Basics in Genetics

Pascale RICHARD, UF Cardiogenetics and Myogenetics, Pitié Salpêtrière Hospital, Paris, FR

Human Genetics

- Inheritance patterns
- Types of mutations
- Approach to genetic testing

Al Jawhara Center, Manama, Bahrain 21-23 March 2016

The two fields of the Genetics

- **The Clinical Geneticist**
  - Draw the pedigree tree
  - Explain the disease to the patient
  - Ask the Lab for molecular tests
  - Give the result of tests to patients
  - Organize the genetic counselling

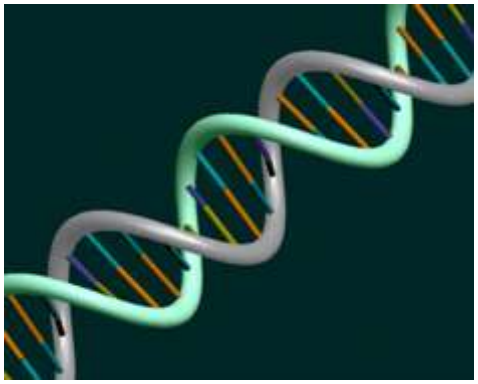
- **The Molecular Geneticist**
  - Validate the prescription (legal and medical aspects)
  - Organise the correct production of molecular tests
  - Interpretation and validation of analyses
  - Deliver a clear and interpretated conclusion of tests to the clinical Geneticist

The aim of molecular genetics

Molecular genetics is now in the current practice of the diagnosis of inherited diseases.

The knowledge of the molecular status of an individual is necessary for a proper diagnosis and genetic counseling

Knowing the genetic basis is useful for:



–Etiology of diseases: confirmation

– Genetic counseling: which individuals are at risk within the family

– Presymptomatic Diagnosis: *what to expect*

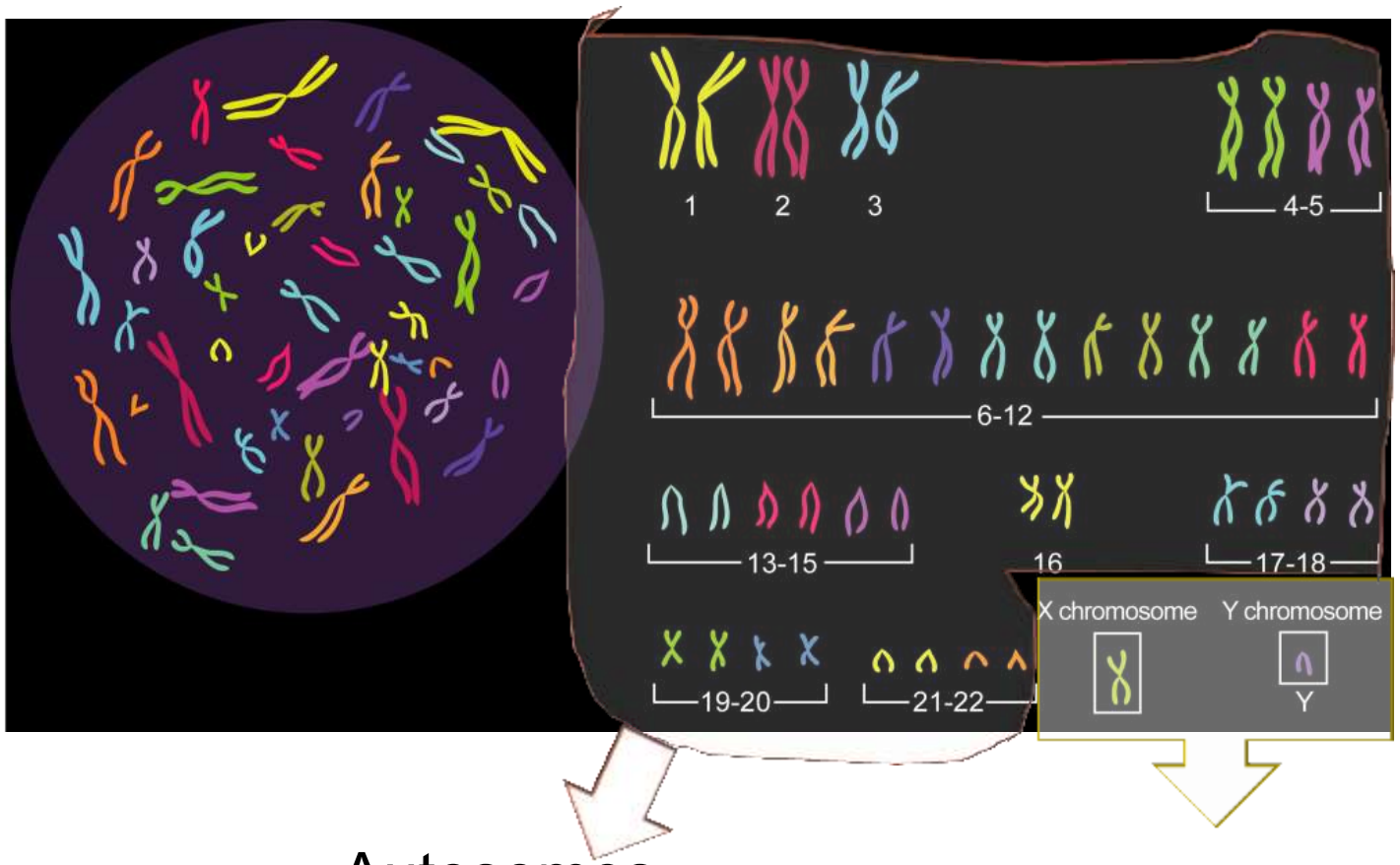
– Clinical care

– Key to understand the disease (pathophysiological)

How to treat

Basics on the genome and genes

Caryotype
(23 pairs of chromosomes)
- 22 autosomal
- 1 sexual

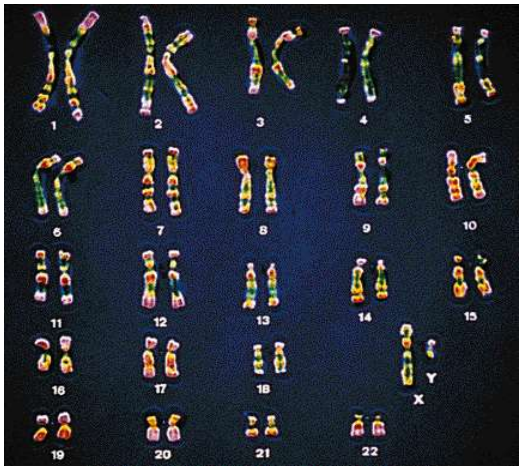


Autosomes

Sexual chromosomes

We inherit one set of chromosomes from each parent

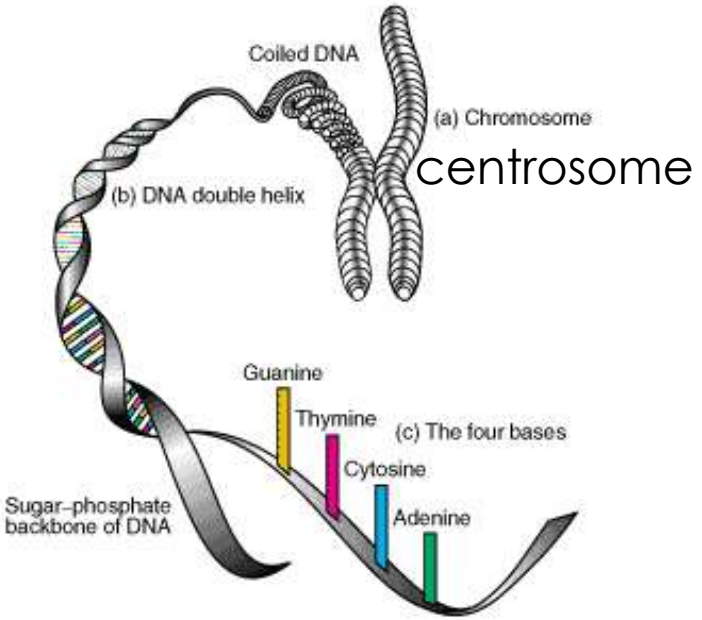
Knowledge of DNA structure is key to understand genetics . . .

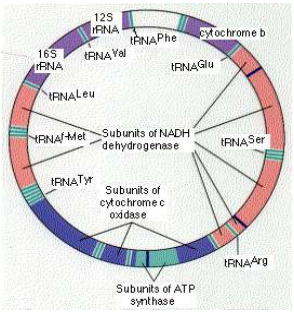


Nuclear genome

30 000 Genes

DNA molecule is composed of 4 nucleotides





Mitochondrial Genome

>> A gene: a section of DNA that codes for a protein

Each individual has two copies of almost all genes

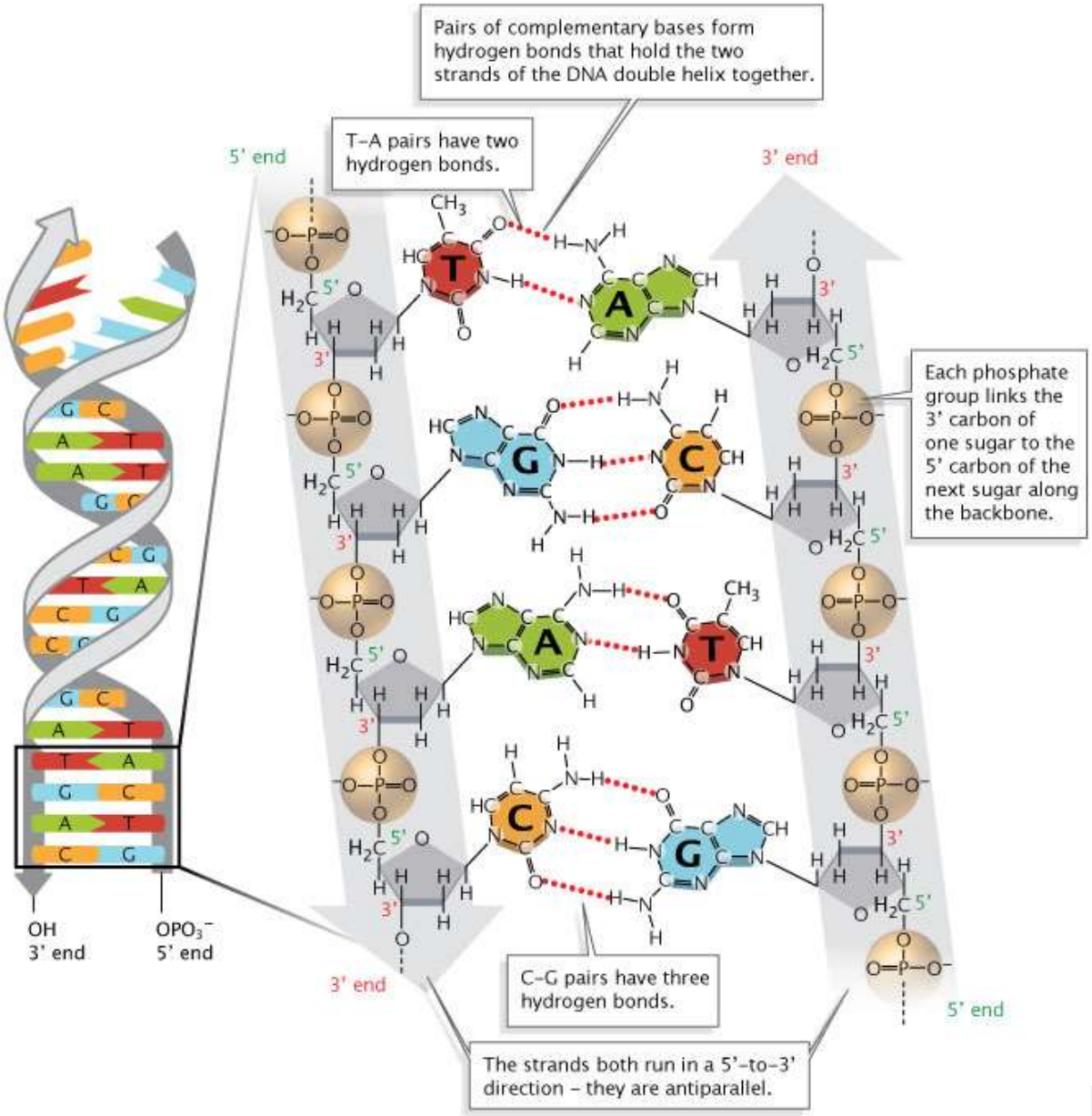
Identity card of our genome

Size of the nuclear genome	3.1 Gb (haploid), 6.4 Gb (diploid)
Euchromatic component	2.9 Gb (93%)
Highly conserved fraction of genome	150 Mb (5%)
Protein-coding DNA sequences	35 Mb (1.1%) (translated) commonly quoted: 1.5-2%
Other highly conserved DNA	115 Mb (3.9%)
Segmentally duplicated DNA	160 Mb (5.5%)
Highly repetitive DNA	1.6 Gb (50%)

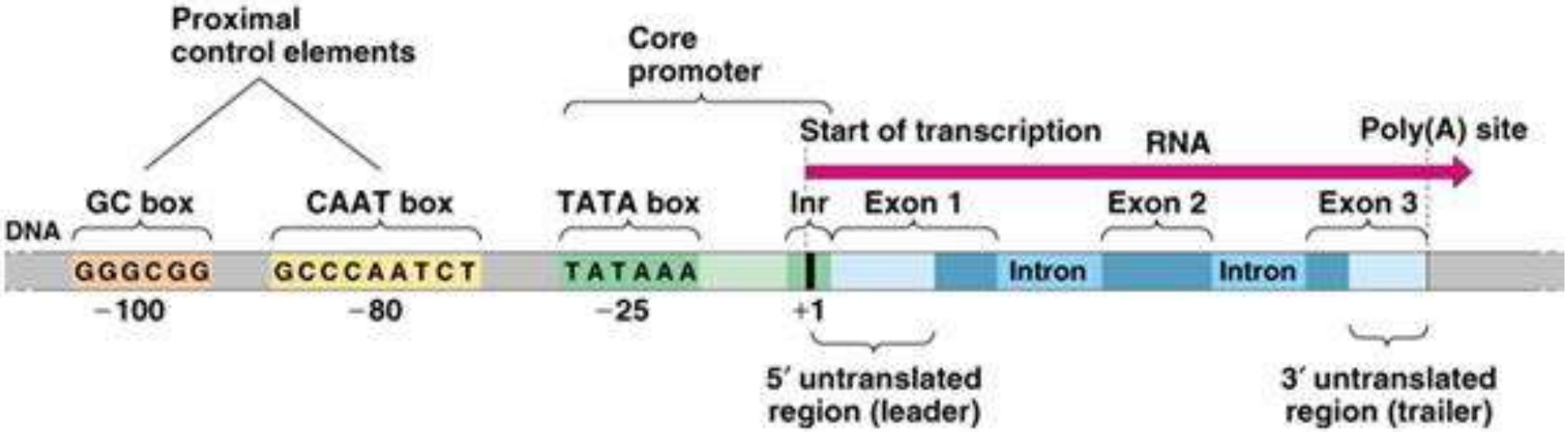
DNA per chromosome	48 Mb - 249 Mb
How many genes in the nuclear genome	> 26,000
How many protein-coding genes are there?	20,000 - 21,000
How many RNA genes are there?	>6000
How many pseudogenes repated to protein-coding genes are there?	>12,000

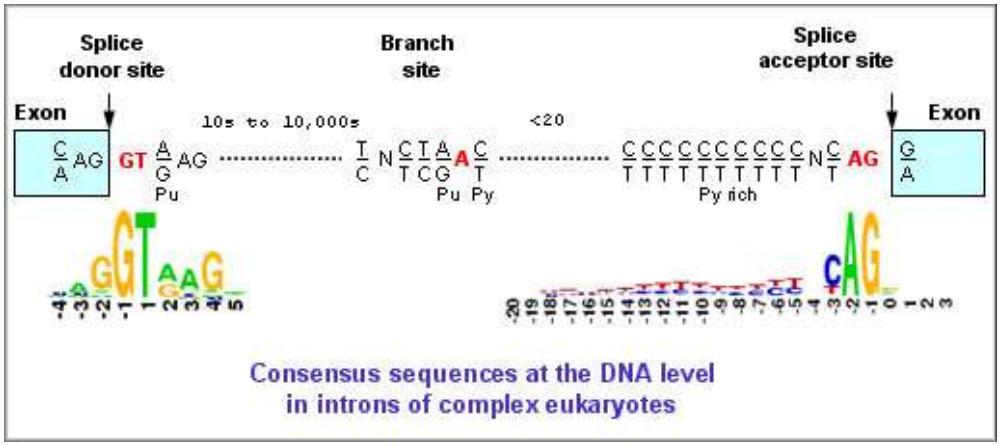
How many genes in the mitochondrial genome	37 genes
Size of the mitochondrial genome	16.6 kb

The DNA molecule



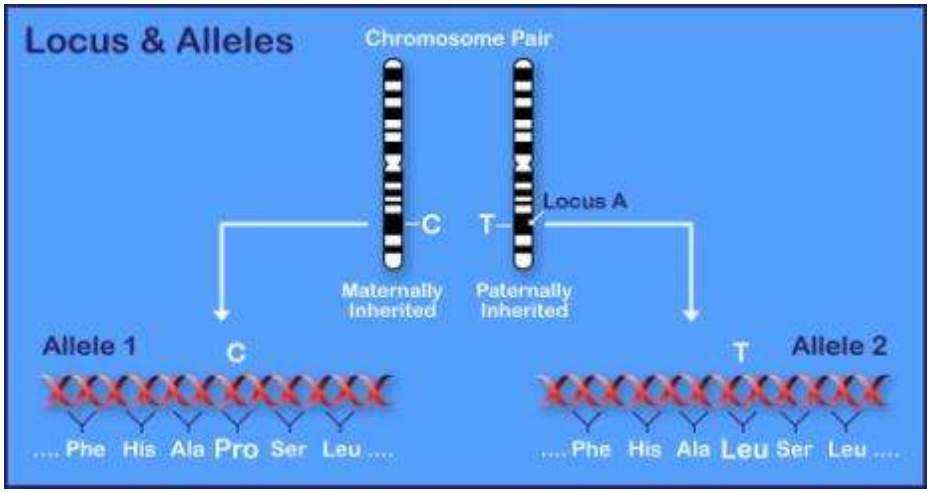
Gene structure: important to understand the rôle of mutations





Important definitions

- **LOCUS:** Precise chromosomal region
- **GENE:** part of a locus containing the genetic information (coding)
- **ALLELE**: Variations of the information in a gene.



- **HOMOZYGOTE**: identical alleles on the 2 chromosomes
- **HETEROZYGOTE**: different alleles on each chromosome
- **HEMIZYGOTE:** in males carrying variants on chromosome X

The gene can exist in several alleles

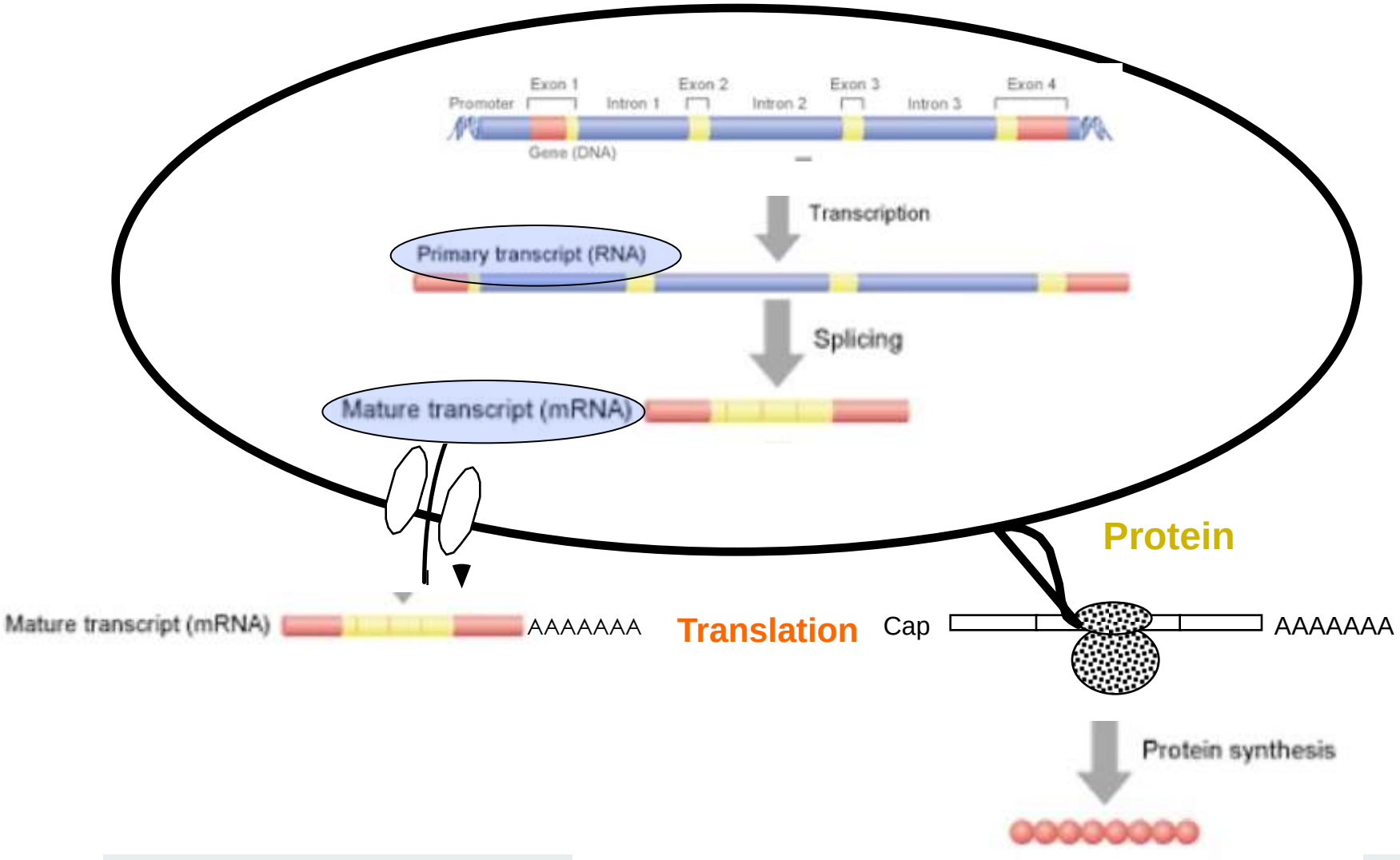
→ « **Wild type allele** »: The most frequent allele in the healthy population

→ **« Mutant Allele»:**

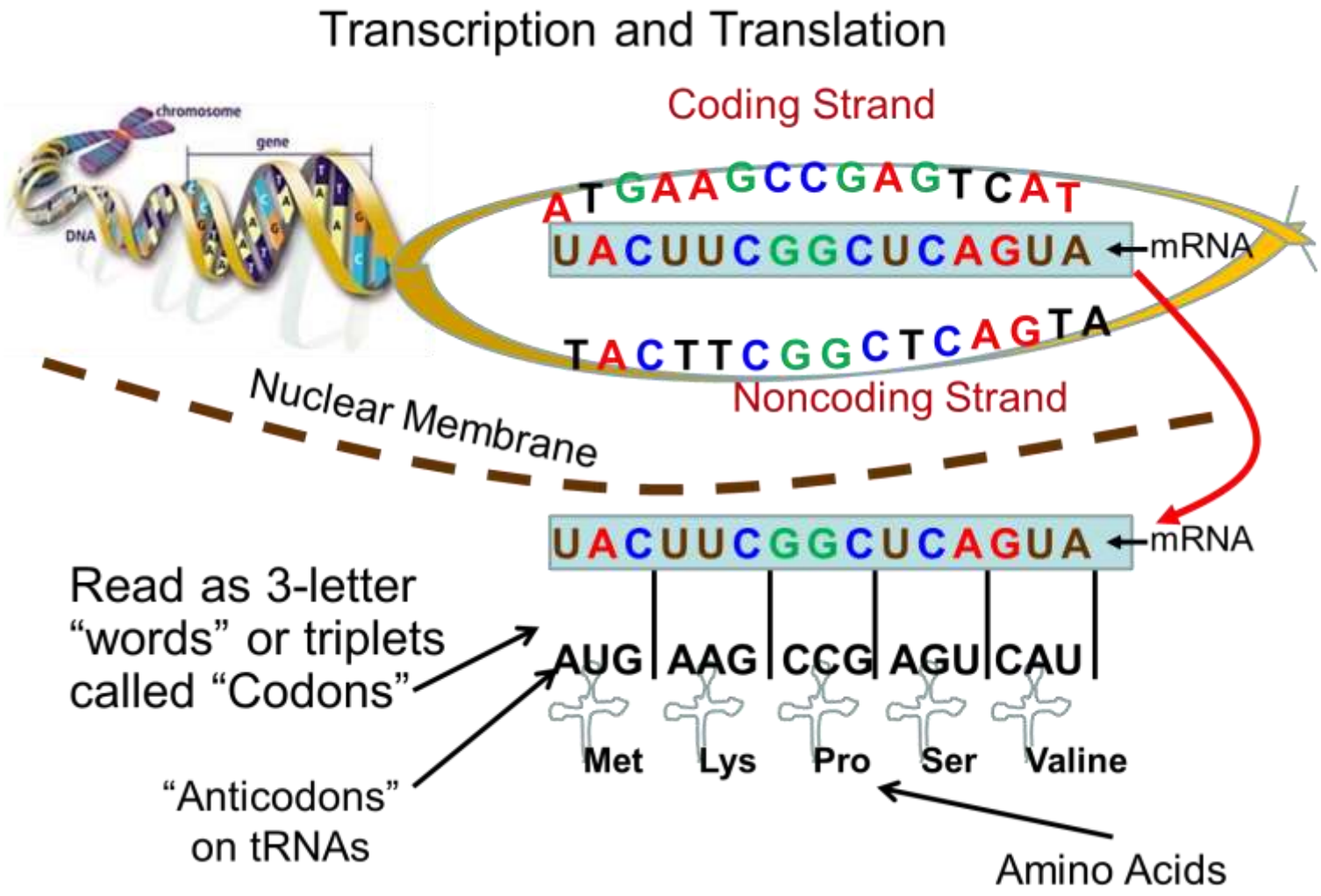
- **« polymorphism »: Alternative alleles from the wild type** but not associated with a disease
- **« Morbid allele»: rare disease causing alleles not found in the healthy population**

From genes to proteins

the basis of genetic diseases



The reading frame



What are the different ways in which a genetic condition can be inherited?

- Mendelian Inheritance (Single gene disorders of the nuclear genome)
  - 1. autosomal dominant (AD)
  - 2. autosomal recessive (AR)
  - 3. X-linked inheritance (XL)

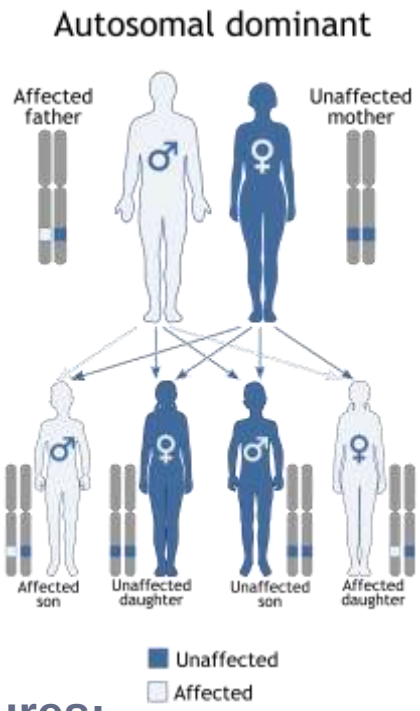
- Non-Mendelian Inheritance (the rest)
  - 4. Mitochondrial (matralineal)
  - 5. Polygenic
  - (6. Chromosomal)

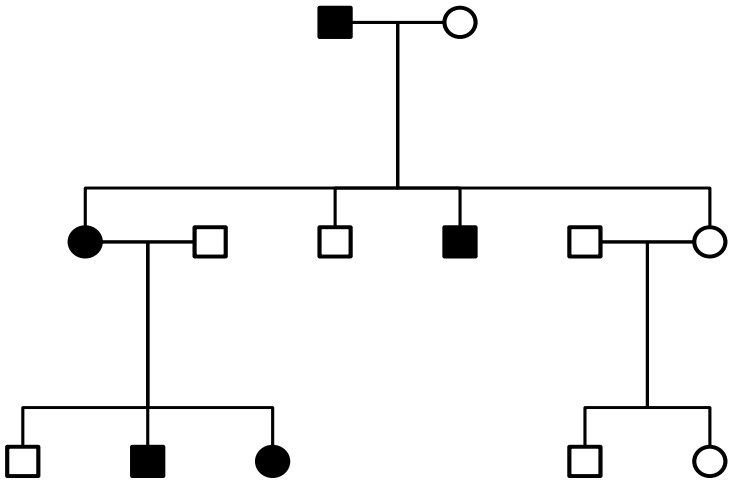
How to determine the possible/likely mode of inheritance in a family ??

- Family history
- Drawing the PEDIGREE

Autosomal dominant (AD) inheritance

A mutation in one allele of a gene (located on an autosome) is sufficient to cause the disease





Main features:

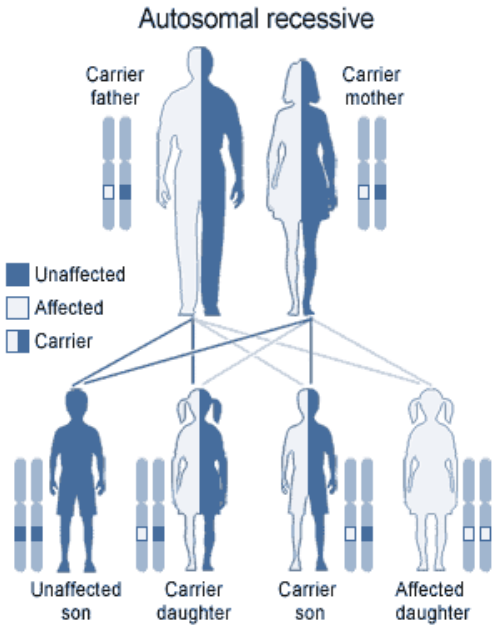
1. Males and females are equally likely to be affected.
2. Each child of an affected person has a 50% risk of inheriting the gene mutation.
3. The condition is seen in sequential generations, affecting 50% of individuals in each generation.
4. AD diseases are characterized by variable expressivity, incomplete penetrance

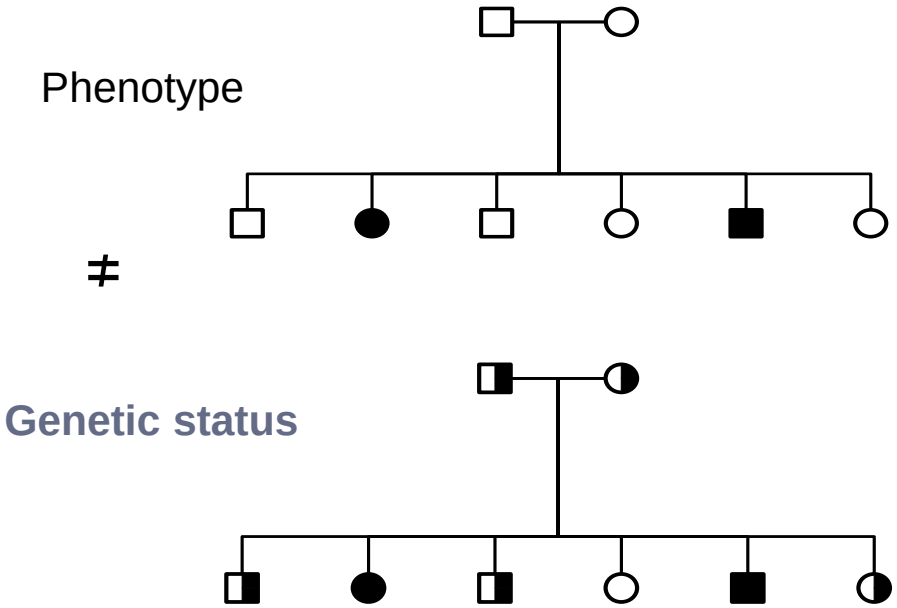
Pitfalls in Recognizing Autosomal Dominant Inheritance

- **Incomplete penetrance.** Some people who have the gene mutation do not express the clinical manifestations (phenotype ≠ genotype).
- **Variable expressivity.** A same mutation may have variable clinical manifestations: the disorder may range from mild to severe phenotype
- ***De Novo* mutation.** An affected person may be the first person in the family with the disease, due to a mutation arising for the first time in sperm, egg, or embryo
- **Germline mosaicism.** A new mutation may arise in testis or ovary. The result is an unaffected parent transmitting the disease to two or more children. The risk of transmission is linked to the degree of mosaicism.

Autosomal recessive (AR) inheritance

A mutation in the two alleles of a gene (located on an autosome) is needed to cause disease



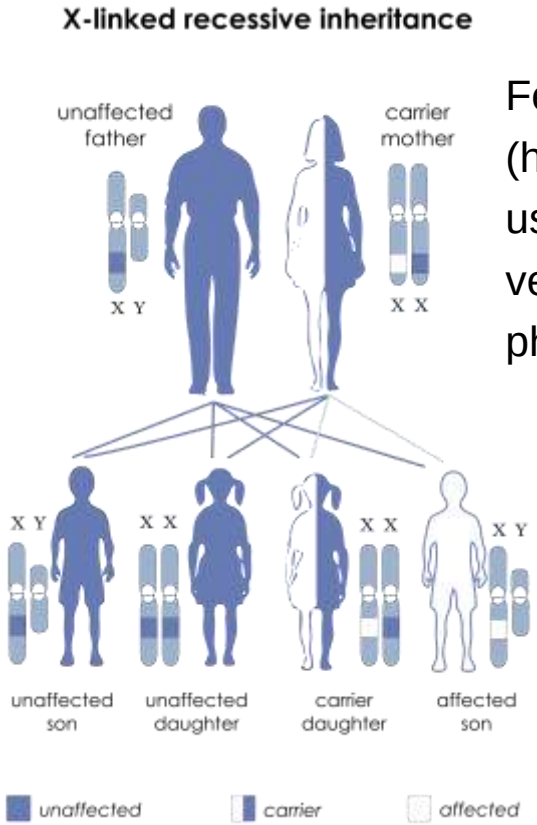


Main features:

1. No difference whether you are male or female
2. Both parents must carry gene mutation(s) in the same gene for children to be at risk
3. In that situation each child has a 25% risk of inheriting the disease
4. Consanguineous parents have an increased chance of children with AR disease

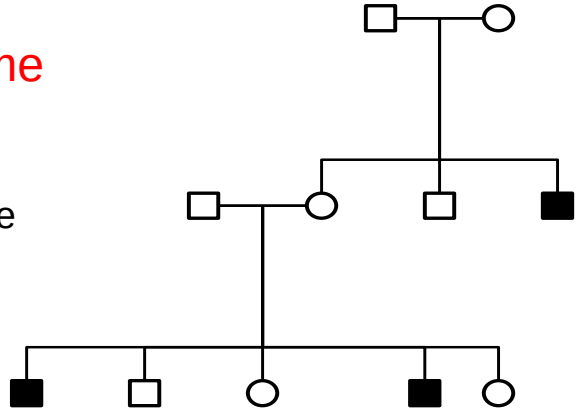
X-linked inheritance

The disease gene is located on the X chromosome

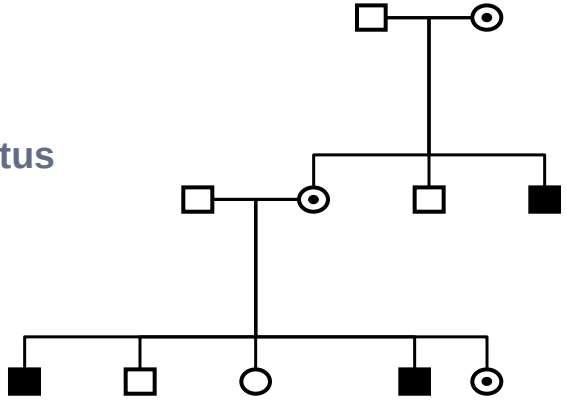


Female carriers (heterozygous) usually show no (or very mild) phenotype

Phenotype



Genetic status



Main features:

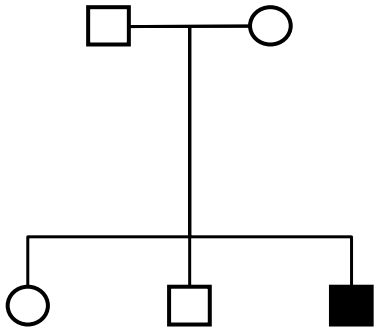
1. Affected males linked by unaffected (or mildly affected) females
2. No male to male transmission
3. Affected boys are HEMIZYGOTES

De novo mutations

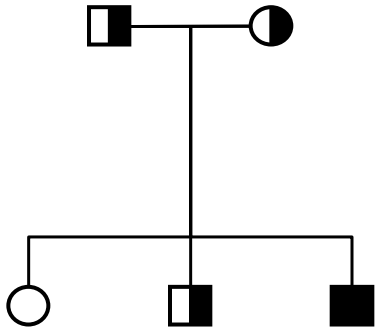
A mutation arises ‘new’ in a family (parents test negative for mutation).

- Non-paternity/non-maternity is another possible explanation
- The “*De Novo*” mutation is dominant, but typically mimics autosomal recessive inheritance – can be confusing

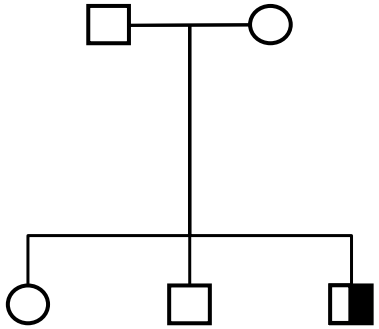
Phenotype



Genotype: AR disease



Genotype: de novo AD

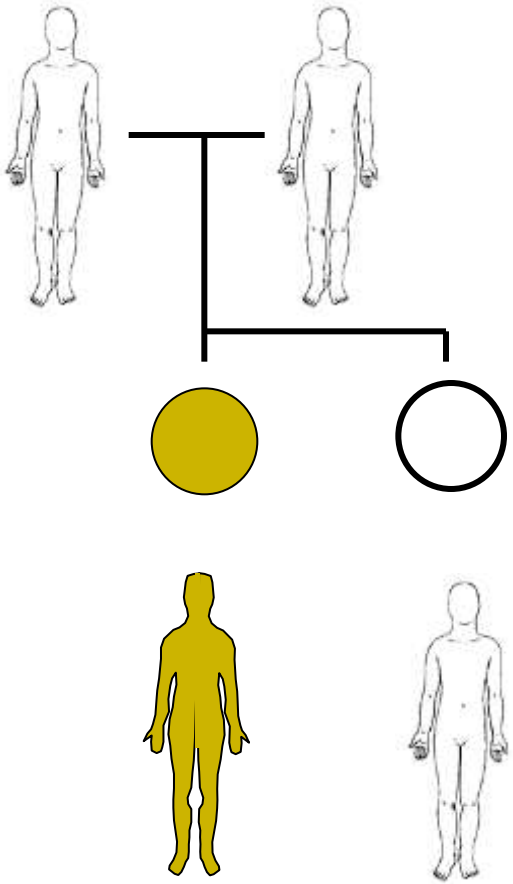


- Rates of neo mutations depend on the gene
- Severe phenotype

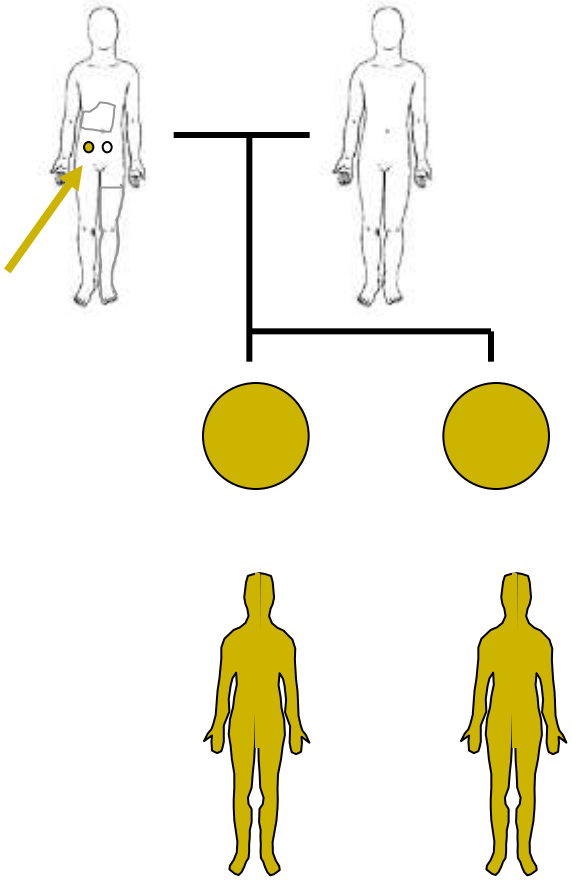
ACTA1 – the majority of mutations
LMNA – 40%
UCMD (COLVI rel.myopathies)- 60%
DMD – 33%
FSHD – 20%

De novo mutation vs gonadal mosaicism

de novo in child



Gonadal mosaicism



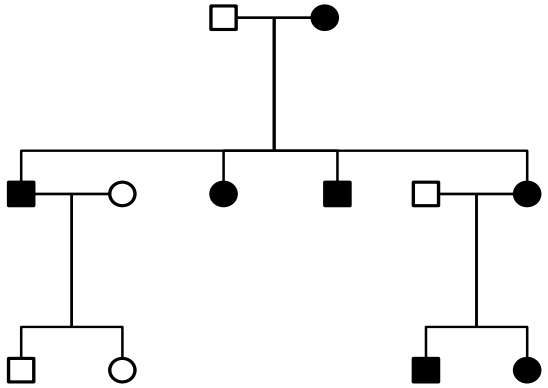
Very difficult to demonstrate. Can never exclude gonadal mosaicism

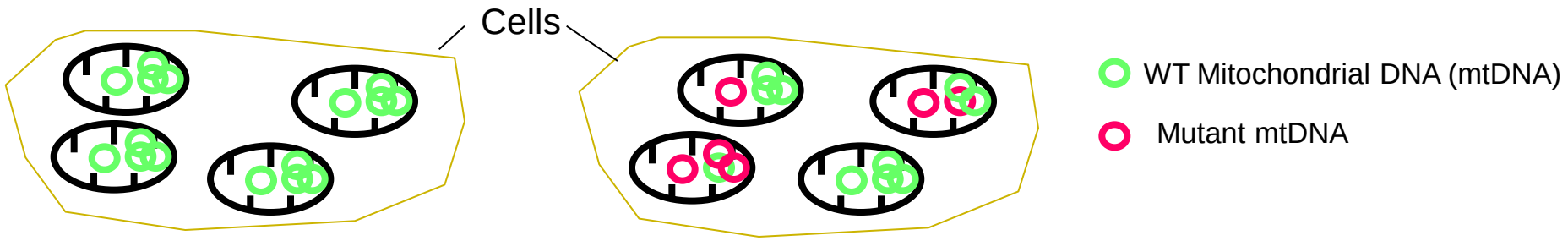
Mitochondrial inheritance

Mutations in the mitochondrial genome follow matrilineal inheritance

Main features:

1. Affected mothers pass on the mutation to all children (though mutant loads may vary)
2. Males cannot pass on the disorder
3. Variable expressivity is common between patients and between tissues





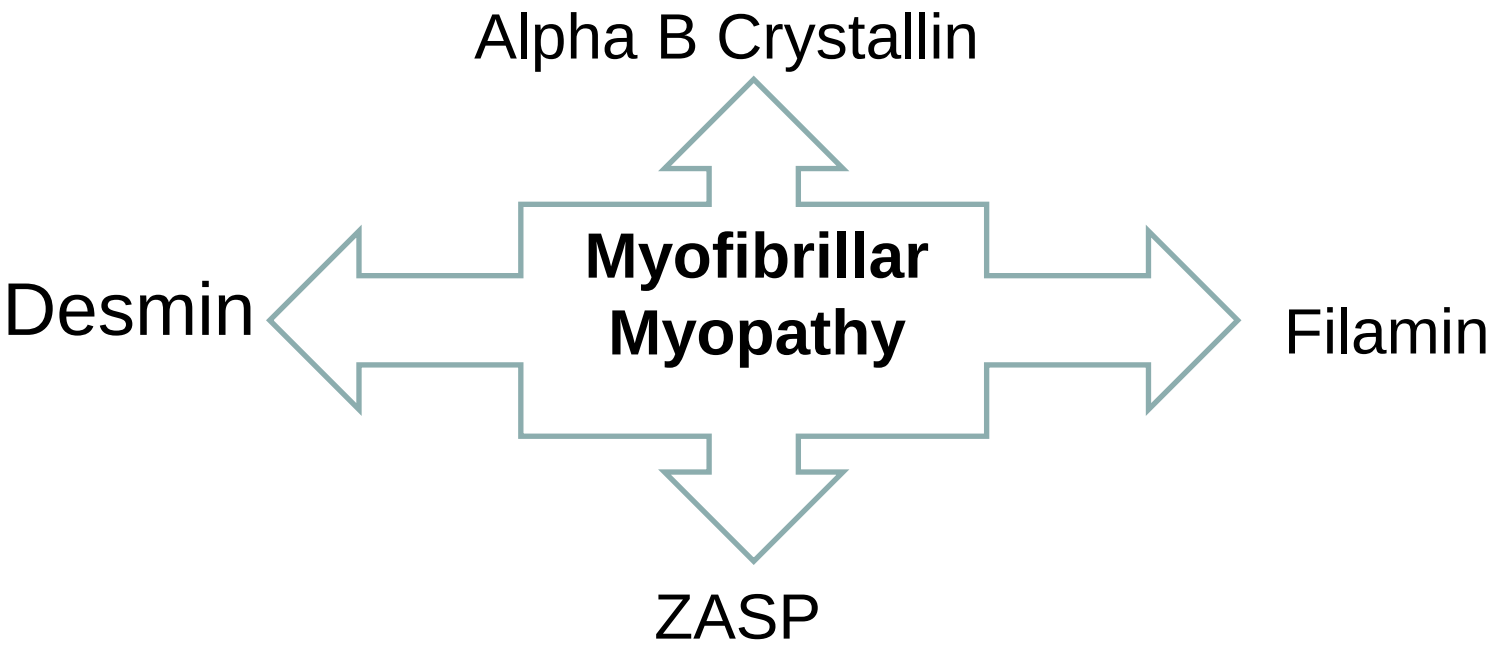
Homoplasmy: only one molecule of DNA (wt or mutated).

Heteroplasmy: mutant mtDNA molecules coexist with normal mtDNA from 0 to 100% load.

The mutant load threshold that will cause tissue dysfunction varies between different tissues. A source of variability in disease effects, severity … pb counselling

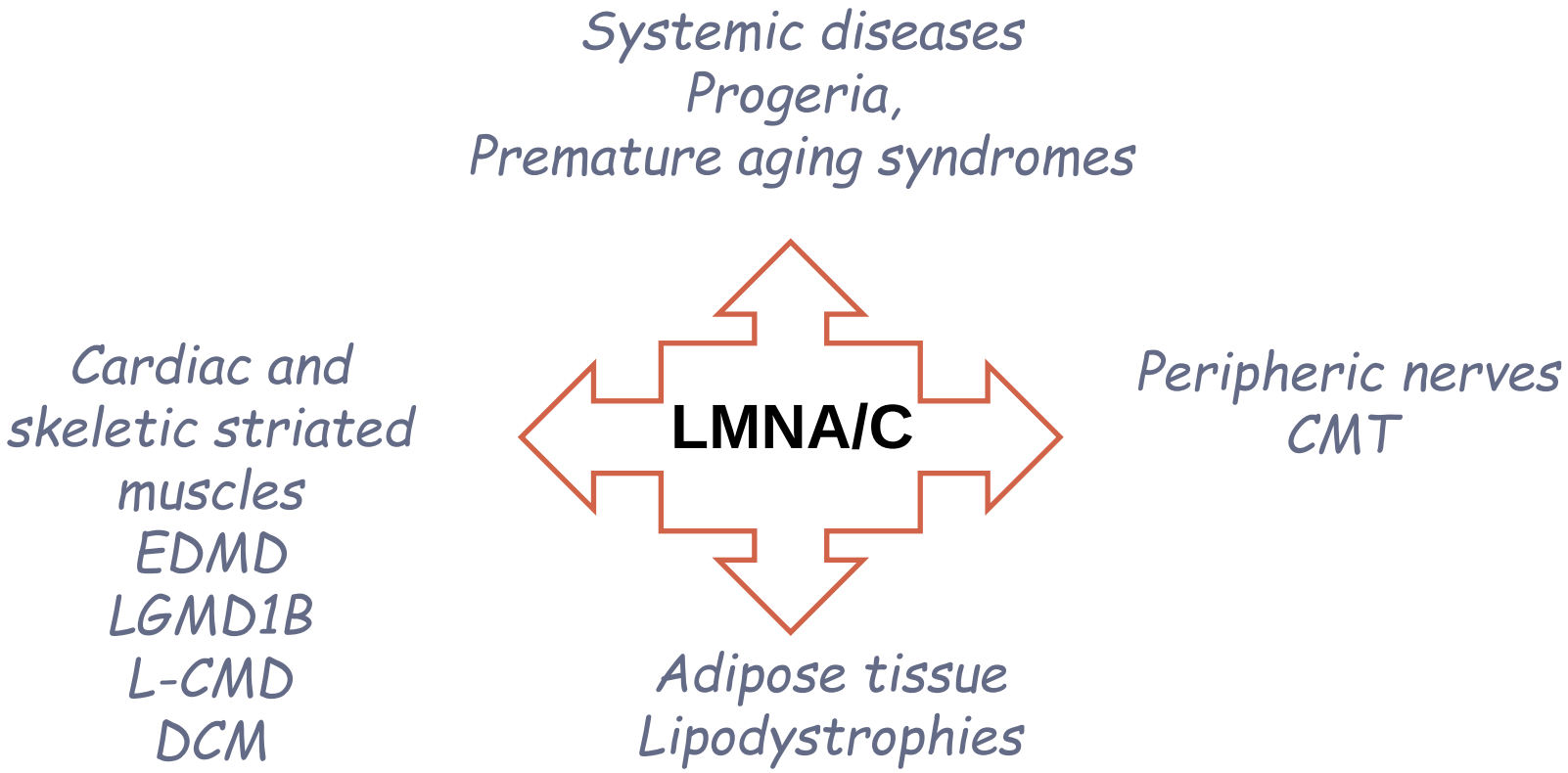
Genetic heterogeneity

The same monogenic disease can be due to mutations in different genes



Phenotypic heterogeneity

The same gene can cause different diseases

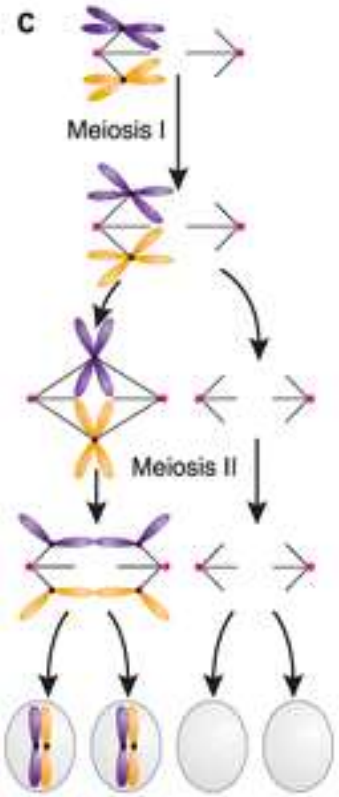


The genetic diseases : three groups

somatique or germ line cell

Genomic Mutations

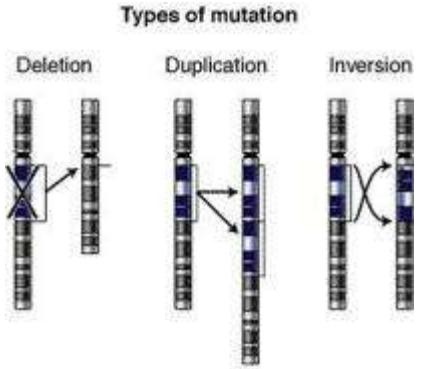
abnomalities in the number of chromosomes



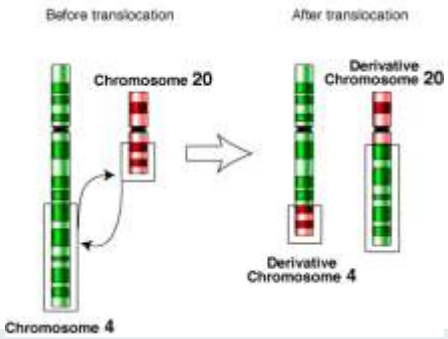
Chromosomic Mutations

Chromosomal rearangments

On a single Chromosome



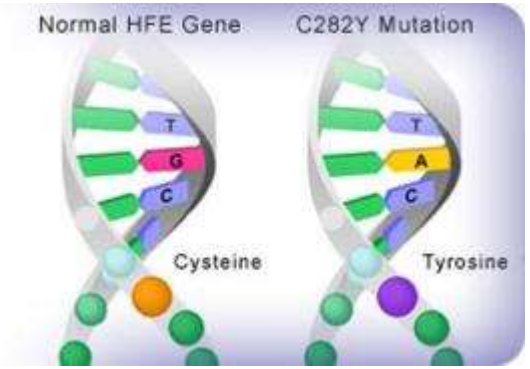
Between Chromosomes



Genic Mutations

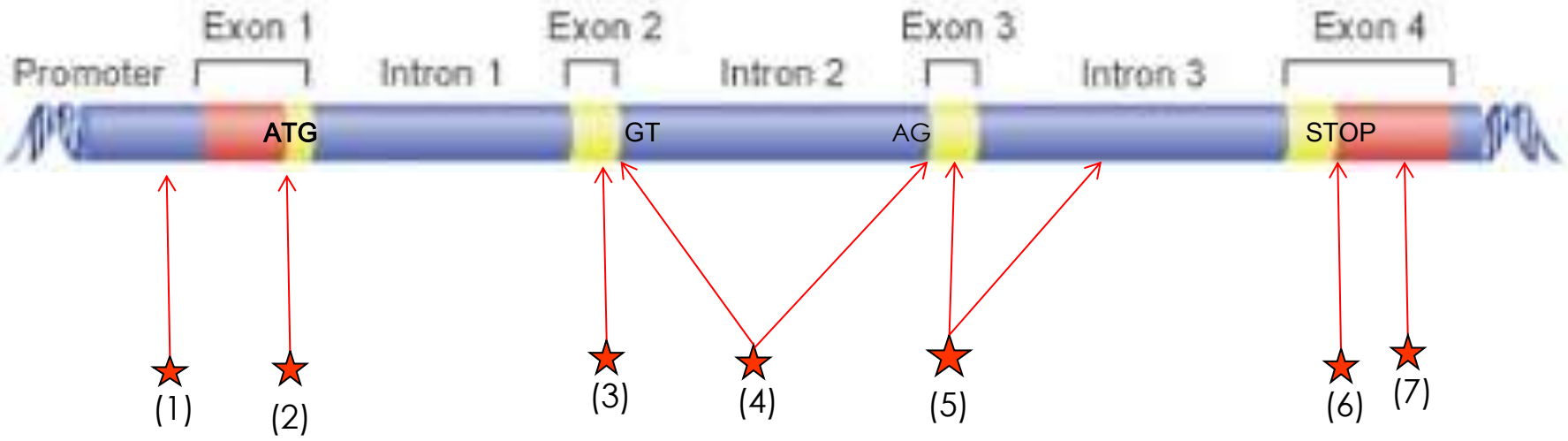
Abnormality in a gene at nucleotide level :

exonic, intronic



Genic Mutations : allelic heterogeneity

Mutations can be present all along the gene sequence



(1) Quantitative expression of the gene

(2) Abnormality in the initiation of traduction

(3) Substitution or insertion/deletion in exonic sequences (disease causing or polymorphisms)

(4) Splicing mutations (donor or acceptor sites, exons skipping, intronic sequences retention)

(5) Creation of cryptic splice sites in exonic or intronic regions

(6) Abolition of stop codon

(7) Abnormality in the mRNA polyadenylation -> stability of mRNA

Genetic diseases are caused by defects in the sequence of genes

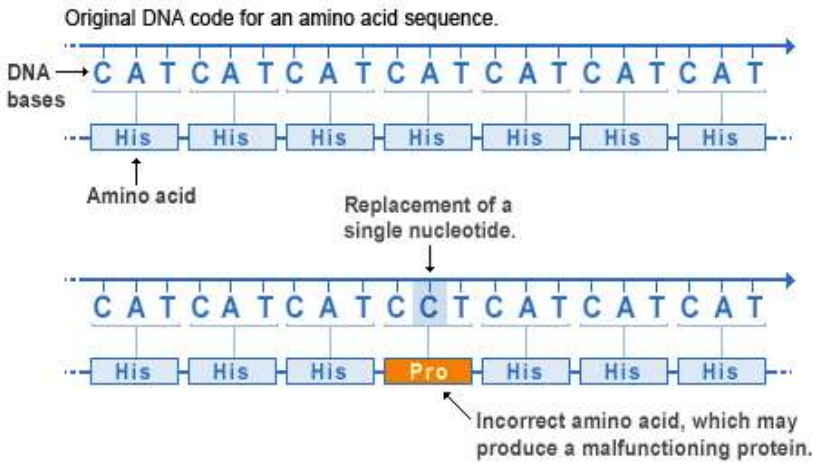
- in **one** gene : **mono**genic disease
- in **several** genes : **multi**genic (polygenic) diseases

- **Types of genic mutations:**
  - **POINT MUTATIONS**
    - Missense (polymorphism or disease causing)
    - Non sens (often disease causing)
    - Silent (often polymorphism)
  - **INSERTIONS OR DELETIONS**
    - In frame (polymorphism or disease causing)
    - Disrupting the reading frame (often disease causing)

Point mutations

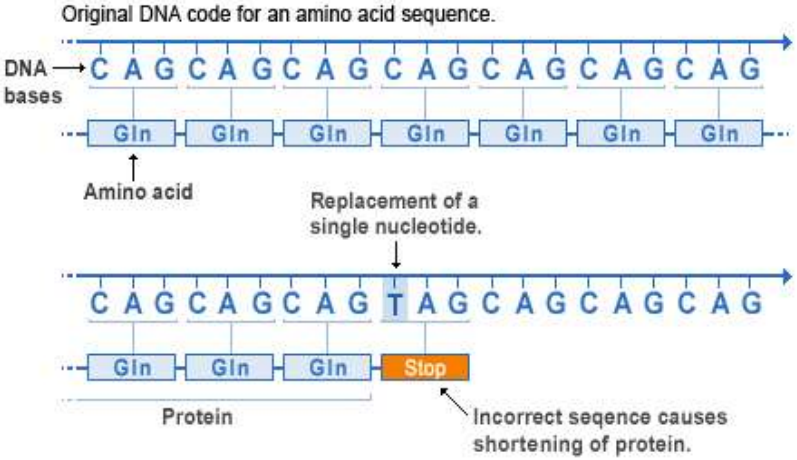
- Missense mutation : This type of mutation is a change in one DNA base pair that results in the substitution of one amino acid for another in the protein made by a gene.
- Nonsense mutation : A nonsense mutation a change in one DNA base pair. The altered DNA sequence prematurely signals the cell to stop building a protein.
  - This type of mutation results in a shortened protein or no protein

Missense mutation



U.S. National Library of Medicine

Nonsense mutation

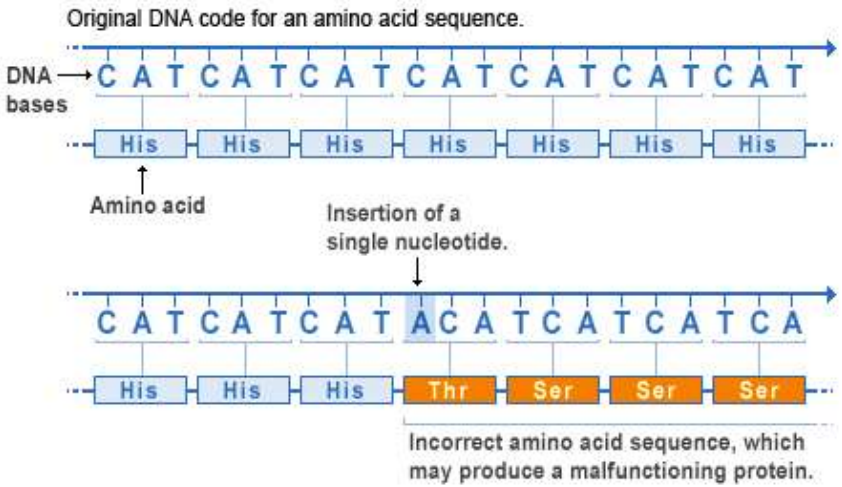


U.S. National Library of Medicine

Types of genic mutations: InDel

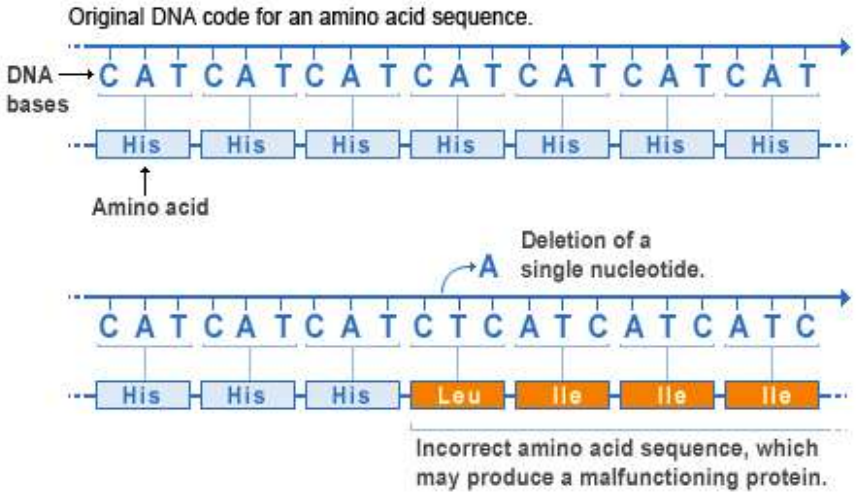
- Insertion/ Duplication : It changes the number of DNA bases in a gene by adding nucleotides in DNA.
- Deletion : A deletion changes the number of DNA bases by removing nucleotides in DNA sequence. Small deletions may remove one or a few base pairs within a gene, while larger deletions can remove an entire gene or several neighboring genes.

Insertion mutation



U.S. National Library of Medicine

Deletion mutation

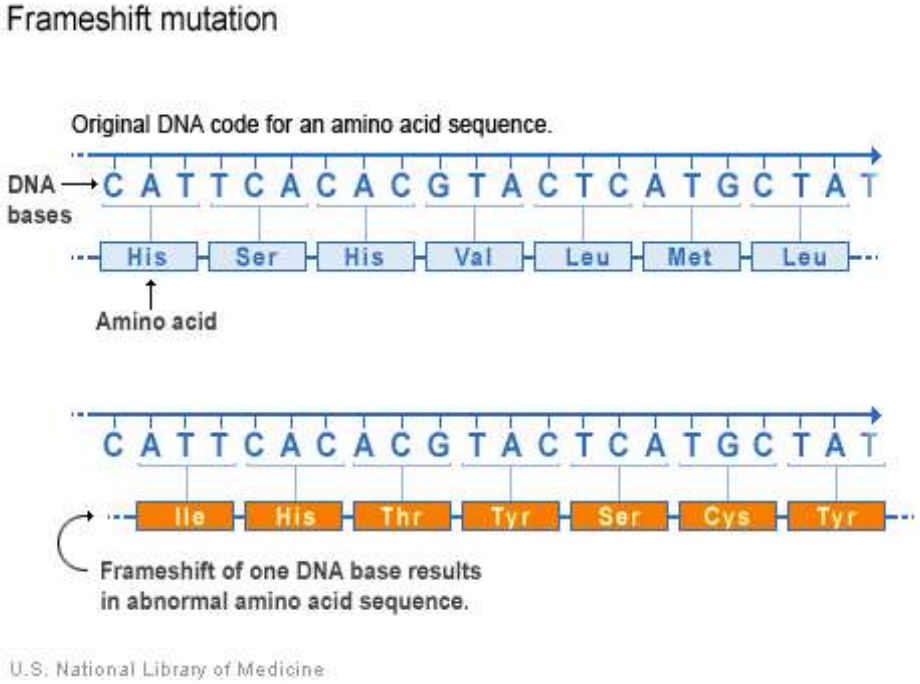


U.S. National Library of Medicine

Frame shift mutations

- Frameshift mutation : This type of mutation occurs when the addition or loss of DNA bases changes a gene's reading frame.

A frameshift mutation shifts the grouping of the 3 bases of the reading frame and changes the code for amino acids. The resulting protein is usually nonfunctional. Insertions, deletions, and duplications can all be frameshift mutations.



Small Deletions/Duplications

‘Reading frame hypothesis’ Monaco et al., 1988

THE RED CAR WAS BIG AND HOT

In-frame deletion (3 Nt)

THE RED CAR WAS BIG AND HOT

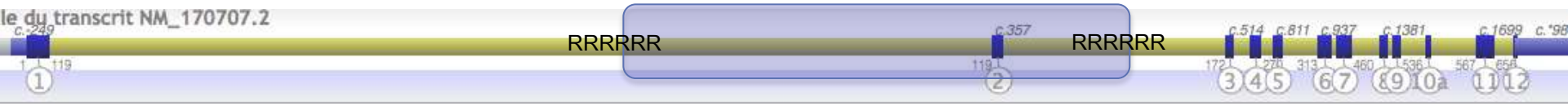
THE RED CAS BIG AND HOT

Out of frame deletion (2Nt)

THE RED CAR WAS BIG AND HOT

THE RED CAA SBI GAN DHO T

Chromosomal Microdeletion (CopyNumberVariation)



- In some patients, the disease causing mutation consist in an intragenic recombination during replication between intragenic repeated sequences on a single allele.
- The consequence often lead to one or several exon deletion (entire gene)
- The deletion can be in frame or out of frame depending on the deleted exons

The classical tools of molecular biology cannot detect these mutations. Only the remaining allele will be analysed

How to define the pathogenicity of a mutation found

- Does the clinical Story fit?
- Segregation in family,
  - Finding the mutation in other affected patients (nor always possible in recessive or rare diseases)
  - If absent in an affected relative: Good at excluding pathogenicity.
  - de novo changes - good evidence of pathogenicity
- Knowledge on the protein
  - Structure (functional domains, catalytic sites…)
  - Rôle (structure, enzyme…)
  - Conservation among species and evolution
- Bioinformatics analysis
  - In silici prediction sites (Polyphen, SIFT, Mutation Tester, GVGD)
  - Frequency of the variants in the databases (ExAC, EVS, 1000g). Must compare the same population as patient (age, gender, ethnicity)

This part of the interpretation is the job of the molecular geneticist

DNA Variation:
pathogenic vs polymorphism

The question is not to find « a variant » in the gene but is to link the variation with the disease

- A synonymous substitution is not always a polymorphism
- A stop codon is not always disease causing
- Missense changes can also account for non pathogenic variation in humans

Important to get right

Mistaking a polymorphism for a mutation could cause

1. Patient: wrong diagnosis and prognosis, treatment
2. Wrong diagnosis in other family members….
3. Incorrect genetic counselling: prenatal diagnosis,
4. It Stops you thinking for other diagnoses

Different type of genic mutations -> different mechanisms and consequences

Missense Mutations

(substitution of an amino acid for another one)

- Effects on the protein secondary structures
- Abolition of a catalytic enzyme site
- Acts as a poison polypeptide (dominant negative effect)
  - 50% Normal allele
  - 50% Mutant allele

Nonsense Mutations

(Stop, out of frame, splice and ins/del mutations)

- Effects on the stability of mRNA which are eliminated via the mRNA mediated decay pathway (NMD)
- Acts as an haploinsufficiency mechanism
  - 0% mutant allele
  - 50% nomal allele

If homozygous, corresponds to a knock out of the gene

Identification of a genetic defect
Interest of diagnosis

To confirm a clinical diagnosis

1. to orientate the follow up of the patient or his/her treatment
   Eg. - *RYR1* – malignant hyperthermia precautions
   - *LMNA* – cardiac monitoring essential

2. to provide accurate genetic counselling
   - Enable genetic testing if requested by families

3. to understand the molecular basis of diseases – physiopathology and evolution

In the future

- Essential for future gene- or mutation-specific therapy

All genetic testing must be carefully guided by clinical assessment

A good clinical diagnosis

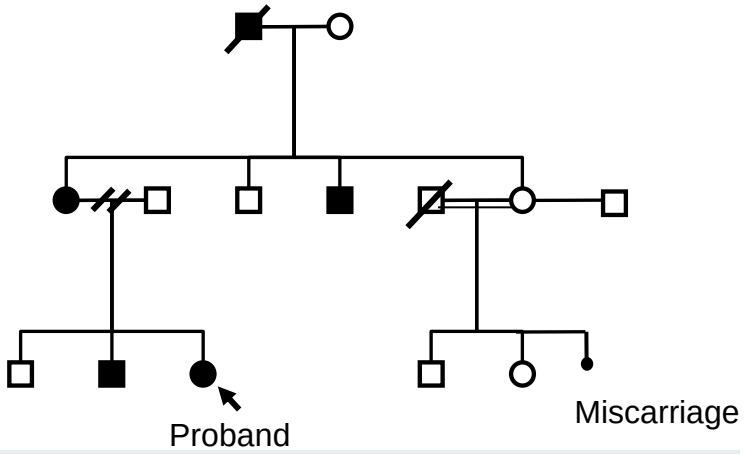
- clinical examination
- complementary examinations (CK, EKG, MRI, EMG…)
- careful analysis of the muscle biopsy (Morphology, IHC Staining)

Information from inheritance: Make pedigree trees

- Quick to draw, informative, helps thought processes
- A universal record of family structure
- Careful family history

Ask about

- Consanguinity
- Distant relatives with similar disease



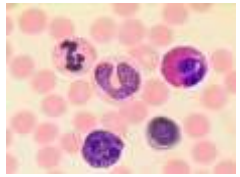
After the diagnosis… The genetic counselling

- Inform the Index case: Organized with a multi-disciplinary consultation (separate)
  - The neurologist
  - The geneticist
  - The psychologist
- The patient must inform its relative of the test result.
- Dominant diseases: difficult because of the delayed penetrance
  - Major relatives: can ask for presymptomatic diagnosis after a consultation with a geneticist.
  - Minor relatives: depending on the benefits of knowledge.
- Recessive disease:
  - Testing relatives is systematic
  - The genetic couselling is usefull to avoid a new affected child (PND)

Material for molecular biology

DNA: Standard gene sequencing, Southern blot….

- Blood samples/ **EGTA** sent at room temperature within 48 hours
  DNA is extracted from Lymphocytes
- Buccal cells, biopsies,…



mRNA: Sequencing coding regions of large gene,
Allow splicing analysis +++

- Muscle biopsy: Snap frozen
- Transport by dry ice – requires careful planning/care

Jellyfish of DNA molecules

Biopsy- derived cell cultures: Increasingly useful

- Skin biopsy for fibroblast culture (IHC, mRNA)**Collagen VI**
- Myoblast cell lines: specialist research centres

Find out what samples are needed before hand

- Collagen VI (fibroblast cell line)
- RYR1 – frozen muscle (for mRNA extraction)

What legislation for what test?

This discipline is framed by a rigorous legislation different from the one used for the other biological tests

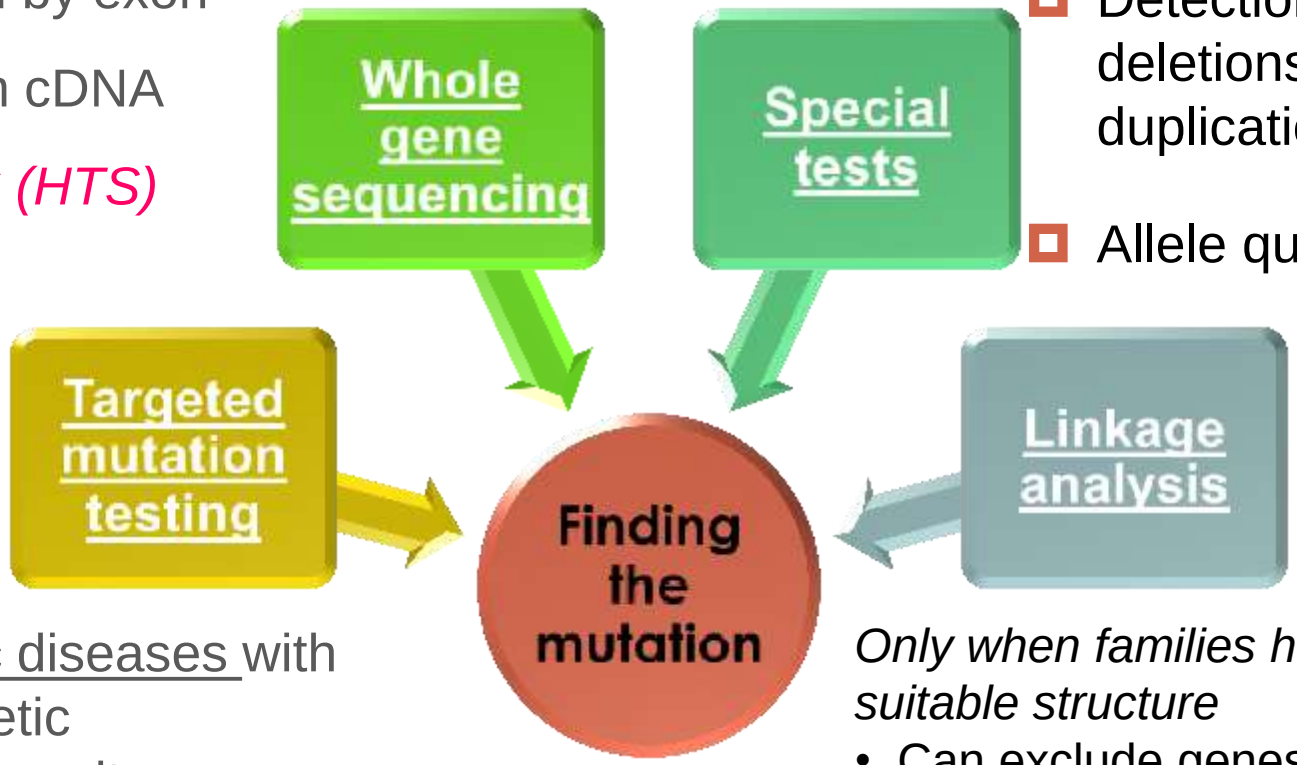
- **Diagnosis in an affected patient**: the **neurologist** can ask for the genetic test and inform the patient of the result.
- **Pre-symptomatic diagnosis**: **Geneticist**
  - In an unaffected relative in a genotyped family
    - Predict the future clinical status
    - Adaptation of the medical follow up
- **Prenatal diagnosis**: **Geneticist**
  - Multidisciplinary consultations (neurologist, geneticist, psychologist)

What forms must come with the sample?

- Informed written and **SIGNED** consent
  - Signed by the patient **and** the physician
  - Signed by the 2 parents in minors patients
- Detail of clinical examination
- Detail of paraclinical informations
- Precision in the request (what disease, which gene…)
  - Index case or relative
  - Genetic analyses are highly specialized tests and each lab has developed a panel of test in a group of diseases
- Samples and forms must be send in appropriate authorized lab in which biologists and technicians have an accreditation

Diagnosis strategies

- Exon by exon
- From cDNA
- *NGS (HTS)*



- Detection of large deletions or duplications (MLPA)
- Allele quantification

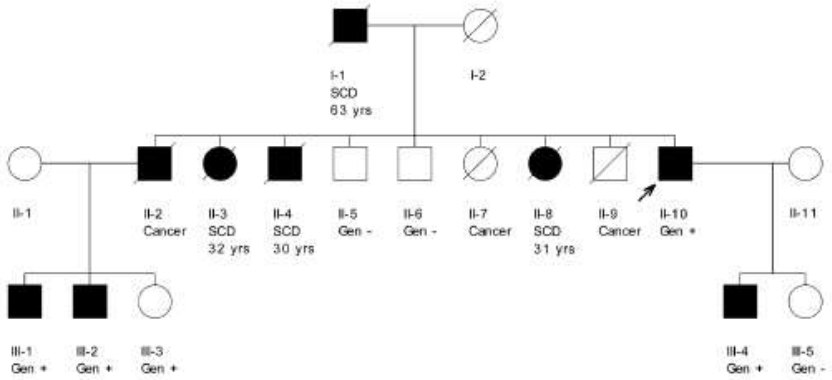
- specific diseases with no genetic heterogeneity

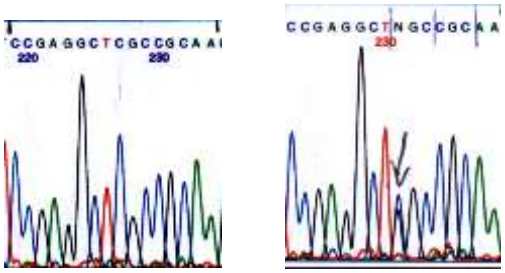
Only when families have a suitable structure

- Can exclude genes without sequencing
- Central to identifying new genetic causes

Usual cases

Dominant transmission

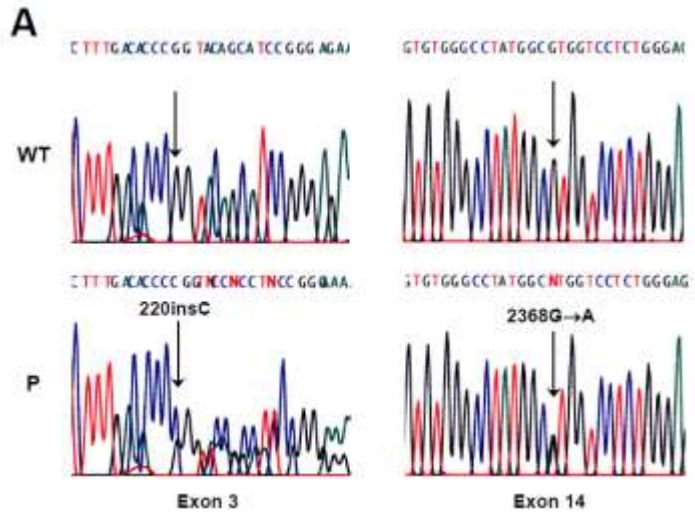


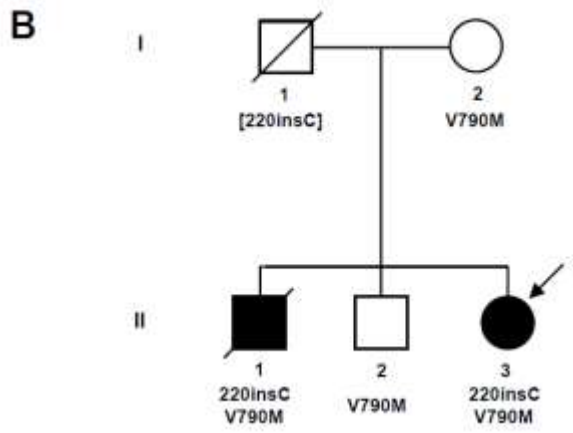


Heterozygous
CGC: Arg
GGC:Asp

Co-segregation of the mutation with the disease

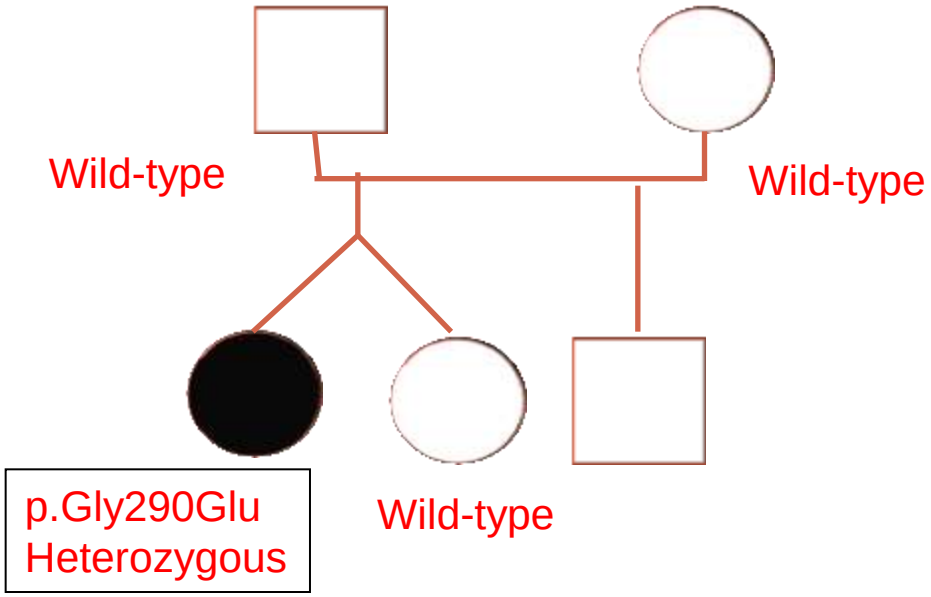
Recessive transmission



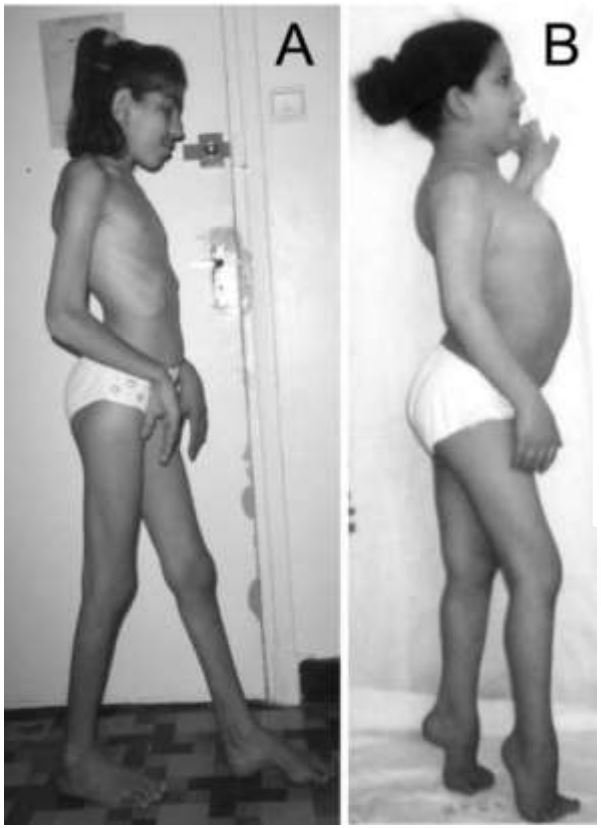


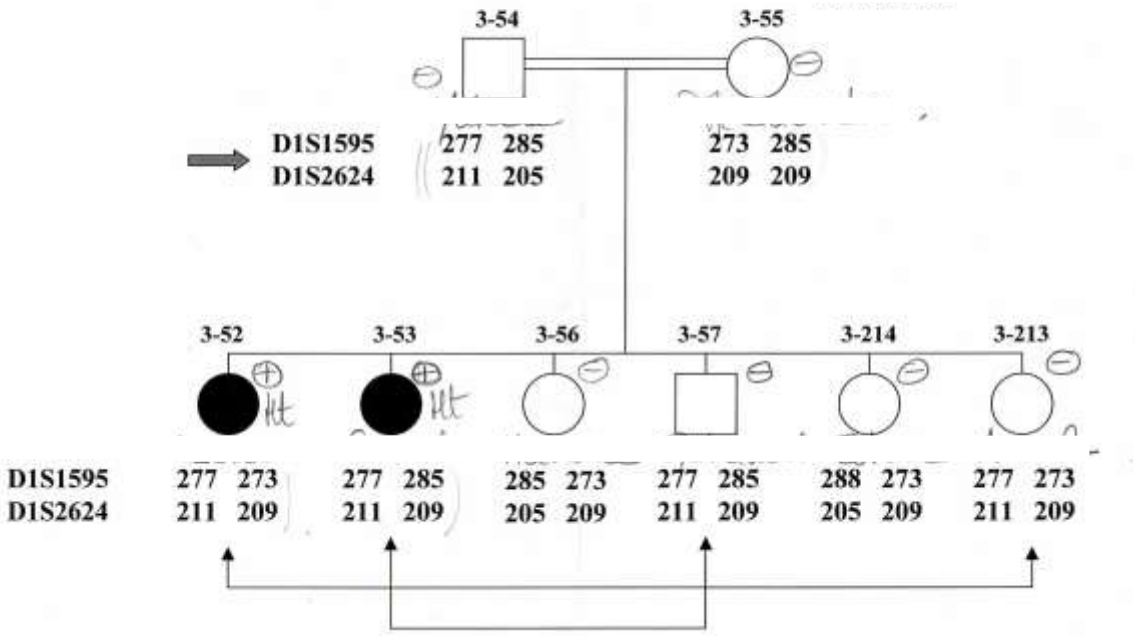
Neomutation (*de novo*)

UCMD: COL6 gene
* autosomal recessive
* autosomal dominant (60 % neomutations)



Germ line Mosaicism





LMNA/C, p.Arg527Pro

A technological revolution
Next generation Sequencing (NGS)

High throughput sequencing of a very large amount of sequences

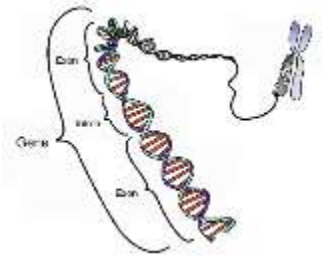
- Sanger sequencing
  - Genes are sequenced individually on a sequential approach
  - The sequencing of a gene is done after amplification of each exons separately (500 bp)
  - Long and Tedious
  - Not exhaustive
  - Expensive

- NGS technology
  - All the genome of an individual can be sequenced in a single run
  - The exome (all coding sequences) of several patients can be sequenced together (60Mb)
  - Several targeted genes of numerous patients can be sequenced in a run (20Gb)

NGS : What is the problem to solve??

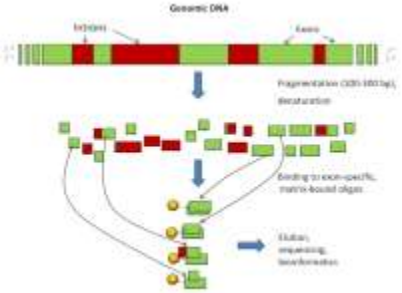
- **Genome** : Lot of sequences including 98% non coding
- **Exome**: 30 000 genes including non coding (but 2% missing!!)
- **Clinicome**: 1000 genes known to be involved in diseases
- **Targetted genes:** **First step of selecting genes and sequences of interest**

Gene



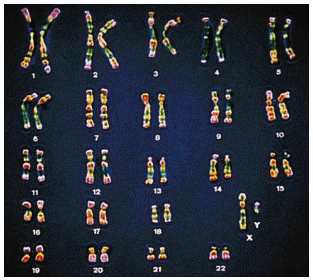
- 1Kb à 100Kb
- **1 to 100 exons**
- **(max 360 exons)**

Exome



- 30 megabases (Mb)
- **180 000 exons**
- **1% of genome**

Genome



- 3500 Mb (**3,5 milliards of nucleotides) – haploïde**
- **27000 à 30000 genes**

PGM: 10 Mb up to 1 Gb
Ion Proton: 10Gb >120Gb

MySeq: 120 Mb –7.5 Gb
HiSeq: 600 Gb

Nothing is perfect

- The library preparation is tricky
- The bio-informatic analysis needs competences
- Large Amount of sequences → numerous variants: distinguish disease causing from polymorphisms
- Although this technology is a real progress
  - Not suitable for analysis of relatives within a family
  - Not suitable for urgent diagnosis
  - Essential to correlate the variants found with the phenotype

Internet ressources

PubMed® http://www.ncbi.nlm.nih.gov/pubmed/

OMIM Online Mendelian Inheritance in Man  http://www.ncbi.nlm.nih.gov/omim

orphanet http://www.orpha.net/consor/cgi-bin/index.php

GENEReviews

http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=gene

GENETests http://www.ncbi.nlm.nih.gov/sites/GeneTests/

e!Ensembl http://www.ensembl.org/index.html