consequences: single nucleotide variants, structural variants and aneuploidies

Further reading

- Ohno, M. (2019). Spontaneous de novo germline mutations in humans and mice: rates, spectra, causes and consequences. *Genes Genet. Syst.* 94, 13–22.
- Acuna-Hidalgo, R., Veltman, J.A., and Hoischen, A. (2016). New insights into the generation and role of de novo mutations in health and disease. *Genome Biol* 17.
- Weckselblatt, B., and Rudd, M.K. (2015). Human structural variation: mechanisms of chromosome rearrangements. *Trends Genet* 31, 587–599.
- Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., et al. (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421.
- Zhang, F., Gu, W., Hurles, M.E., and Lupski, J.R. (2009). Copy Number Variation in Human Health, Disease, and Evolution. *Annu Rev Genomics Hum Genet* 10, 451–481.
- Jackson, M., Marks, L., May, G.H.W., and Wilson, J.B. (2018). The genetic basis of disease. *Essays Biochem* 62, 643–723
- Belyeu, J.R., Brand, H., et al. (2021). De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families. *The American Journal of Human Genetics* 108, 597–607.