GENETIC VARIATION

By changing one allele to another, mutations modify the structure or amount of gene's protein product, and these modifications in protein structure or amount can influence phenotype

p.s. Some mutations don't have an impact on gene function and thereby don't influence phenotype

MUTATION = an event that produces a heritable (stable) change in DNA sequence
source of genetic variation and phenotypic variability
Provide the basis for adaptation to environmental changes and evolution
Genetic changes can also lead to disease
Provide the primary tools for genetic analysis

Often the term mutation is used to describe the outcome of the process, the DNA variant, especially when it's associated to altered phenotypes

DNA variants
- population frequency > 1% → DNA polymorphisms
- population frequency < 1% → rare variants

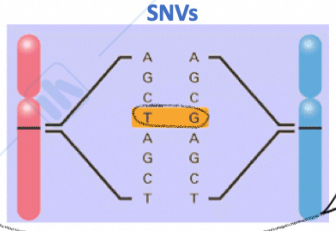
- Germline mutation: occurs in germ cells (egg o sperm), passes from progeny for reproduction
- Somatic mutation: occurs in somatic cells, do not pass to progeny; if the mutation occurs at an early stage of development, the same zygote will give rise to an individual with two or more genetically different types of cells (mosaic)

○ Inherited mutations - are always transmitted through the ~~gremlin~~ germline, although a parent may also have a mosaic mutation
○ De novo mutations - are defined as genotypes observed in a child but not in either parents, they may have originated in parental germ cell or postzygotically
○ Somatic mutations - may occur relatively early in development or at any later time throughout the lifespan, generally affecting fewer cells

Size of genetic variation from single base to chromosome:

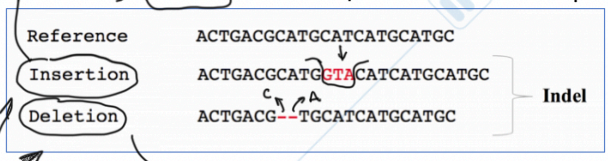
- Point mutation or short genetic variation > small (changes of a single or few nucleotides)
  1) single nucleotide variants (SNV)
  2) insertions/deletions (indel)
  - microsatellite (short tandem repeat) variants

Transitions, transversions, small deletions or insertions

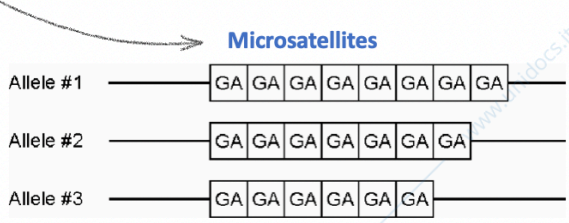


addition of one or more nucleotide pairs

2) INDELs insertion/deletion of 1 or few bp



a block of one or more nucleotide pairs is lost from a DNA molecule



- tandem repetitions of short sequence units
- 1-2-3 nt
- mostly outside coding regions
- higher mutations rate compared to SNV because of DNA replication slippage

- structural variation (SV)
  - copy number variants
  - balanced SV (inversion, translocations)
  - chromosome rearrangements (cytogenetically visible)
  - submicroscopic

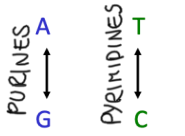
- variation in chromosome number > large

SNVs are the most **common type** of DNA variation

divide into

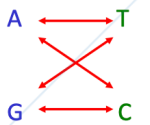
1) (SNVs) SUBSTITUTIONS: a base at a certain position in one strand of the DNA molecule is replaced by one of the other three bases

Transition
Purine <--> purine
Pyrimidine <--> Pyrimidine



Transitions

Transversion
Purine <--> Pyrimidine
Pyrimidine <--> Purine



Transversions

common SNVs are generally called:
SNPs Single Nucleotide polymorphisms

DNA replication errors

Most DNA replication errors are caused by mispairings of bases

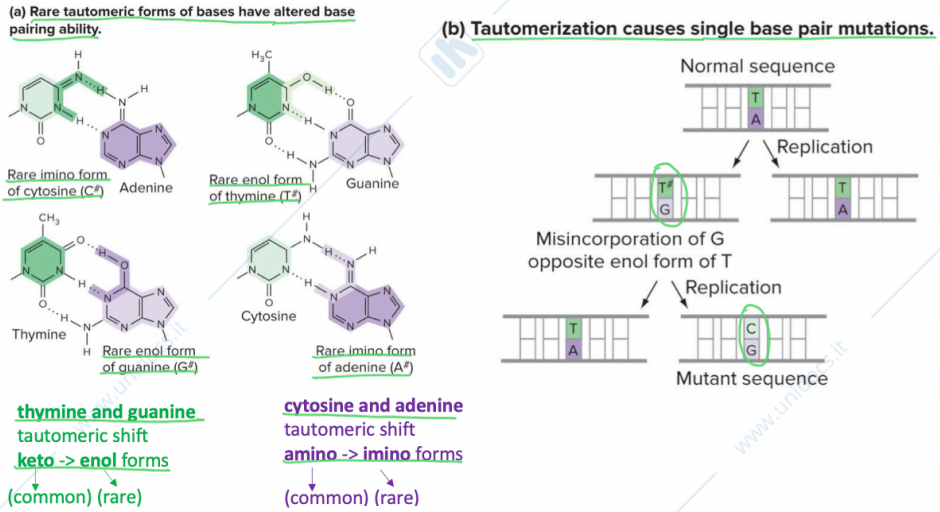
- if not corrected by DNA repair

→ àpoint mutation

- Misincorpora;on of nucleo;des by DNA polymerase during DNA replica;on (rare,< 10−9 in bacteria and humans)

DNA polymerase has proofreading ac;vity, called 3'-to-5' exonuclease, which recognizes and excises mismatches

DNA replication errors

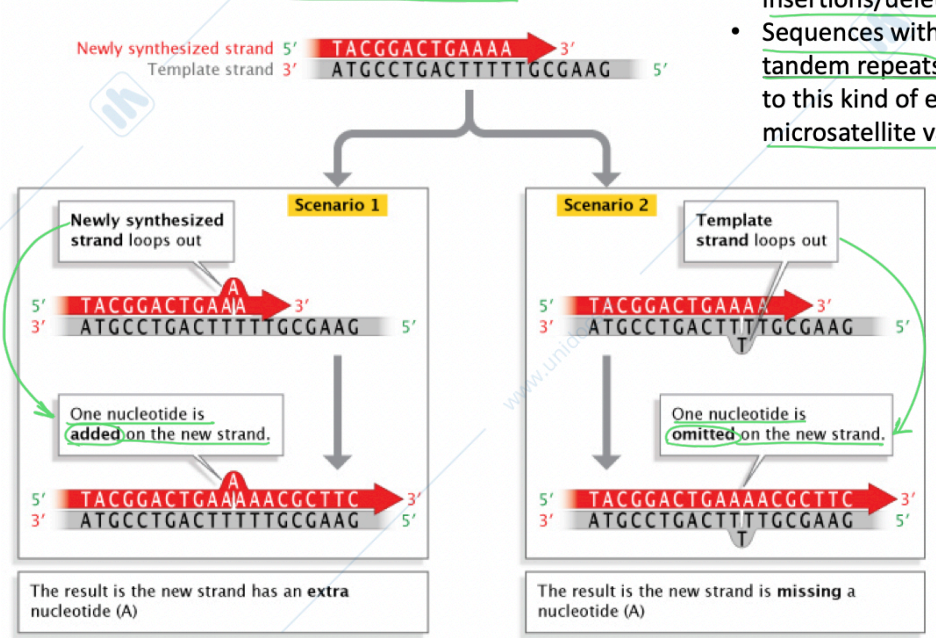


thymine and guanine
tautomeric shift
keto -> enol forms
(common) (rare)

cytosine and adenine
tautomeric shift
amino -> imino forms
(common) (rare)

Abnormal pairing:
T-G **C-A**

Replication slippage

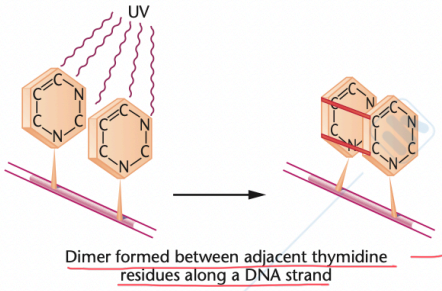


Strand slippage

- Causes small insertions/deletions
- Sequences with short tandem repeats are prone to this kind of error → microsatellite variants

Natural processes cause mutations through DNA damage (2)

(2) ➤ Ultraviolet (UV) radiation



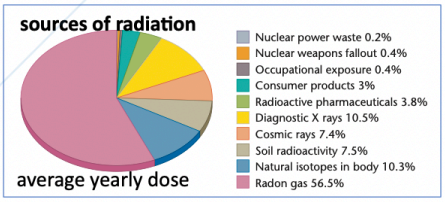
- Ultraviolet (UV) light causes adjacent thymines to form abnormal covalent bonds (thymine dimers)
- dimers distort the DNA conformation and inhibit normal replication

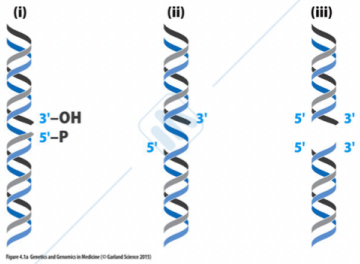
FIGURE 15–8 Induction of a thymine dimer by UV radiation, leading to distortion of the DNA. The covalent crosslinks occur between the atoms of the pyrimidine ring.

(3) ➤ Ionizing radiation (X rays)

break the sugar - phosphate backbone

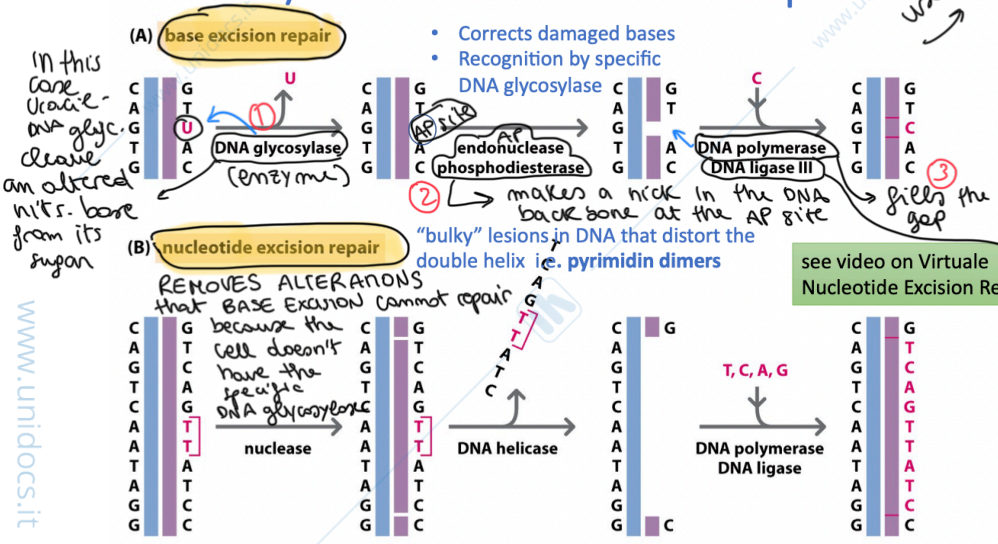
- Penetrate more deeply into tissues
- Cause ionization of the molecules, and generate free radicals
- free radicals can directly or indirectly damage DNA
  - purines and pyrimidines modifications
  - phosphodiester bonds breaks
  - **double strand breaks** → chromosomal aberrations, such as deletions and translocations





DNA REPAIR MECHANISMS

Base/ nucleotide excision repair



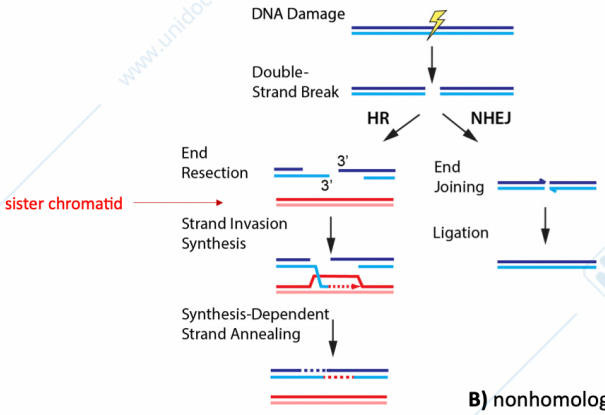
see video on Virtuale Nucleotide Excision Repair

Strategy of homology dependent repair: first remove a small region from the DNA strand that contains the altered nucleotide and then use the other strand as a template to resynthesize the region removed and recreate the original sequence

④ seals up the backbone of the newly repaired DNA strand

2 MECHANISMS OF Repair of double strand breaks

NER and Xeroderma Pigmentosum
- NER is defective in individuals with XP
- Aut. Rec. disease characterized by extreme sensitivity to UV, skin cancer and neurological symptoms
- mutation in at least 9 different genes involved in the NER system

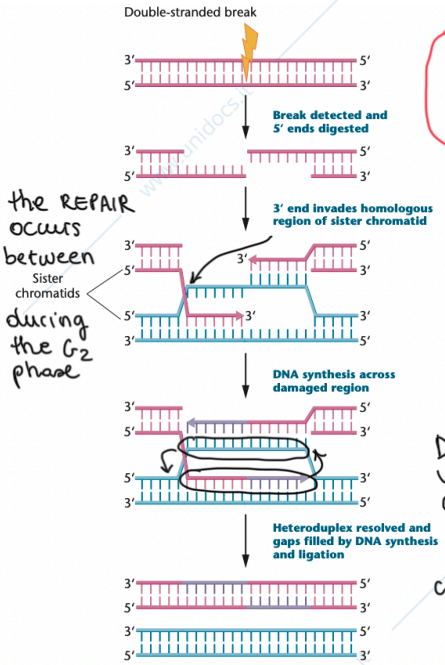


A) homologous recombination repair
- during the late S or early G2 phase of the cell cycle, after DNA replication
- Accurate repair

B) nonhomologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ)
- **in G1**
- **error prone**

Homologous recombination repair (HR)

uses complementary base pairing to repair breaks accurately with no loss or gain of nucleotides

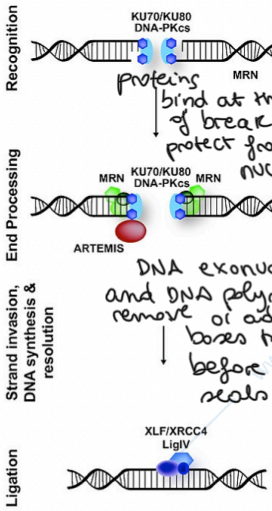


- **homologous recombination repair**
- during the late S or early G2 phase of the cell cycle, after DNA replication
- Accurate repair

Nonhomologous end joining (NHEJ)

Doesn't involve DNA homologies → can join together any DNA ends

can bring together even DNA ends that were not previously adjacent to each other, and a few base pairs can be lost or added improperly in the process



- Activated in G1, prior to DNA replication.
- Protein complex bind to the free ends of the broken DNA, trim the ends, and ligate them back together.
- Some nucleotide sequences may be lost in the process of end joining, it is an error-prone repair system.
- If more double-strand breaks are present, the wrong ends could be joined together, leading to abnormal chromosome structures

Genotype and allele frequencies

A locus with two alleles (*A* and *a*) in a population of 20 individuals

Genotypes	AA	Aa	aa
	12	4	4

Genotype frequencies:

AA = 12/20 = 0.6
Aa = 4/20 = 0.2
aa = 4/20 = 0.2

→

Allele frequencies:

In 20 people, there are a total of 40 alleles

p Frequency of *A* alleles = (24 + 4)/40 = 0.7
q Frequency of *a* alleles = (8 + 4)/40 = 0.3

OR

p = f(AA) + ½ f(Aa) = 0.6 + 0.1 = 0.7
q = f(aa) + ½ f(Aa) = 0.2 + 0.1 = 0.3

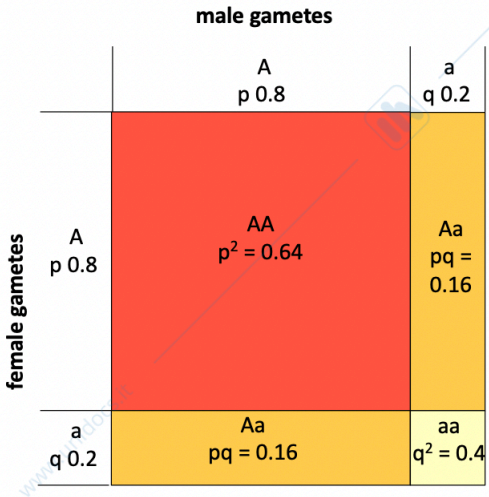
From genotype to allele frequencies:

Genotype	*AA*	*Aa*	*aa*	Total
Number of individuals	12	4	4	20
Genotype frequency	0.6	0.2	0.2	1.0

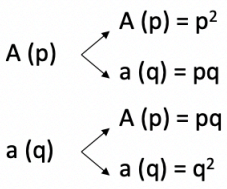
Allele	*A*	*a*	Total
Number of chromosomes	28	12	40
Allele frequency	0.7	0.3	1.0

The Hardy-Weinberg equilibrium (HWE)
correlates allele and genotype frequencies
Five simplifying assumptions:

- The population has a large number of individuals
- Individuals mate at random
- No new mutations (negligible)
- No migration (negligible)
- No selection (genotypes have no effect on ability to survive and transmit alleles to the next generation)



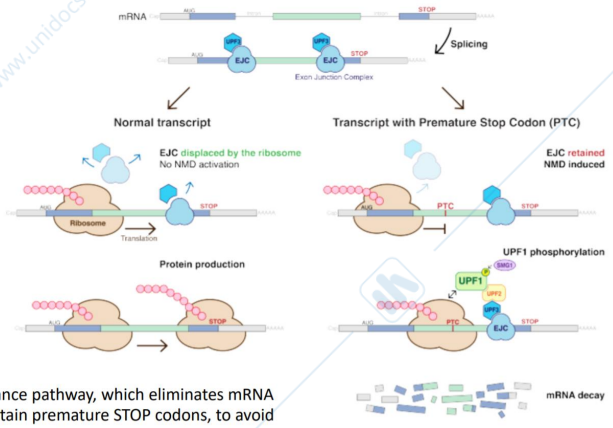
In a large population of randomly breeding individuals with no new mutations, no migration, and no genotype-dependent differences in fitness:



Genotype	AA	Aa	aa
Frequency	p^2	2pq	q^2

$p + q = 1$ $p^2 + 2pq + q^2 = 1$ 52

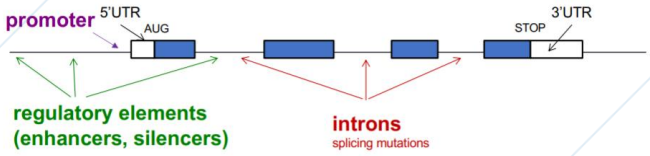
Nonsense-mediated decay



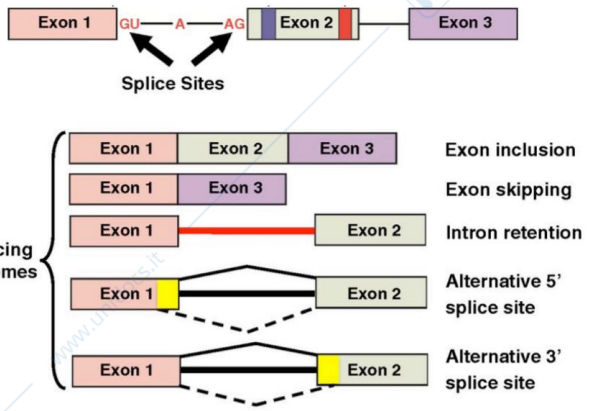
Surveillance pathway, which eliminates mRNA that contain premature STOP codons, to avoid production of aberrant truncated protein

At the end the complex is removed and the ribosome catches the mRNA with premature stop codon. Some factors degrade the wrong mRNA. There's a total loss of function.

Mutations in non-coding sequences

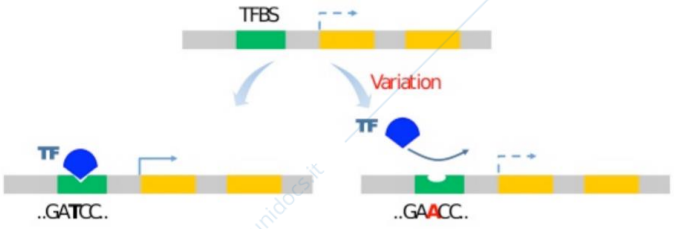


Splicing mutations produce a non functional protein



Regulatory mutations:

mutations within noncoding regions may alter the process of gene expression, or mRNA stability

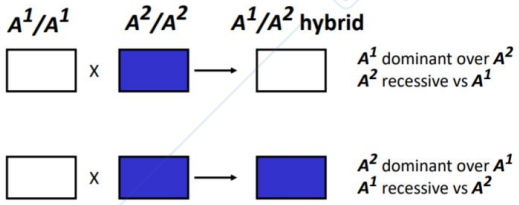


Functional consequences of alleles on phenotypic expression: **dominance or recessiveness**

The terms dominant and recessive have a phenotypic basis

- the dominance of one allele over another is influenced by the effect of that allele on the protein product

- the overall phenotype is the consequence of the activities of the protein products of the alleles of the gene

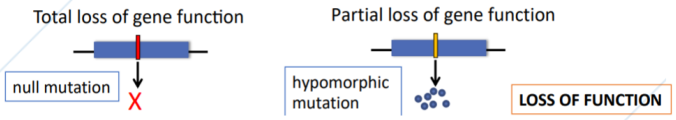


Functional consequences of mutations: **loss of function**
Examples:

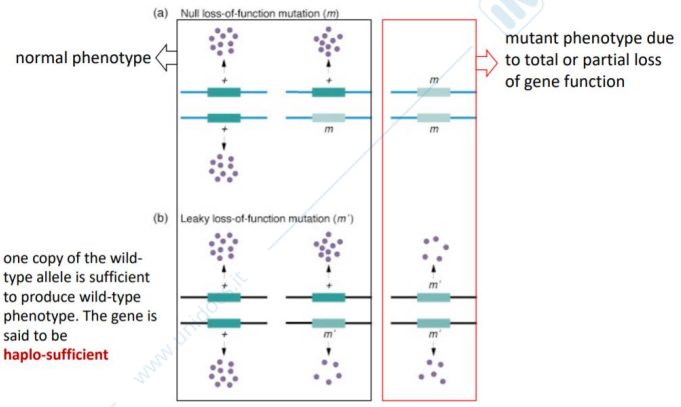
- Nonsense (Stop gain)→introduction of a premature stop-codon
- Frameshift
- Damaging missense→damaging when aa replaces another aa
- Alteration of splicing→splicing errors that produce non functional proteins

- Total or partial gene deletion
- Disruption of a gene by a chromosomal rearrangements
- Disruption of promoter/regulatory elements→elimination of regulatory sequences or of the promoter

We can have a total loss of gene function or a partial loss of gene function:



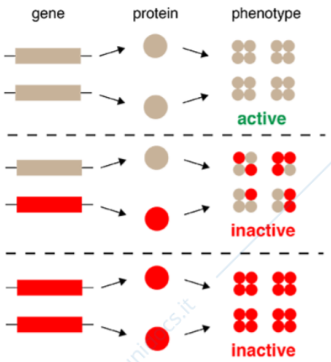
loss of function may be recessive



In this case in the heterozygous individual there are enough proteins that assure the normal phenotype.

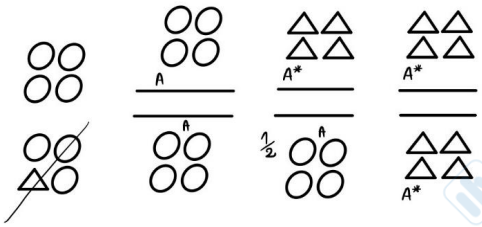
Dominant-negative mutations → dominant phenotype

The mutant gene product acts antagonistically to the wild-type allele Sometimes proteins are composed by subunits, and called multimeric proteins.
Multimeric proteins (composed of two or more polypeptides that join together to form a functional protein) are particularly subject to dominant negative mutations.



in the heterozygote state, the mutant allele interferes with the function of the wild-type allele by specifying polypeptides that inhibit, antagonize, or limit the activity of the wild-type polypeptide

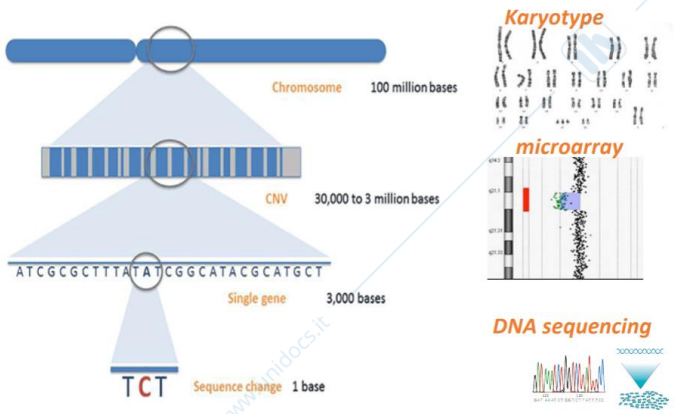
In the second case we have 50% of the normal product, so: 1/2 x 1/2 x 1/2 x 1/2 = 1/8 probability of having the right protein function



Functional consequences of mutations: dominance and recessiveness

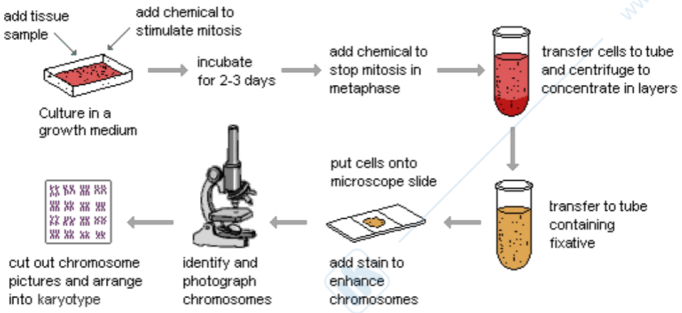
RECESSIVE	DOMINANT
• Loss of function	• Gain of function • Haploinsufficiency • Dominant-negative

VARIATION IN CHROMOSOME NUMBER and STRUCTURAL VARIATION The range of genetic variation, from single base to chromosome



Sequence change: 1 base→DNA sequencing Single gene: 3000 bases→microarray
CNV and chromosome: million bases→karyotype

Karyotype analysis: way of studying the structure of chromosomes

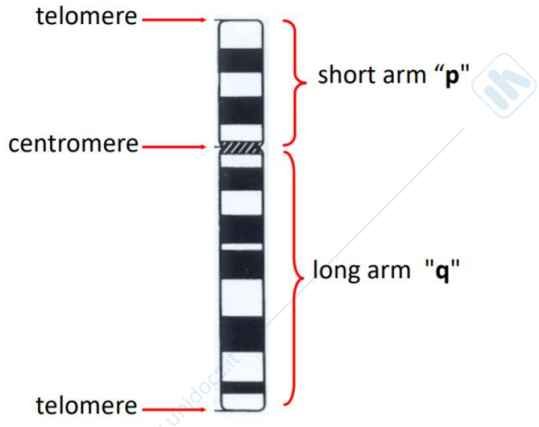


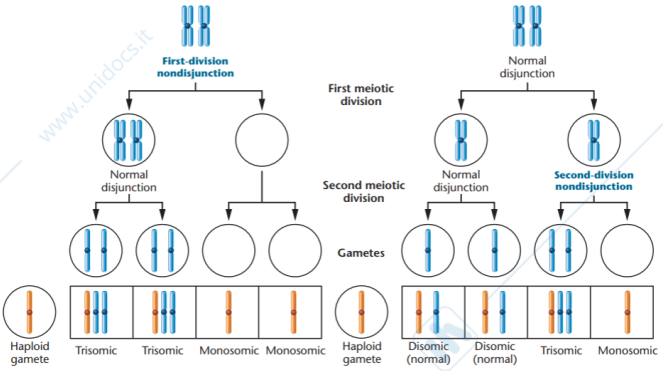
Karyotype describes the chromosome count of an organism and what these chromosomes look like under a light microscope.
Different banding techniques are used to visualize the chromosomes.

G-banding is obtained with Giemsa stain following digestion of chromosomes with trypsin. The dark regions tend to be heterochromatic, late-replicating and AT rich. The light regions tend to be euchromatic, early-replicating and GC rich. This method will normally produce 300–400 bands in a normal, human genome. We can see a pattern that allows us to see all the different homologous pairs in chromosomes.

Human male karyotype

Metaphasic chromosome





Consequences of aneuploidy
Monosomy for any of the autosomes is not usually tolerated in humans or other animals. (mortal)

Trisomy for an autosome is slightly more tolerated in humans or other animals (i.e. chr 21 in human), but otherwise lethal. In plants, trisomic individuals are usually viable, but their phenotype may be altered.

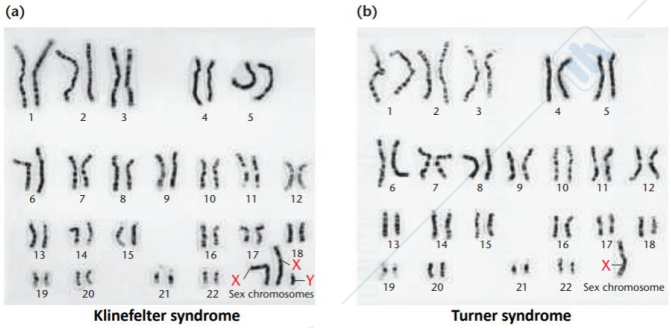
In humans, a large proportion of spontaneously aborted fetuses demonstrate some form of chromosomal imbalance→normal embryonic development requires a normal diploid complement

Trisomy 21- down syndrome

The only human autosomal trisomy not associated to early lethality
• 1 / 800 live births.
• associated with mild-moderate intellectual disability, a characteristic facial appearance, and weak muscle tone (hypotonia) in infancy.
• The large majority of cases derive from a non-disjunction error during female meiosis
• Down syndrome is usually sporadic
• Rare cases involve a translocation of chromosome 21 and may be inherited

Down syndrome is correlated to maternal age

Sex chromosome aneuploidies in humans



Klinefelter: XXY (maschio) Turner: X0 (femmina)

Genomic structural variation

Variation in structure of an organism's chromosome that involves relatively large DNA segments.

It's generally defined as a structure variation of a region of DNA of about 1 Kb or larger (to whole chromosome regions).
submicroscopic SV (from 1 kb to ~ 3-5 Mb)

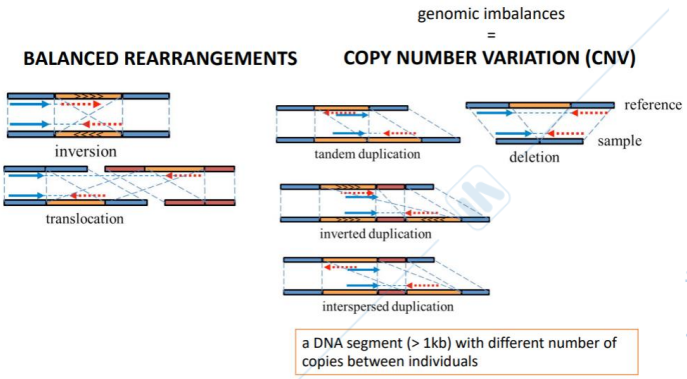
microscopic SV (> 3-5 Mb) à visible with karyotype

Represent an important mutational source shaping genome evolution and function, and a significant contributor to disease.

Main classes of SV
- Insertions
- Deletions
- Inversions
- Translocations

} **COPY NUMBER VARIANTS (CNVs)**

Main classes of SV:

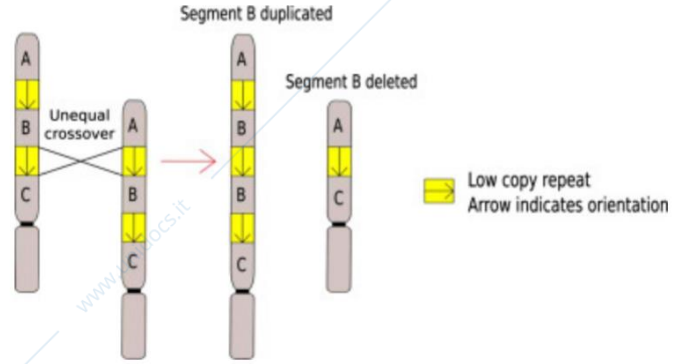


Inversion: rotate a segment of DNA of 180°
Translocation: segment of DN moved from one chromatid to another Deletion or duplication: of a segment of DNA

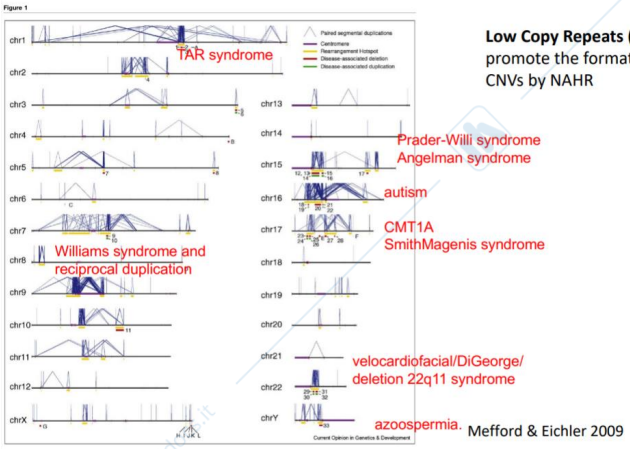
Deletions or duplications

In many cases, CNVs are formed through nonallelic homologous recombination, facilitated by the presence in the human genome of repetitive DNA elements, low copy repeats, that cause chromosomes to misalign during recombination

CNVs created via this mechanism can mutate relatively frequently and have similar breakpoints. They are called recurrent CNV



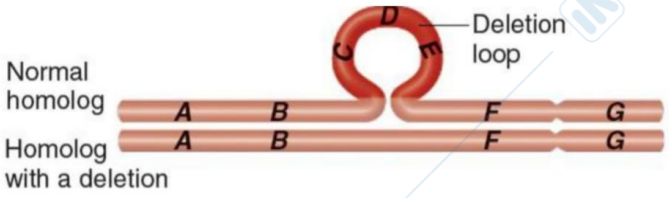
Genomic hotspots:



Low Copy Repeats (LCR) promote the formation of CNVs by NAHR

Blue lines: repeated sequences→can frequently give rise to insertion or deletions 20

Deletions/ duplications: pairing at meiosis



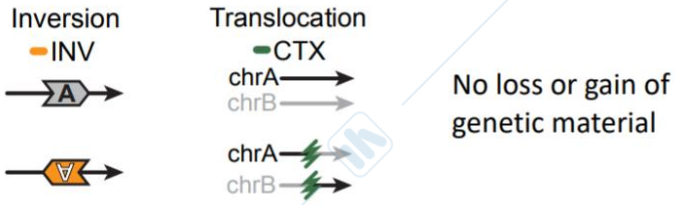
Copy number variations in the human genome

Provide a major contribution to genome variability among individuals

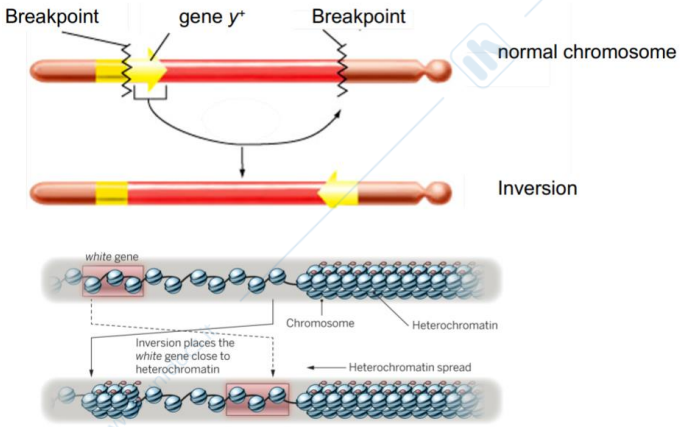
May have no phenotypic effect, account for adaptative traits or can underlie disease

Circa 9% of the genome contributes to common CND

Phenotypic number of balanced SV



- Can alter the phenotype if the breakpoint interrupts the coding or functional sequences of a gene causing loss of function
- Loss/disruption of regulatory elements (i.e. promoter, enhancer)
- Position effect (a gene is placed in a different chromatin or regulatory environment)
- Translocations may result in a fusion (chimeric) gene→gain of function alleles whose protein products have an altered structure or level of expression à important role in cancer



Philadelphia chromosome translocation

In chronic myelogenous leukemia a translocation brings together exons of the BCR gene from chromosome 22 and the ABL1 gene from chromosome 9.
- One product of the translocation is the Philadelphia chromosome (Ph1), containing a chimeric BCR– ABL1 fusion gene.

- This encodes a constitutionally active tyrosine kinase that does not respond to normal controls. - translocation occuring in somatic cells→abnormalities only in the affected cell line

Structural variants: impact on meiotic pairing and crossing-over

- if an individual is heterozygous for a chromosomal rearrangement, unusual pairing configuration occurs during meiotic synapsis.
- If the rearrangement is balanced (does not cause loss

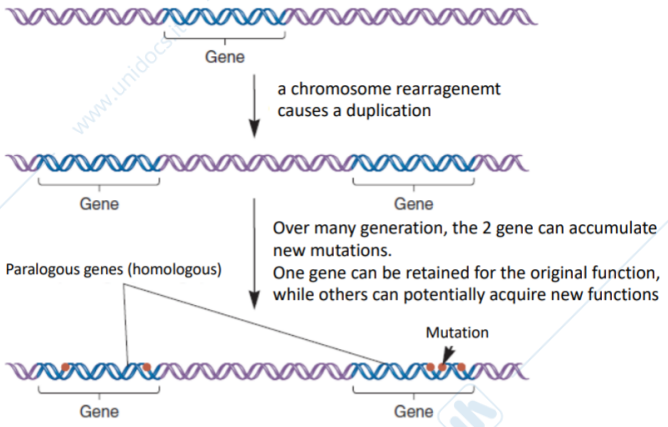
Variation in gene copy number

1) Gene duplications→provide a means to amplify gene dosage for products that are needed in great abundance
• i.e. genes coding for rRNA

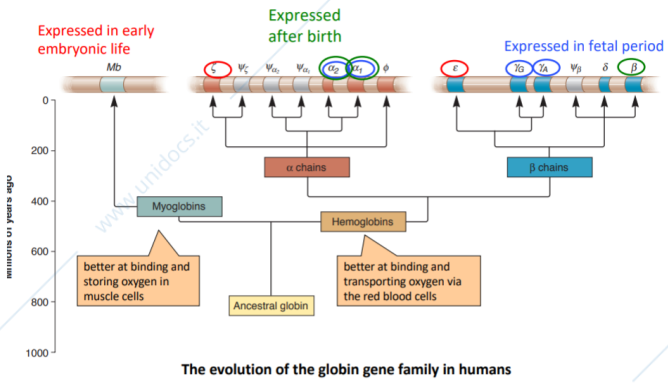
2) Adaptations to environmental changes
• i.e. number of copies of the AMY1A gene, encoding for salivary amylase, an enzyme required to metabolize starch

3) A source of new genes in evolution

Duplication in evolution

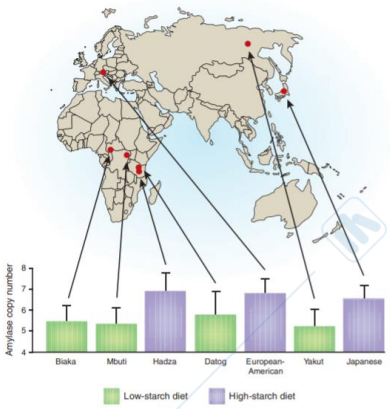


Gene families:



The evolution of the globin gene family in humans

Salivary amylase copy number:



- The human *AMY1A* gene is normally present in multiple copies and is variable (chimpanzees have a single copy of this gene)
- There is a correlation between higher copy number in population with with high-starch diets vs lower copy number in populations with low-starch diets, even when they are geographically near
- Beneficial adaptation in presence of human diets that increasingly became rich in insoluble starch.

Perry et al, Nat Genet 2007

BIOCHEMICAL GENETICS

- Haploid fungi represent the ideal genetic system to study the basic biochemical and molecular-cellular functions
- Easily and inexpensively maintained in the lab
- They are haploid so ideal for mutational analysis because mutant alleles are always expressed directly in the phenotype.
- They can grow on a minimal medium (a source a carbon, nitrogen, phosphorus, inorganic salts, water), from which they can chemically synthesize all the essential molecules
- Can also grow on rich medium and break down complex macromolecules

Analysis of neurospora nutritional mutants

1940s George Beadle and Edward Tatum (Nobel prize in 1958)

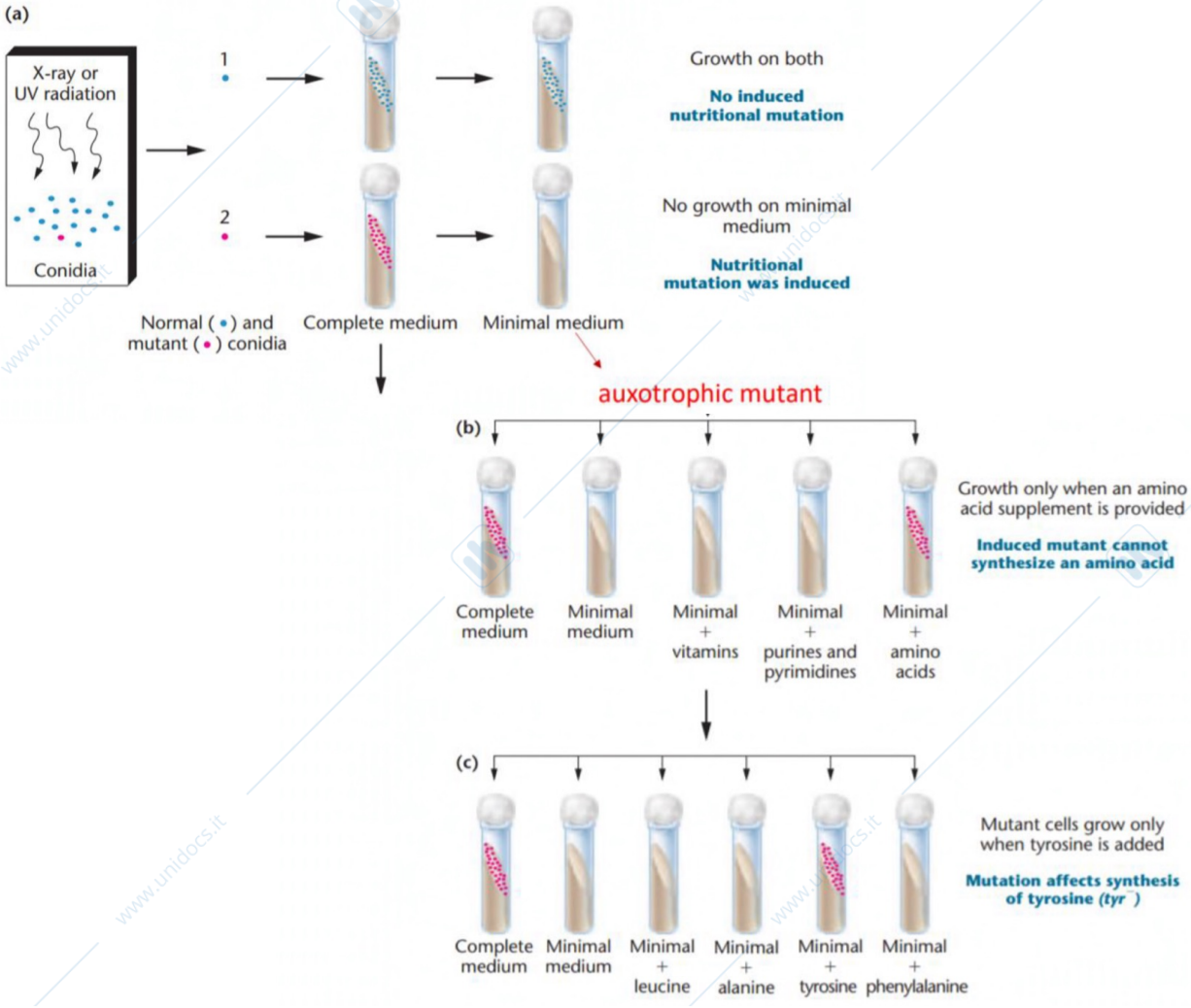
They isolated from yeast many different mutants that were lacking the ability to synthesize different molecule, for example mutants that were not able to synthesize a specific vitamin, or amino acid, or that require a specific nitrogen base. They carried out a systematic work in order to isolate many different mutants from yeast.

- Neurospora is capable of synthesizing all the essential metabolites
- The biosynthesis of these metabolites must be under genetic control
- Mutations in genes whose products are involved in biosynthesis of a metabolite would produce mutant strains with nutritional requirements

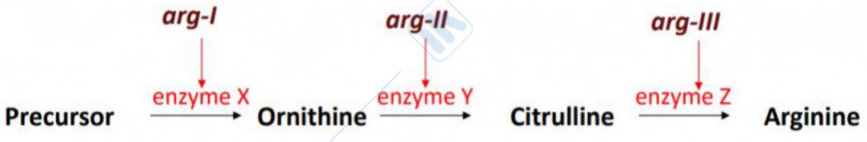
They used x-ray or UV radiations that can increase the mutation rate, in order to increase the probability to isolate mutations. So they irradiated cells to create a lot of mutants, and then let them grow on complete mediums. The complete medium contains all the necessary molecules to grow. They transferred each mutant strain on a minimal medium in order to isolate some strain that lacked some kind of ability. They characterized which was the specific pathway that was mutated in each single mutant strains.

For example they isolate a mutant strain that does not grow on minimal medium. They study which pathway is defective in the mutant. So they transfer the mutant in the minimal medium in which it will grow, and then in many different minimal medium with the addition for example of different vitamins, or purines and pyrimidines, or the specific aminoacid.

Then you can go further and see which specific requirement is needed in the strain by adding to the minimal the single aminoacid. For example you see that the mutant cell can grow only when tyrosine is added. So the mutant has a defect in the pathway, that is fixed by adding tyrosine in the minimal medium.



mutant	No suppl.	Ornithine	Citrulline	Arginine
Wt	+	+	+	+
Class I	-	+	+	+
Class II	-	-	+	+
Class III	-	-	-	+



One gene-one enzyme hypothesis
→later modified in one gene-one protein

EXAMPLES:
• Isolation of several mutant strains, all requiring "E" for growth.
• Each mutant was tested for the ability to grow on minimal medium in presence of different supplements (A, B, C, D)
• Try to determine the biochemical pathway leading to the synthesis of product "E" and which step is blocked in each mutant strain

	Supplements to minimal medium				
Mutant	A	B	C	D	E
1	+	0	+	+	+
2	+	0	0	0	+
3	+	0	0	+	+
4	0	0	0	0	+

Maybe there is a biochemical pathway of different steps required for the synthesis of E, and all these mutants are mutants of different steps. We isolate different compounds, A B, C and D.
Mutant 1 can grow when you add A, C, D and E, but not with B.
The 1st observation is that all mutants cannot grow when you add B. all the mutations are following B in this pathway, so it must be the 1st precursor, and all the mutants are after it.
You can order the different mutants according to the different nutrients that are needed.

	Supplements to minimal medium				
Mutant	B	C	D	A	E
1	0	+	+	+	+
3	0	0	+	+	+
2	0	0	0	+	+
4	0	0	0	0	+