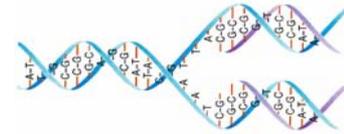
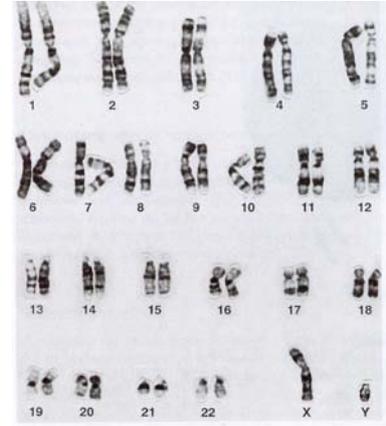
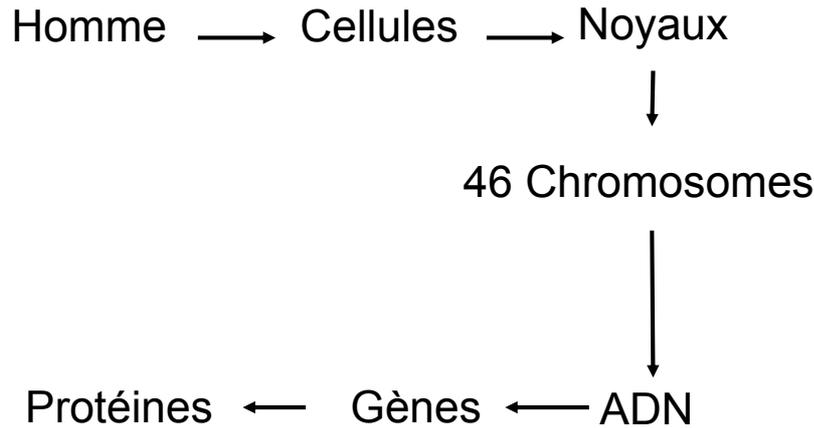
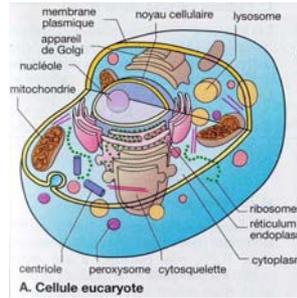


Génome humain
Polymorphismes
Marqueurs génétiques

Dr Chrystelle COLAS
Département de Génétique
GH Pitié Salpêtrière

Génome humain

Génome Humain (1)



Génome Humain (2)

➤ **1 copie de l'ADN situé dans une cellule humaine**

➤ **46 chromosomes**

▪ **22 paires d'autosomes homologues**

▪ **2 chromosomes sexuels XY**

➤ **4 bases**

▪ **Adénine / Thymine**

▪ **Cytosine/ Guanine**

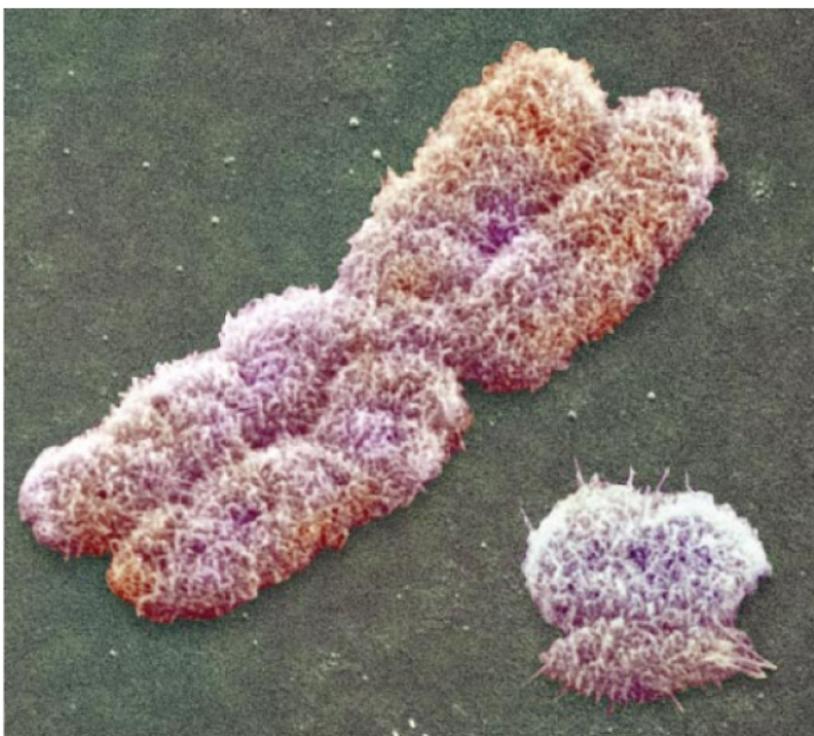
➤ **Taille : 3.2 milliard de pb**

Il faudrait environ 9,5 ans pour lire à haute voix (sans s'arrêter) les plus de trois milliards de paires de bases dans la séquence du génome d'une personne ». [Source: Human Genome Projects Information].

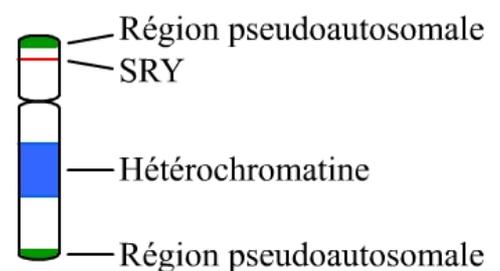
Génome Humain (3)

- >30 000 gènes (<1.5% du génome !)
 - Portion d'ADN codant pour une protéine
 - Répartis sur les chromosomes
 - Régions plus ou moins riches en gènes
- Il existe beaucoup de zones non codantes !
 - >40% séquences répétées

Toutes les séquences d'ADN existent en double exemplaire car les chromosomes vont par paires homologues



**Exception
Hommes : XY**



Régions pseudo-autosomales :

- 5% du chromosome Y
- Permet l'appariement à la méiose
- Contient 29 gènes identiques à l'X

Autres régions

- 95% du chromosome Y
- Contient 78 gènes dont certains ont un homologue sur l'X et d'autres non

Polymorphismes

Définitions

- ***Polymorphisme:***
 - Toute variation de séquence génomique entraînant l'existence au même locus d'au moins 2 formes différentes de la séquence, appelés allèles, dans la population
 - Fréquence au moins = à 1%

- ***Locus:*** emplacement d'un segment d'ADN sur un chromosome

- ***Types de polymorphisme:***
 - Substitution d'un nucléotide (**SNP**: single nucleotide polym)
 - Polymorphisme de répétition (VNTR, CA repeat...)
 - Insertion/délétion ou inversion

SNP

- **SNP: single nucleotide polymorphism**
 - Substitution d'un nucléotide par un autre
 - Stable
 - Très abondants dans le génome
 - **1SNP / 100 à 1000 pb selon les régions**
 - Répartis non uniformément

.....atcgttccgtaacggtgctaacgtgatcgtgac.....
.....atcgttccgtaacggtgataacgtgatcgtgac.....

Pour un locus : 2 allèles possibles = biallélique

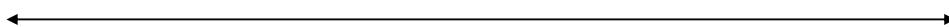
Diversité des séquences humaines

Diversité nucléotidique π :

Nombre de différences, rapporté à la longueur de la séquence examinée, entre 2 séquences prises au hasard dans la population

AGTCCTA.....GGAAACTG.....TTTACGT....

AGT**G**CTA.....GG**T**AACTG.....TT**G**ACGT....



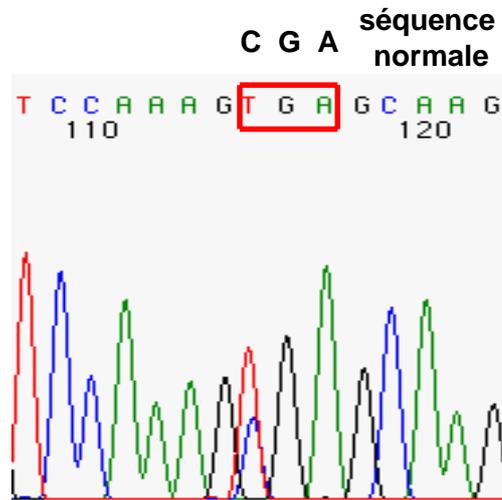
7500 bp

$$\pi = 3/7500 = 1/2500$$

Détection des SNPs (1)

➤ Séquençage direct

- Après PCR de la région d'intérêt
- Nécessite de connaître la séquence normale mais pas nécessairement le SNP



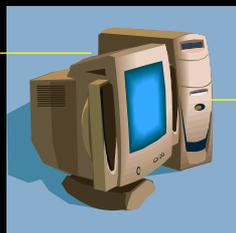
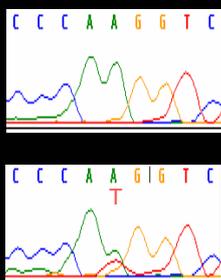
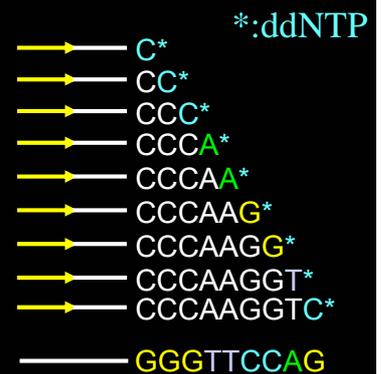
Séquençage direct sur produit de PCR



PCR



Produit amplifié par PCR



Laser
Caméra CCD



Electrophorèse capillaire

Détection des SNPs (2)

- **RFLP (Restriction Fragment Length Polymorphism)**
 - Si le SNP modifie un site de restriction (création ou abolition)
 - Digestion enzymatique
 - Southern blot et révélation par sonde spécifique
 - Coupure ou non = différence de taille des fragments

- **Autres techniques de recherche de SNP connus**
 - ASO
 - Extension d'amorces

- **A plus grande échelle pour SNP inconnus : Puces à ADN**

Localisations possibles des SNP

Situation:

- Dans une zone codante
 - Synonyme = silencieuse car redondance code génétique
 - Non synonyme = changements d'acide aminé

- Dans une zone non codante

1 ^{er} nucléotide (en 5')	2 ^e nucléotide				3 ^e nucléotide (en 3')
	U	C	A	G	
U	Phe:F	Ser:S	Tyr:Y	Cys:C	U
	Phe:F	Ser:S	Tyr:Y	Cys:C	C
	Leu:L	Ser:S	STOP	STOP	A
	Leu:L	Ser:S	STOP	Trp:W	G
C	Leu:L	Pro:P	His:H	Arg:R	U
	Leu:L	Pro:P	His:H	Arg:R	C
	Leu:L	Pro:P	Gln:Q	Arg:R	A
	Leu:L	Pro:P	Gln:Q	Arg:R	G
A	Ile:I	Thr:T	Asn:N	Ser:S	U
	Ile:I	Thr:T	Asn:N	Ser:S	C
	Ile:I	Thr:T	Lys:K	Arg:R	A
	Met:M	Thr:T	Lys:K	Arg:R	G
G	Val:V	Ala:A	Asp:D	Gly:G	U
	Val:V	Ala:A	Asp:D	Gly:G	C
	Val:V	Ala:A	Glu:E	Gly:G	A
	Val:V	Ala:A	Glu:E	Gly:G	G

Diversité de séquence des gènes humains

	Nb gènes	Nb SNPs	Pop	Codant	Syn	Non syn	Non codant
Halushka (1999)	75	726	Européens	4.5	9.0	3.1	5.4
			Africains	6.3	12.9	4.2	6.8
Cargill (1999)	106	560	Européens	5.6	9.8	3.7	4.9
Onishi (2000)	41	187	Japonais	2.4	1.2	1.2	4.3
Tiret (2001)	50	228	Européens	3.0	5.6	2.3	4.3

➤ Mutations des séquences non codantes plus fréquentes que sur les séquences codantes

➤ Dans séquences codantes: silencieuses plus fréquentes que non silencieuses

Polymorphismes de répétition

➤ Répétitions en tandem d'une séquence déterminée

- Le nombre de répétitions varie d'un sujet à l'autre :
- Tous les 6 kb (plus fréquent en subtélométrie)
- Sans conséquence fonctionnelle

➤ Mini-satellites

- VNTR : variable number of tandem repeats = 11 à 16pb

(CAGCATCAGGTT)_n

➤ Micro-satellites

- VNDR : variable number of dinucleotides repeats = 2 à 4pb
=STR : Short Tandem repeat

(CA)_n

*Pour un locus : très nombreux allèles possibles
= multiallélique*

Les minisatellites (1)

- Le polymorphisme d'un minisatellite est généré par le nombre variable de répétitions du motif qui lui est propre:
 - ✓ explorés le + souvent par la technique du southern blot ou PCR
 - ✓ révélé soit par une sonde monoclocus spécifique d'un VNTR donné;
 - ✓ soit par une sonde multilocus - séquence commune à différents minisatellites dans le génome (image complexe).

- Avec un taux d'hétérozygotie supérieur à 90 %

- Certains minisatellites sont associés à des maladies:
 - ✓ Diabète et minisatellite en 5' du gène de l'insuline
 - ✓ Troubles du comportement et VNTR de 48 bp du HLA DRD4

Exemple de minisatellite

```

5'---- CGCCTCGGCCTCCCAGAGTGCTGAGATTACAGGCGTGAACCACCATGCCTAGCCGTTAGCTCCCA
CTTATGAGTGAGAACAGGTGATGTTTGGTTTTCCATTCTGAGTTACTTTACCCAGAATTGTTGT
CTCCAATCTCATCGAGGTCCTGCGAATGCCAGTAAATTCATTCCTTTTTATGGCTAAGTAGTATT
CCATCGTATATATACATATACATATATATGTATACACACATATACATATATATGTATACACACAT
ATACATATATATGTATACACACATATACATATATATGTATACACACATATACATATATATGTATA
CACACATATACATATATATGTATACACACATATACATATATATGTATACACACATATACATATAT
ATGTATACACACATATACATATATATGTATACACACATATACATATATATGTATACACACATATA
CATATATATGTATACACACATATACATATATATGTATACACACATATACATATATATGTATACAC
ACATATACATATATATGTATACACACATATACATATATATGTATACACACATATACATATATATG
TATACACACATATACATATATATGTATACACACATATACATATATAGTATACACACATATACATA
TATATGTATACACACATATACATATATATGTATACACACATATACATATATATGTATACACACAT
ATACATATATATGTATACACACATATACATATATATGTATACACACATATACATATATACATATA
TACACAAACACACATGCACCGCACTTTTTTTTTTTTTTTTTTTTTTGGAGATGGGAGTCTCACTCTAT
CACCAGGCTGGAGTGCAGTGGTGTGATCTTGGCTCACTGCAACCTCTGTCTCCTGGGTTCAAGCT
ATTCTCCTGCTTCAGCCTCCTGAGTAGCTGGGATTACAGGTGCTCACCACCATGCCAGCTAATT
TTTGTATTTTTACCATGTTGGCCAGGATGATCTCCATCTCCTGACCTCCTGATCCTCCTGCCTTG
GGTCCCAAAGTGCAGGGATTGCAGGCATGAGCCACTGCATGTGGCCACACCACACTTCCTTTAT
CCACTTGTGATGGATGGCATTGTGTTGAAAAGAAGTCATCTCTGTTTTTCTTTCCCTGAACG----3'
    
```

← CCTTTTTATGGCTAAGTAGTATTCATCGT →
← TTCAGTAGAGCAAAAAGAAAGGGACTTGC →
} Amorces

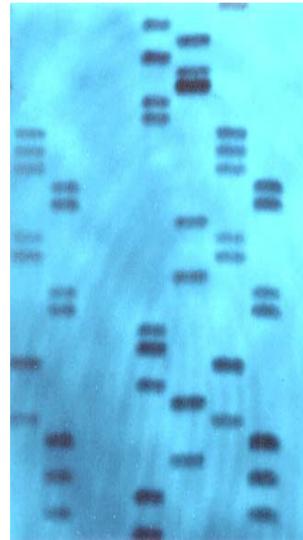
Méthodes d'étude

révélé soit par une sonde **monocus** spécifique d'un VNTR donné;



soit par une sonde **multilocus** - séquence commune à différents minisatellites dans le génome (image complexe).

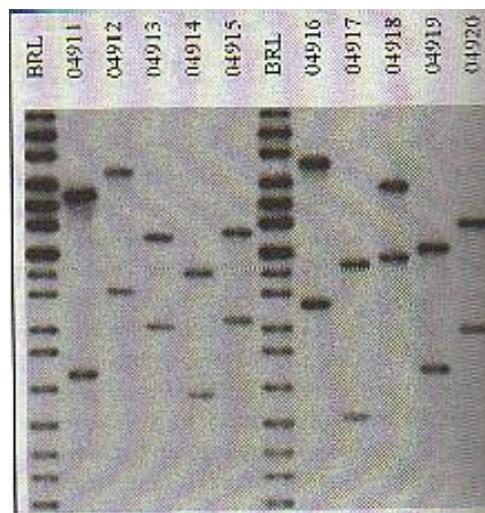
Ind 1
Ind 2
Ind 3
Ind 4



Les minisatellites (2): empreinte Génétique

➤ Profil génétique (très) spécifique d'un individu obtenu grâce à une combinaison de marqueurs de type VNTR très informatifs, principalement les minisatellites, ou en associant plusieurs microsatellites.

➤ Intérêt en médecine légale



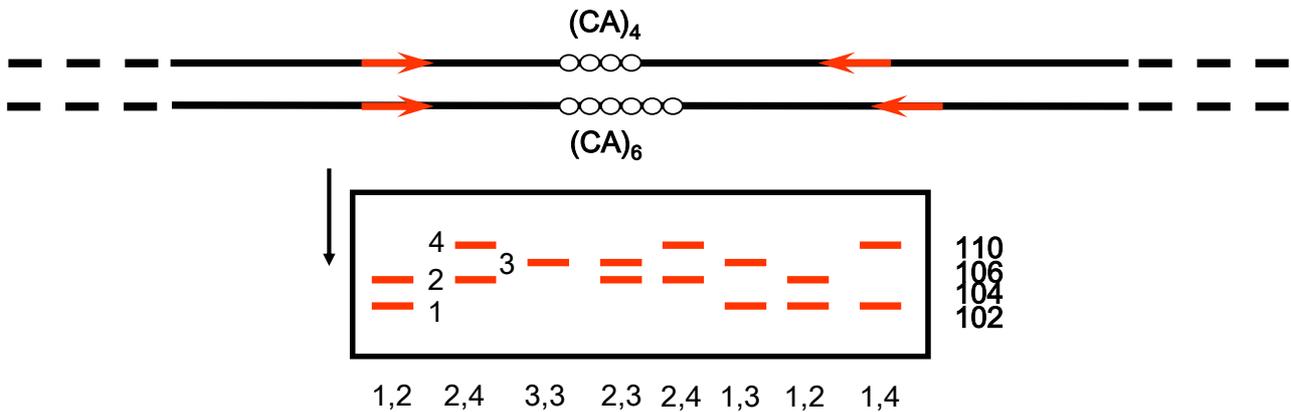
Minisatellite monocus

➤ Nécessite de l'ADN en qualité et en quantité suffisante au moins pour l'amplification par PCR.

Les microsatellites (1)

Répétition de motifs de 1 à 4 nucléotides

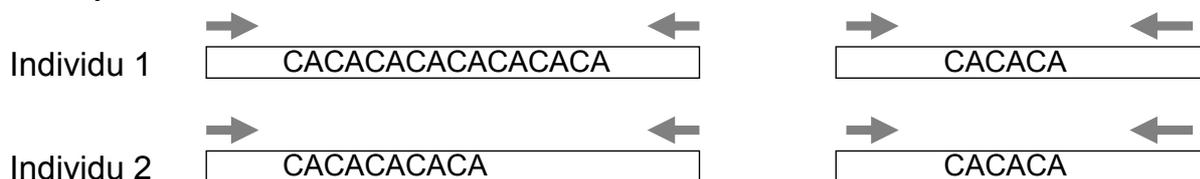
- Les plus fréquents et les plus utilisés : dinucléotides (CA)/(GT) répartis sur tout le génome et très informatifs.
- Taux d'hétérozygotie souvent supérieur à 70%



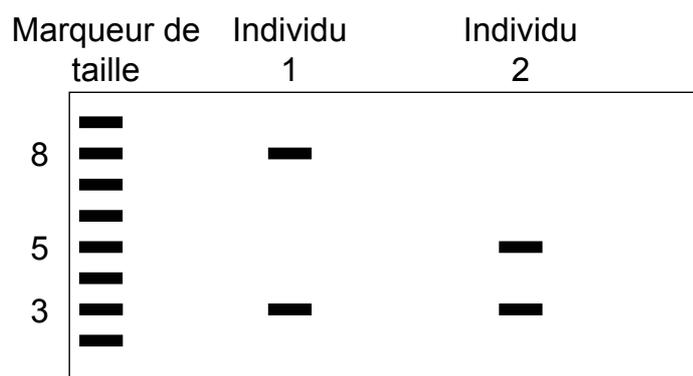
Méthodes d'étude

• PCR

amplification

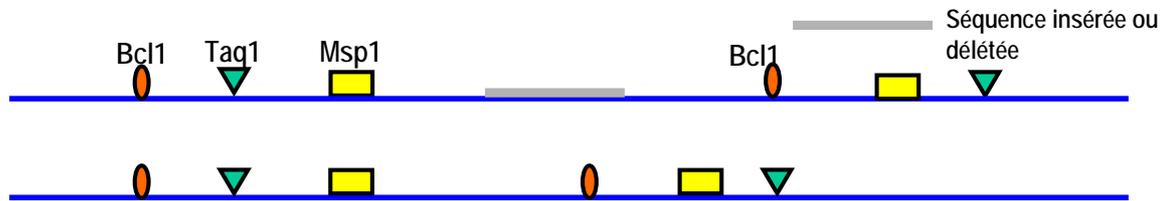


migration sur gel



Autres Polymorphismes

➤ Délétions / insertion



➤ Inversion



Conséquence des polymorphismes

- **Aucune conséquence pathologique directe**
- **Mais peut entraîner des variations phénotypiques**
 - couleur de yeux,
 - prédisposition à certaines pathologies,
 - réponse différents à des médicaments...
- probablement <1% des <SNP

Différence polymorphisme / mutation

Mutation = changement délétère

Conséquences des mutations sur le fonctionnement du gène:

- Délétion plus ou moins complète du gène
- Altération de la phase de lecture : troncature par codon stop prématuré

AUG ~~GCC~~ UCU AAC CAU GGC → AUG CUC UAA
Met Ala Ser Asn His Gly Met Leu **STOP**

- Codon faux-sens : interaction, adressage, fonction catalytique, modification post-traductionnelle
- Codon stop, mutation du codon stop, codon initiateur
- Epissage du gène
- Expression et régulation du gène

Marqueurs génétiques

Marqueurs génétique (1)

- Le marqueur génétique sert à repérer ou identifier un segment de chromosome ou un allèle transmis

- Les marqueurs définissent le **génotype** (constitution génétique d'un individu);

- On appelle marqueurs génotypiques:
 - ✓ les variations identifiables de la séquence du génome
 - ✓ de localisation connue
 - ✓ d'informativité estimée par le *taux d'hétérozygotie*
 - ✓ et codominants : les 2 allèles sont détectables chez les hétérozygotes.

Marqueurs génétiques (2)

Un marqueur génotypique est caractérisé par:

- ses allèles correspondant à chaque variation possible de la séquence
- son *Taux d'hétérozygotie* $H=1 - (\sum p^2)$, qui est donc fonction du nombre et de la fréquence de chacun des allèles (p), par exemple:

• 3 allèles de fréquence 0,2 / 0,3 / 0,5

$$H=1- (0,04+0,09 + 0,25) = 0,62$$

• 2 allèles de fréquence 0,2 / 0,8 :

$$H=1- (0,04 + 0,64)= 0,32$$

Rq: Un marqueur biallélique ne pourra pas avoir un taux d'hétérozygotie supérieur à 0,5 qui sera atteint en cas d'allèles de fréquence égale.

Les différents types de marqueurs

	SNP (RFLP)	VNTR (Minisat) [Variable Number of Tandem Repeat]	STR (Microsat) [Short Tandem Repeat]
Polymorphisme	de substitution, délétion/insertion	de répétition > 10 nucléotides	de répétition 1 à 5 nucléotides
Marqueur génétique	Biallélique	Multiallélique	Multiallélique
Informativité	+	+++	+++
Détection	Southern PCR (hybridation, RFLP)	PCR Southern	PCR uniquement

Utilisation des marqueurs génétiques

➤ **Analyse de liaison génétique:** suivre la transmission des allèles d'un marqueur par rapport à un trait morbide.

Il existe 2 cas: °en recherche: on teste s'il y a **co-ségrégation** entre un allèle du marqueur et la maladie. Si oui, il y a liaison génétique: le marqueur est alors génétiquement proche de la maladie.

°lors d'un **diagnostic indirect:** on suit la mutation dans *une famille* à travers la transmission des allèles d'un marqueur qui est génétiquement proche du gène responsable

➤ **Étude d'association** chez des malades non apparentés: établir une différence de distribution des fréquences alléliques d'un marqueur chez un groupe de malades non apparentés par rapport à un groupe témoin.(Etude ApoE et Alzheimer)

➤ **Identifier un individu** avec certitude : médecine légale, paternité

➤ **Rechercher un événement génétique somatique** dans une cellule

- recherche d'une perte d'allèle ou instabilité

Liaison génétique et cartographie

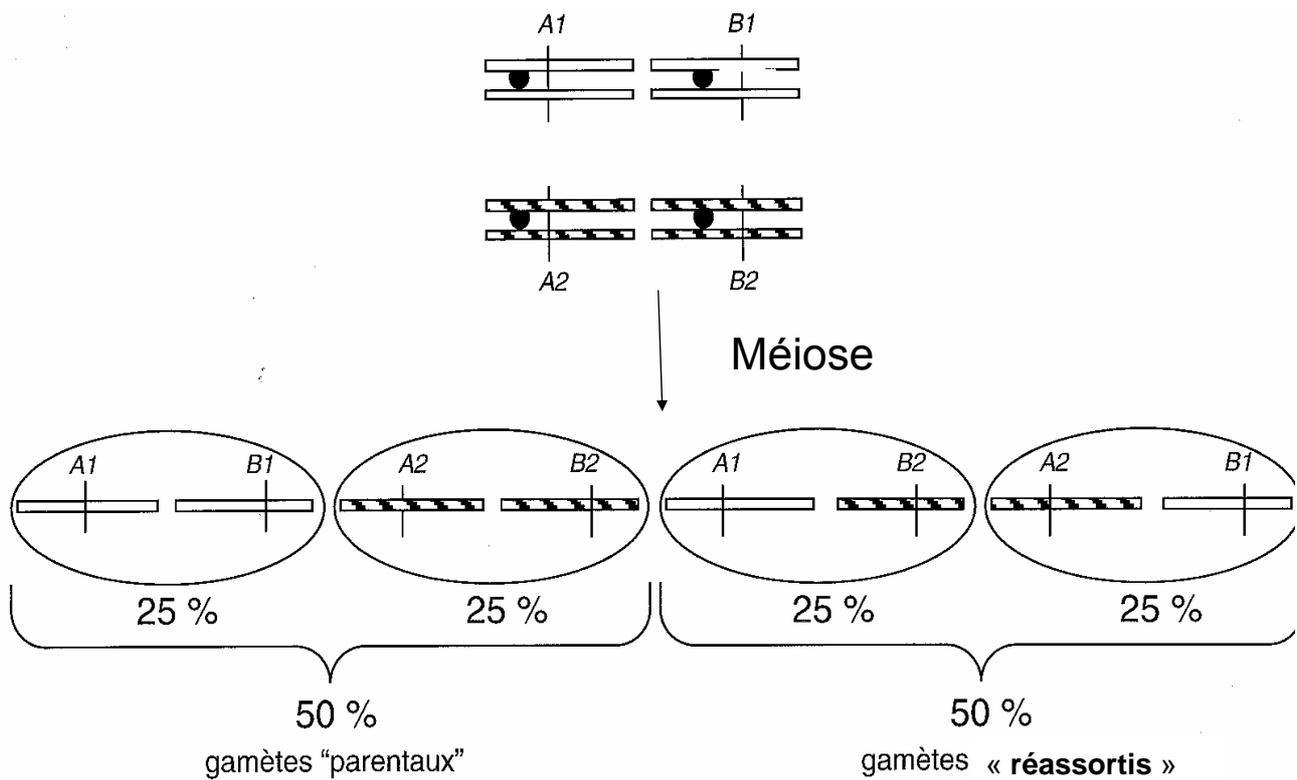
Liaison Génétique (1)

➤ Si 2 marqueurs sur 2 chromosomes différents : transmission indépendante;

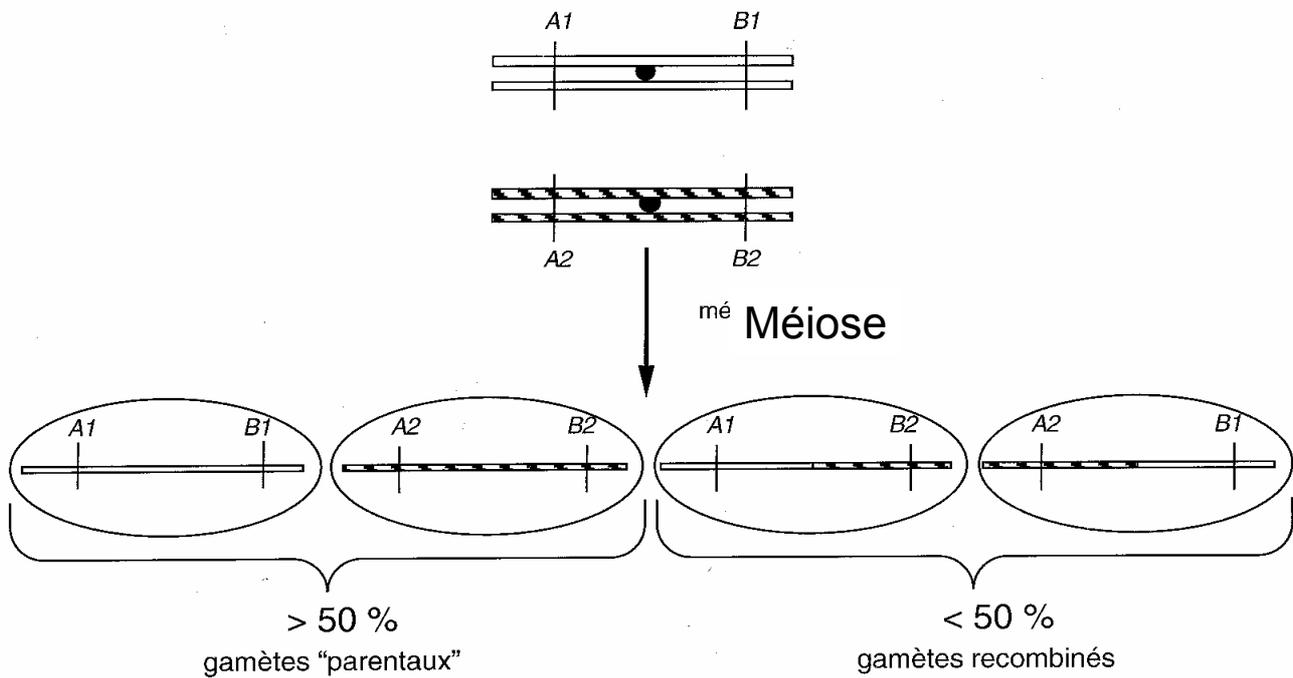
➤ Si 2 marqueurs proches sur le même chromosome → tendent à être transmis ensemble et non de façon indépendante : liaison génétique

Conséquences : les allèles de marqueurs situés sur des loci génétiquement liés auront tendance à être transmis dans la même combinaison que chez les parents : conservation des haplotypes parentaux

Marqueurs indépendants



Marqueurs liés



Liaison Génétique 2

- La fréquence de recombinaisons entre 2 gènes dépend de leur localisation relative sur le chromosome : plus l'intervalle qui les sépare est grand, plus grande est la probabilité pour qu'un crossing-over survienne dans cet intervalle;
- Entre 2 loci (gènes ou marqueurs), on peut ainsi mesurer la fréquence de *recombinaison* θ

Notions de distance génétique

- A partir de la fréquence de recombinaison θ , on peut calculer la distance génétique exprimée en centiMorgan (cM)

Notions de distance physique

- En moyenne, une distance génétique d'1cM correspond à 1 Mb de distance physique

Définition de l'haplotype

- Un haplotype est constitué des allèles de plusieurs marqueurs génétiques ordonnés sur un même chromosome.
- L'haplotype est déduit de l'observation de la transmission des allèles de chaque marqueur sur au moins 2 générations (parents-enfants).
 - Suppose de connaître la localisation précise de chaque marqueur;
 - Suppose de connaître **la phase**, c'est à dire le chromosome sur lequel se situe les allèles de chaque génotype.

Les haplotypes ne sont pas toujours conservés d'une génération à l'autre. On peut observer dans la descendance :

- Les haplotypes **parentaux**
- Haplotypes **recombinés** dus à recombinaison méiotique

Analyse de liaison

➤ *But* : mettre en évidence une liaison génétique entre un marqueur génétique (de localisation connue) et le locus de la maladie (inconnu)

➤ *Type d'études*: familiales

➤ *Marqueurs très informatifs* : microsatellites

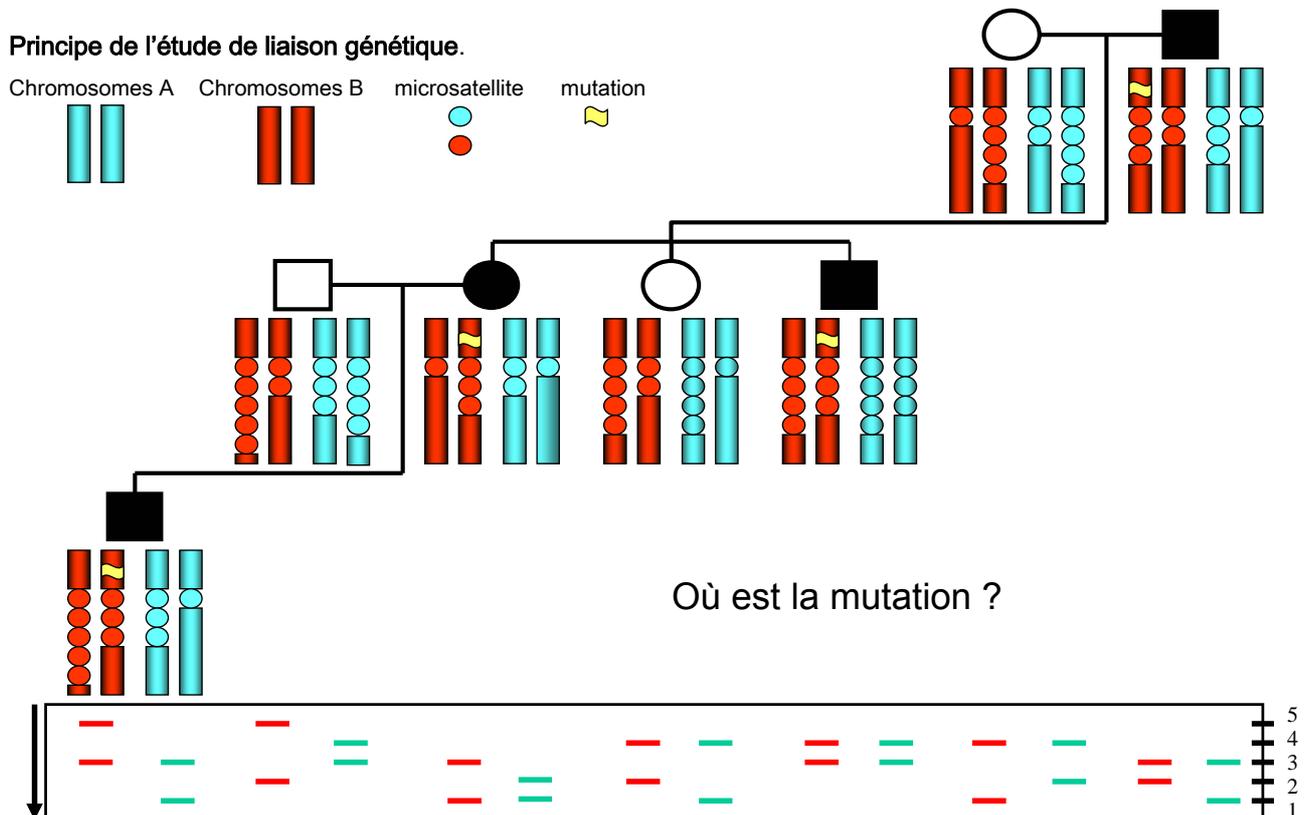
➤ *Type d'analyse* : analyse paramétrique avec modèle de la maladie

➤ *Liaison testée* par méthode de vraisemblance : Lod-score :
ratio de la vraisemblance de l'hypothèse de liaison sur celle de non liaison.

➤ les seuils de significativité:

✓ Liaison entre la maladie et le marqueur: $Z > 3.00$

✓ Rejet de la liaison quand $Z < -2.00$



Comment localiser le gène responsable d'une maladie héréditaire ?

➤ 3 circonstances favorables:

1- Dans la famille étudiée, la maladie co-ségrège avec une anomalie Chromosomique, délétion, duplication, translocation,....., décelable au caryotype classique

2- L'anomalie protéique ou la voie cellulaire touchée est connue: hypothèse fonctionnelle → gène(s) candidat(s)

3- Il existe un modèle animal de la maladie (rongeur):

a- Localisation du gène chez l'animal

b- Consultation des cartes de syntenie animal-homme;

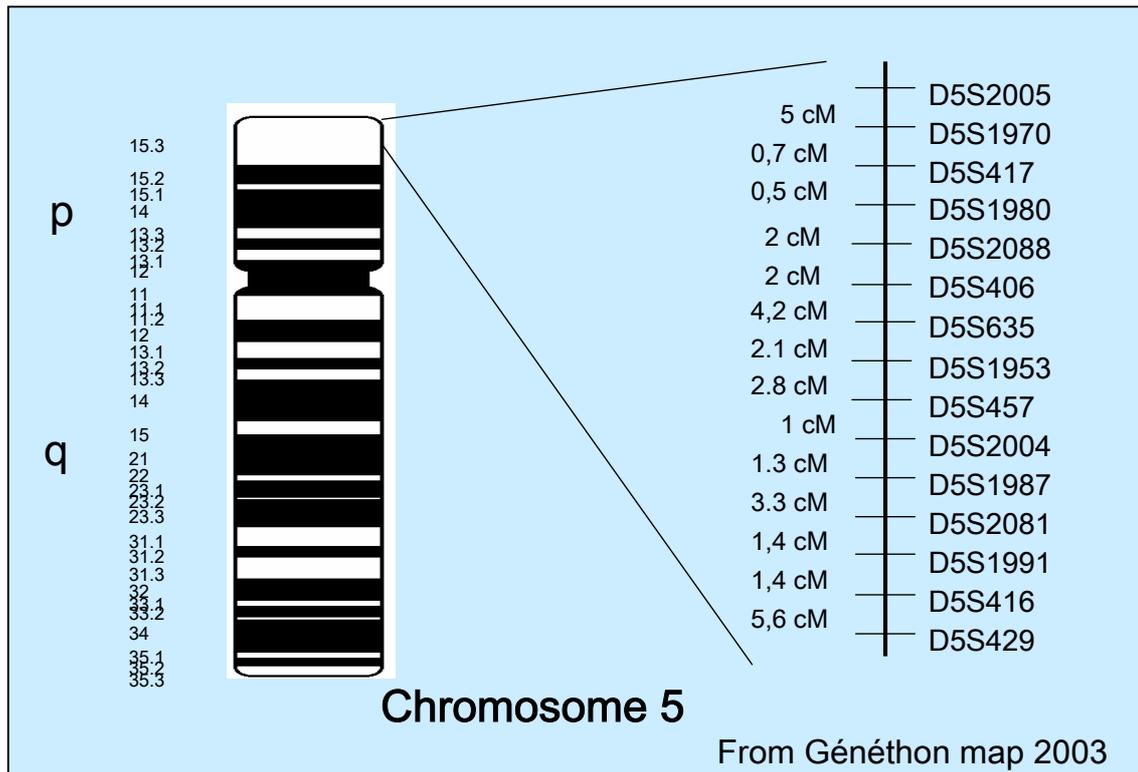
c- utilisation de microsatellites de la région pour valider le locus dans la famille étudiée.

Si ni 1 ni 2 ni 3, cartographie primaire

CARTOGRAPHIE PRIMAIRE: en pratique.....

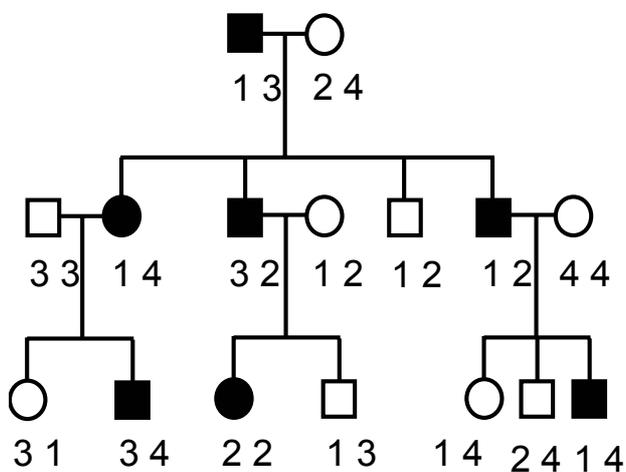
- Etape 1: **étude clinique** et paraclinique complète de chaque membre de la famille: sujet à risque et *conjoint*s;
- Etape 2: prélèvements, signature du consentement et extraction d'ADN;
- Etape 3: exclusion des gènes et loci *déjà connus*;
- Etape 4: **genome-scan** (400 marqueurs répartis le long du génome):
- Etape 5: interprétation des résultats du genome-scan
 - calcul des lod-scores,
 - haplotypage,
 - Genotypage de nouveaux marqueurs dans les régions candidates (15 à 30 régions)
- Etape 6: **précision de la région candidate** par l'étude:
 - de nouveaux marqueurs et
 - de *nouvelles familles*
- Etape 7: recherche de gènes candidats dans les banques de données in silico

CARTE HAUTE DENSITE DE MICROSATELLITES

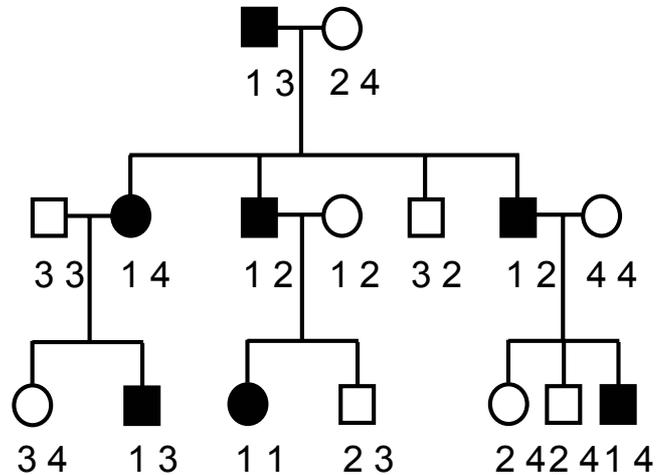


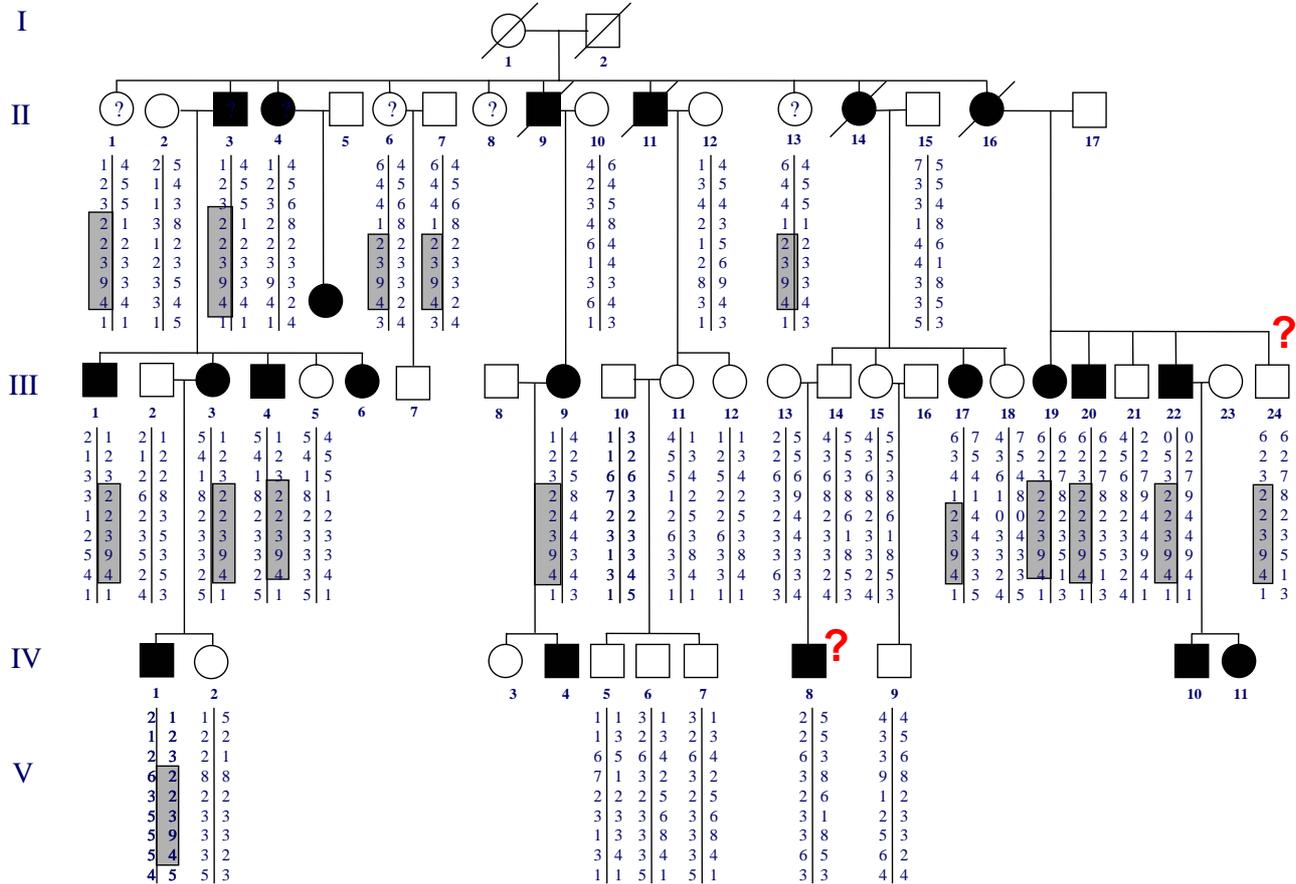
Exemple maladie dominante

Microsatellite non lié à la maladie



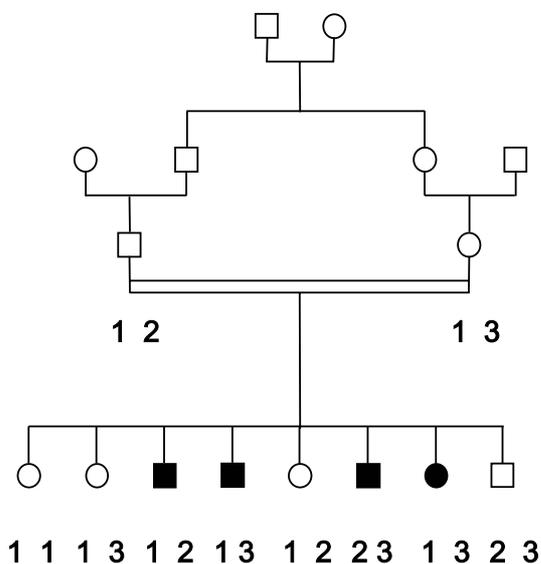
Microsatellite lié à la maladie



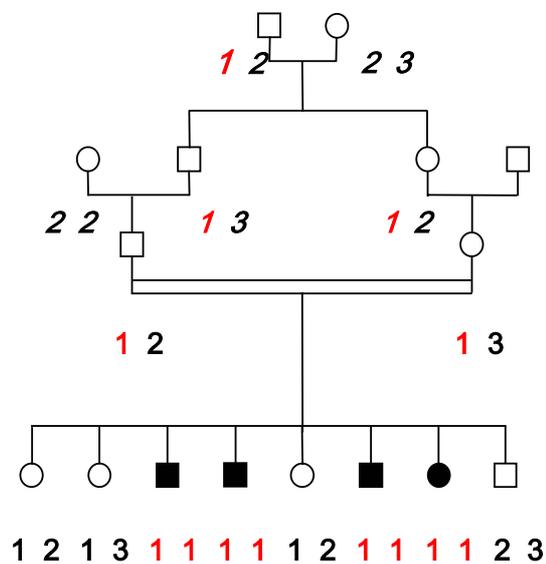


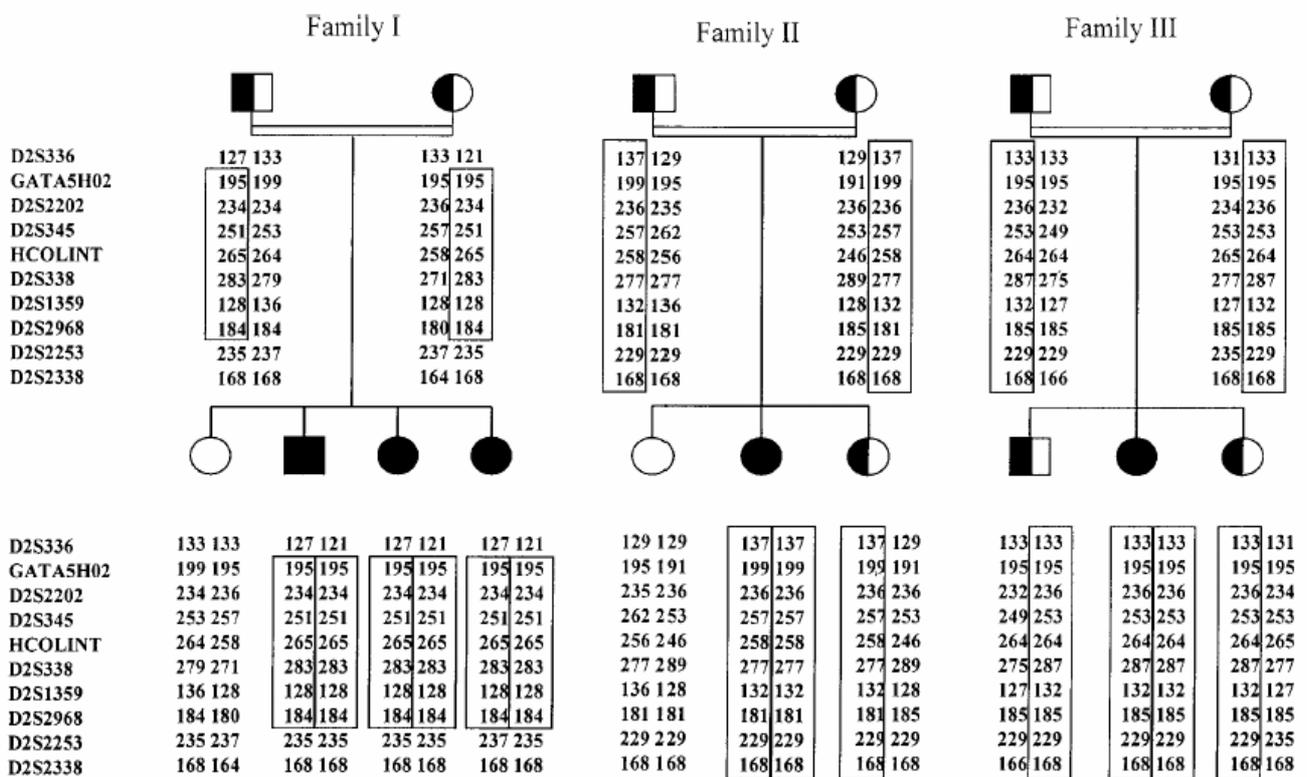
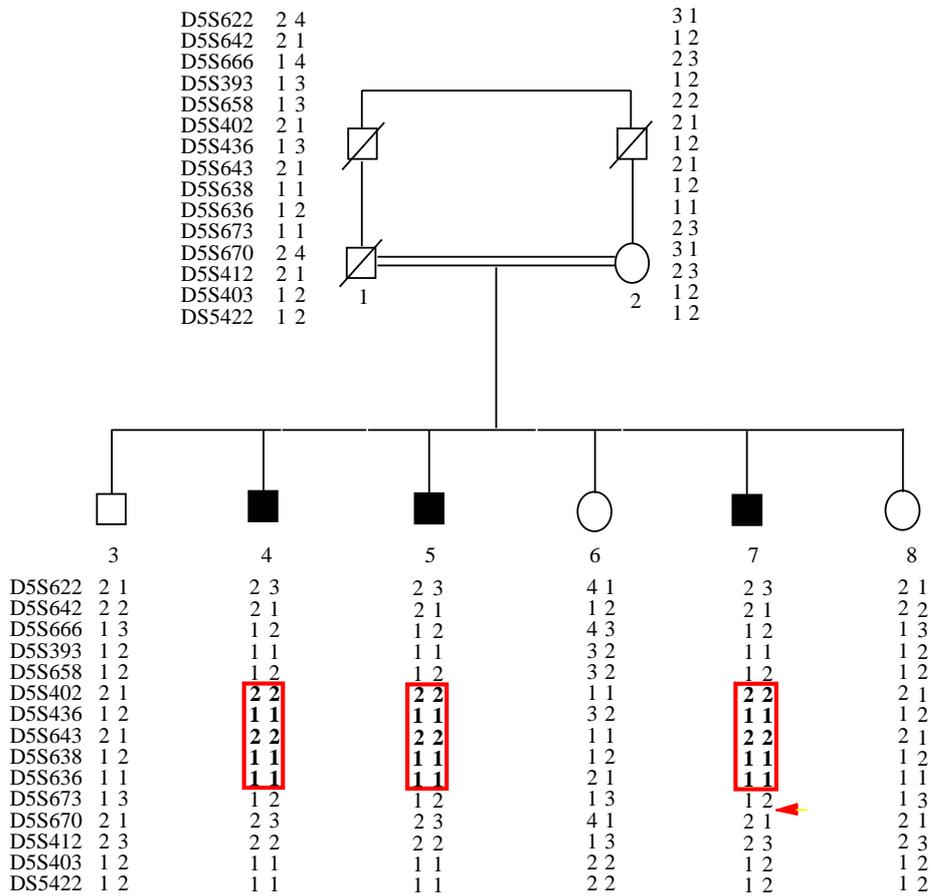
Exemple maladie récessive

Microsatellite non lié à la maladie



Microsatellite lié à la maladie





Le diagnostic génétique indirect

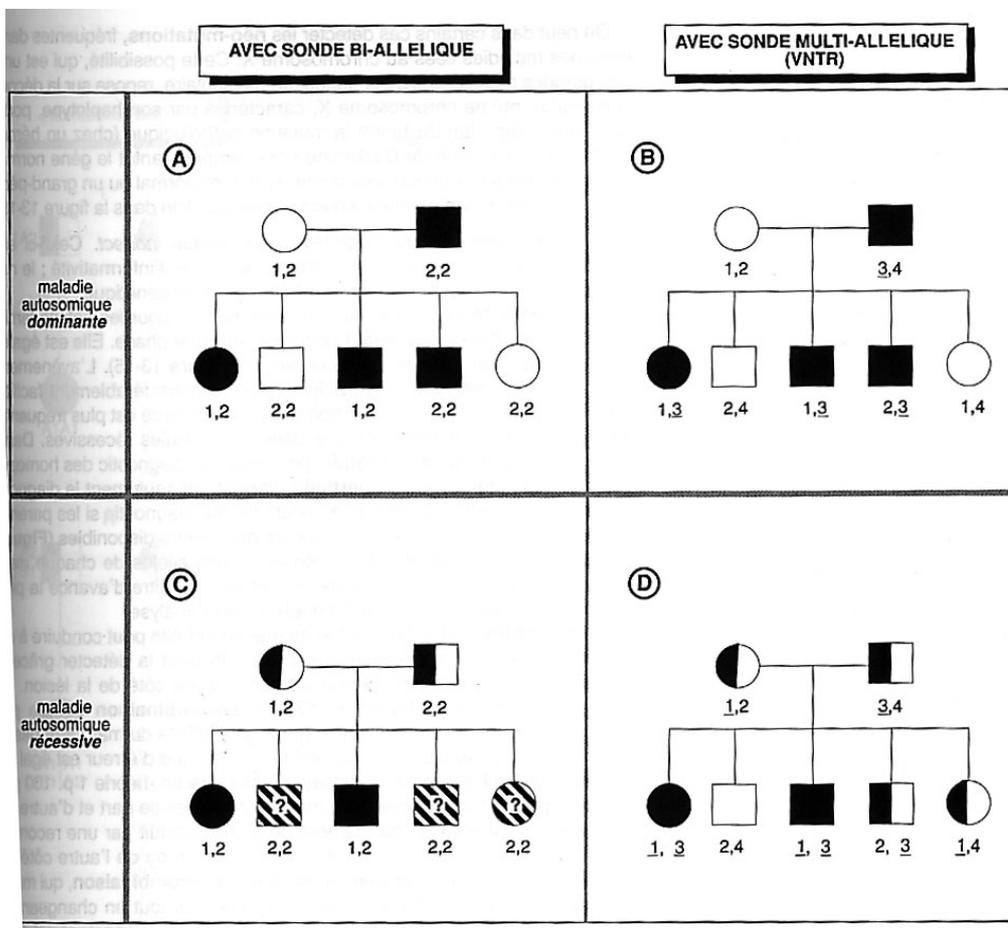
1- **Distinguer** les 2 chromosomes du (des) parent(s) susceptibles d'avoir transmis le gène de la maladie

- Polymorphismes pour lesquels les parents sont **hétérozygotes**
- Plusieurs polymorphismes pour les grands gènes
- microsatellites car très informatifs

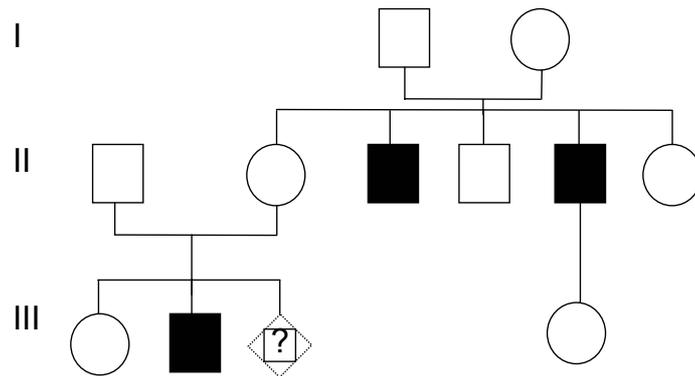
2- Déterminer la **phase** : déterminer quel allèle du marqueur ségrège avec l'allèle de la maladie (il faut au moins l'ADN d'un sujet atteint).

- Plus facile pour les maladies liées à l'X car le père a 1 seul chr X

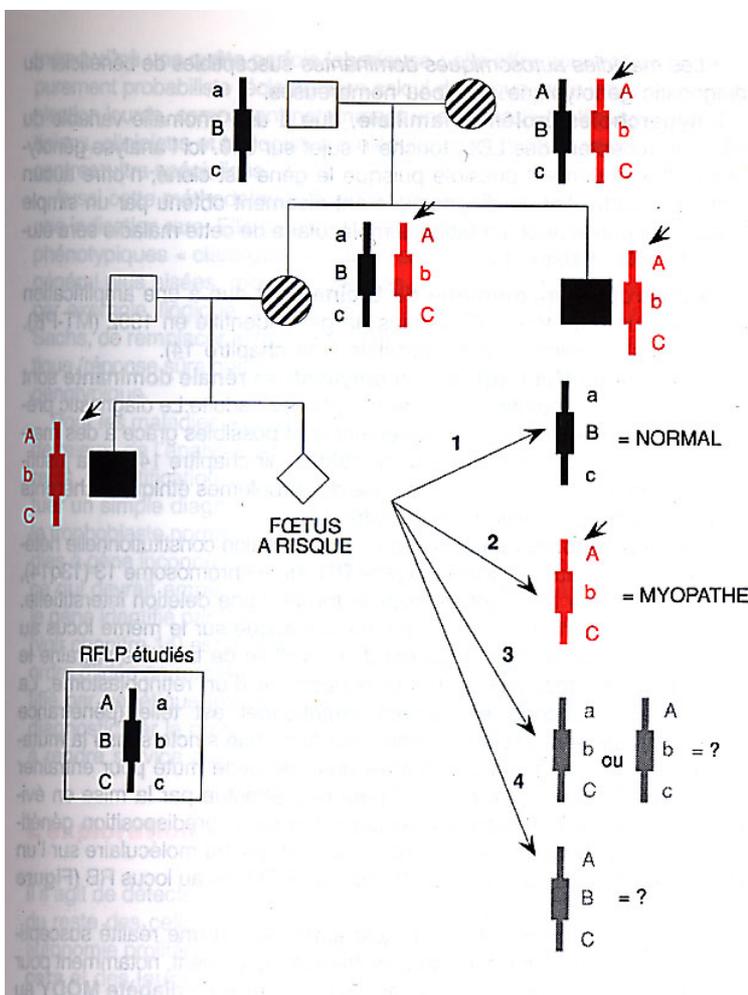
3- Typer le ou les marqueurs chez le sujet à tester pour identifier l'haplotype qui a été **transmis** au sujet.



Exemple : hémophilie A



patient	Marqueur A	Marqueur B
III1	2	5
II2	1-3	3-6
II3	3	6
III1	1-2	3-5
III2	3	6
III3	1	3



Diagnostic prénatal indirect de la myopathie de Duchenne :

difficultés liées aux recombinaisons

Recombinaison simple ?

Double recombinaison ?

Le diagnostic génétique indirect

A- **Les limites** du diagnostic indirect :

- Disponibilité du cas index
- Le manque d'informativité des marqueurs: parent homozygotes
- la survenue d'une **recombinaison** entre le marqueur et le site de la mutation:
 - toujours utiliser au moins 1 marqueur de part et d'autre du gène
 - *quand gène de grande taille (DMD):* nécessité de marqueurs intra-géniques et de marqueurs flanquants le gène pour dépister les recombinaisons

Rq: La fraction de recombinaison entre les différents marqueurs doit être connue

B- **Les risques** associés au diagnostic indirect:

- Quand 1 seul marqueur informatif, le risque est égale à la fraction de recombinaison entre le marqueur et le gène;
- Quand 2 marqueurs flanquants informatifs, le risque est celui d'une double recombinaison

Études d'associations

Études d'associations

- Comparaison des fréquences alléliques pour un marqueur entre un groupe de patients et un groupe de témoins
- Les témoins doivent être appariés aux patients pour l'âge, le sexe et l'origine ethnique (géographique)
- Test non paramétrique car le mode de transmission n'est pas défini (tests statistiques : chi²)
- Test peu puissant qui nécessite de grands échantillons
- Marqueurs utilisés : SNP.

ApoE et maladie d'Alzheimer (1)

L'apoE est le seul transporteur du cholestérol en fonction dans le cerveau. Il interagit avec des récepteurs cellulaires impliqués dans le métabolisme des lipides.

Il existe 3 allèles de ce gène : **e2**, **e3** et **e4**, codant pour 3 isoformes de la protéine : apoE2, apoE3, apoE4.

Méta-analyse regroupant **6262** sujets témoins et **5107** patients avec Alzheimer d'origine caucasienne:

Génotypes	Patients	Témoins
e3/e3	36%	61%
e3/e4	41%	21%
e4/e4	15%	1%
e2/e3	8%	17%

ApoE et maladie d'Alzheimer (2)

- La présence de e4 accroît le risque de développer une maladie d'Alzheimer :
 - ✓ risque relatif de **3,7** pour les hétérozygotes **e3/e4**,
 - ✓ risque relatif de **30,1** pour les homozygotes **e4/e4**.

- Propriétés des facteurs de susceptibilité:
 - ✓ Un grand nombre de patients Alzheimer ne sont nullement porteurs d'un allèle e4,
 - et
 - ✓ la présence de cet allèle e4 ne signifie pas que le porteur sera forcément atteint d'Alzheimer.

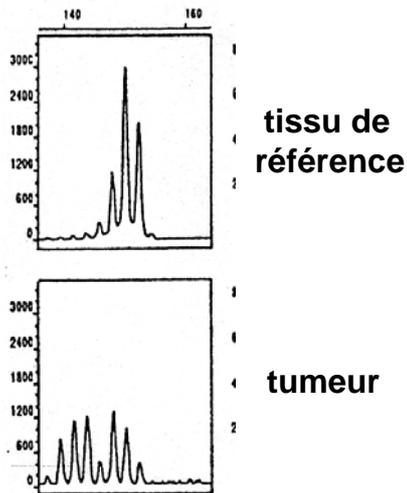
D'autres facteurs jouent donc un rôle important.

Difficulté d'utiliser ces résultats en médecine

Génétique somatique

Exemple d'applications des polymorphismes de répétition

➤ Instabilité des microsatellites et cancer



Syndrome HNPCC : les tumeurs coliques sont instables par anomalie du système MMR

Rechercher une instabilité permet de repérer les patients porteurs d'une prédisposition au cancer

Pertes d'hétérozygoties (LOH)

