GENETICS

Modified no. 12

Writer : Eman Amjad

correcter : Noor Abu Hantash

Doctor : Mohammad Alsboul

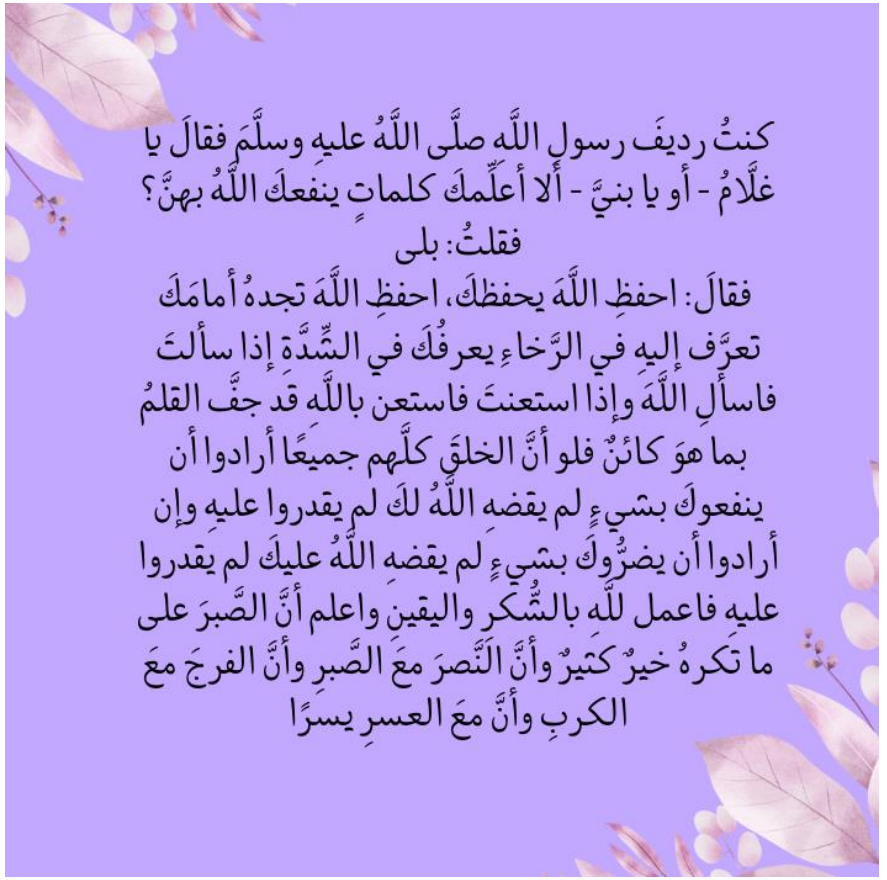
بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

Green color: doctors notes , Pink: extra from us

- Lecture 9- Genetic variation 1- outline:
- Describe different types of genetic variations
- Give examples of genetic variations

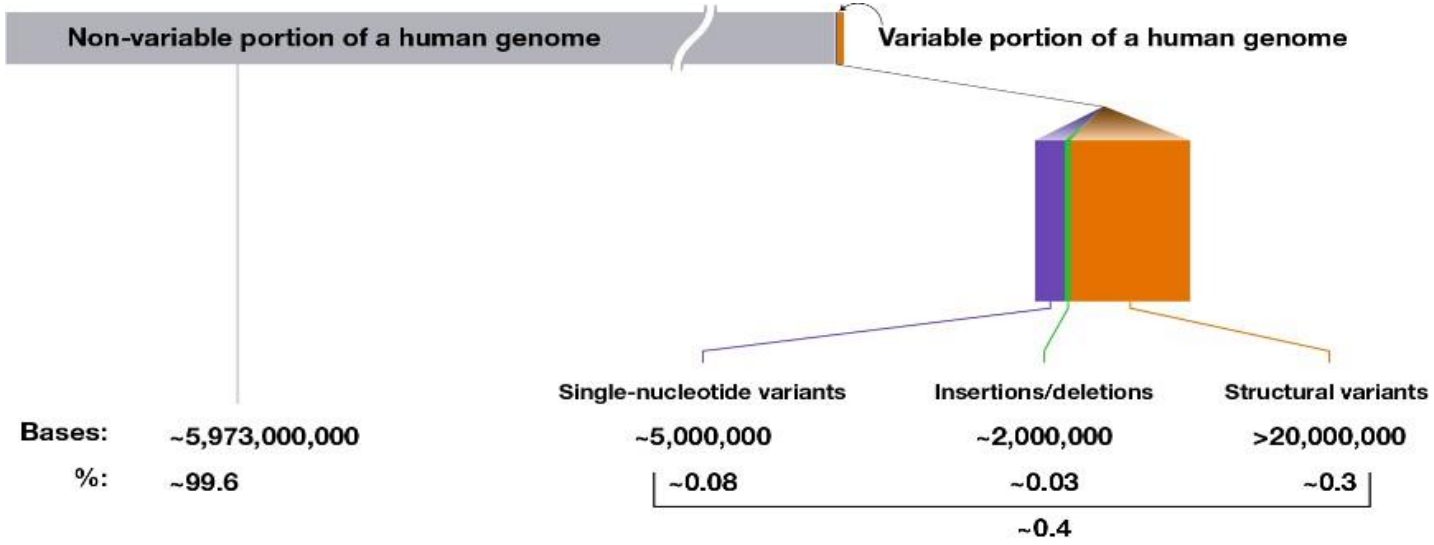
A mini-detailed-outline from us:

- Variants concepts & types
- Types of DNA damages & repair
- Mutations types (small scale vs large scale)
- Regulatory mutation
- Nomenclature of DNA variants



Similarity of DNA

- One copy of the human genome contains about 3 billion nucleotides (in 23 chromosomes)
- Two random people are 99.6% similar and the rest 0.4% as the pic indicates is due to: single-nucleotides variants & insertion or deletion & structural ones (mostly on chromosomes level). **Note**: majority of this 0.4% exists in non-coding regions introns besides the exons, but it doesn't mean that all mutations in non-coding regions are harmless, indeed some mutations there can cause diseases. (This note is mentioned in 2nd part video)



The Nature of Mutations

- A **mutation** is a change in a DNA sequence (is rare in a population and typically affects the phenotype)
  - Relatively rare, with a population frequency less than 1%. So between 100 people if we find a certain variance with 2%, we conclude that's not a mutation (it should be less than 1%)
  - They occur both in the nuclear and in mitochondrial genomes.
  - The term mutant refers to the phenotype
  - An agent that causes a mutation is called a mutagen, it can happen due to environmental factors chemical or physical = changes in genetic material that can lead to cancers
  - This variation leads to a disease or an infected patient
- A **polymorphism** ("many forms," describing multiple alleles at a locus): is a genetic change that (is frequent in a population>1%) any percentage above 1%
- - a polymorphism concept means: it doesn't lead to a disease

- Nowadays, these both terms are excluded. So they don’t call any variation as mutation or polymorphism
- Why? Because some polymorphisms are associated with common diseases as diabetes or Parkinson’s, and it’s a normal variance, but having it increases the risk of having the disease

GJB2:c.109G>A

chr13-20763612 C>T | p.Val37Ile | NM_004004.6 |

Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
East Asian	1665	19952	96	0.08345
Ashkenazi Jewish	83	10342	0	0.008026
Other	31	7212	0	0.004298
Latino/Admixed American	95	35428	1	0.002681
European (Finnish)	42	25104	0	0.001673
European (non-Finnish)	179	128578	1	0.001392
African/African-American	25	24964	1	0.001001
South Asian	12	30584	0	0.0003924
XX	1083	129104	53	0.008389
XY	1049	153060	46	0.006854
Total	**2132**	**282164**	**99**	**0.007556**

-One example of an autosomal recessive variation associated with deafness is caused by mutations in the GJB2 gene, which encodes a protein called connexin 26. Mutations in this gene are a common cause of non-syndromic autosomal recessive deafness where the frequency of guanine is higher than adenosine at a particular genomic position linked to the GJB2 gene

-From the tablel if we divided 99byn 2132, the percentage will be more than 1% but we can't say it is polymorphism (which means normal no disease variance) because it's highly pathogenic so we replace them by variance type 👍

- **We replace them by variance types Which are as following:**
- (disease causing classification):

- 1) pathogenic variance 2) diseases causing 3) likely pathogenic

- pathogenic = many individuals are infected & carrying this variance, & by functional test it affects proteins structure or activity, & by animal models it gives same phenotypes.

- **Or we depend on** another classification for (normal variation, or for some genes we don't know if they are associated with common/ complex diseases):

- 1) Benign. 2) Likely benign variance

- **Or depending on** (**variants of unknown significance**) classification: after "in silico" tool is used to test the strength of unknown variation. In case of an evidence for causing a protein defect, we call them "variant of unknown significant" instead of "likely benign". However, it can be reclassified by time for the another types (حسب نسبة ظهورها و نتيجتها)
- Common diseases/comlpex: not a monogenic variation, many genes can cause them.

- e.g. Autosim is associated with more than 300 gene, in aids of environmental factors…

- In labs they classify unknown variance by doing functional test or testing other family members, also as we mentioned before the animal model test.

Mutation, polymorphism and variant

- "A **mutation** is defined as a permanent change in the nucleotide sequence with a frequency below 1%
- A **polymorphism** is defined as a variant with a frequency above 1%
- The terms "mutation" and "polymorphism," however, which have been used widely, often lead to **confusion** because of incorrect assumptions of pathogenic and benign effects, respectively.
- Thus, it is recommended that both terms be replaced by the term "variant"" ACMG 2015 guidelines (American collage of medical genetics)

Types of Mutations

- Human DNA variants can be classified as:

1) large scale versus small scale, 2) common versus rare, and 3) pathogenic versus nonpathogenic.

- Human genetic variation ranges from single nucleotide changes through to gains or losses of whole chromosomes.
- Small-scale variants normally have their primary effect, if they have any effect, on a single gene on gene level, **whereas** large-scale variants usually affect several or many genes on chromosomal rearrangement or alterations level.

All mutations fall into two basic categories:

- Those that produce changes in a single gene are known as **gene mutations.**
- Those that produce changes in whole chromosomes are known as **chromosomal mutations.**

Gene mutations

- Gene mutations, including base pair substitutions, insertions, and deletions, can originate by either of two basic mechanisms:
  - Errors introduced during the normal process of DNA replication, 99% of errors will be corrected by proofreading by different polymerase (The Doctor said we will study polymerase later)
  - Mutations arising from a failure to repair DNA after damage and to return its sequence to what it was before the damage. DNA damage or by environmental agents or fail of repairing, all of this happen in cell division during interphase basically.

DNA DAMAGE AND REPAIR MECHANISMS

- **Variants arise from 1+2+3**
- Some DNA variants arise from **1)** errors **in** DNA replication or recombination in meiosis 1 which results in exchange genetic materials, if it's not correct will result in permanent mutations , but a major source is a failure to repair DNA damage.
- **2)** (induced) Chemical reactions and some physical processes constantly damage genomic DNA.
  - The majority are corrected using the undamaged **strand as a template.** In interphase-S phase (between S & G2) so I can have a sister chromatide to compare the variant with normal gene
  - Some base changes escape repair, and an incorrect base serves as a template in replication. By templates humans' proteins can distinguish if the base pair are correct or not.
  - The human genome contains genes for > 130 repair proteins: more than 99.9% of errors of DNA replication are corrected.
- The agents that damage DNA can either be external to the cell or they may arise as undesired effects of internal cell chemistry.
- **3)** spontaneous inside cells

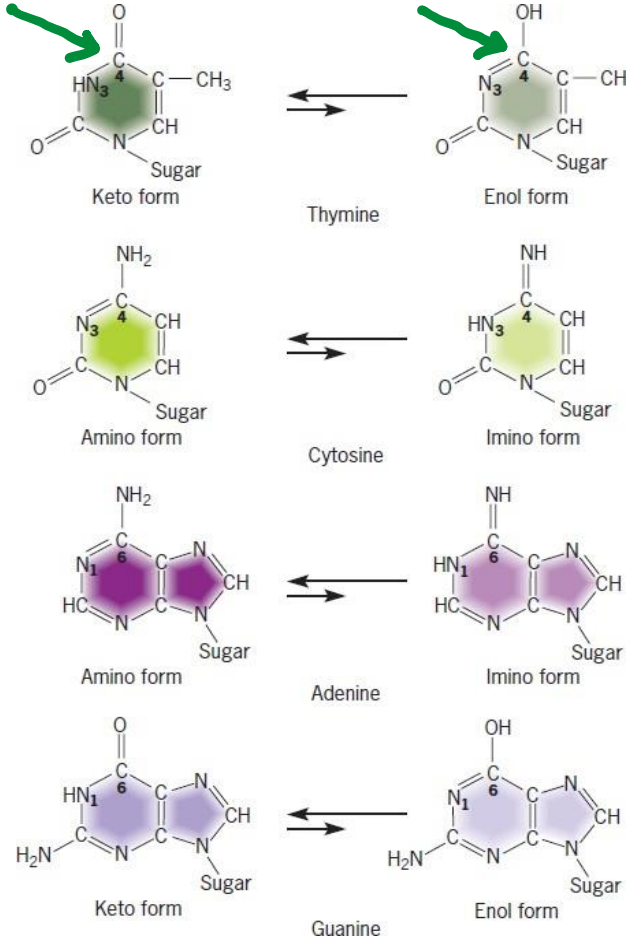
Causes of gene mutations

- **Spontaneous:**
  - A spontaneous mutation is one that occurs as a result of natural processes in cells, for example can arise from 1) DNA replication errors. These can be distinguished from induced mutations;
  - 2) slipped strand mispairing can occur at homopolymeric regions runs (mono, di, or trinucleotide repeats) (in next slides it will be explained)
    - Homopolymeric regions: sequence of consecutive identical bases. As "ACACACA" repetitive can cause a misalignment for the starting point of replication, which results in looping out strands = inducing deletion in genetic material.
- Induced: Exposure to mutagens
  - Ionizing radiation—gamma rays and X-rays can cause single- strand or double-strand breaks in the sugar–phosphate backbone
  - Ultraviolet radiation—(sunlight that can penetrate the ozone layer).
  - Environmental chemicals—these include hydrocarbons (for example, in cigarette smoke/ barbecue). Attack by reactive oxygen species—highly reactive superoxide anions (O_2^-) and related molecules are generated as a by-product of oxidative metabolism in mitochondria.

- Spontaneous
  - ATCG bases can exist normally in cells in two forms depending on the position of atoms & the bonds between atoms:
  - ✓ 1) common form (keto-amino form)
  - ✓ 2) Rare form (enol, imino)
  - Exist in alternating forms called **tautomers**
    - Tautomer pairs = ketone-enol isomer
    - Chemical modification of bases followed by mispairing
    - When the bases are present in their rare imino or enol states, they can form adenine–cytosine and guanine–thymine base pairs
    - Usually corrected by proofreading

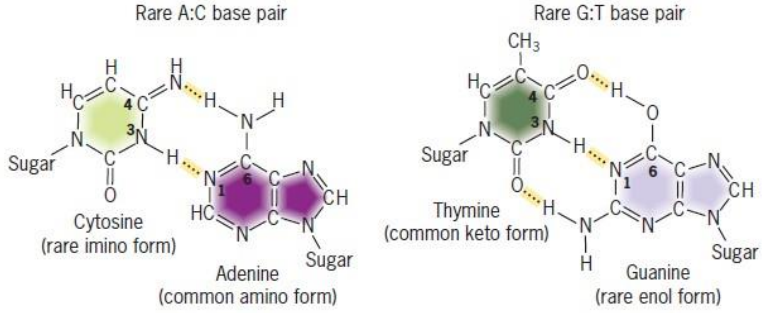
This note is for induced variants

- Those that occur as a result of interaction of DNA with an outside agent or mutagen that causes DNA damage.
- Moreover, some sites on chromosomes are “hotspots” where mutations arise at a higher frequency than other regions of the DNA.
- Notice that in enol forms, the base pair will be T with G,
- In imino: A with C



انتبه للفرق

Hydrogen-bonded A:C and G:T base pairs that form when cytosine and guanine are in their rare imino and enol tautomeric forms.



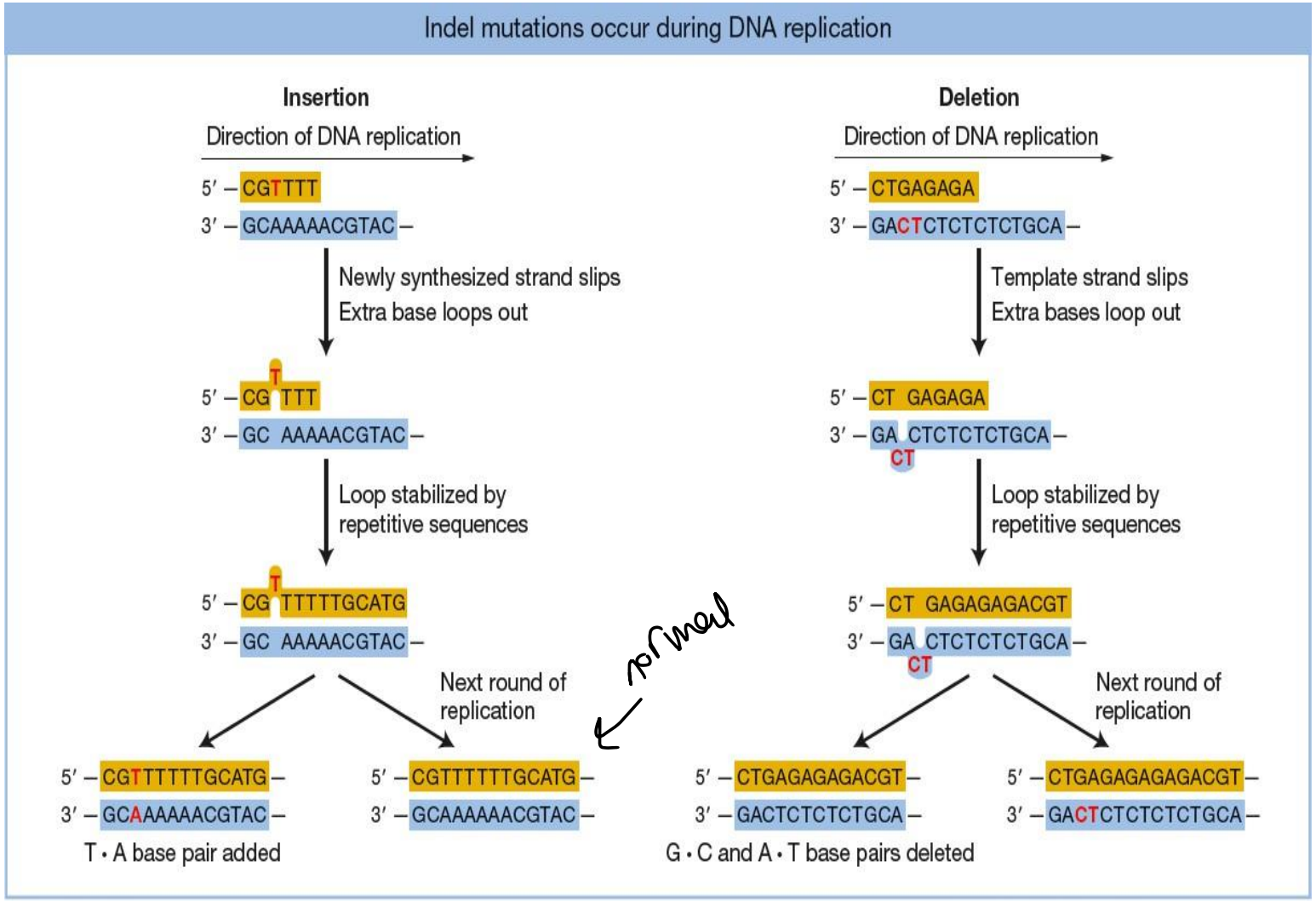


FIGURE 15-7 In the course of DNA replication, base insertions and deletions (indel mutations) are formed through the slipped mispairing of repeated sequences.

ANIMATED ART SaplingPlus
Molecular mechanism of mutation

https://youtu.be/E18Jl-LP4TU
a short nice video

Slipping strands :
Notice the pics: in original-blue strand there's a repetitive sequence (homopolymeric regions AAAAA)

when polymerases start functioning they get confused due to the repetition, and bind to the site of replication (by skipping adenosine), This results in looping out in the growing complementary strand = having an insertion or may lead to deletion which occurs due to looping out in the original template strand.

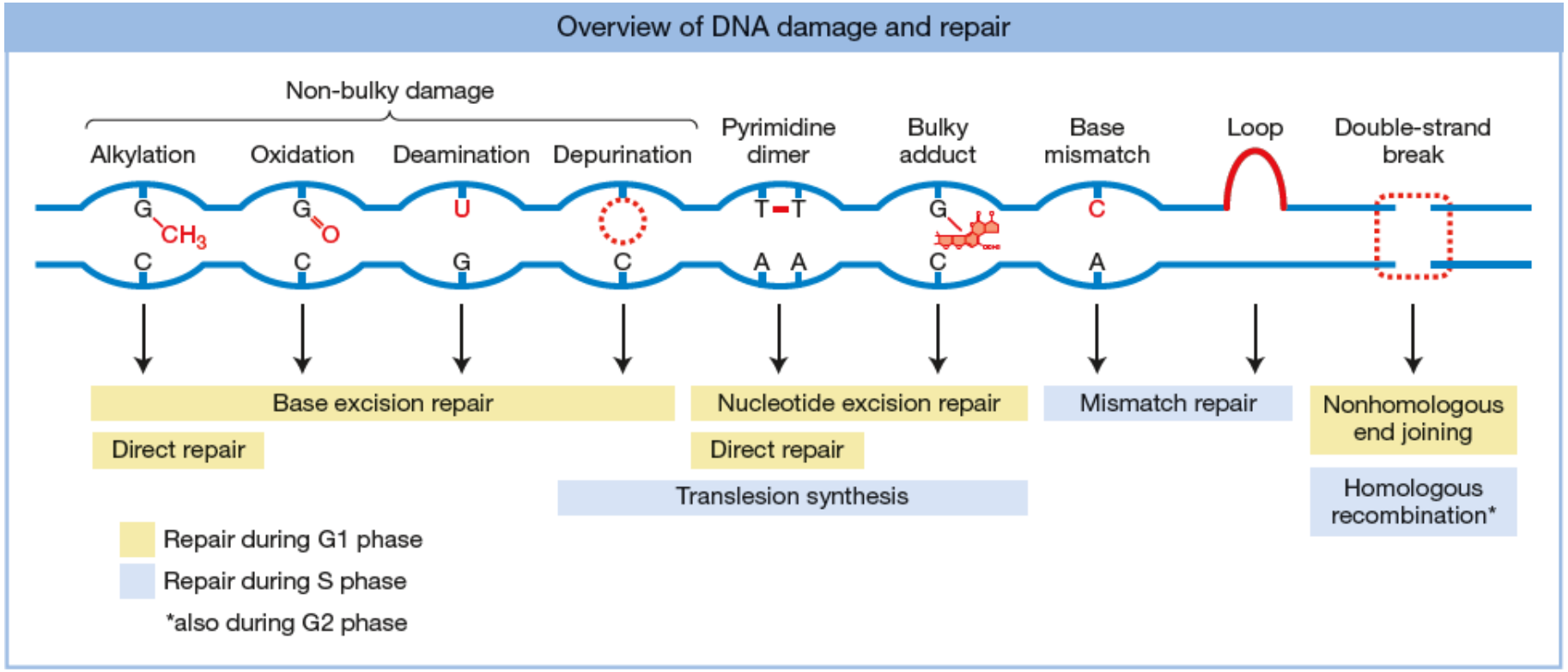
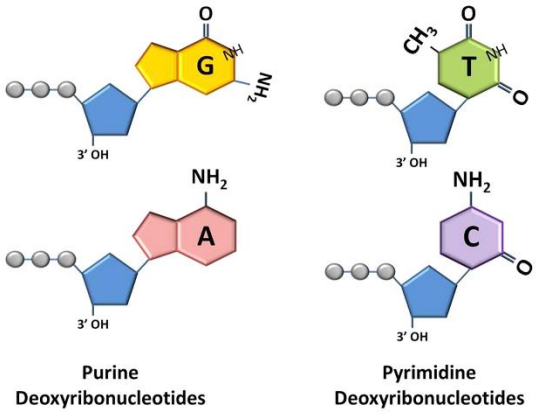


FIGURE 15-13 DNA repair mechanisms are paired with the types of DNA damage they act on. DNA damage is indicated in red. DNA repair mechanisms highlighted in tan function during G1 phase of the cell cycle, and those highlighted in blue function during S phase of the cell cycle. Homologous recombination also functions in G2 phase of the cell cycle. **The figure is important**

Notes are in the next slide ✅

Extra: remember the groups
Pure As Gold



- Even if the damage is recognized and excised, the repair machinery may not read the complementary strand accurately and, as a consequence, will create mutations by introducing incorrect bases.
- Thus, in contrast to replication-related DNA changes, which are **usually corrected through proofreading mechanisms,** nucleotide changes introduced by DNA damage and repair often result in permanent mutations.

DNA damage types in brief:

1) Spontaneous occurs inside cells:
   a) Alkylation : CH_3 or CH_2 binding to nitrogen bases resulting in defects
   b) oxidation : binding of Oxygen
   c) deamination : removing amine group ($R\text{-}NH_2$) from cytosine results in recognizing it as uracil (RNA), but usually it's corrected. Another e.g. adenosine deamination results in guanine like structure "hypoxanthine" binding to cytosine = transition mutation. Also Guanine deaminates to "xanthine" molecule,
   d) depurination: in which the N-glycosidic bonds are cleaved to release the corresponding purine or pyrimidine from DNA.
2) Pyrimidine dimer: due to UV-lights
3) Bulky adducts: DNA is exposed to certain chemicals as aflatoxins binding to guanine = inhibition of replication
4) Base mismatch: due to looping
5) Double DNA breakage: due to ionized radiation
6) looping

💡Note:

Transition mutation: is changing in genetics between same groups of nitrogenous bases AG/CT

Transverted mutation: purine to pyrimidine and vice versa

DNA repair types:

1) Base excision: detects wrong base-pairs
2) Nucleotide excision: more than one base-pairing correction.
3) Direct repair for thymine dimmers: certain enzymes absorb lights & break the linkages in between.
4) Mismatch
5) Homologous recombination vs nonhomologous end joining: both for breakage in DNA. The homologous occur in S-G2 phase (templates= repairing well & less errors), while the nonhomologous happens in G1 phase (no templates & before duplication).
   Nonhomologous = not well aligned ends breaking = overhanging of strands, so the repair system cuts off these overhung ends & blunts the rest, then ligase enzyme gets both strands together. Errors happen because of losing DNA material.

Nice table for summarizing mutations

We'd already taken them & the rest will be discussed

TABLE 12.1

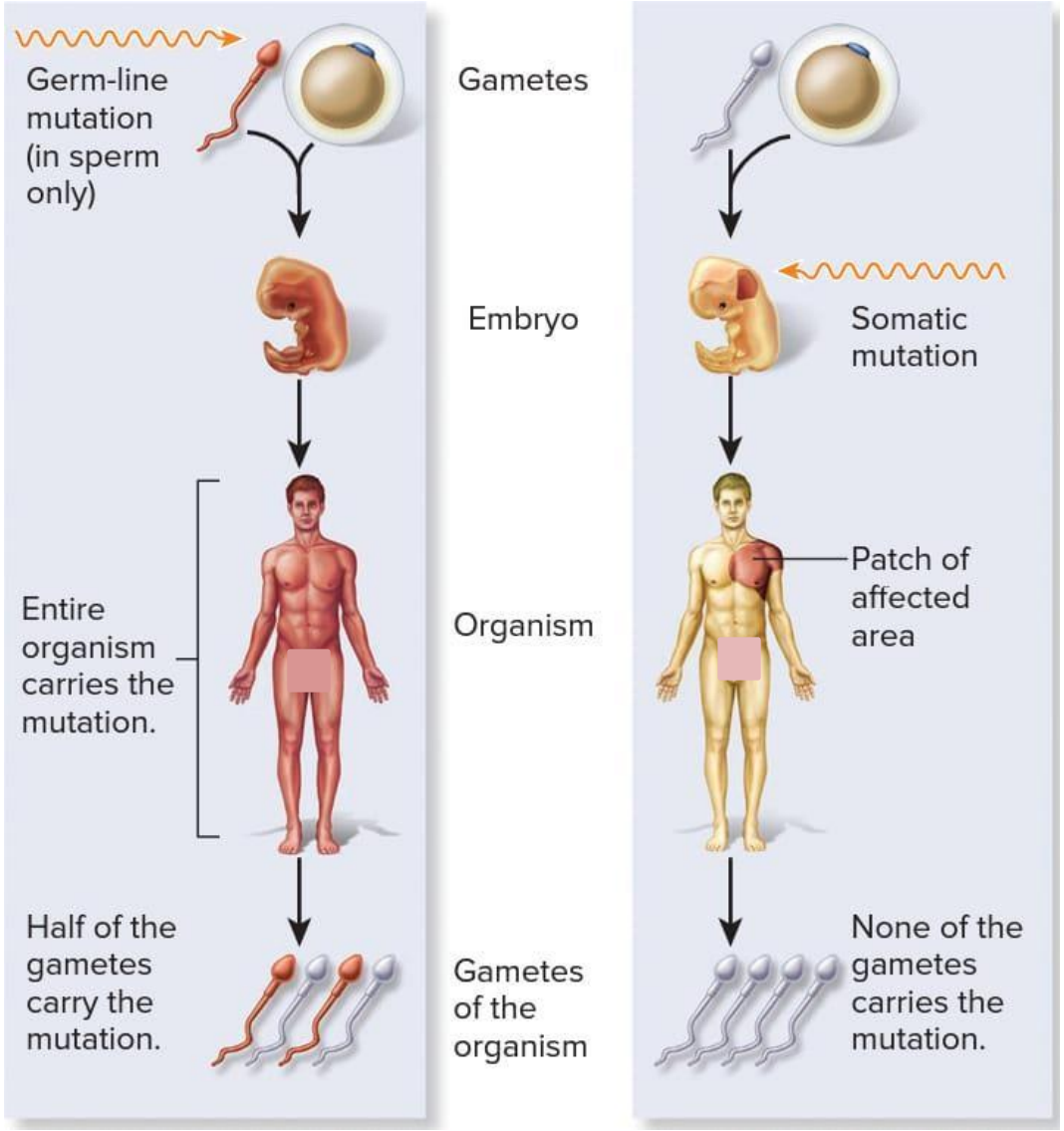
Major Types of Mutations and Their Distinguishing Features

Basis of classification	Major types of mutations	Major features
Origin	Spontaneous	Occurs in absence of known mutagen
	Induced	Occurs in presence of known mutagen
Cell type	Somatic	Occurs in nonreproductive cells
	Germ-line	Occurs in reproductive cells
Expression	Conditional	Expressed only under restrictive conditions (such as high temperature)
	Unconditional	Expressed under permissive conditions as well as restrictive conditions
Effect on function	Loss-of-function (knockout, null)	Eliminates normal function
	Hypomorphic (leaky)	Reduces normal function
	Hypermorphic	Increases normal function
	Gain-of-function (ectopic expression)	Expressed at incorrect time or in inappropriate cell types
Molecular change	Base substitution	One base pair in duplex DNA replaced with a different base pair
	Transition	Pyrimidine (T or C) to pyrimidine, or purine (A or G) to purine
	Transversion	Pyrimidine (T or C) to purine, or purine (A or G) to pyrimidine
	Insertion	One or more extra nucleotides present
	Deletion	One or more missing nucleotides
Effect on translation	Synonymous (silent)	No change in amino acid encoded
	Missense (nonsynonymous)	Change in amino acid encoded
	Nonsense (termination)	Creates translational termination codon (UAA, UAG, or UGA)
	Frameshift	Shifts triplet reading of codons out of correct phase

Table 12.01: Major types of mutations and their distinguishing features.

Two categories of mutations

- A mutation is change in a DNA sequence is rare in a population and typically affects the phenotype
- Somatic Mutations
  - Occur in cells of the body that do not form gametes
  - Occurs in mitosis: can lead to mosaicism (some cells are normal & others are not)
  - Is not transmitted to future generations
  - Most cause of Tumor : uncontrolled growth
- Germ-line Mutations
  - Occur in cells that produce gametes (ova & sperms)
  - Occurs during meiosis
  - Monogeneic mutation
  - Transmitted to future generations - inherited



Types of mutations and their phenotypic consequences

Mutations are of fundamental importance in molecular biology for several reasons:

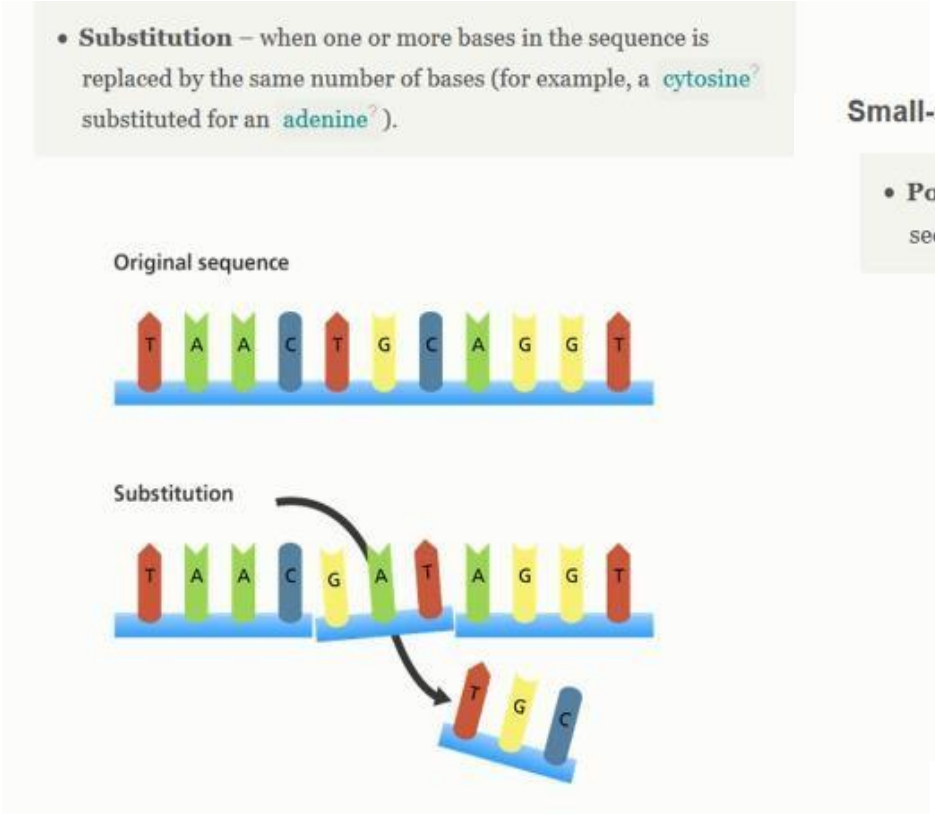
1) As noted above, mutations are important as the major source of genetic variation. Variations are between individuals or in the same person, each person can have (30-100) variants (70 on average) as unique variants mostly associated with severe diseases, but the person himself is a carrier.

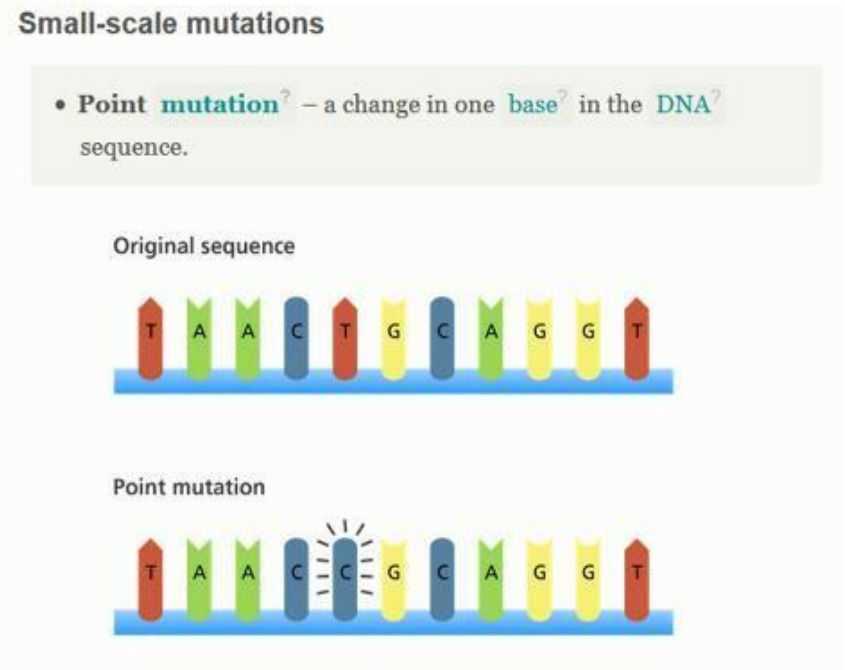
2) Mutations may have deleterious or (rarely) advantageous consequences to an organism or its descendants. Mutations in germ cells can lead to heritable genetic disorders, while mutations in somatic cells may lead to acquired diseases such as cancer or neurodegenerative disorders. Some unique deleterious mutations as autosomal dominant can be de novo and having it increases between non-relative parents (sheet 10). Most of these denove mutations are associated with neurological defects

3) Mutant organisms are important tools for molecular biologists in characterizing the genes involved in cellular processes. To prove the function of a variance on animal models as: frogs & zebra fish (we can see the phenotype within weeks), but mice (need months)

Small- scale mutations/ variations

- The simplest type of mutation is a nucleotide (base-pair) substitution.
- Mutations that alter a single nucleotide are called point mutations.





Mutations
Small scale

Point mutations by transition or transversion:
Missense مخطئة التعبير
Silent صامتة
Nonsense غير معبّرة

Frameshift:
Insertion
Deletion

Note:
when we say a small variant it means, in general, on gene level
Large scale variant: a change on chromosome level, or on many genes
Frameshift is small scale variant

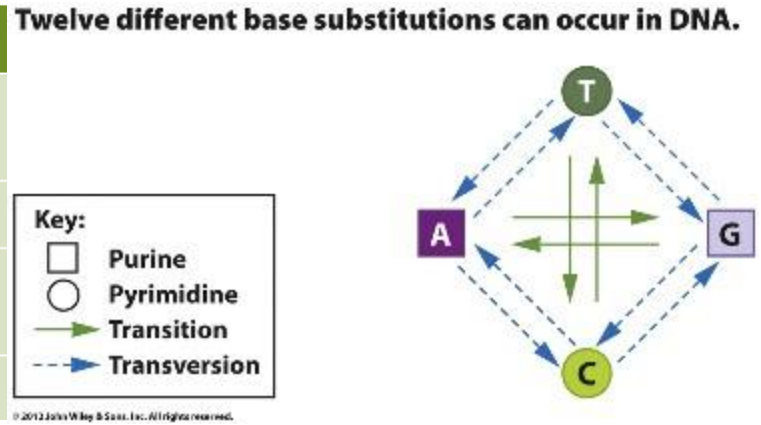
Types of mutations

At the Nucleotide Level (Point mutations):

- A base-pair substitution replaces one nucleotide and its partner with another pair of nucleotides is called base-pair substitution mutations.
- Two forms:
  - Transition: replaces a pyrimidine with another pyrimidine or a purine for another purine, 4 possibilities
  - Transversion: replaces a pyrimidine with a purine or or vice verse, 8 possibilities

Table 7.1 Types of nucleotide substitutions.

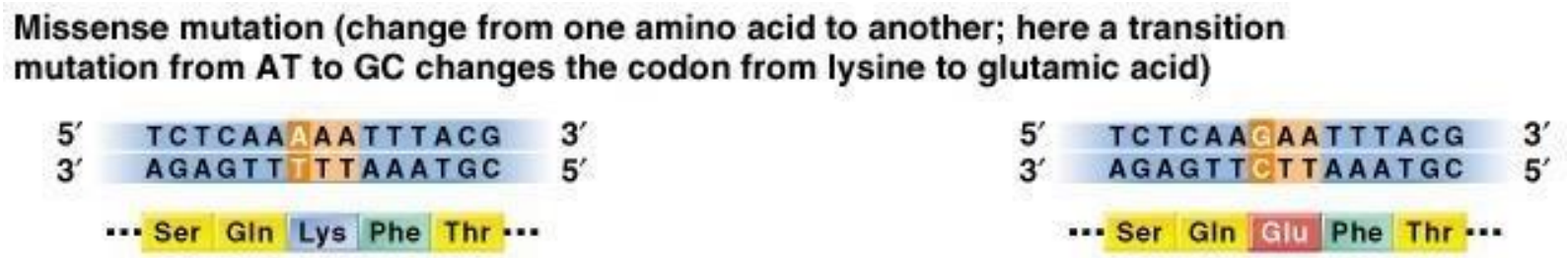
Nucleotide substitution	Mutation
Transition mutation Pyrimidine → pyrimidine	T → C or C → T
Purine → purine	A → G or G → A
Transversion mutation Pyrimidine → purine	T → A, T → G, C → A, or C → G
Purine → pyrimidine	A → T, A → C, G → T, or G → C



Theoretically: من الأسهم في الصورة
Tranversion : transition
2:1
In reality:
1:2
Because tranversion is corrected more than transition, more recognized by the repair system

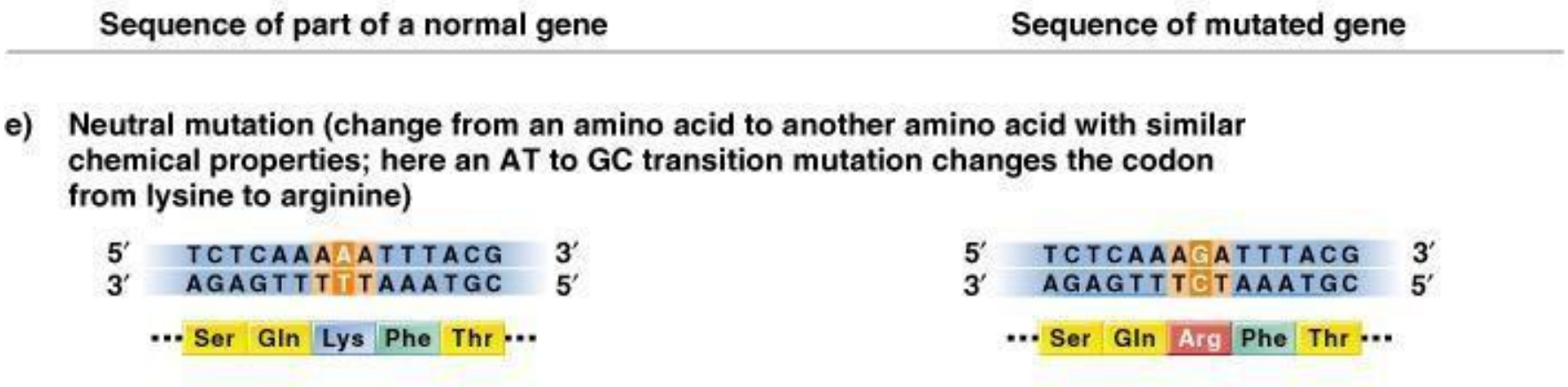
Missense mutations

- Nucleotide substitutions in protein-coding regions that do result in changed amino acids are called nonsynonymous mutations or missense mutations.
- May alter the biological properties of the protein.
- A classic example of a phenotypic effect of a single amino acid change is **phenylketonuria**, a disorder that can be caused by a base substitution in the **phenylalanine hydrolase gene**
- The mutation that leads to PKU is a **transversion** from a **G to a C at codon 413**
- This mutation is also a missense mutation in that it will result in an amino acid substitution in the corresponding protein from Pro413 to Arg 413
- This mutation will result in a PAH enzyme that is no longer functional, and will not be able to metabolize phenylalanine
- This leads to a variety of developmental defects including severe mental retardation
- As we know after transcription, evert 3 base pairs are called as Codon at mRNA level, and it's translated to an amino acid.

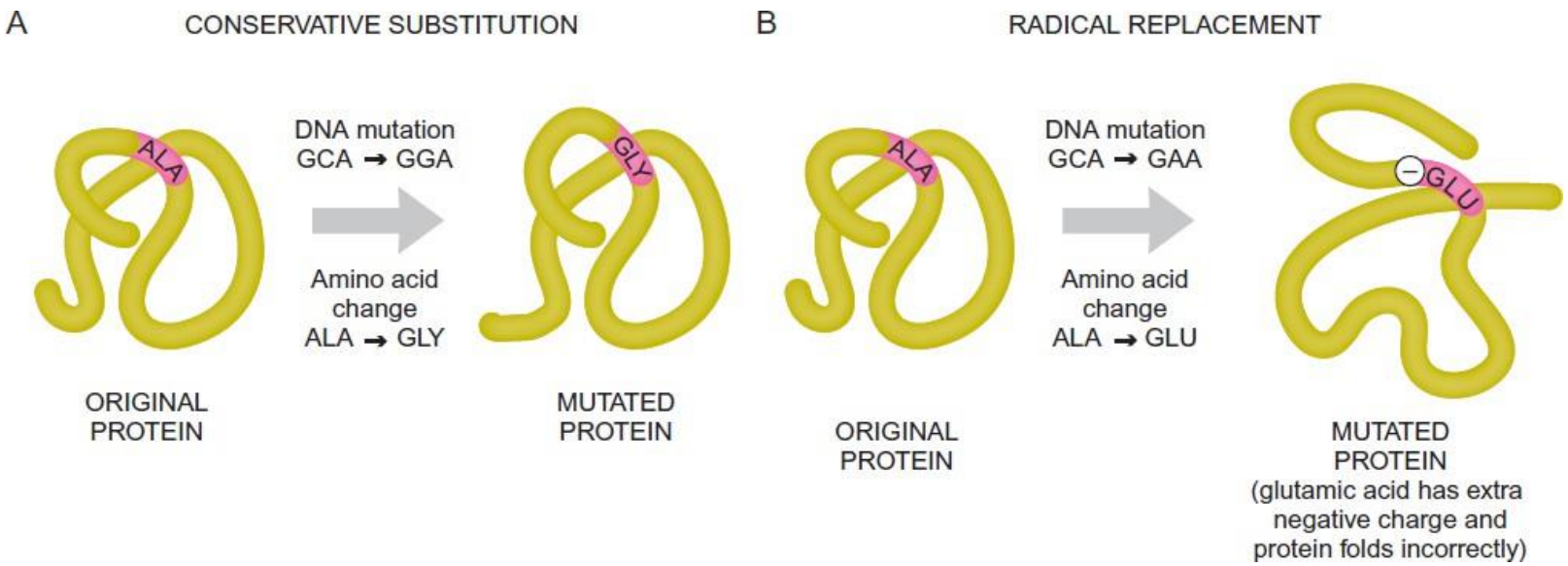


Missense mutations

- Neutral nonsynonymous mutation (same conservative) : Base pair substitution results in substitution of an amino acid with similar chemical properties (protein function is not altered). In other words: replacement of a polar a.a with another polar one from same group= mild effect.



- Conservative – chemical properties of mutant amino acid are similar to the original amino acid
  - e.g. aspartic acid [(-)charged] → glutamic acid [(-)charged]
- Nonconservative (missense) – chemical properties of mutant amino acid are different from original amino acid
  - e.g. aspartic acid [(-)charged] → alanine (uncharged)

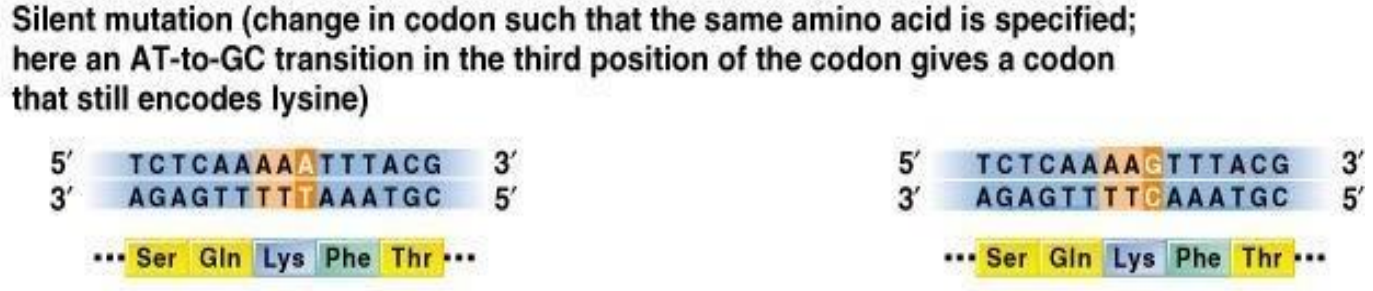


Conservative:
Changing amino acid by another from same a.a. group= no protein structural defect

nonconservative:
Different a.a & different chemical features = a phenotype will appear

Silent mutations

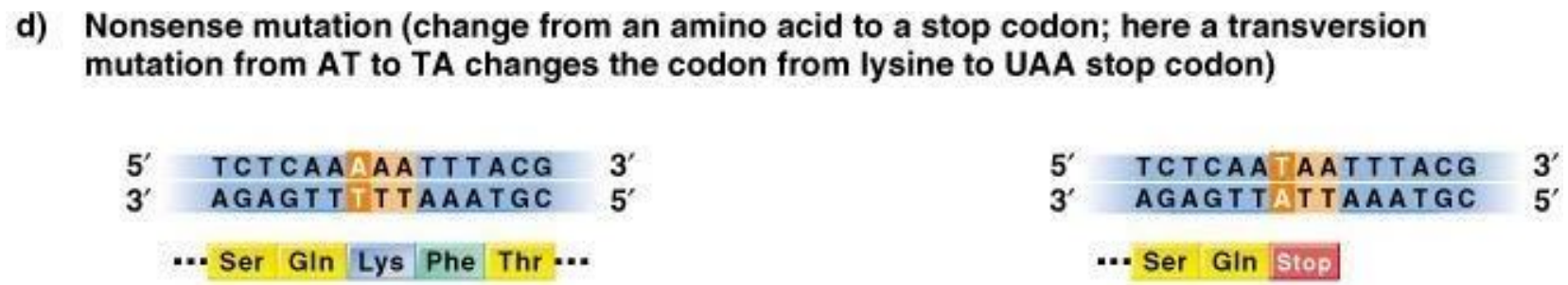
- Nucleotide substitutions in a protein-coding gene may or **may not change** the amino acid in the encoded protein. Usually happens in 3rd base-pair of the codon
- Mutations that change the nucleotide sequence without changing the amino acid sequence are called synonymous mutations or silent mutations.
- Mutational changes in nucleotides that are outside of coding regions can also be silent. However, some noncoding sequences do have essential functions in gene regulation and, in this case, mutations in these sequences **would have phenotypic effects**.
- We have 4 bps, and the codon needs only 3 bps, so the possibles codons in number are 64 (4*4*4). If we excluded Stop codons: UAG / UGA / UAA, we will have 61 codons to make 20 amino acid. Therefore, the single amino acid can be recognized by 3 codons.
- Example: AUU to AUC still codes for Isoleucine (Ile)

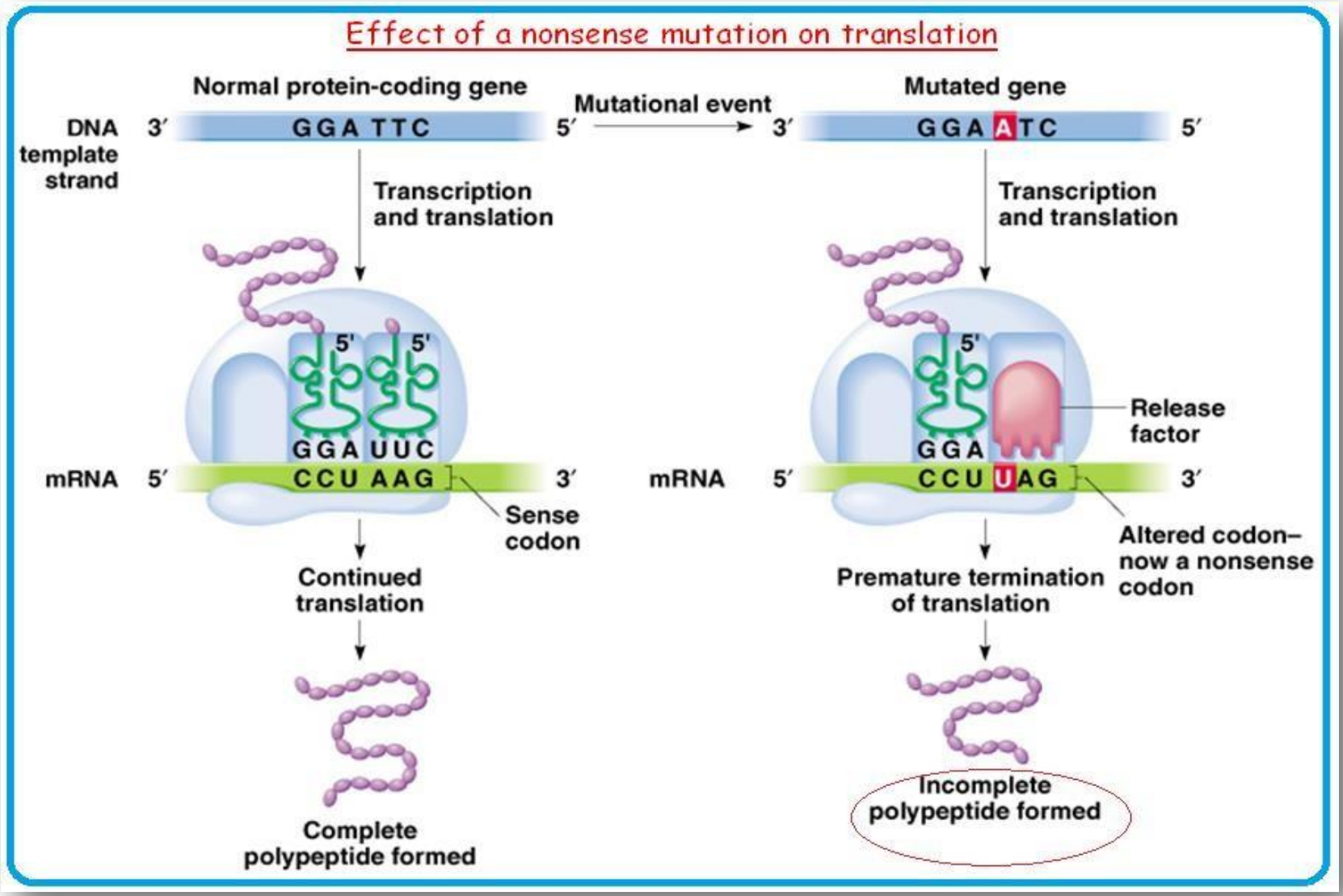


Q: Is silent mutation really silent(harmless)?
A: in exons-introns boundaries, cells try to remove introns by large ribonucleoprotein (RNP) complex "RNA splicesome". This complex functions by recognizing certain sequences at the beginning/end of introns or exons. So any mutation at these positions **can cause a defect in splicing /disease even if it's silent mutation.**

Nonsense mutations

- Nonsense mutations change codon that encodes an amino acid to a **stop codon on RNA level (UGA, UAG, or UAA)**
- Nonsense mutations (genes containing **premature** termination codons) often cause production of the **truncated protein: abnormal function protein** that might be predicted.
- Cells have a mechanism, **nonsense-mediated decay (NMD)**, that detects **mRNAs** containing premature termination codons and **degrades them**. Thus, the usual result of a nonsense mutation is to prevent any expression of the gene
- Nearly always a nonfunctional product.
- Severe point mutation.
- Results: truncated proteins or no proteins production



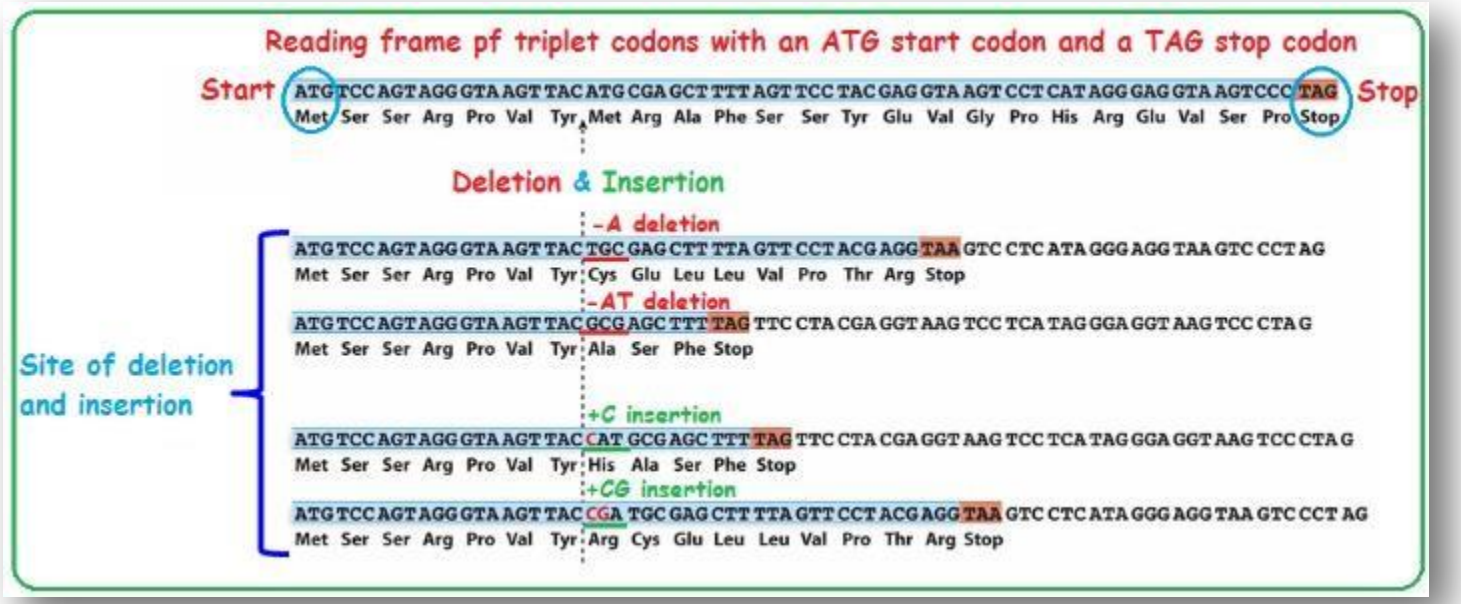


1) Mutation
2) Results in stop codon
3) The rest of strand won't be recognized by tRNA
4) **Releasing factors** will replace empty spaces
5) halting of translation
6) Premature termination
7) Incomplete protein
8) Abnormal protein function or degradation

Insertions or deletions can cause frameshift mutations

- Insertions or deletions of nucleotides can also occur in DNA, but at a rate considerably lower than that of nucleotide substitution.
- **Out-frameshif**: In single bp or anything else **but** not a 3 bps or multiple of 3.
- Insertion and deletion mutations are collectively referred to as indels.
- The DNA sequence from the start codon to the stop codon is referred to as a **reading frame.**
- Because nucleotides are decoded in triplets, an indel mutation of only one or two base pairs in the coding sequence of a protein throws off the reading frame after the mutation, resulting in a frameshift mutation.

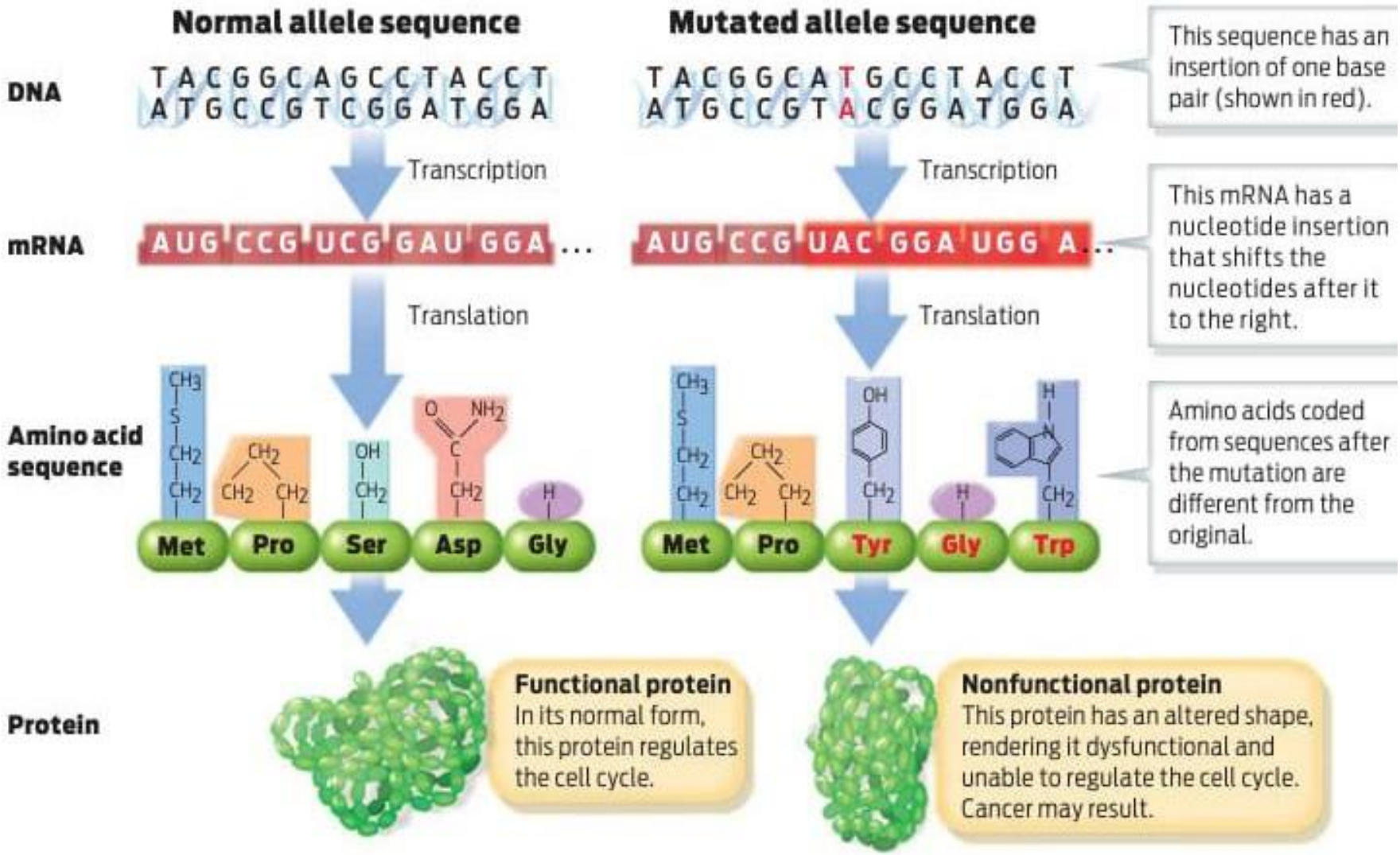
❖ Insertion or deletion of one bps or two, results in burrowing a bp from the next codon which causes shifting. Therefore translation is processed till a stop codon is achieved, so shifting results in indirectly a stop codon & abnormal protein. Also, in-frameshift will not lead to premature termination and often doesn't affect phenotypes but not an absolute rule.



Continuing to the previous slide:
Q: inserting of 3 bps or 6, what will cause?
A: the change will occur only on the added or deleted codon. 3 bps = 1 codon
This is in-frameshift mutation type: usually doesn't develop a phenotype, but as silent mutation not 100% harmless.

Just an e.g. of variants results from inframeshift

- A classic example of a phenotypic effect of a **small deletion is** **the change responsible for the human hereditary disease cystic fibrosis**.
- The deletion of three base pairs in the nucleotide sequence of **the cystic fibrosis transmembrane conductance regulator (CFTR)** gene results in **the loss of the codon for phenylalanine .** (Delta-F508 del)
- If the length of an **insertion or deletion is not an exact multiple of three nucleotides, the mutation shifts the phase in which the ribosome reads the triplet codons and, consequently, alters all of the amino acids downstream from the site of the mutation.**
- Such **mutations are called frameshift mutations** because they "shift" the reading frame of the codons in the mRNA.
- Usually leads to production of a nonfunctional protein.


Normal allele sequence
Mutated allele sequence
DNA
T A C G G C A G C C T A C C T
A T G C C G T C G G A T G G A
T A C G G C A T G C C T A C C T
A T G C C G T A C G G A T G G A
This sequence has an insertion of one base pair (shown in red).
Transcription
Transcription
mRNA
AUG CCG UCG GAU GGA ...
AUG CCG UAC GGA UGG A...
This mRNA has a nucleotide insertion that shifts the nucleotides after it to the right.
Translation
Translation
Amino acid sequence
Met
Pro
Ser
Asp
Gly
Met
Pro
Tyr
Gly
Trp
Amino acids coded from sequences after the mutation are different from the original.
Protein
Functional protein
In its normal form, this protein regulates the cell cycle.
Nonfunctional protein
This protein has an altered shape, rendering it dysfunctional and unable to regulate the cell cycle. Cancer may result.

Regulatory Mutations

- Occur in noncoding regions of genes: 5’ UTR (untranslated regions), 3’ UTRs, splice end regions, at end/ beginning of introns regions (i.e. splicing acceptor / splice donor)
- Results in:

  1) introns retention (part of introns will be included In proteins)
  2) exons skipping (deleted from the protein)
  3) cryptic splice sites (usually due to silent mutations, so causing a confusion in recognizing boundaries)

Three types are commonly recognized: promoter mutations, Polyadenylation, splicing mutations, and cryptic splice sites.

Splice Site Mutations

Two general classes of splicing mutation have been described:

- A mutation can also alter splice junctions in eukaryotes (splicing mutations)
  - Leads to splicing abnormalities by changing the donor and acceptor site of splicing sites-> interfere with normal RNA splicing at that site.
- A second class of splicing mutations involves intron base substitutions (cryptic splice sites), such mutations create alternative donor or acceptor sites that compete with the normal sites during RNA processing.
  - Cryptic donor or acceptor are new sites of splicing donors or acceptors.

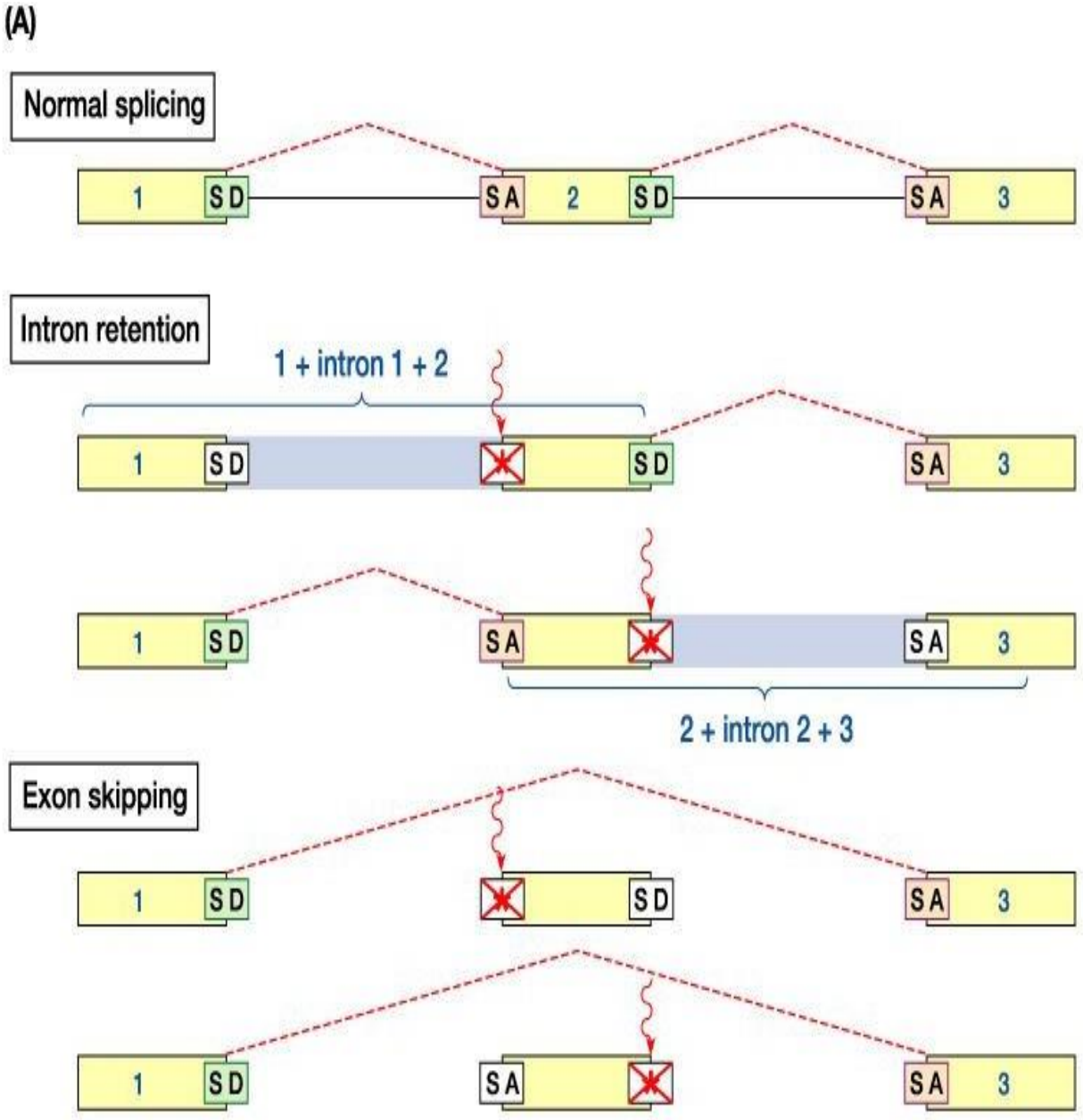


Figure 11-11 part 1 of 2 Human Molecular Genetics, 3/e. (© Garland Science 2004)

Donor Site Mutations: These mutations affect the 5' splice site at the beginning of the intron. The consensus sequence at this site is typically "GU" (G for guanine and U for uracil). Mutations that disrupt this sequence or its surrounding nucleotides can interfere with the recognition of the splice site by the spliceosome, leading to aberrant splicing.

Acceptor Site Mutations: These mutations affect the 3' splice site at the end of the intron. The consensus sequence at this site is typically "AG" (A for adenine and G for guanine). Mutations that disrupt this sequence or its surrounding nucleotides can similarly interfere with proper splicing.

32

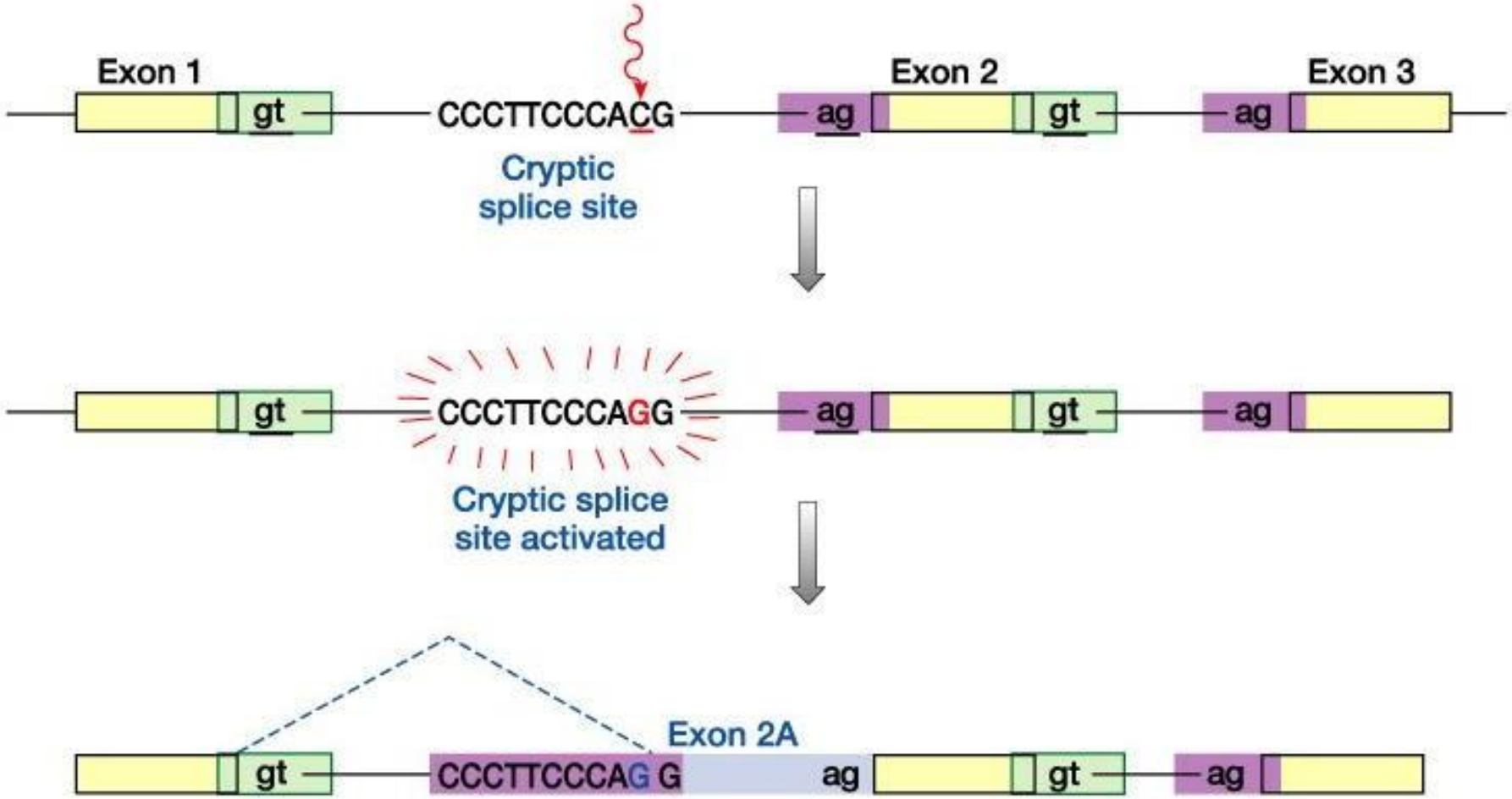


Figure 11-13 Human Molecular Genetics, 3/e. (© Garland Science 2004)

33

Not required pic as doctor mentioned, but you can read it nothing is new

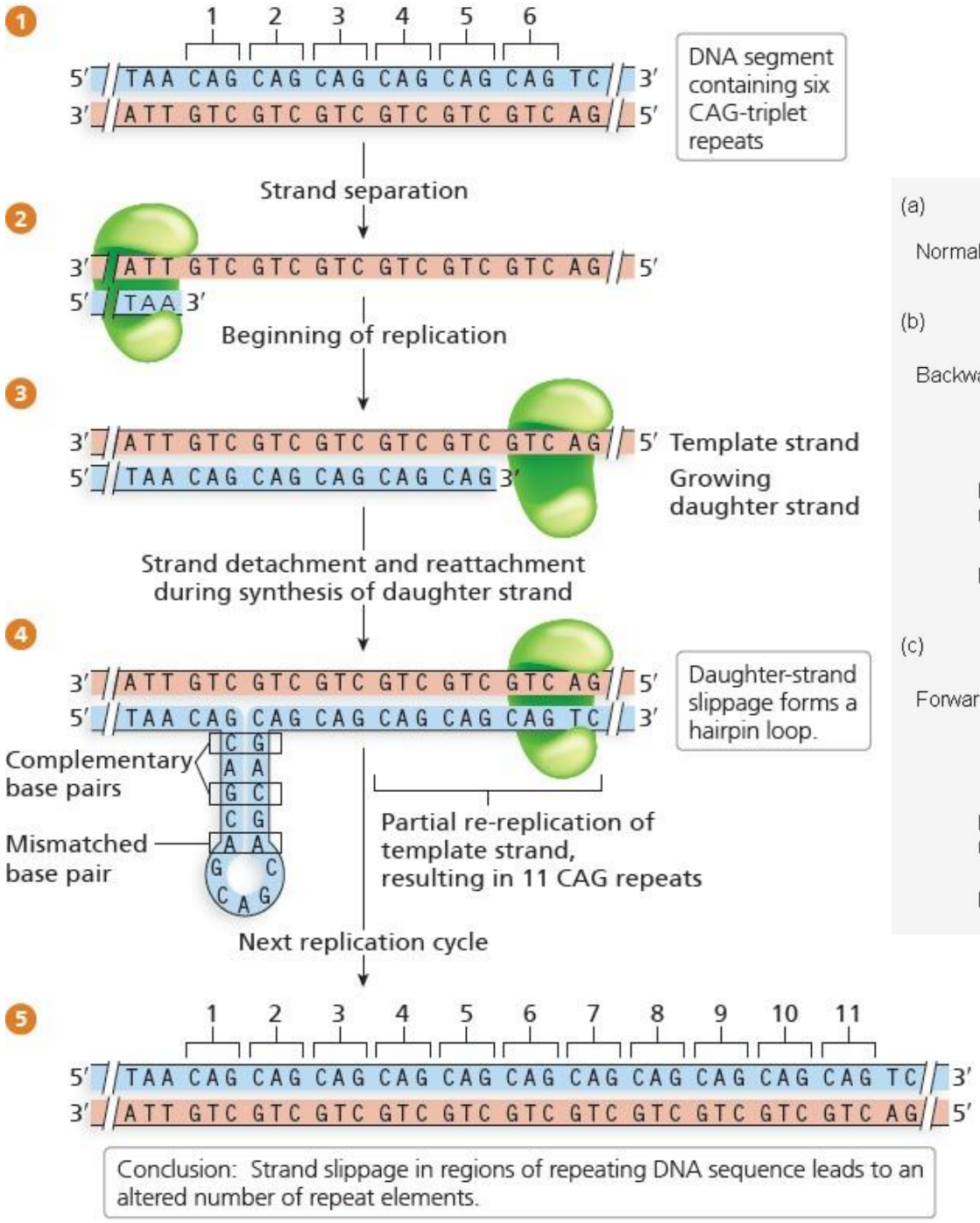
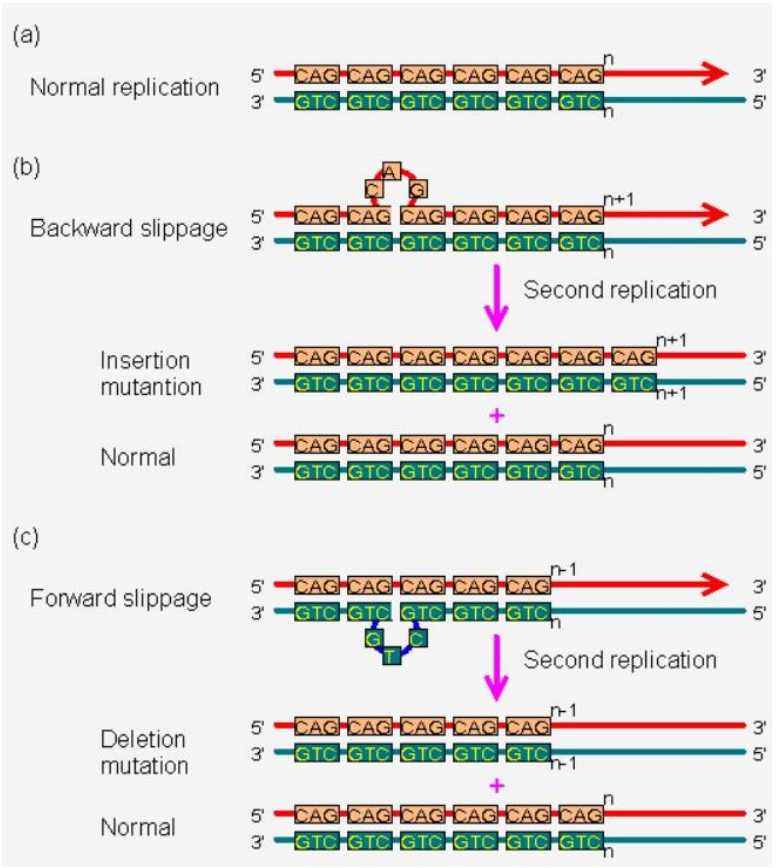


Figure 11.7 Strand slippage during DNA replication.



These slides about numbering variances

Gene and Variant nomenclature

Genes: https://www.genenames.org/
Variant: https://varnomen.hgvs.org/

Extra sites from the doctor

- The HGVS nomenclature guidelines are used worldwide for genetic variant interpretation

- The Human Genome Variation Society (HGVS) nomenclature standard was developed to prevent the misinterpretation of variants in DNA, RNA, and protein sequences. The HGVS nomenclature standard is used worldwide, especially in clinical diagnostics, and is authorized by the Human Genome Organisation (HUGO).

Reference Sequence Types

- Depending on the variants to be reported, different reference sequence files are used at the **DNA level is most**, RNA or protein level. It is mandatory to indicate the type of reference sequence file using a**prefix**preceding the variant description. Approved reference sequence types are**c.**,**g.**,**m.**,**n.**,**o.**,**p.**and**r.**:
- •DNA
- •**g.**=linear **genomic** reference sequence, (including both introns & exons) (so at level bps 501 can be in intron or exon)
- •**o.**=circular genomic reference sequence (bacterial)
- •**m.**=mitochondrial reference(special case of a circular genomic reference sequence)
- •**c.**=coding DNA reference sequence(based on a protein coding transcript) (exons only)
- •**n.**=non-coding DNA reference sequence(based on a transcript not coding for a protein), introns
- We commonly depend on g. on genomic level & cDNA & p. On protein level to examine at amino acids (position of codon).

Variant nomenclature: cDNA

- Nucleotide **1** is the A of the **ATG initiation codon** (there is no c.0)
- The nucleotide **5'** of the ATG initiation codon is **-1, the previous -2**, etc. regions before the gene
- The nucleotide **3'** of the stop codon is ***1**, the next *2, etc. after the gene is ended
- Intronic nucleotides
- 5' end of the intron: the number of last coding nucleotide of the preceding exon, a plus sign and the position within in the intron,

➢ e.g., c.36+1G, c.36+2T. Explaination, c: codon level, position 36 of coding regions(exon no. 36), +: enter introns (after the 36th exon), at 1st bps of introns, G: a guanine exists (variance). +2T: in the next intron, at the 2nd bps, a Thymine variant.

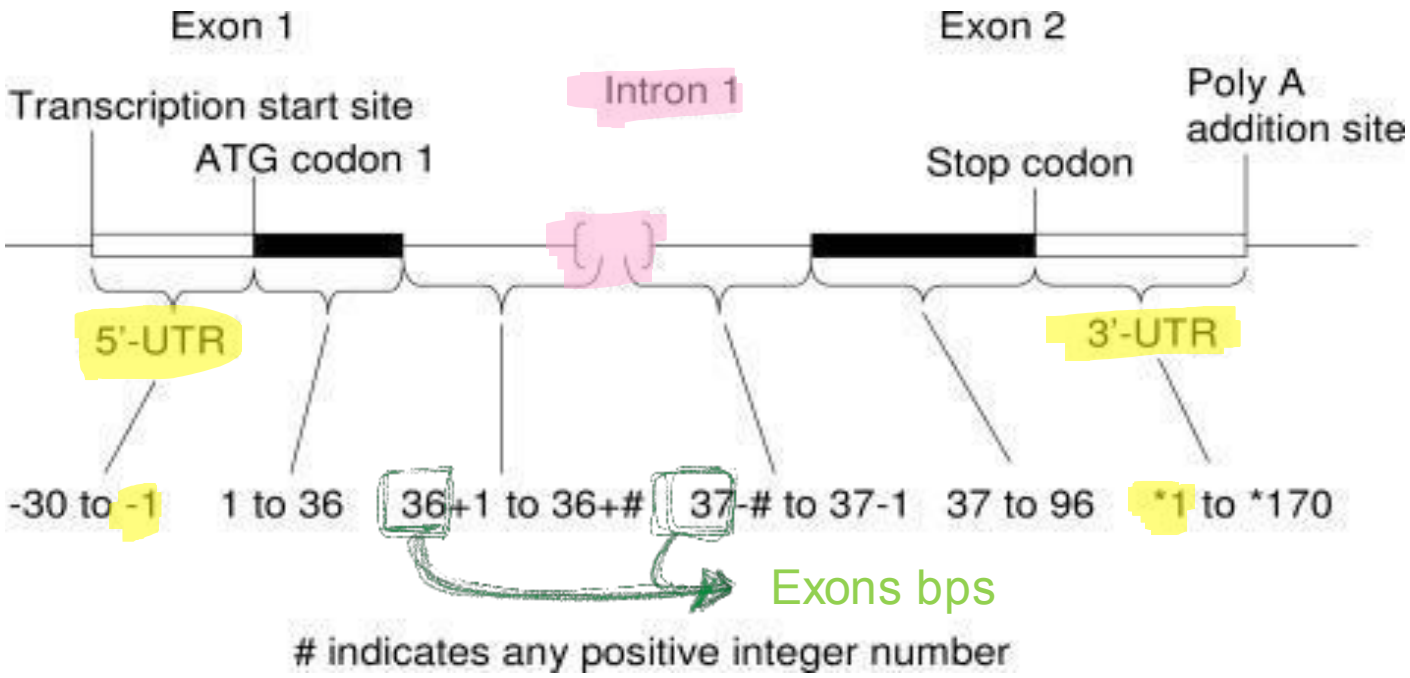
- 3' end of the intron: the number of the first coding nucleotide of the following exon, a minus sign and the position upstream in the intron,

➢ e.g. c.37-1G, c.37-2A.

- Notice that if the gene has many exons it could be shorter than a gene has less exons, because it depends on introns length mainly.
- As we learnt in previous slides, introns has a splice donor beginning & splice acceptors end
- If we want to name a variant on position 450 in an intron of 500bps, we **don't** write c.36+450G !! **Instead**, we search for the next exons & count backwards (c,37-50G) مثلًا
- So in other words, counting form splice donor (+), counting from splice acceptors (-)

To sum up rules from the previous method:

1) Start with a letter to indicate the reference, most common is (c.) for only coding regions, then (g.) for whole genome
2) No 0 number
3) First Numbering 1,2,3,4… in (c.) means bps in exons, from the ATG codon.
4) Putting (+) or (-) & another numbers to indicate the bps in introns, either after the exon or before the exon.
5) The letter (A,C,T,G) indicates the variance.



6) If we start by (g.) = all genome bps, we just count from The beginning of the gene till the variant position.
7) e.g.: g.486G **vs** c.37-50G: a variance in position 486 of a whole genome, or in position of 50 bps of an intron before 37 bps of the exon.

We assumed that the intron in the pic length is 500, and a variant is on position 450 of the intron.

Note that (in c. reference type) we start from ATG codon at level of a protein coding DNA. But in (g.) it's based on genomic / chromosomal nucleotides sequence.

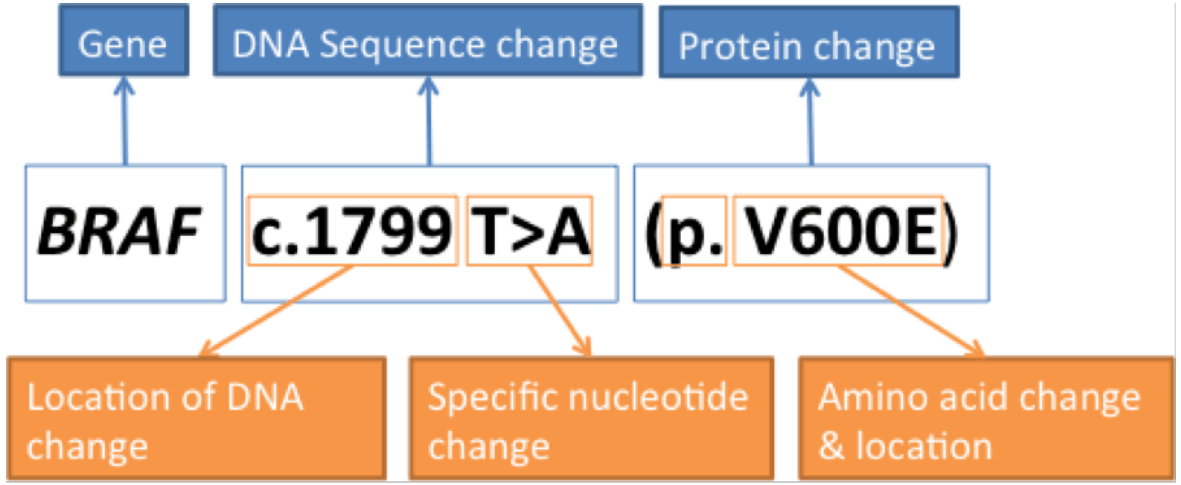
Genomic reference sequences can be based on an entire chromosome or cover small sequence only a specific gene or specific genomic segment

Symbols and abbreviations

>	c.4375C>T	**Substitution** of the C nucleotide at position c.4375 with a T
del	c.4375_4379del or c.4375_4379delCGATT	Nucleotides from position c.4375 to c.4379 **deleted**
dup	c.4375_4385dup or c.4375_4385dupCGATTATTCCA	Nucleotides from position c.4375 to c.4385 **duplicated**
ins	c.4375_4376insACCT	ACCT **inserted** between positions c.4375 and c.4376
delins	c.4375_4376delinsACTT or c.4375_4376delCGinsAGTT	Nucleotides from position c.4375 to c.4376 (CG) are **deleted** and **replaced** by ACTT

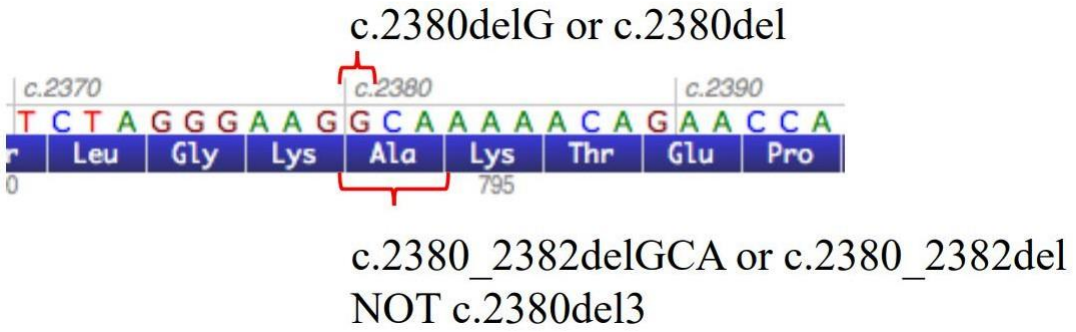
(_): the between bps
(>): replacing substitution
Note: from less number is counting

1) Gene name: italics & upper case
2) DNA level
3) Protein level. (P. One letter or 3 of the Amino acid. Variant Position. To new a.a name)



"c." prefix denotes standard variant nomenclature based on coding DNA reference sequences
"p." prefix denotes standard variant nomenclature based on protein-level amino acid sequences

- Format: **"prefix""position(s)_deleted""del",** e.g. g.123_127del



- For all descriptions the **most 3' position** possible of the reference sequence is arbitrarily assigned to have been changed (3'rule)

Variant nomenclature: protein

- **3-letter amino** (preferred)acid code is preferred to describe the amino acid residues (Lys vs. K for lysine)
- For all descriptions the most C-terminal position possible is arbitrarily assigned to have been changed
- **Methionine** encoded by the **translation initiation site** (start codon) is numbered as residue 1 ("Met1" or " M1") (1^{st} -3^{rd} bps)
- "**Ter" or "*"** designating a translation **termination stop codon** (some labs use X)

Protein nomenclature

- **Silent changes:** p.Leu54Leu or p.=
- **Substitutions:** p.Trp26Cys
- Nonsense variant: p.Trp26Ter or p.Trp26*
- **No-stop change:** p.Ter110GlnextTer17 or p.*110Glnext*17 (**just skip this as the doctor mentioned**)
- **In-frame deletions:** p.Gln8del or p.Cys28_Met30del
- **Duplications:** p.Gly4_Gln6dup
- **Insertions:** p.Lys2_Met3insGlnSerLys
- **Frameshifts:** short description: p.Arg97fs

  long description: p.Arg97Profs*23 (on position 23 before the actual stop codon, a variant has developed a new stop codon, so 23 is new stop codon, the variant on position 97,) /Note: that (fs) is for out frameshift variant.

where the “Arg97Pro” describes the substitution of Arg for Pro at position 97, “fs” indicating the frameshift and the “*23” describes the position of the translational termination (stop) codon in the new reading frame (starting with proline as amino acid #1)

Questions: past-papers

تم بحمد الله

1) A frameshift mutation can be caused by?
a) 4 bases being inserted
b) 2 bases being deleted
c) 3 nucleotides being added
d) 3 nucleotides are removed
e) A & B
(outframeshift)

2) An apyrimidinic or apurinic (AP) mutation forms, in which type of DNA repair ?
(Actually from an extra note I've written)
a . Photoreactivation
b . Mismatch repair
c . Base excision repair
d . Direct repair
e . Nucleotide excision repair

3) In mammals , double - strand breaks in DNA are primarily repaired through:
A . Direct repair
B . Mismatch repair
C . Non homologous end joining
d . Nucleotide excision repair

4) Homologous recombination occur in G1 phase?
a. True
b. False

5) Which type of DNA lesion is caused by UV light?
a. Large deletion
b. Depurination
c. Mismatch
d. Pyrimidine dimer

6) Which one results from deamination adenosine?
a. Uracil
b. Xanthine
c. Hypoxanthine
d. Cytosine

Ans: e – c – c - b – d - c

V1

V2 : highlight

V3 (nomenclature)