

**UE optionnelle: Maladies Héritaire du Métabolisme**

**CM: OUTILS MOLECULAIRES POUR LE DIAGNOSTIC  
DE POLYMORPHISMES/MUTATIONS DU GENOME  
HUMAIN**

**Bernard CARCY (bernard.carcy@umontpellier.fr)**

**Laboratoire de Biologie Cellulaire et Moléculaire**

**EA 4558 « Vaccination antiparasitaire »,**

**Faculté de Pharmacie, Université Montpellier**

## A- INTRODUCTION

**1- Quelques généralités sur le génome humain et les maladies génétiques héréditaires monogéniques**

**2- Quelques exemples de maladies héréditaires (et leurs mutations) impliquant un défaut de métabolisme**

**3- Quelques généralités sur les polymorphismes**

**4- Mutations rapportées dans la littératures**

**5- Position des mutations sur un gène et conséquences phénotypiques**

## B- OUTILS MOLECULAIRES D'IDENTIFICATION DES MUTATIONS

### 1- Basés sur l'amplification génique

- Rappels sur la méthode de base (PCR classique)
- La PCR en temps réel (Real Time PCR)
- - la PCR-RFLP (détection mutation touchant un site de restriction)
- SSCP (Single Strand Conformational Polymorphism)
- DGGE (Denaturing Gel Gradient Electrophoresis)
- DHPLC (Denaturing High Performance Liquid Chromatography)
- la HRM (Analyse des courbes de fusion à haute résolution)
- la MLPA (Multiplex Ligation dependant Probe Amplification)

### -2- Basés sur le séquençage d'ADN

- Rappels sur la méthode de base de séquençage d'un ADN (Sanger & Coulson)
- Le pyroséquençage et la SMRT (Single Molecule Real Time)

### 3- Basés sur l'hybridation moléculaire

- Southern blot
- puces à ADN

## 1.1 - LE GENOME HUMAIN

Tableau D. ADN mitochondrial et nucléaire

### ADN nucléaire (ADNg)

Un ADN dans le noyau

Transmission par les deux parents

**3 milliards de pb**

Organisation promoteur, région 5' non codante, exons/introns, région 3' non codante

≈ 25 000 gènes

**23 paires de chromosomes:**

**-22 autosomiques**

**de 249 (ch1) à 54 (ch21)Mpb**

**-1 sexuelle (141Mbp pour X et**

**60Mpb pour Y)**

**-0,1% de variation entre 2 individus**

### ADN mitochondrial (ADNmt)

Environ 1000 ADNmt/cellule

Transmission maternelle

16,5 kb

Pas d'intron

37 gènes codés par l'ADNmt :

- si pathologie : hétéroplasmie mélange d'ADNmt normal et pathologique (ADNmtN et ADNmtM, respectivement)

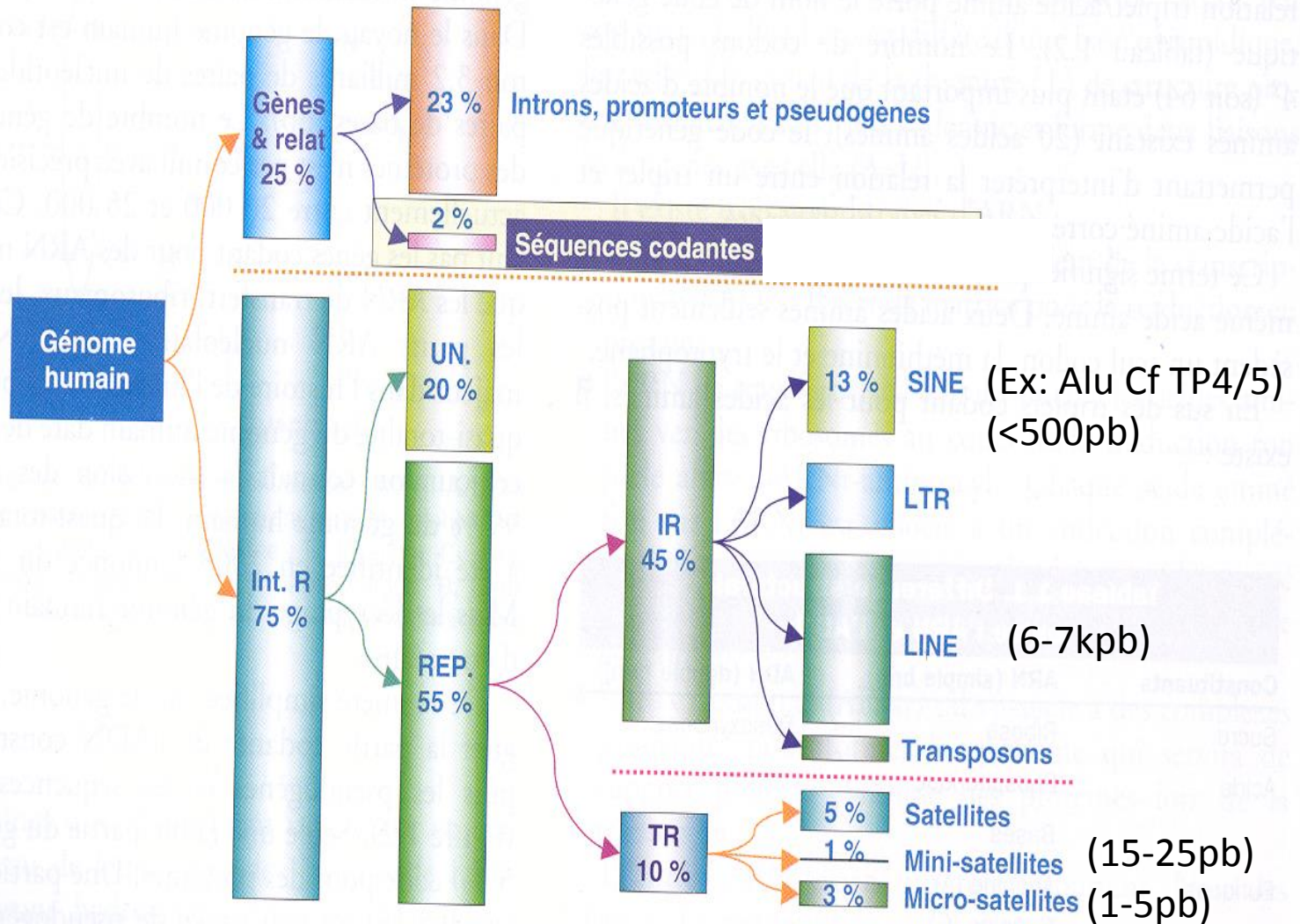
- effet seuil : dans une cellule, le rapport ADNmtN/ADNmtM doit dépasser un seuil pour que le phénotype se déclare (notamment défaut de la chaîne respiratoire)

- variabilité du nombre d'ADNmt muté intracellule et/ou intra-tissu et/ou intra-organe et/ou intrafamiliale

Interaction ADNg et ADNmt



# LE GÉNOME HUMAIN



**Figure 1.2** Composition générale du génome humain. Le pourcentage représente la quantité de séquences par rapport à la séquence totale connue du génome. Gènes & relat : gènes et séquences associées ; Int. R : régions intergéniques ; Un. : séquences intergéniques uniques ; Rep. : séquences intergéniques répétitives ; IR : séquences intergéniques répétitives dispersées ; TR : séquences intergéniques répétées en tandem.

SINE: petit élément nucléaire intercalé; LINE: grand élément nucléaire intercalé



## 1.2- MALADIES HEREDITAIRES

### DEFINITION

- **La maladie héréditaire est une maladie qui est transmise par le spermatozoïde ou par l'ovule.** Il s'agit d'une affection présente dès le début de la vie intra-utérine. Celles-ci sont le résultat de la présence, au niveau des chromosomes, de gènes anormaux. On parle également de génopathie ou d'hérédoopathie.

\*Une anomalie chromosomique (aberration chromosomique) peut également être à l'origine d'une maladie héréditaire.

\*Le terme autosomique caractérise le fait que l'affection héréditaire est transmise par des chromosomes non sexuels.

\*On distingue habituellement les maladies héréditaires autosomiques dominantes des maladies héréditaires autosomiques récessives.

\*La différence essentielle entre les **maladies récessive ou dominante autosomiques** est que dans le premier cas il est nécessaire que les deux parents portent l'anomalie génétique pour que la descendance ait la maladie.

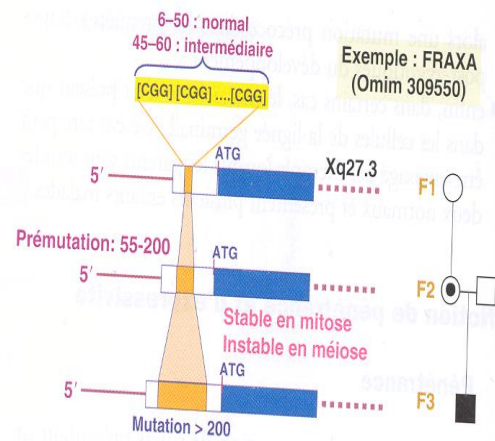
- On oppose le terme de maladie acquise à celui de maladie héréditaire ou de maladie génétique. **Les maladies acquises sont des affections apparaissant après la fécondation**, soit pendant la vie intra-utérine (à l'intérieur de l'utérus pendant la grossesse) soit après la naissance. Les premières portent le nom d'embryopathie (maladie de l'embryon) et de foetopathie (maladie du fœtus).

# Les maladies héréditaires monogéniques

- Recensement de ces maladies regroupé dans le **catalogue Omim** accessible sur Internet (On line Mendelian Inheritance in Man)
- Sont **nombreuses (>4000)** mais rares. Cependant, réel problème humain et de santé publique
- **Les plus fréquentes ont une incidence faible:**
  - mucoviscidose (incidence 1/3000)
  - déficit en alpha 1- antitrypsine (incidence 1/1700)
  - hypercholestérolémie familiale (incidence 1/500)
  - neurofibromatoses (incidence 1/3000)
- **Plusieurs modes de transmission:**
  - transmission héréditaire **mendélienne**
    - \* récessif lié à l'X (Ex: hémophilie A; omim 306700; **maladie de Fabry**)
    - \* dominant lié à l'X (Ex: formes de la maladie de Charcot-Marie-Tooth; omim 302800)
    - \* autosomique récessif (Ex: mucoviscidose, omim 219700; **maladies de Mc Ardle et de Tarui = glycogénoses de type V et 7; maladie de Brody; maladies de Gaucher et de Pompe**)
    - \* autosomique dominant (Ex: porphyrie aigüe intermittente, omim 176000)
    - \* codominant (Ex: groupes sanguins ABO)
  - transmission héréditaire **non mendélienne** (phénomène d'anticipation; extension de séquences répétées en nombre anormal; Ex: **maladie de Steinert** (dystrophie myotonique, Omim 160900, autosomique dominant)
  - transmission héréditaire **mitochondriale** = selon les lois de l'hérédité dite maternelle (Ex: myopathies mitochondriales; **syndrômes MELAS, MERRF**)

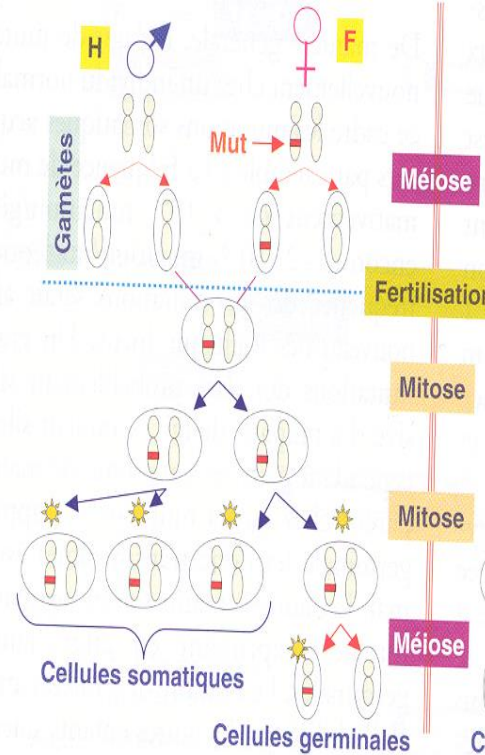
# Transmission des MALADIES HEREDITAIRES

## Transmission non mendélienne

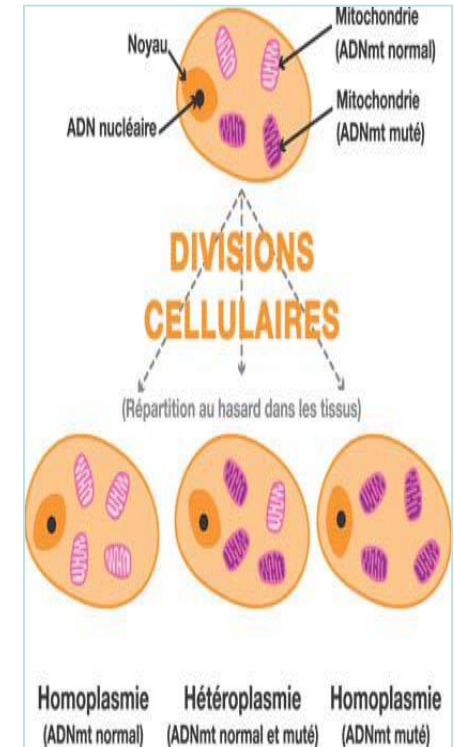


**Figure 2.5** Le phénomène d'anticipation. Dans le syndrome de l'X fragile (Oimim 309550), la trinuécléotide CGG est répétée entre 6 et 50 fois chez le sujet normal. En cas d'anticipation, d'une génération à l'autre, cette répétition de microsatellite augmente. Lorsque le nombre de répétitions du microsatellite est au-dessous du seuil de 200, la maladie n'est pas déclarée encore, c'est le stade de pré-mutation. Au stade de pré-mutation, il peut y avoir une instabilité de microsatellite aboutissant à une expansion du triplet (phénomène dynamique). Et au-delà d'un certain nombre de répétitions (> 200 en moyenne), la maladie apparaît et s'aggrave au fil des générations, apparaissant de plus en plus précocement : c'est le phénomène d'anticipation.

## Transmission mendélienne



## Transmission mitochondriale





## 2- LES Maladies Héréditaires du Métabolisme

### 1- Glycogénose

#### Maladie de McArdle (glycogénose de type V).

La maladie de McArdle (glycogénose de type V) est une myopathie métabolique de transmission **autosomique récessive** causée par une mutation du gène de la **myophosphorylase**.

Le gène *PYGM* est codé en 11q13 On dénombre plus de **65 mutations différentes**. Elles sont variées, **de type insertion, délétion et inversion, donnant des anomalies de type faux-sens, non-sens ou des anomalies d'épissage**.

**La plupart des mutations de *PYGM*** modifient la zone régulatrice de la myophosphorylase en N-terminal ou la zone catalytique en C-terminal. Elles induisent des modifications de structure ou des modifications de fonction majeures responsables d'une **réduction drastique de l'activité enzymatique**. Une nouvelle **mutation faux-sens de *PYGM* (S814N)** entraîne la substitution d'une sérine par une asparagine au codon-814 (exon-20) ce qui modifie l'hydrophobicité d'un site interne de la protéine et a un donc un effet délétère sur sa fonction.

Aucune étude n'a trouvé de relation entre la sévérité de la maladie de McArdle et le génotype de *PYGM*

**La mutation la plus fréquente (60%) est la mutation non-sens R50X de l'exon 1.** La mutation R50X est aussi bien présente dans des formes avec simple intolérance à l'effort que dans des formes sévères de l'adulte ou à début dans l'enfance.

L'étude en biologie moléculaire de ce gène reste cependant précieuse car elle permet de confirmer le diagnostic de maladie de McArdle et elle permettrait dans certains cas caractéristiques de porter un diagnostic sans réaliser de biopsie musculaire.

# 1- Glycogénose

## Maladie de Tarui (glycogénose de type 7).

Numéro Orphanet	ORPHA371
Prévalence des maladies rares	<1 / 1 000 000
Hérédité	• Autosomique récessif
Âge d'apparition	Enfance
Code CIM 10	• E74.0
numéro MIM	• <a href="#">232800</a>
Synonyme(s)	GSD type 7 Glycogénose type 7 Maladie de Tarui

Le déficit en **phosphofructokinase** musculaire (maladie de Tarui), ou glycogénose de type 7, est une forme rare de maladie de stockage de glycogène caractérisée par une fatigue à l'effort et une intolérance à l'exercice musculaire.

La maladie **est due à des mutations du gène *PFKM* (12q13)** codant l'isoenzyme musculaire de la phosphofructokinase ;

## 2- Deficit de la chaîne respiratoire

### MYOPATHIES MITOCHONDRIALES

Autres appellations : **Syndrome de MELAS** (Mitochondrial Encephalopathy Lactic Acidosis Stroke), **Syndrome de MERRF** (Myoclonic Epilepsy Ragged Red Fibers),

**La gravité** des maladies mitochondriales est très variable selon les individus. Dans les maladies mitochondriales dues à une anomalie de l'ADN mitochondrial, l'atteinte d'une cellule (donc d'un tissu, donc d'un organe...) **dépend de la proportion de mitochondries "malades"** qu'elle contient : car dans chaque cellule coexistent des mitochondries normales et des mitochondries malades (cette coexistence se nomme hétéroplasmie). Il existe un effet **seuil**.

### Que sont les myopathies mitochondriales ?

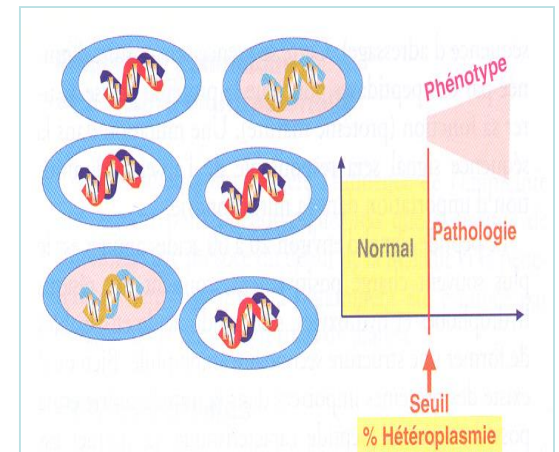
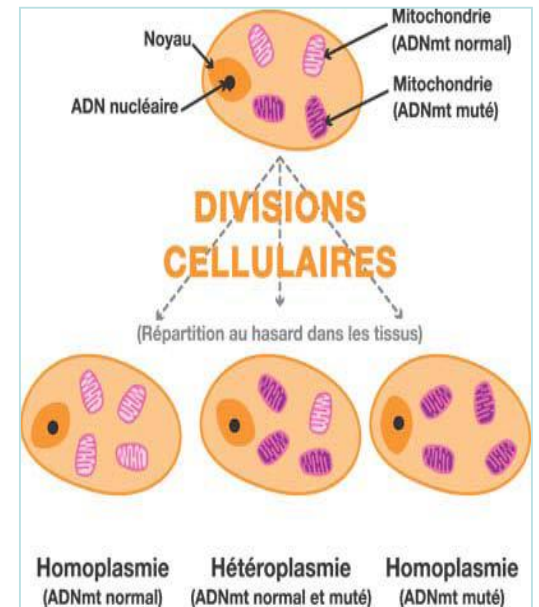
Les maladies mitochondriales sont des maladies qui ont en commun des anomalies de la chaîne respiratoire mitochondriale.

Les mitochondries sont comme des centrales produisant l'énergie nécessaire au bon fonctionnement de la cellule. Chaque cellule de l'organisme en contient plusieurs centaines, voire plusieurs milliers.

Lorsqu'une majorité de mitochondries est "malade", la cellule n'a plus assez d'énergie pour assurer correctement sa fonction. Ces mitochondries "malades" peuvent se trouver dans n'importe quelle cellule de l'organisme : n'importe quel organe peut donc être touché dans les maladies mitochondriales (cytopathies mitochondriales ou mitochondriopathies).

**Les myopathies mitochondriales touchent cependant plus particulièrement les muscles. Elles comprennent notamment les syndromes de MELAS et de MERRF.**

Ces maladies peuvent être héréditaires. Leur transmission est complexe. Les formes liées à une altération primitive de l'ADN mitochondrial n'obéissent pas aux lois classiques de l'hérédité mendélienne mais sont **transmises selon les lois de l'hérédité dite maternelle**. Dans d'autres cas, il peut s'agir d'anomalies de gènes de l'ADN nucléaire (l'ADN « classique ») contrôlant indirectement la chaîne respiratoire et se transmettant sur le mode autosomique dominant, soit sur le mode autosomique récessif. **Mise à jour** : juin 2006, Tuy Nga Brignol, (M.D.), J.Andoni Urtizbera, (M.D), copyright AFM



**Figure 2.7** Notion d'effet seuil dans la pathologie mitochondriale. La répartition aléatoire entre le nombre de mitochondries normales et pathologiques est variable. Lorsque le nombre de mitochondries pathologiques atteint une certaine quantité (seuil), les phénomènes pathologiques se déclarent notamment dans les tissus dépendant en grande partie du métabolisme mitochondrial (par exemple, le cerveau).



## MYOPATHIES MITOCHONDRIALES

Autres appellations : **Syndrome de MELAS** (Mitochondrial Encephalopathy Lactic Acidosis Stroke), **Syndrome de MERRF** (Myoclonic Epilepsy Ragged Red Fibers),

# Syndrome MELAS

Numéro Orphanet	ORPHA550
Prévalence des maladies rares	1-5 / 10 000
Hérédité	<ul style="list-style-type: none"><li>Sporadique</li><li>Hérédité mitochondriale</li></ul>
Âge d'apparition	Enfance
Code CIM 10	<ul style="list-style-type: none"><li>G71.3</li></ul>
numéro MIM	<ul style="list-style-type: none"><li><a href="#">540000</a></li></ul>
Synonyme(s)	Encéphalomyopathie mitochondriale, acidose lactique, pseudo-épisodes vasculaires cérébraux Myopathie mitochondriale, encéphalopathie, acidose lactique, pseudo-épisodes vasculaires cérébraux

Le syndrome de MELAS est provoqué par des mutations de l'ADN mitochondrial. Au moins **10 mutations différentes** ont été identifiées mais **80 % des patients sont porteurs de la mutation 3243A>G dans le gène de l'ARN de transfert de la leucine (ARNt Leu).**

La recherche de la mutation causale doit tenir compte de son hétéroplasmie c'est-à-dire de sa coexistence avec une population résiduelle d'ADN mitochondrial normal. La proportion de la mutation peut varier considérablement d'un tissu à l'autre mais, **dans le syndrome de MELAS, elle est très élevée (>90% de la population d'ADN mitochondrial)** dans tous les tissus et peut donc être cherchée dans le sang.

## MYOPATHIES MITOCHONDRIALES

Autres appellations : **Syndrome de MELAS** (Mitochondrial Encephalopathy Lactic Acidosis Stroke), **Syndrome de MERRF** (Myoclonic Epilepsy Ragged Red Fibers),

# Syndrome MERRF

Numéro Orphanet	ORPHA551
Prévalence des maladies rares	1-9 / 1 000 000
Hérédité	• Hérédité mitochondriale
Âge d'apparition	Enfance
Code CIM 10	• G40.3
numéro MIM	• <a href="#">545000</a>
Synonyme(s)	Epilepsie myoclonique avec 'ragged-red-fibers' Epilepsie myoclonique à fibres rouges en lambeaux Syndrome de Fukuhara

L'expression clinique est très variable d'un malade à l'autre dans une même famille et d'une famille à l'autre. Le syndrome de MERRF résulte de mutations de l'ADN mitochondrial. La mutation 8344A>G dans le gène (MTTK) de l'ARN de transfert de la lysine (ARNt Lys) est responsable de plus de 80% des cas de la maladie.

La recherche de la mutation causale doit tenir compte de son hétéroplasmie c'est-à-dire de sa coexistence avec une population résiduelle d'ADN mitochondrial normal. La proportion de la mutation peut varier considérablement d'un tissu à l'autre mais, **dans le syndrome MERRF, elle est très élevée (>90% de la population d'ADN mitochondrial)** dans tous les tissus et peut donc être cherchée dans le sang.

## 3- Anomalie du métabolisme calcique

### **MALADIE DE BRODY**

#### **Autres appellations :**

**Mise à jour :** Juin 2006. Rédaction : TN Brignol (M.D.), validation : J. Andoni Urtizbera (M.D.), copyright AFM

#### **Qu'est-ce que la maladie de Brody ?**

La maladie de Brody est une maladie génétique rare du muscle squelettique décrite par Brody en 1969. Elle se manifeste par des crampes musculaires non douloureuses et une impossibilité de relâchement musculaire après une contraction musculaire due à un exercice physique.

La maladie de Brody est due à un trouble du transport des ions  $Ca^{2+}$  au niveau du réticulum sarcoplasmique des fibres musculaires de type II ou fibres rapides.

Le gène lié à la maladie est le gène *ATP2A1* localisé sur le chromosome 16 (en 16p12.1-12.2). Ce gène code la protéine SERCA1 (abréviation de sarco(endo)plasmic reticulum  $Ca^{2+}$ -ATPase) ou enzyme  $Ca^{2+}$ -ATPase du réticulum sarcoplasmique jouant un rôle dans la contraction rapide du muscle strié.

**Trois mutations du gène *ATP2A1* ont été identifiées (1996) dont deux sont liées à des codons stop**, conduisant à la formation d'une protéine SERCA1 tronquée dans trois familles dont le mode de transmission est autosomique récessif.



## 4- Dystrophies

### Wikipédia

Dystrophie myotonique de Steinert	
Référence MIM	160900
Transmission	<b>Dominante</b>
Chromosome	19 q13.2-q13.3
Gène	DMPK
Empreinte parentale	Non
Mutation	<b>Expansion de triplet</b>
Mutation de novo	Rare
Nombre d'allèles pathologiques	Sans objet
Anticipation	Oui maternelle
Porteur sain	Sans objet
Prévalence	<b>1/20 000</b>
Maladie génétiquement liée	Aucune
Diagnostic prénatal	Possibe
<a href="#">Liste des maladies génétiques à gène identifié</a>	
<a href="#">modifier</a>	

**La mutation est une expansion instable d'un triplet CTG du gène DMPK.** Ce triplet est répété plus de 37 fois chez les personnes atteintes.

**Le nombre de répétitions du triplet CTG est généralement associé à la sévérité** de la maladie ainsi qu'à l'âge d'apparition des symptômes.

La **dystrophie myotonique de Steinert** ou **maladie de Steinert** est une maladie génétique **autosomique dominante**, à pénétrance incomplète et marquée par l'anticipation, qui **affecte plusieurs organes** : le squelette, les muscles lisses, l'oeil, le coeur, le système endocrinien et le système nerveux central.

Les signes de cette maladie sont variés allant **d'une forme légère à grave.**

**1- La forme légère** est caractérisée par une cataracte et une myotonie modérée. L'espérance de vie est normale.

**La forme classique** est caractérisée par une faiblesse musculaire généralisée et une myotonie généralisée, une cataracte et des troubles de la conduction cardiaque. L'adulte peut perdre son autonomie et **l'espérance de vie est réduite** si le patient n'est pas suivi pour le coeur.

**La forme congénitale** avec une hypotonie musculaire, souvent associée à une insuffisance respiratoire avec **décès précoce**. Le retard mental est fréquent dans cette forme.

**La certitude du diagnostic est obtenue par PCR (Cf diapos 55) avec un prélèvement sanguin (Le consentement de la personne prélevée est indispensable):**

- **sujets sains (nombre de CTG inférieur à 37 pour les 2 allèles sur le gène DMPK)**

- **maladie si les individus ont un allèle à moins de 37 répétitions, et un allèle à plus de 37 (et généralement plus de 50 CTG).**

## 5- Maladies de surcharges lysosomales

**La maladie de Gaucher** est une affection génétique, transmise sur le mode autosomique récessif dont la prévalence en France est estimée entre 1 cas pour 40.000 et un cas sur 100.000 personnes. Sa prévalence est plus importante dans la population juive Ashkénaze (1 cas pour 850 naissances vivantes). **C'est la plus fréquente des maladies de surcharge du lysosome.** Il est essentiel d'en faire un diagnostic précoce, en raison de l'existence de traitements spécifiques permettant de limiter voire d'éviter la survenue de complications graves.

### Aspects génétiques

Les manifestations cliniques de la maladie de Gaucher, transmise de façon héréditaire sur le mode autosomique récessif, sont hétérogènes. Cette affection est liée à la **mutation du gène codant pour la glucocérébrosidase**, situé sur le bras long du chromosome 1. On retrouve plus de **150 mutations** différentes connues actuellement, **Les cinq mutations les plus fréquentes sont : N370S, 84GG, L444P, IVS2 et Rec.** Il n'y a généralement pas de corrélation entre génotype et phénotype, mais **certains génotypes sont associés à des formes plus graves de la maladie (mutations L444P et D409H par exemple).**

### Maladie de Pompe

<b>Transmission</b>	<a href="#">Récessive</a>
<a href="#">Chromosome</a>	<a href="#">17 q25.2-q25.3</a>
<a href="#">Gène</a>	<a href="#">GAA</a>
<a href="#">Empreinte parentale</a>	Non
<a href="#">Nombre d'allèles pathologiques</a>	Plus de 150 connus 1 sur 14 000 chez les américains d'origine africaines à 1 sur 100 000 en Europe
<a href="#">Incidence</a>	
<b>Maladie génétiquement liée</b>	Aucune

[Dia](#) **Un article de Wikipédia**

[Liste des maladies génétiques à gène identifié](#)

Si les examens biologiques montrent une disparition ou une diminution de l'activité de la protéine alpha-glucosidase acide, **l'analyse génétique est toujours pratiquée pour confirmer le diagnostic.** Le test se fait sur un échantillon sanguin duquel on extrait l'ADN des cellules pour analyser les mutations du gène GAA.

La **maladie de Pompe** est une **maladie génétique** progressive et souvent fatale en rapport avec une anomalie de fonctionnement de l'*alpha-1,4-glucosidase acide*, une enzyme **lysosomiale** qui hydrolyse le **glycogène** en **glucose**.

## 5- Maladies de surcharges lysosomales

### Diagnostic moléculaire et conseil génétique

La détermination du statut des femmes à risque dans les familles où ségrège la maladie est pourtant essentielle pour délivrer un conseil génétique optimal, soulignant l'intérêt du diagnostic moléculaire qui permet d'assurer le dépistage des hétérozygotes.

Le diagnostic moléculaire peut être indirect ou direct.

### Diagnostic moléculaire indirect

L'étude des polymorphismes de restriction intragéniques ou flanquants du gène *GLA* et des séquences polymorphes liées au gène *GLA* au locus Xq22 peut permettre d'identifier indirectement le chromosome X porteur de l'allèle *GLA* mutant. Lorsqu'ils sont informatifs, ces marqueurs anonymes peuvent être utiles pour le dépistage des hétérozygotes et le diagnostic prénatal par analyse moléculaire indirecte, dans les familles pour lesquelles la mutation spécifique est inconnue [50]. Une limitation de cette méthode est le risque d'erreur diagnostique liée à un événement de recombinaison (*crossing-over*) [51].

### Diagnostic moléculaire direct

La détection de la mutation du gène *GLA* apparaît donc comme la méthode de référence pour l'identification précise des conductrices (figure 4) et permet de délivrer aux familles un conseil génétique clair [51] avec, en outre, la possibilité d'affirmer le diagnostic dans les cas où aucun sujet masculin n'est disponible pour l'étude. La mutation responsable du déficit enzymatique et du phénotype clinique peut être mise en évidence par séquençage du gène [52] précédé ou non d'une analyse par une méthode de criblage [14], [16] and [53].

**Chez les hommes** la façon la plus fiable de faire le diagnostic est la mesure de l'activité enzymatique de l'alpha-galactosidase dans les leucocytes.

**La détection des femmes porteuses** de la mutation par mesure de l'activité enzymatique de l'alpha-galactosidase n'est pas fiable. Seule la recherche de la mutation du gène *GLA* permet de détecter les porteuses.

La détection de la mutation est possible dans près de 100 % des hommes malades.

### Maladie de Fabry

Référence MIM	301500
Transmission	Récessive à l'X
Chromosome	Xq22
Gène	GLA
Empreinte parentale	Non
Mutation	Ponctuelle
Mutation de novo	Rare
Nombre d'allèles pathologiques	Plus de 300 connus
Anticipation	Non
Incidence	1 sur 50000 naissances de garçons
Maladie génétiquement liée	Aucune
Diagnostic prénatal	Possible

Liste des maladies génétiques à gène identifié

**La maladie de Fabry**, ou **syndrome de Rüter-Pompen-Wyers** est une **maladie lysosomale** génétique, liée au **chromosome X**, résultant d'un **déficit enzymatique de l'alpha-galactosidase lysosomale** avec accumulation de globotriaosylcéramide et de digalactosylcéramide dans les cellules.



### 3- LES POLYMORPHISMES OU VARIATION DE SEQUENCES (OU MUTATIONS)

#### MULTI-ALLELIQUES

- Short Tandem Repeat (STR) = microsatellites (1-5pb)
- Variable Number of Tandem Repeat (VNTR) = minisatellites (15-70pb)

#### BI-ALLELIQUES

- Les SNP (Single Nucleotide Polymorphism)
- les RFLP (Restriction Fragment Length Polymorphism)
- Insertion, délétion, insertion/délétion

EPIGENETIQUES (ex: méthylation)

DE L'ADN MITOCHONDRIAL

*Ce sont des variations de séquence d'un locus donné, d'un individu à l'autre au sein d'une même espèce.*

# LES POLYMORPHISMES BI-ALLELIQUES

(Exemples de relation SNP/maladie)

# LES POLYMORPHISMES MULTI-ALLELIQUES

(Exemples de relation STR/maladie génétique monogénique=omim)

Tableau 2.1. Exemples de polymorphismes bi-alléliques (SNP) associés à des maladies

Gène	Variant(s)	Pathologie associée
Gène de susceptibilité :		
- apolipoprotéine E	Allèle ε4	Maladie d'Alzheimer
- NRAMP1 ( <i>natural resistance associated macrophage protein 1</i> )	Asp543Asn	Tuberculose
- PPARG ( <i>peroxisome proliferator-activated receptor γ</i> )	G>C IVS4 Pro12Ala	
Pronostic :		
- MBL2 ( <i>mannose binding lectin</i> )	Variants structuraux B, C et D	Mucoviscidose
- TNF ( <i>tumor necrosis factor</i> )	-376G>A -308G>A dans le promoteur	Neuropaludisme
Pharmacogénétique :		
ADRB2 ( <i>β2 adrenergic receptor</i> )	Gly16Arg	Catécholamines (albuterol)
CYP2C9 ( <i>cytochrome P450 IIC, polypeptide 9</i> )	Arg144Cys	Dosage de la warfarine

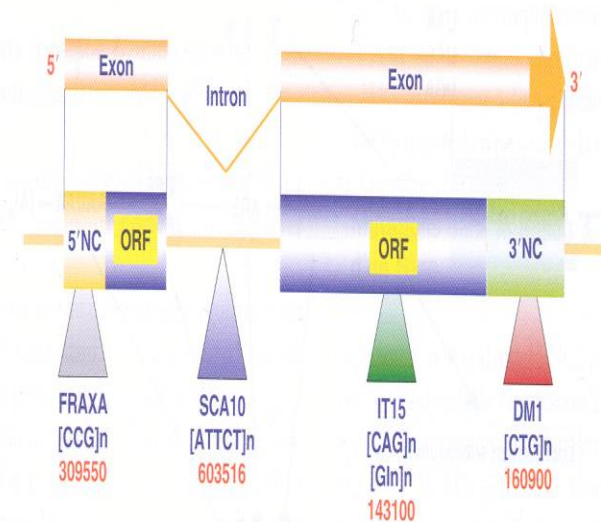


Figure 2.24 Exemples de localisation d'expansion de séquences répétées dans le génome. Le nombre indiqué en dessous de chaque abréviation de gène est le numéro Omim. 5'NC : région 5' non codante ; 3'NC : région 3' non codante ; ORF (*open reading frame*) : partie codante d'un gène ; FRAXA (*Fragile site mental retardation 1*) : syndrome de l'X fragile (le gène muté est FMR1) ; IT15 (appelé aussi huntingtine) : gène muté dans la chorée de Huntington ; SCA10 (*spinocerebellar ataxia*) : gène muté dans l'ataxie spinocérébelleuse ; DM1 (*dystrophia myotonica 1*) : gène muté dans la maladie de Steinert (*dystrophie myotonique*).

# -LES EPIMUTATIONS: pathologies de l'épigénétique

Exemple du syndrome *immunodeficiency, centromeric instability and facial anomalies*(ICF, Omim 242860):

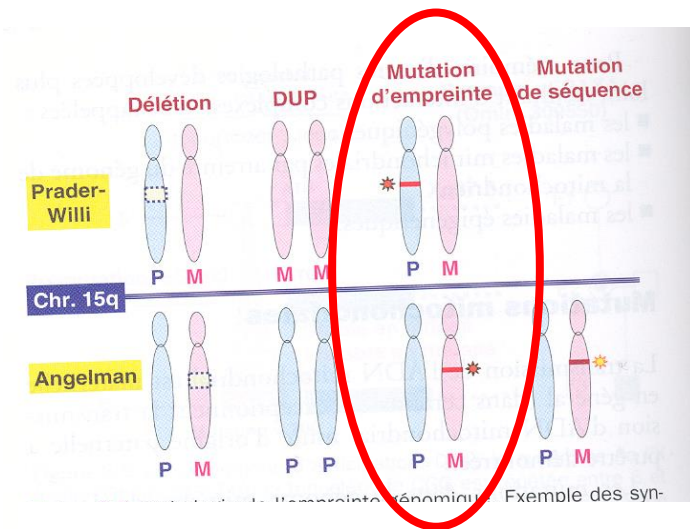
syndrome rare, de transmission autosomique récessif.

caractérisé par un immunodéficit général associé à des anomalies de développement.

Dans les lymphocytes, des anomalies chromosomiques sont observées (hétérochromatine péricentromérique allongée sur le chromosome 1 ou 16 au niveau de régions riches en séquences répétées). **Ces régions sont normalement fortement méthylées.**

**Dans le syndrome ICF, elles sont hypométhylées:** conséquence d'une mutation sur le gène codant une ADN méthylase (DNMT3B).

La méthylation différentielle de gènes parentaux peut être importante dans le phénomène **d'empreinte génomique** (expression différentielle d'un gène selon son origine maternelle ou paternelle).



Certaines **anomalies de méthylation provoquent un dérèglement de l'empreinte génomique.** C'est le cas par exemple dans le syndrome de Beckwith-Wiedemann (BWS, Omim 130650) responsable d'anomalies du développement et de risque élevé de tumeurs dans l'enfance.

# Mutations et pathologies mitochondriales

## MYOPATHIES MITOCHONDRIALES

Autres appellations : **Syndrome de MELAS** (Mitochondrial Encephalopathy Lactic Acidosis Stroke), **Syndrome de MERRF** (Myoclonic Epilepsy Ragged Red Fibers),

**La gravité** des maladies mitochondriales est très variable selon les individus. Dans les maladies mitochondriales dues à une anomalie de l'ADN mitochondrial, l'atteinte d'une cellule (donc d'un tissu, donc d'un organe...) **dépend de la proportion de mitochondries "malades"** qu'elle contient : car dans chaque cellule coexistent des mitochondries normales et des mitochondries malades (cette coexistence se nomme hétéroplasmie). Il existe un effet **seuil**.

## Que sont les myopathies mitochondriales ?

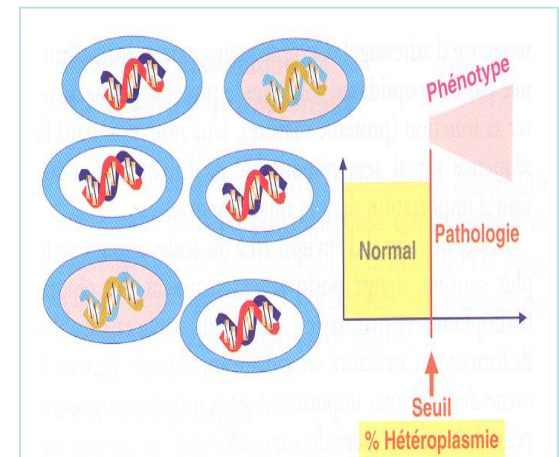
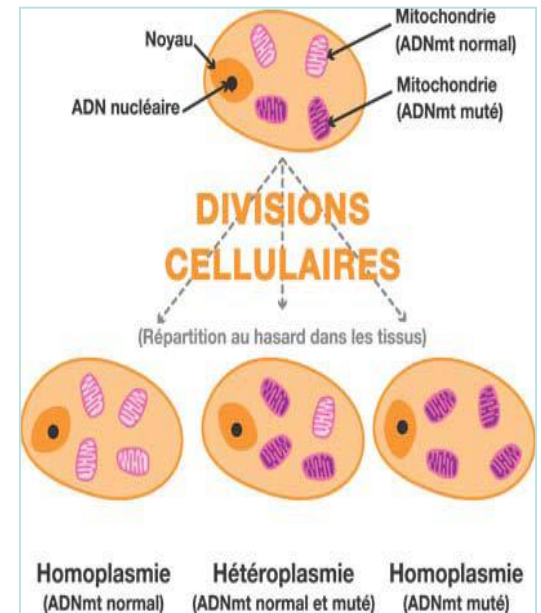
Les maladies mitochondriales sont des maladies qui ont en commun des anomalies de la chaîne respiratoire mitochondriale.

Les mitochondries sont comme des centrales produisant l'énergie nécessaire au bon fonctionnement de la cellule. Chaque cellule de l'organisme en contient plusieurs centaines, voire plusieurs milliers.

Lorsqu'une majorité de mitochondries est "malade", la cellule n'a plus assez d'énergie pour assurer correctement sa fonction. Ces mitochondries "malades" peuvent se trouver dans n'importe quelle cellule de l'organisme : n'importe quel organe peut donc être touché dans les maladies mitochondriales (cytopathies mitochondriales ou mitochondriopathies).

Les myopathies mitochondriales touchent cependant plus particulièrement les muscles. Elles comprennent notamment les syndromes de MELAS et de MERRF. Ces maladies peuvent être héréditaires. Leur transmission est complexe. Les formes liées à une altération primitive de l'ADN mitochondrial n'obéissent pas aux lois classiques de l'hérédité mendélienne mais sont **transmises selon les lois de l'hérédité dite maternelle**. Dans d'autres cas, il peut s'agir d'anomalies de gènes de l'ADN nucléaire (l'ADN « classique ») contrôlant indirectement la chaîne respiratoire et se transmettant sur le mode autosomique dominant, soit sur le mode autosomique récessif.

**Mise à jour** : juin 2006, Tuy Nga Brignol, (M.D.), J.Andoni Urtizbera, (M.D.), copyright AFM

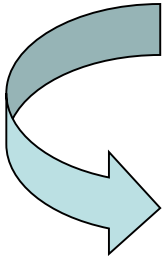


**Figure 2.7** Notion d'effet seuil dans la pathologie mitochondriale. La répartition aléatoire entre le nombre de mitochondries normales et pathologiques est variable. Lorsque le nombre de mitochondries pathologiques atteint une certaine quantité (seuil), les phénomènes pathologiques se déclarent notamment dans les tissus dépendant en grande partie du métabolisme mitochondrial (par exemple, le cerveau).



# NOMENCLATURE

- Selon les auteurs, **polymorphisme** peut signifier :
  - un changement de séquence sans conséquence pathologique
  - un changement de séquence dont la fréquence est  $>1\%$  dans la population générale
- Selon les auteurs, **mutation** peut signifier :
  - un changement de séquence, quelle que soit la conséquence
  - un changement induisant une pathologie

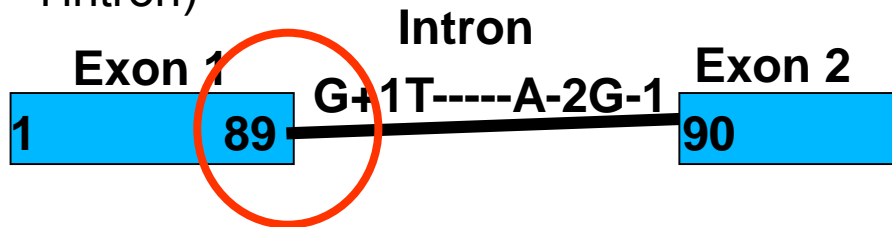


Les auteurs de la **nomenclature utilisent donc le terme variation de séquence** pour éviter toute confusion avec les termes polymorphisme et mutation (**voir nomenclature**)

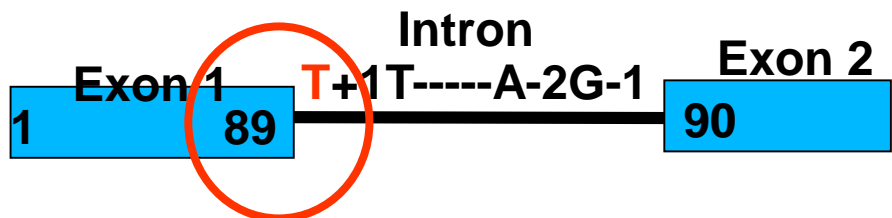
## Nomenclature au niveau des introns:

**-Au début de l'intron:** on donne le numéro du dernier nucléotide de l'exon précédent puis le signe + suivi de la position dans l'intron

-Ex: **89+1G**, 89 est le numéro du dernier nucléotide de l'exon précédant l'intron et G est en 1ere position sur l'intron)

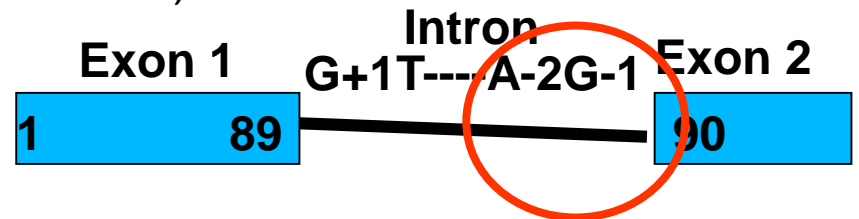


**Donc 89+1G>T signifie:**

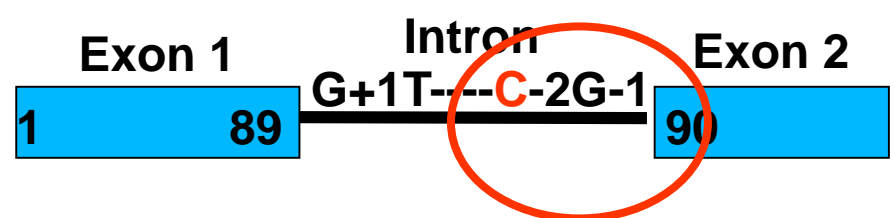


**-A la fin de l'intron:** on utilise la numérotation du premier nucléotide de l'exon suivant l'intron suivi d'un signe - et de la position du nucléotide en amont dans l'intron

-Ex: **90-1G**, 90 est le numéro du premier nucléotide de l'exon suivant l'intron et G est en dernière position de l'intron)



**Donc 90-2A>C signifie:**



## Nomenclature des modifications de nucléotides:

**-Les substitutions (mutations ponctuelles):** désignées par le caractère

« > »

\* **76A>C** signifie qu'au niveau du nucléotide 76, le A est substitué en C.

\* **88+1G>T** signifie qu'au niveau de l'intron, le G (en position +1 dans l'intron) a été substitué en T (la position 88 étant celle du 88ème nucléotide dans l'ADNc)

\***89-2A>C** (ou IVS2-2A>C) signifie qu'au niveau de l'intron, le A (en position -2 dans l'intron) a été substitué en C (la position 89 étant celle du 89ème nucléotide dans l'ADNc)

## Nomenclature des modifications d'acides aminés:

### Les substitutions:

\* modification **faux-sens: W26C** (ou trp26Cys) signifie qu'au niveau de l'acide aminé 26, le W (tryptophane) a été changé en C (cystéine).

\* modification **non-sens: W26X** (ou trp26X) signifie qu'au niveau de l'acide aminé 26, le W (tryptophane) a été changé en **codon-stop (X)**.

## Nomenclature des modifications de nucléotides:

**-Les délétions:** désignées par l'abréviation « **del** »

\* **76\_78delACT** (ou 76\_78del) signifie une délétion de 3 nucléotides (ACT) des positions 76 à 78.

**-Les insertions:** désignées par l'abréviation « **ins** »

\***76\_77insT** signifie l'insertion du nucléotide T entre les nucléotides 76 et 77.

## Nomenclature des modifications d'acides aminés:

**Les délétions:** désignées par l'abréviation « **del** »

\* **K29del** (ou Lys29del) signifie une délétion de l'acide aminé lysine (Lys ou K) en position 29.

\* **C28\_M30del** (ou Cys28\_Met30del) signifie une délétion de 3 acides aminés de la cystéine en position 28 à la méthionine en position 30.

**Les insertions:** désignées par l'abréviation « **ins** »

\***K29\_M30insQSK** (ou Lys29\_Met30insGlnSerLys) signifie l'insertion de 3 acides aminés (QSK ou glutamine/lysine/serine) entre les acides aminés lysine (K) en position 29 et méthionine (M) en position 30.

Ainsi la séquence ...**CKM**GH... devient ...**CKQSKM**GH...

## Nomenclature des modifications de nucléotides:

### Variabilité des séquences répétées courtes:

Soit la séquence 5'-**AC TG TG TG** CC avec une séquence répétée TG (ici 3 fois) et dont le nombre de variation va de 3 à 6.

Si le nucléotide **A** est en position 1991, la séquence répétée variable sera désignée **1993(TG)3-6**.

## Nomenclature des modifications d'acides aminés:

### Variabilité des séquences répétées courtes:

Soit la séquence ...MKMGHQQQCC... avec l'acide aminé Q (Gln) en position 6 et dont le nombre de variation va de 3 à 6 fois.

La variabilité sera indiquée **Gln6(3-6)**.



## 4- LES MUTATIONS DU GENOME HUMAIN

**Tableau 2.2. Mutations rapportées dans la littérature (2093 gènes analysés)**  
(référence : <http://www.hgmd.cf.ac.uk/docs/hohoho.html>)

Mutation	Nombre	%
<b>Microlésions :</b>		
- mutations ponctuelles (non-sens, faux-sens)	31 251	57,3
- épissage	5188	9,5
- zones de régulation de l'expression	700	1,3
- petites délétions	9014	16,5
- petites insertions	3626	6,6
- petites insertions/délétions	533	1
<b>Macrolésions :</b>		
- variation de séquences répétées	143	0,2
- grandes insertions ou duplications	470	0,9
- réarrangements complexes (incluant les inversions)	639	1,2
- grandes délétions	2998	5,5
Total	54 562	100

**>90%**

**<10%**

*Mutation= variation allélique anormale pouvant aboutir au développement d'une maladie*

*Les SNP sont les variants les plus fréquents dans le génome humain*

*2 types de mutation:  
-Macrolésions (grande taille)  
- microlésions (petites taille)*

## LES MACROLESIONS (<10% des mutations)

- Ce sont des réarrangements géniques dont la taille peut aller de **quelques kpb** à **plusieurs Mpb**.

Plusieurs **mécanismes** possibles:

\*Variation de **séquences répétées** (mutations de réplication) **(0,2% des mutations)**

\* **grande insertion, duplication, inversion, délétion**  
=> réarrangements géniques complexes **(environ 8% des mutations avec majorité de grandes délétion)**

# MACROLESIONS: grande insertion, duplication, inversion, délétion

-Ce sont des mécanismes de réarrangements géniques **complexes**

- les réarrangements peuvent être soit **interchromosomiques** ou intrachromosomiques (**inter ou intrachromatides**)

- via **mécanismes de recombinaisons (homologue ou non):**

\* au sein de famille de gènes regroupés en groupe (cluster)

\* au sein de séquences répétées

- **La duplication d'un gène (tout ou en partie)** peut se faire dans la même orientation ou en orientation inversée

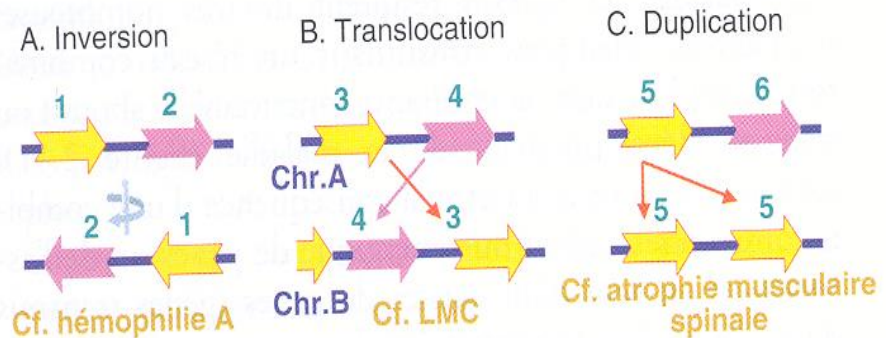
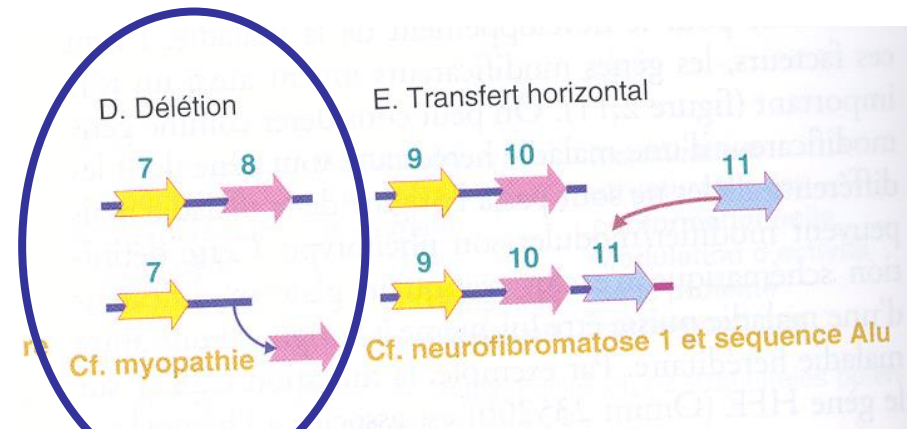


Figure 2.12 Réarrangements géniques -Exemples.



**>5% des mutations**



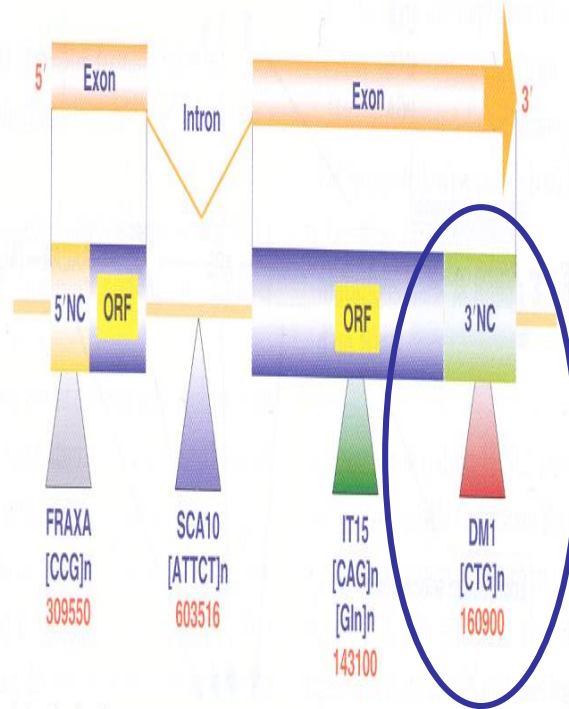
# MACROLESIONS: Variation de séquences répétées = Mutations de réplication ou dynamiques (=> n'importe où dans gène)

-Ce sont en général des erreurs de réplication mitotique ou préméiotique.

- Elles résultent d'un **dérage (slipped strand mispairing) dans certaines zones instables du génome** (comme les zones répétées).

-Ce mécanisme est retrouvé dans des maladies génétiques au cours desquelles le phénomène d'anticipation observé est le résultat de l'instabilité de la répétition de trinuécléotides ou de dinuécléotides (STR=microsatellites). A partir d'un certain seuil (nbre de trinuécléotides donné), il y a pathologie.

**Figure 2.24** Exemples de localisation d'expansion de séquences répétées dans le génome. Le nombre indiqué en dessous de chaque variation de gène est le numéro Omim. 5'NC : région 5' non codante ; 3'NC : région 3' non codante ; ORF (open reading frame) : partie codante d'un gène ; FRAXA (Fragile site mental retardation 1) : syndrome de l'X fragile (le gène muté est FMR1) ; IT15 (appelé aussi huntington) : gène muté dans la chorée de Huntington ; SCA10 (spinocerebellar ataxia) : gène muté dans l'ataxie spinocérébelleuse ; DM1 (dystrophia myotonica 1) : gène muté dans la maladie de Steinert (dystrophie myotonique).



## Dystrophie myotonique de Steinert (Omim 160900)

-**Maladie héréditaire du métabolisme de transmission autosomique dominant**

-Conséquence d'une expansion anormalement élevée du triplet CTG en 3'NC du gène DMPK (dystrophia myotonica protein kinase)= **mutation de réplication ou dynamique**

En condition normale (5-30 triplet en 3'NC du gène DMPK), une protéine se fixe sur les triplets de l'ARNm pour réguler l'expression et la localisation cellulaire de la protéine.

-chez sujet malade jusqu'à 1000 triplets. Hypothèse: Il en résulterait Une modification de structure de l'ARNm qui induirait un **effet négatif** sur l'expression de la protéine DMPK et **sur son exportation hors du noyau**.

## LES MICROLESIONS (>90% des mutations)

- Ce sont des **substitutions, suppressions ou additions d'un ou de quelques nucléotides**:

1- Quand une base est substituée en une autre, on parle de **mutation ponctuelle**:

- **Les mutations ponctuelles** (non-sens et faux-sens) **sont les plus courantes** des mutations rapportées dans la littérature (environ **57%**).

- De nombreux **mécanismes** sont responsables des mutations ponctuelles:

\* la dépurination

\* la désamination notamment des cytosines méthylées (metC en T)

\* des erreurs de réplication ou de réparation

\* transition: changement d'une base purique (ou pyrimidique) en une autre (G en A ou C en T)

\* transversion: changement d'une base purique en pyrimidique (G en T) (ou réciproquement par exemple C en A)

2 - délétion, insertion et délétion/insertion de quelques nucléotides sont des **mutations non ponctuelles (environ 25% des mutations)**



## -Mutations **ponctuelles** dans les régions codantes:

### - mutations faux-sens

\*induisent un **changement d'AA** sur la protéine

\*conséquence sur le **fonctionnement de la protéine** est variable

\*dans certains cas peut induire une **anomalie d'épissage** (saut d'exon par abolition fonctionnelle d'un site ESE ou modification structure 2aire ARNm)

### -mutations silencieuses

\***modifie le codon mais pas l'AA**

\*pas de conséquence sur la protéine

\*Cependant peuvent être responsables d'un **épissage anormal** de l'exon (saut d'exons si mutation située sur site favorisant [ESE] ou inhibant [ESS] l'épissage).

### -mutations non-sens

\*aboutissent à la **formation d'un codon stop** (UAA, UAG ou UGA)

\*responsables :

- d'une **protéine tronquée** (le + souvent non fonctionnelle)

- de la **dégradation de l'ARNm** correspondant (et donc absence de synthèse de la protéine anormale)

- dans certains cas **d'épissage anormal** d'exon (« saut d'exon »)

## -Mutations **non ponctuelles** dans les régions codantes:

- L'insertion ou la délétion d'une ou plusieurs bases dans la région codante :

\*peut **décaler le cadre de lecture ouvert** (open reading frame).

\*provoque l'apparition d'un codon stop prématuré plusieurs bases (une dizaine à une centaine en général) en aval de la mutation.

-Principales conséquences:

\*formation **protéine tronquée** (plus petite)

\***dégradation de l'ARNm** correspondant

**MICROLESIONS dans les régions codantes**

# MICROLESIONS perturbant l'épissage

=> Représentent 10-15% des anomalies responsables de maladies génétiques.

## Mutations d'épissage (splicing mutations)



### A- CAUSES :

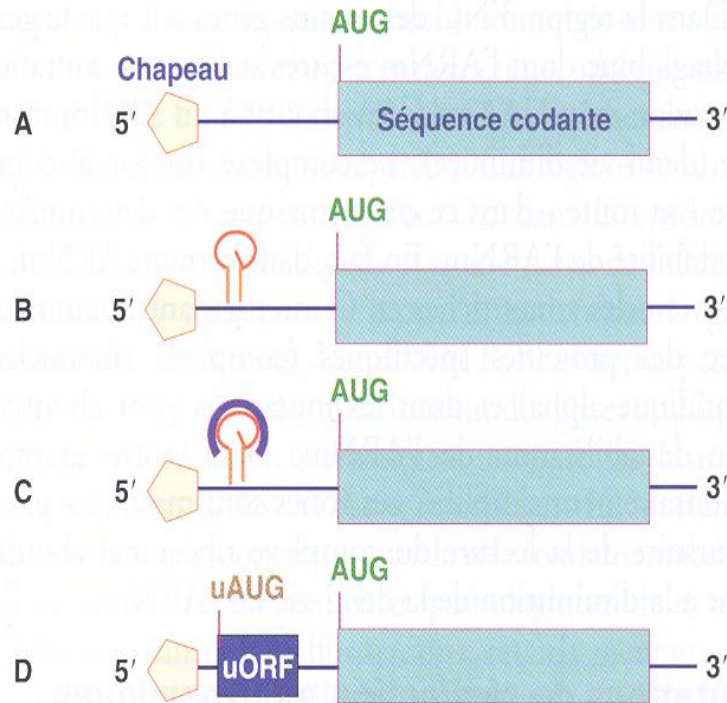
- Sont le plus souvent des mutations ponctuelles
- Microlésion dans site accepteur 3'ag ou donneur 5'gt sera responsable de **saut d'exon** (60% des cas de saut d'exons)
- microlésion dans site de branchement peut induire une **rétenction d'intron** (une portion d'intron sera retrouvée sur le transcrit mature)
- microlésion dans un exon peut créer (et activer) un site cryptique (atypique) d'épissage (SCE) ou inactiver un ESE. Il en résultera un saut d'exon

### - CONSEQUENCES:

- Le plus souvent, aboutissent à la formation d'un codon stop prématuré (**protéine tronquée**) et à la **dégradation de l'ARNm** par nonsense mediated mRNA decay (NMD).

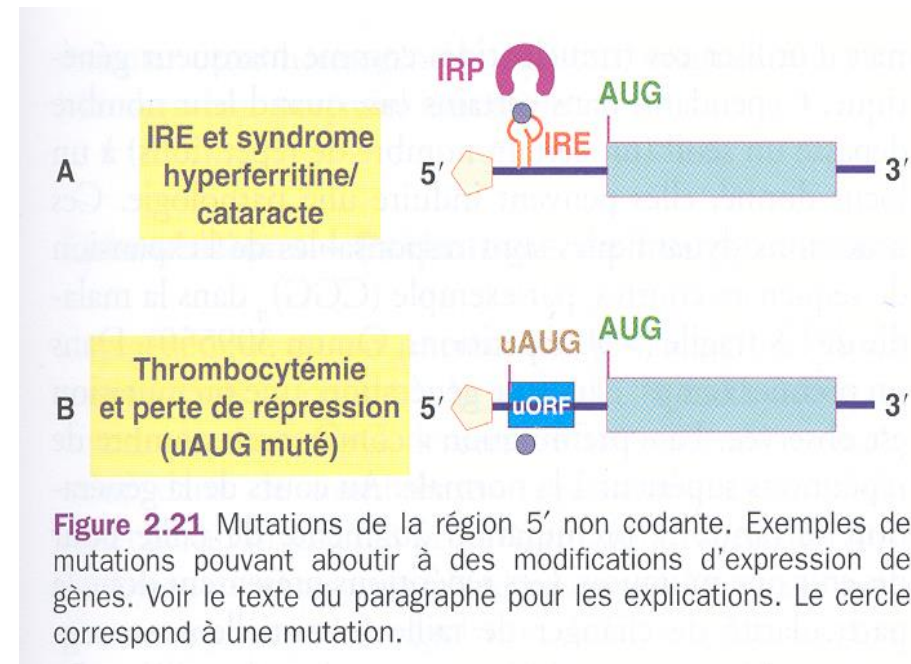
# MICROLESIONS dans la région 5'NC (non codante)

## 1- Régulations négatives de la traduction



**Figure 2.22** Régulation négative de la traduction. Exemples de mécanismes dans la région 5' non codante de l'ARNm. **A.** ARNm sans régulation négative. **B.** Séquence en boucle stable (*hairpin*) située entre l'extrémité 5' (chapeau) et le codon AUG. Cette boucle interfère dans l'assemblage du complexe de pré-initiation et/ou le scanning du ribosome à la recherche de l'AUG. Cette séquence en boucle peut être suffisamment stable pour empêcher l'activité de déroulement de l'hélicase et inhiber le scanning ribosomal. **C.** Sur certains ARNm, la séquence en boucle peut être stabilisée par des protéines. **D.** La partie 5'NC peut posséder des codons AUG alternatifs (uAUG) en amont du codon AUG principal ou physiologique.

## 2- Exemples de pathologies génétiques



**Figure 2.21** Mutations de la région 5' non codante. Exemples de mutations pouvant aboutir à des modifications d'expression de gènes. Voir le texte du paragraphe pour les explications. Le cercle correspond à une mutation.

**Environ 1,5% des mutations**

# MICROLESIONS (ET MACROLESIONS) dans la région 3'NC

Une mutation dans cette région ( environ 1,5% des mutations) peut *modifier l'expression des gènes via différents phénomènes*

Surexpression protéine CCND1 (Omim 168461) dans certaines formes de lymphomes

Mutation impliquant disparition codon stop (peut diminuer la 1/2 vie de l'ARNm; ex: alphaglobine)

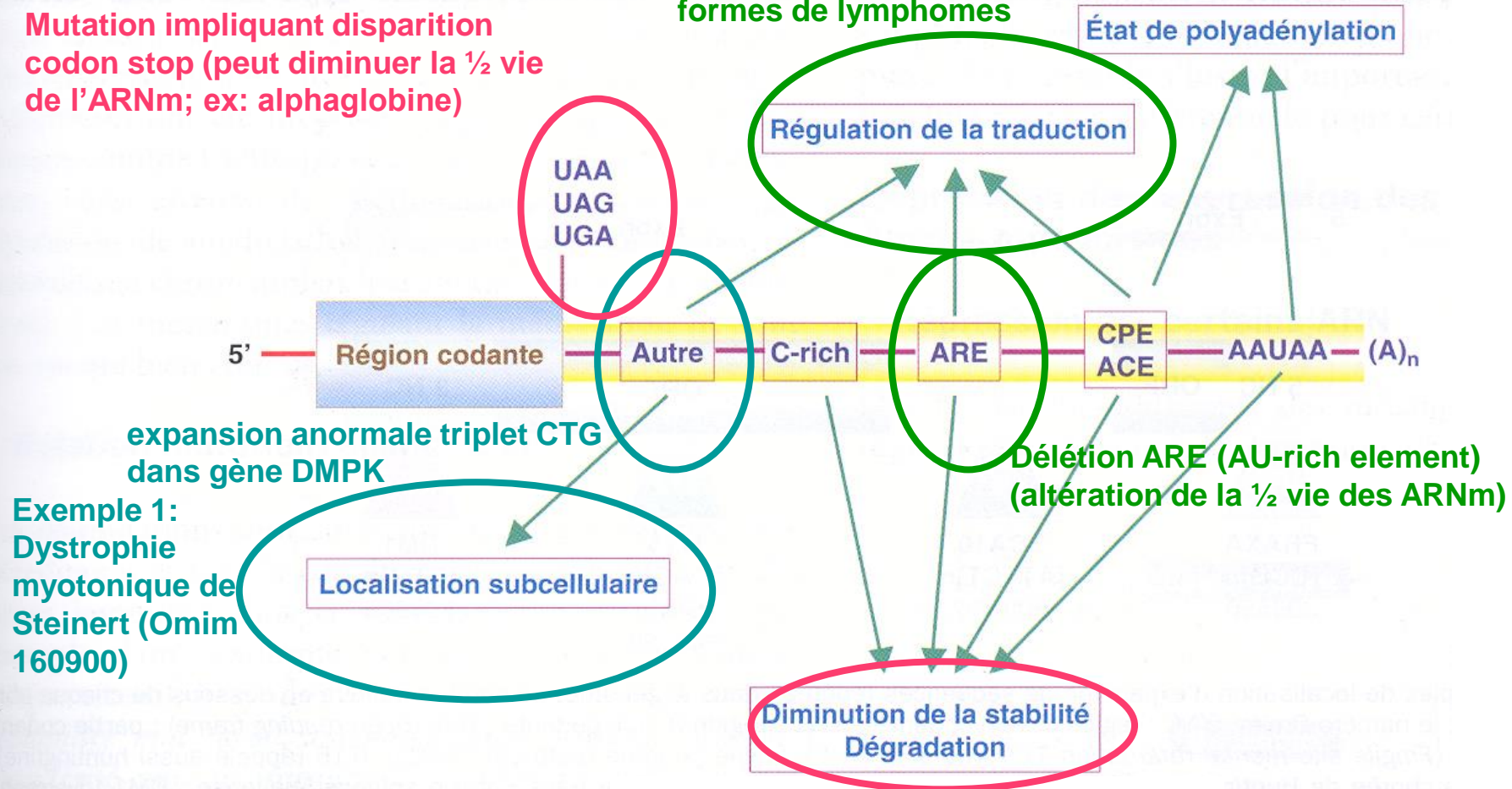


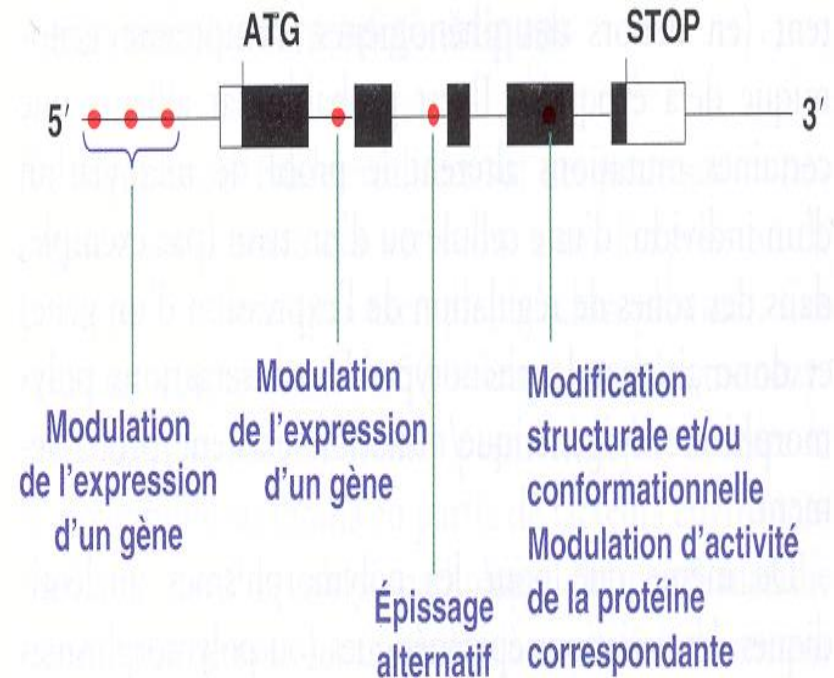
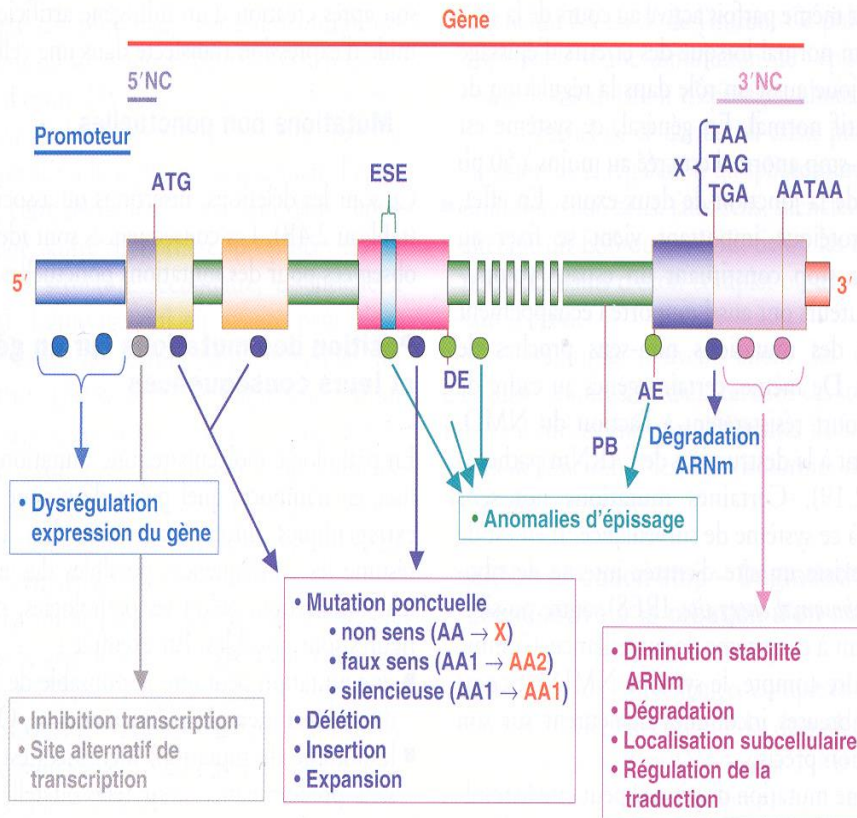
Figure 2.23 Mutations de la région 3' non codante.



## 5- POSITION DES MUTATIONS (macro et microlésions) SUR UN GÈNE ET LEURS CONSÉQUENCES

-En pathologie moléculaire, une mutation peut être localisée en n'importe quel point d'un gène ou des régions extragéniques interagissant avec lui.

- Les conséquences de ces mutations selon leur localisation sont variables



**Figure 2.10** Conséquences fonctionnelles et/ou structurales potentielles des polymorphismes fonctionnels.



## **RESUME DES Mutations induisant la création d'un codon stop prématuré (CSP):**

**Nombreux mécanismes génèrent l'apparition d'un CSP:**

- mutations ponctuelles non-sens
- décalage du cadre de lecture
- réarrangement génique (macrolésions)
- épissage anormal (rétention d'intron, saut d'exon, activation site cryptique d'épissage)
- erreur de réplication, de transcription
- utilisation de sites d'initiation mineurs de la traduction (uAUG)

**Un codon stop anormal peut amener à la formation d'une protéine (tronquée) non fonctionnelle ou aux conséquences délétères**

**-Le système de surveillance **NMD** permet de dégrader ces ARNm anormaux**

# RESUME CONSEQUENCES DES MUTATIONS

- **Une mutation est le plus souvent responsable :**
- d'une anomalie de synthèse protéique (diminution/absence)
- d'une modification structurelle (protéine instable)
- d'une anomalie de ciblage (membrane, organelle)
- d'une protéine non (ou peu) fonctionnelle

(Si la mutation induit des modifications mineures au niveau de la protéine, on peut s'attendre à des conséquences modérées. C'est ce que l'on peut observer dans le cas de certaines microlésions [faux-sens, petites délétions ou insertions])

## **-Les mutations en cause sont essentiellement:**

- mutations non-sens (apparition d'un codon stop prématuré)
  - mutations faux-sens (changement d'AA)
  - celles qui induisent décalage du cadre de lecture (frame-shift)
  - celles qui induisent une anomalie d'épissage (splicing defect)
  - celles qui induisent une délétion ou une insertion de grande taille
- Une mutation épigénétique (**épimutation**) peut également induire une modification de la chromatine ou de l'état de méthylation d'un gène.  
Conséquences sur l'expression du gène.

# LA RELATION GENOTYPE-PHENOTYPE EST COMPLEXE

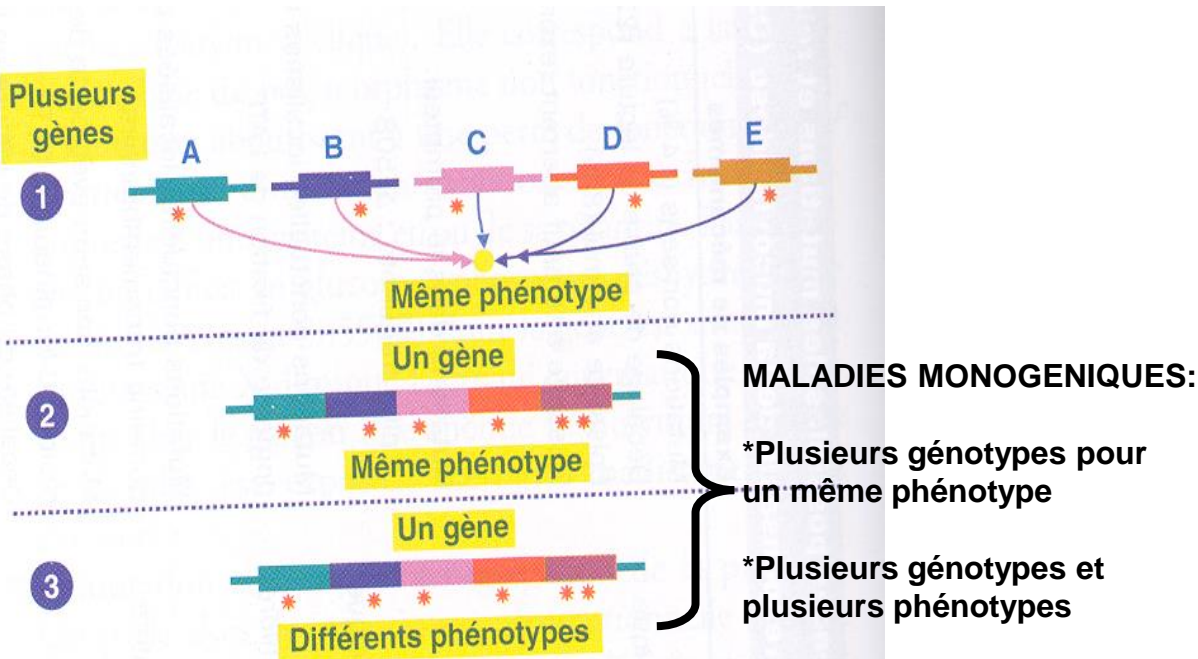
Selon les pathologies, **plusieurs cas de figures de localisation des mutations** sont possibles.

- **une mutation peut être responsable de la quasi-totalité des cas** (ex: drépanocytose, Omim 693903)

- **le nombre de mutations identifiées est très important:**

\* **sans prédominance pour telle ou telle mutation.** Il n'y a pas de point chaud (*hotspot*) de mutation (ex: porphyrie aiguë intermittente, Omim 176000)

\* **avec prédominance d'une mutation sur les autres.** Il y a un point chaud de mutation (ex: mucoviscidose; sur les >1200 mutations répertoriées, la mutation deltaF508 représente 70% des cas; Omim 219700)



**Figure 2.27** Hétérogénéités génotypique et phénotypique. Selon les gènes et les mutations, les conséquences phénotypiques peuvent être différentes. 1. Une même maladie peut être la conséquence de différents gènes : par exemple, syndrome du long QT (Omim 192500, 152427, 603830, 600163, 176261, 600681, 600919, 603796). 2. Plusieurs mutations sur un même gène sont responsables d'une même maladie (par exemple, porphyrie aiguë intermittente, Omim 176000). 3. Selon la localisation de la mutation, la maladie n'est pas la même. Par exemple, les mutations dans le proto-oncogène RET (Omim 164761). Chaque rectangle représente un gène en 1, un exon en 2 et en 3. Chaque étoile représente une mutation différente.



**Tableau 2.5. Conséquences possibles des mutations et relation génotype/phénotype – Exemple de mutations sur le gène CFTR (Omim 602421) responsable de la mucoviscidose (Omim 219700), maladie de transmission autosomique récessive**

Mutations (> 1000)	Exemples de mécanismes	Conséquence(s) biologique(s)	Relation génotype/phénotype
Classe I Anomalies de la synthèse protéique	Mutations non-sens (14 %) Décalage de phase de lecture (22 %) Épissage anormal (8 %) Épissage alternatif aberrant (exon 9)	Synthèse anormale de CFTR Protéine tronquée ARNm détruit Anomalies d'épissage partiel	Phénotype sévère  Phénotype corrélé au niveau d'exon 9 non épissé
	Mutations dans le promoteur	Diminution de la quantité de CFTR synthétisé Réduction de synthèse du CFTR	
Classe II Anomalies de la maturation protéique	La plus fréquente : $\Delta F508$	Localisation anormale de la protéine	Sévère
Classe III Régulation/Ouverture du canal chlore	Mutations ponctuelles localisées dans le domaine de fixation de l'ATP	Réponse diminuée ou absente après stimulation par l'AMPc (localisation correcte)	
Classe IV Anomalies de conductance du chlore	Mutations ponctuelles localisées dans le domaine transmembranaire	Variations anormales de l'activité globale du canal chlore	Phénotype modéré
Classe V Instabilité de la protéine	Mutations non-sens avec perte des 70–98 résidus C-terminaux du CFTR Décalage du phase de lecture en C-terminal (70–98 résidus)	Protéine labile – Demi-vie diminuée	Phénotype modéré
En général	Grande variabilité y compris intrafamiliale Phénotype corrélé au taux d'épissage anormal (pour un même génotype) Absence d'insuffisance pancréatique (environ 8 % des patients) : au minimum un allèle modérément pathologique (classe IV, classe I avec épissage alternatif variable, classe V) Insuffisance pancréatique : deux allèles sévères (classes I, II et III) Atteinte pulmonaire variable même pour une même mutation		
Cas particuliers	Deux mutations en cis (sur le même chromosome) associées à une mutation sur l'autre chromosome		Phénotype variable, une mutation pouvant moduler l'effet de l'autre mutation
Polymorphismes fonctionnels	Modulation d'épissage Diminution relative de la transcription	Effet synergique avec la ou les mutations présentes	
Gènes modificateurs	Influence du terrain génétique (« fond génétique ») Gènes modificateurs à identifier		
Autres pathologies associées (exemples)	Absence bilatérale congénitale du vas deferens (CABVD)	70 % : une mutation trouvée 10 % : deux mutations trouvées Association fréquente d'une mutation à l'épissage alternatif de l'exon 9 (hétérozygote composite)	Pas de mucoviscidose
	Pancréatite	Mutation CFTR et/ou épissage alternatif de l'exon 9	

# B- OUTILS MOLECULAIRES D'IDENTIFICATION DES MUTATIONS

## 1- Basés sur l'amplification génique

- Rappels sur la méthode de base (PCR classique)
- la PCR-RFLP (détection mutation touchant un site de restriction)
- la SSCP (Single Strand Conformational Polymorphism)
- la DGGE (Denaturing Gel Gradient Electrophoresis)
- DHPLC (Denaturing High Performance Liquid Chromatography)
- La PCR en temps réel (Real Time PCR) et HRM (Analyse des courbes de fusion à haute résolution)
- la MLPA (Multiplex Ligation dependant Probe Amplification)

## -2- Basés sur le séquençage d'ADN

- Rappels sur la méthode de base de séquençage d'un ADN (Sanger & Coulson)
- Le pyroséquençage et la SMRT

## 3- Basés sur l'hybridation moléculaire

- Southern blot
- puces à ADN



**Les polymorphismes (micro- et macrolésions) peuvent être exploités par séquençage, Southern blot et PCR** (à la fois par méthodes de base ou dérivées):

Mutations ponctuelles (microlésion)  
- SNP/RFLP

Polymorphismes de répétition (macrolésion)  
- VNTR/STR

Délétions/insertion  
(micro- ou macrolésions)

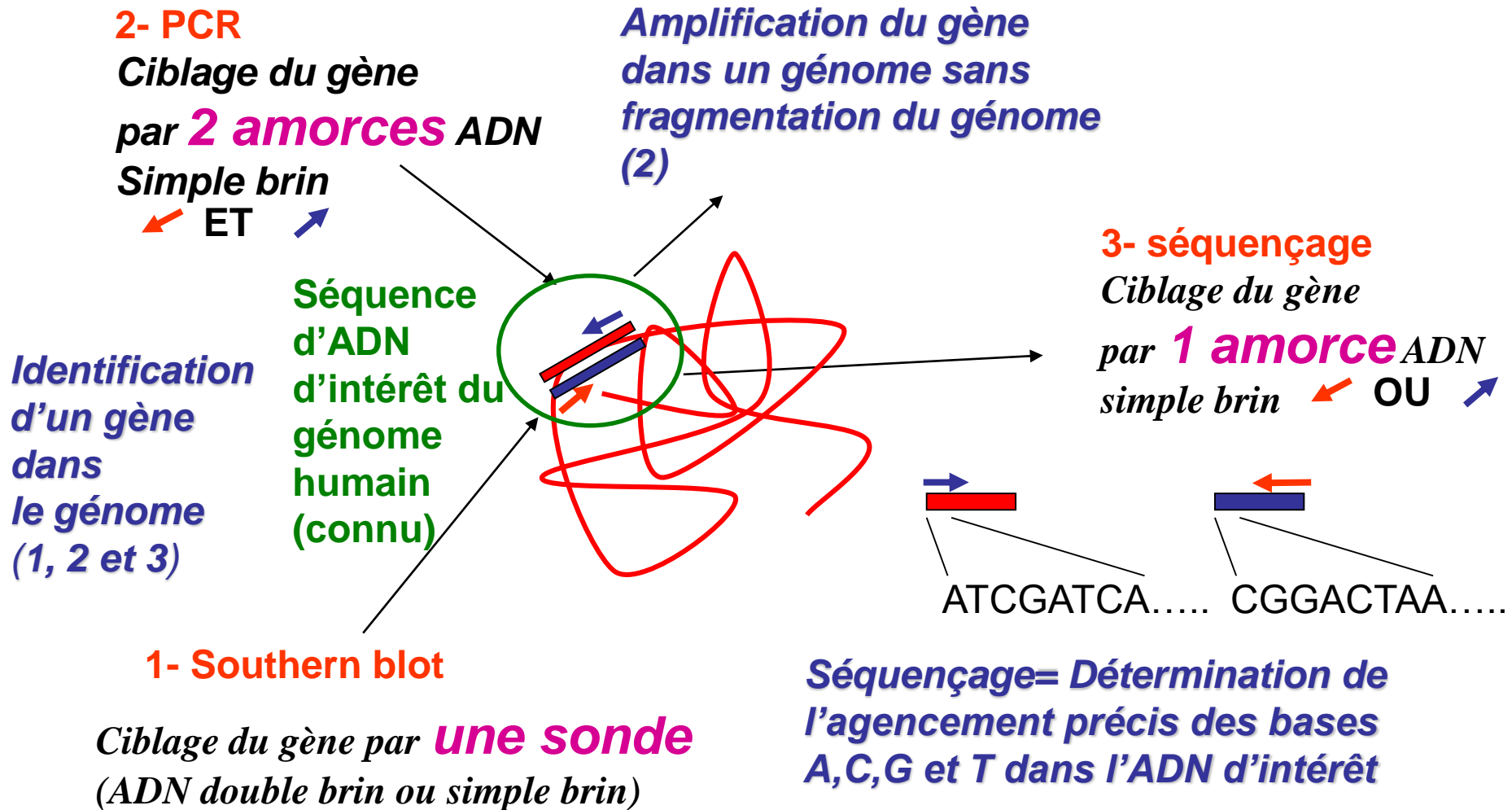
- Pour explorer et cibler une lésion (génique ou non génique) dans le génome diploïde d'individus potentiellement porteurs de cette lésion, il faut disposer d'une sonde (Southern blot) ou d'amorces (PCR et séquençage) spécifique de cette lésion.

- dans le cas d'une exploration de SNP (RFLP ou non) on parle de polymorphisme bi-allélique (2 allèles possibles) et uniquement 3 génotypes (N/N, N/M et M/M) peuvent être détectés dans le génome diploïde

- dans le cas d'une exploration de VNTR ou STR (séquences répétées) on parle de polymorphisme multi-allélique (n allèles possibles) et nombreux (>n) génotypes peuvent être détectés dans le génome diploïde au sein d'individus d'une même espèce.

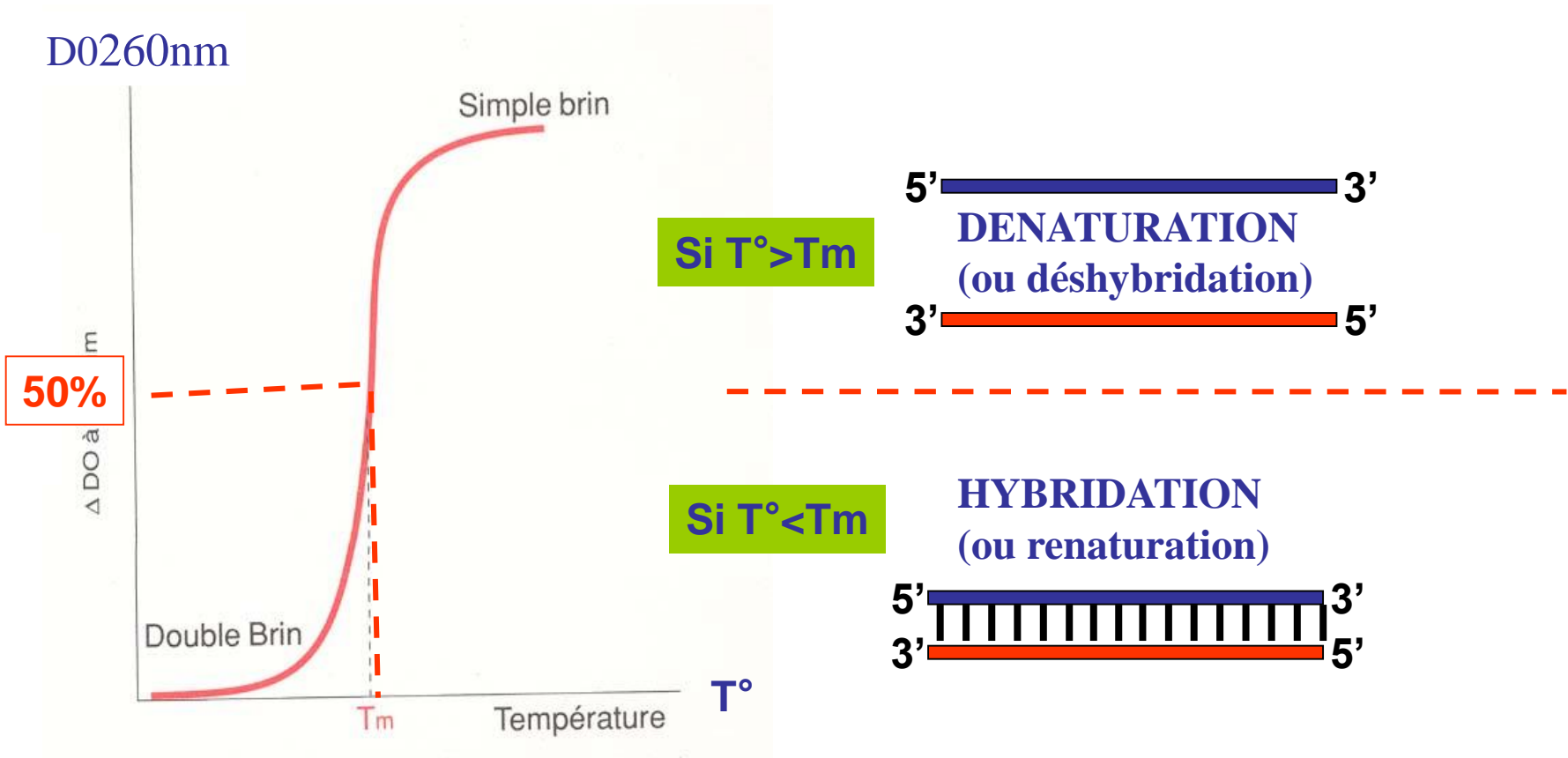
# METHODES DE BASE DE BIOLOGIE

**MOLECULAIRE** (Appliquées à l'identification, l'amplification et le séquençage d'un gène ou d'un génome)



→ **LES 3 APPROCHES FONT APPEL A UNE HYBRIDATION (d'une sonde ou d'une amorce sur l'ADN matrice)**

# Notions d'hybridation moléculaire ( $T_m$ )

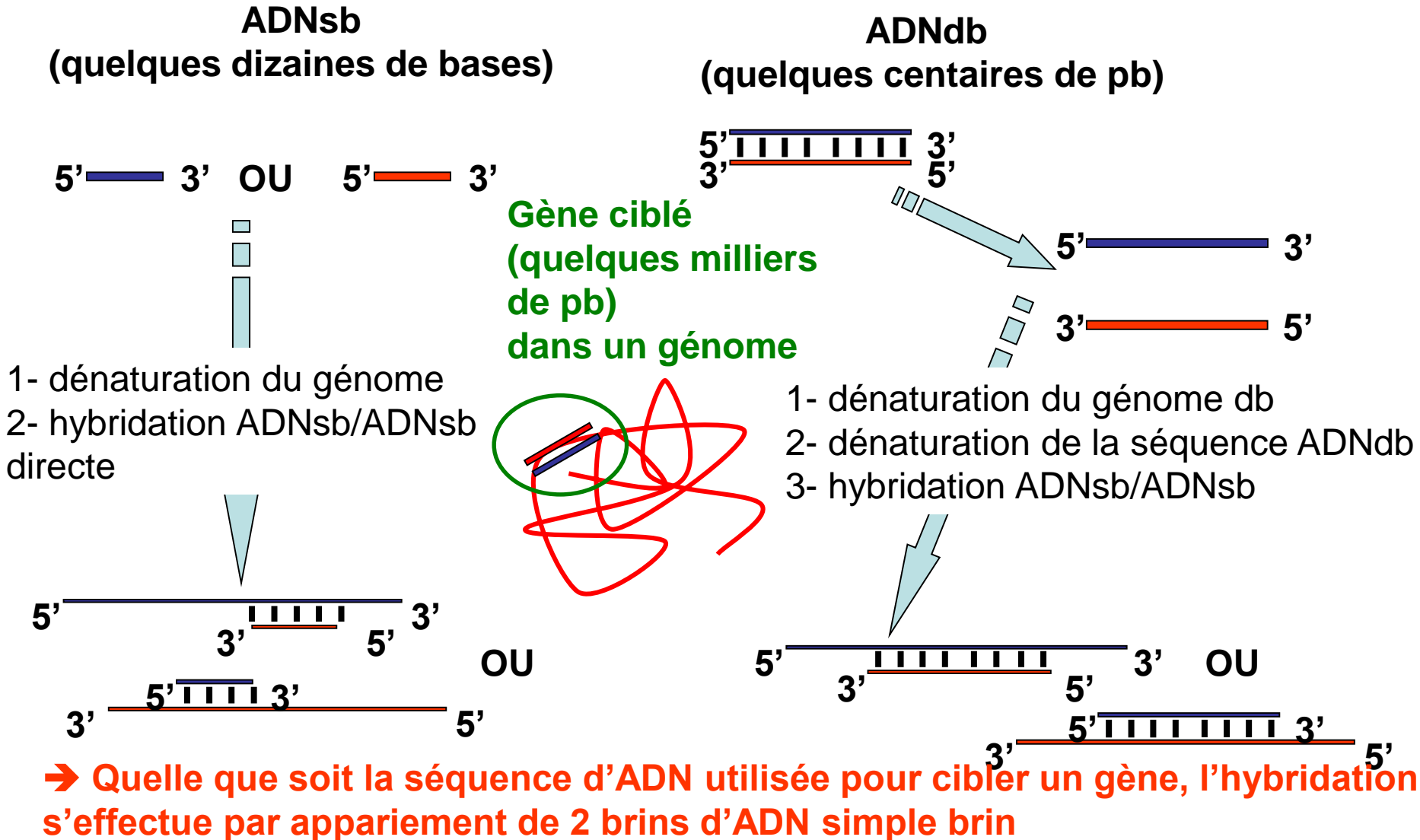


**$T_m$  (melting  $T^\circ$  = point de fusion) d'un génome (ou d'un grand fragment d'ADN)**

**$T_m = 69.4 + 0.41x(\%GC)$**

# Notions d'hybridation moléculaire

Les séquences ADN utilisées pour cibler un gène (ou toute autre séquence d'ADN) :  
=> **Sondes et Amorces**



# B- OUTILS MOLECULAIRES D'IDENTIFICATION DES MUTATIONS

## 1- Basés sur l'amplification génique

- Rappels sur la méthode de base (PCR classique)
- la PCR-RFLP (détection mutation touchant un site de restriction)
- la SSCP (Single Strand Conformational Polymorphism)
- la DGGE (Denaturing Gel Gradient Electrophoresis)
- DHPLC (Denaturing High Performance Liquid Chromatography)
- PCR en temps réel (Real Time PCR) et HRM (Analyse des courbes de fusion à haute résolution)
- la MLPA (Multiplex Ligation dependant Probe Amplification)

## -2- Basés sur le séquençage d'ADN

- Rappels sur la méthode de base de séquençage d'un ADN (Sanger & Coulson)
- Le pyroséquençage et la SMRT

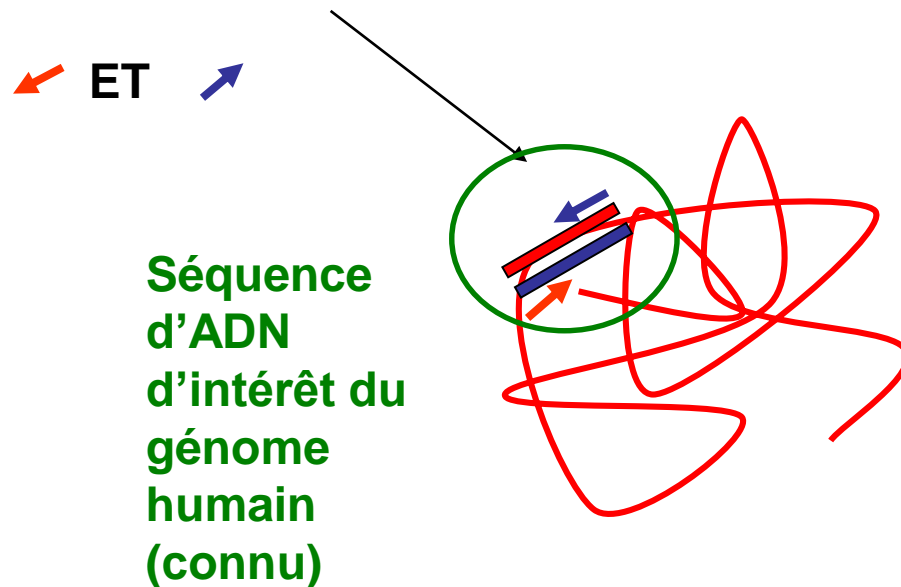
## -3- Basés sur l'hybridation moléculaire

- Southern blot
- puces à ADN



# 1- IDENTIFICATION DE MUTATION PAR PCR (POLYMERASE CHAIN REACTION) ET METHODES DERIVEES

*Ciblage du gène (ou de toute séquence d'ADN)  
par **2 amorces** ADN simple brin*



**Objectif:** Amplification du gène dans un génome sans fragmentation du génome

**Applications multiples:**

- Identification de mutations dans un génome
- Identification d'un agent infectieux chez son hôte
- Empreintes génétiques
- etc

# Notions d'hybridation moléculaire (**les amorces en PCR**)

## UNE AMORCE :

Séquence synthétique d'ADN simple brin courte (20-30 bases le plus souvent)

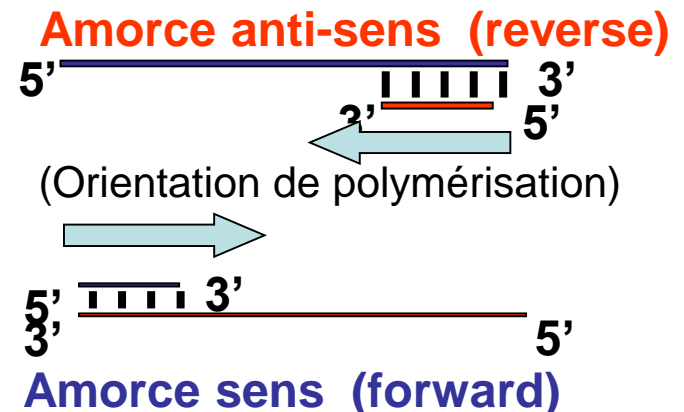
-Moyen de détection très spécifique d'un gène dans un génome (par **PCR** avec 2 amorces).

-Hybridation de l'amorce sur la matrice si et seulement si l'hybridation est effectuée à une  $T^\circ$  inférieure à la  $T_m$  de l'amorce. La  $T_m$  approximative d'une amorce se calcule par la formule  $4GC + 2AT$

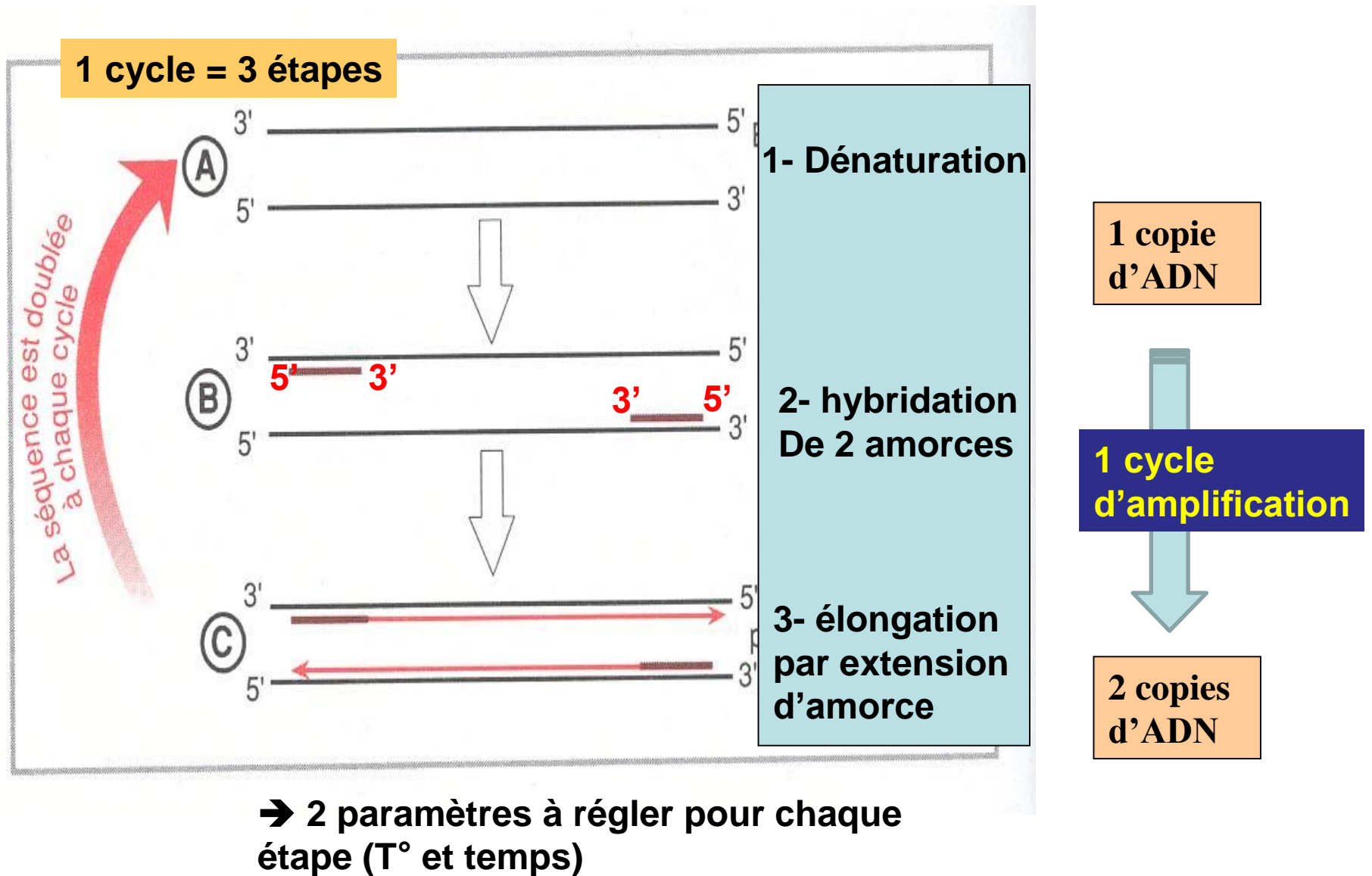
Ex: 5'-CATAGATGT**CGAATATAGGTGCGT**-3'  $\rightarrow T_m = (4 \times 10) + (2 \times 14) = 68^\circ\text{C}$

- Par convention la lecture de la séquence se fait de gauche à droite et correspond à son orientation 5'=>3'

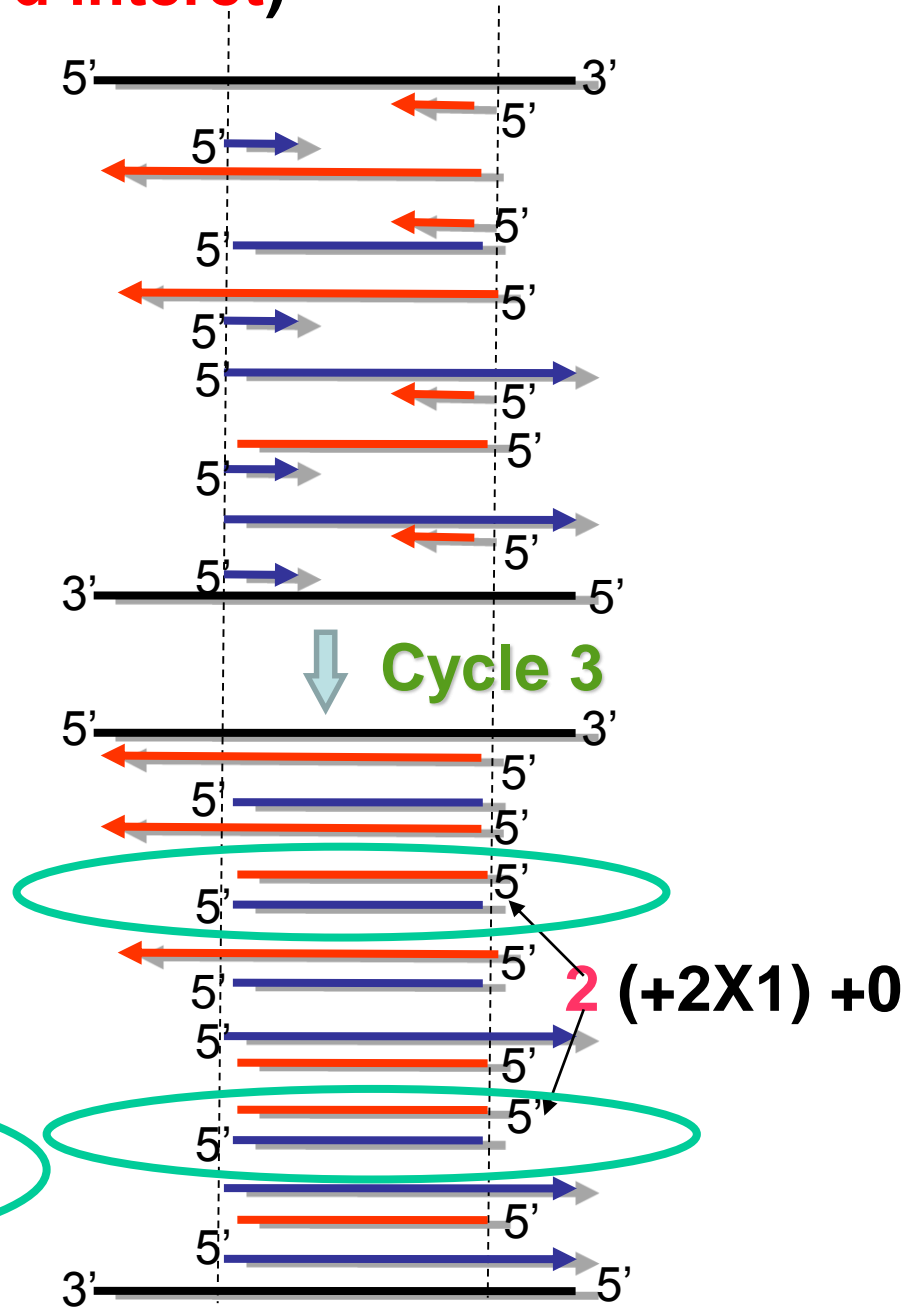
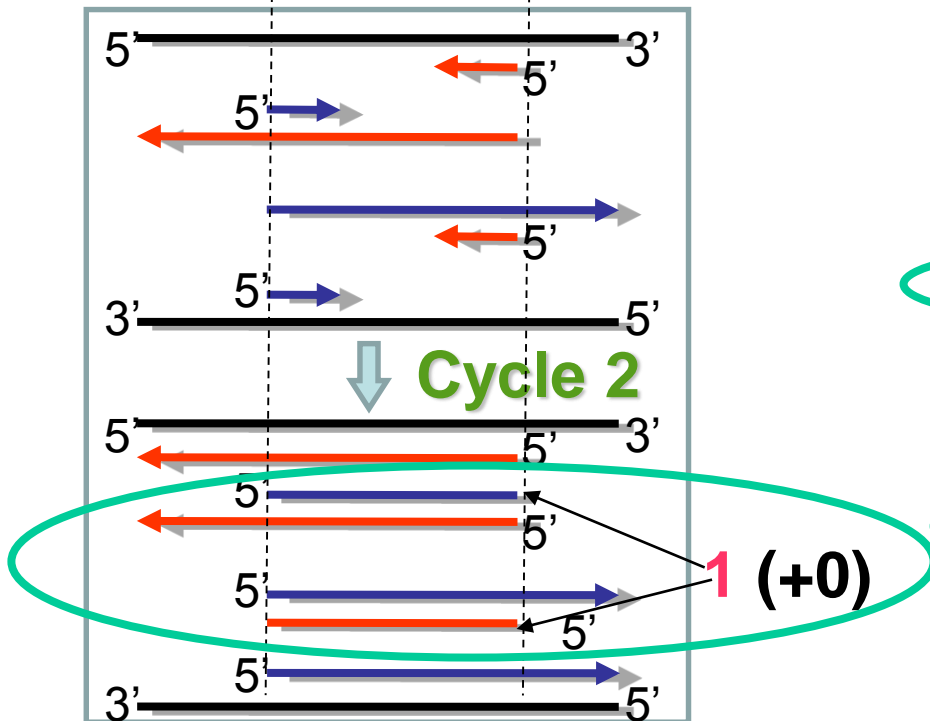
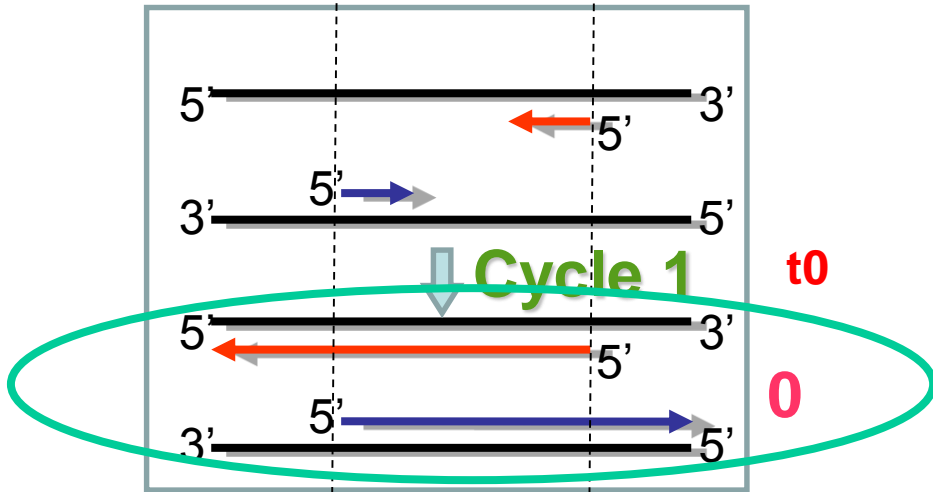
**ATTENTION AU SENS DE LECTURE** (de 5' vers 3') de l'amorce antisens en PCR (notamment lors de sa synthèse)



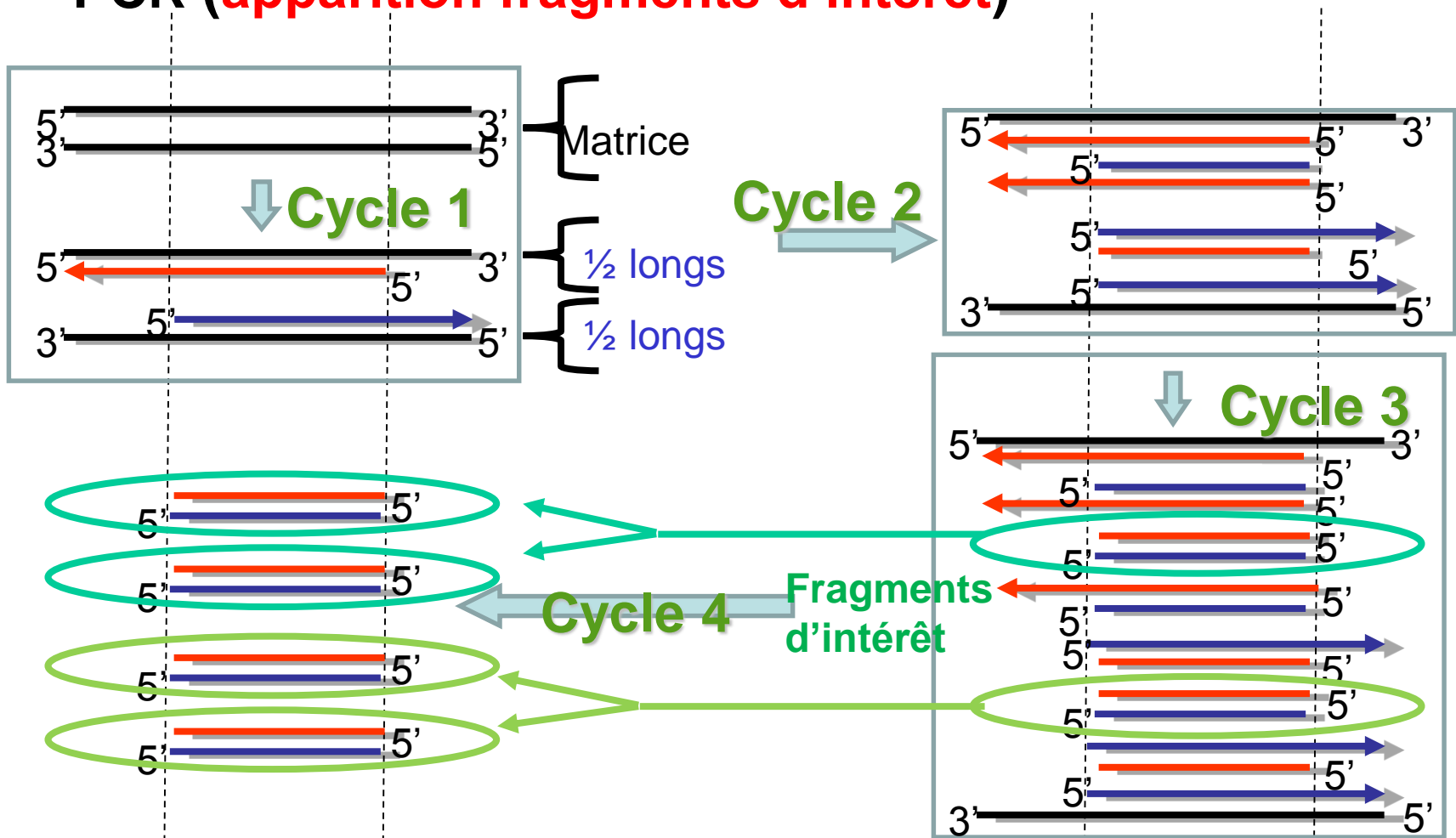
# la PCR (**principe de base**)



# PCR (apparition fragments d'intérêt)



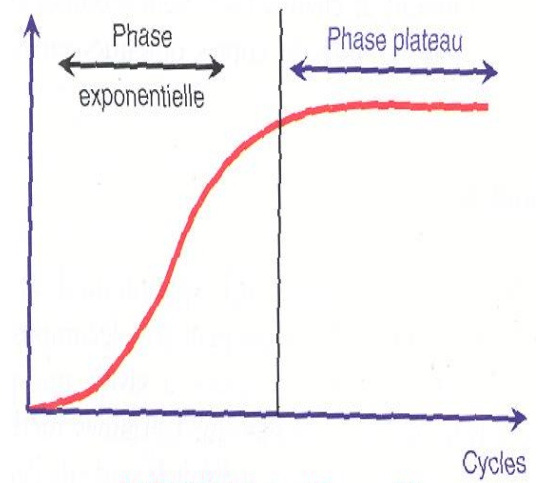
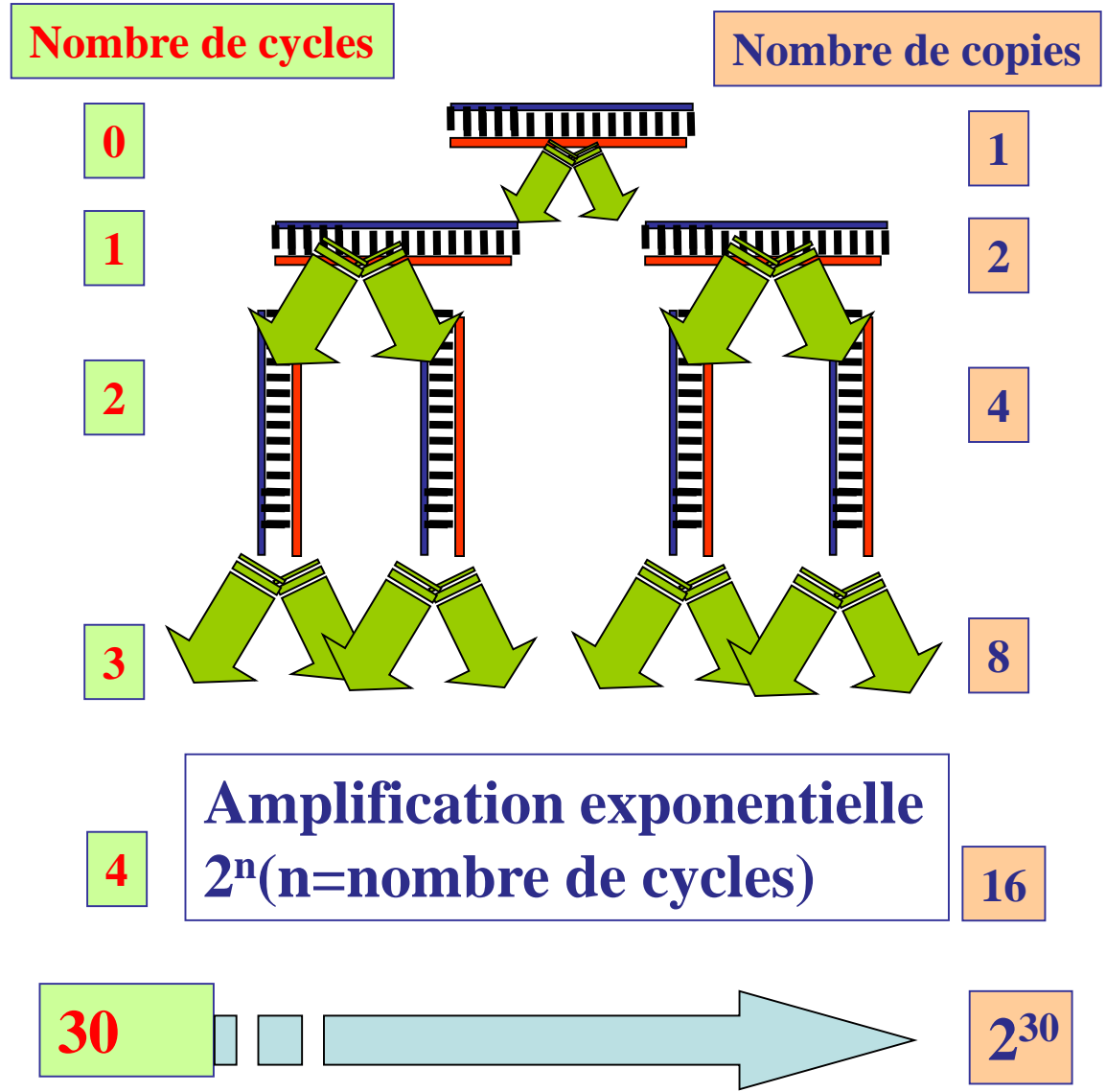
# PCR (apparition fragments d'intérêt)



=> Après quelques cycles, les doubles brins d'intérêt (compris entre les 2 amorces) vont devenir très majoritaires.  
La quantité des brins  $\frac{1}{2}$  longs issus des brins matriciels et les brins matriciels eux-mêmes sera négligeable



# Facteur d'amplification:



## La PCR restriction (PCR-RFLP)

= PCR classique pour l'analyse d'un SNP touchant un site de restriction (RFLP)

### PRINCIPE

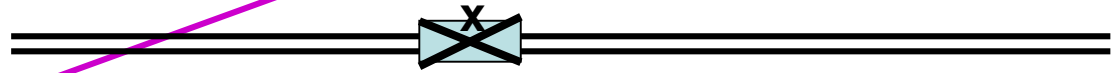
Une mutation peut faire apparaître ou disparaître un site de restriction enzymatique sur l'un des allèles d'une séquence.

exemple : analyse d'un sujet hétérozygote pour une mutation.

Allèle paternel

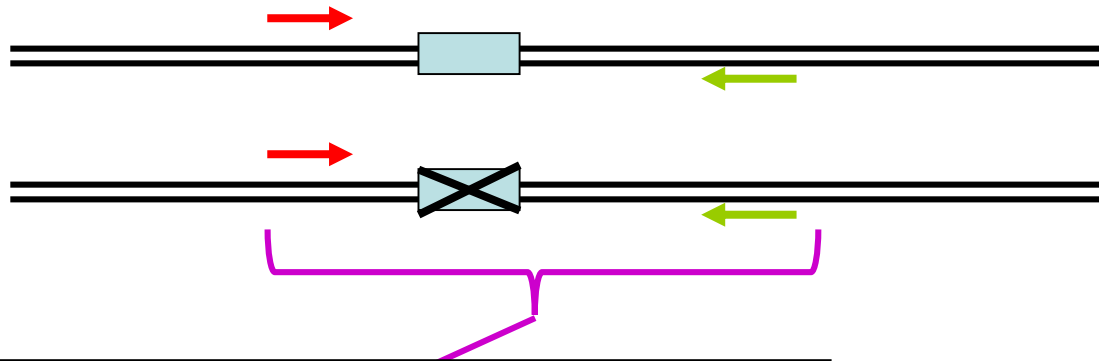


Allèle maternel

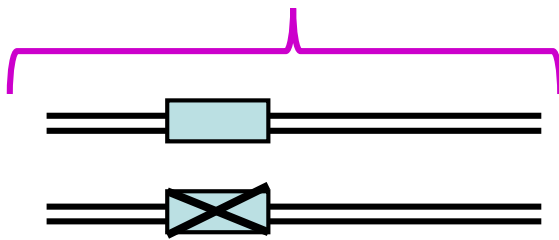


Site de restriction **muté** qui n'est plus reconnu et donc pas coupé par l'enzyme de restriction

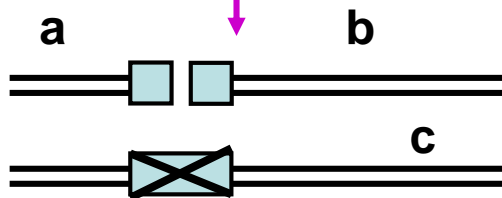
Site de restriction reconnu et coupé par l'enzyme de restriction



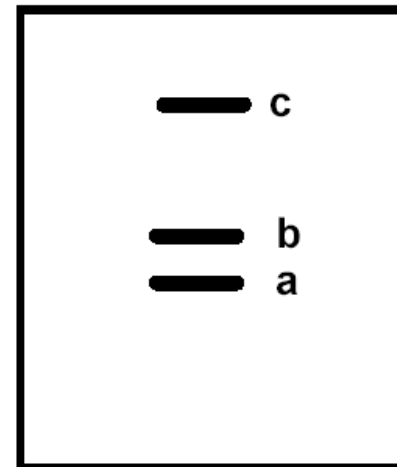
Une zone qui entoure le site de restriction est amplifiée par PCR  
 (les deux allèles seront amplifiés simultanément,  
 = milliards de copies de chaque allèle dans le tube)



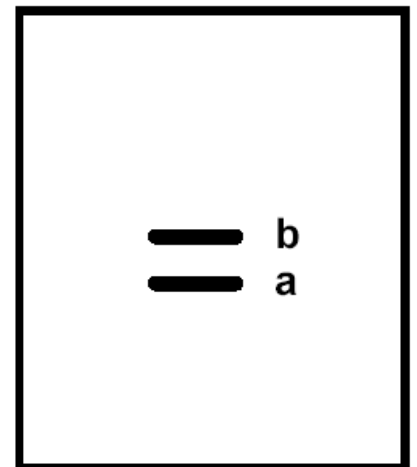
Coupure par  
 l'enzyme de restriction



Migration sur gel



Patient (hétérozygote)  
 (ADN muté/non muté)

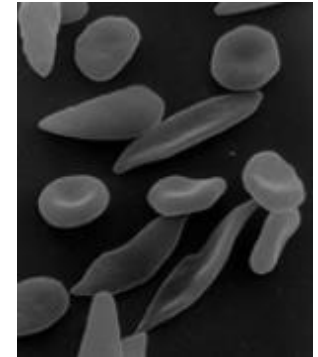
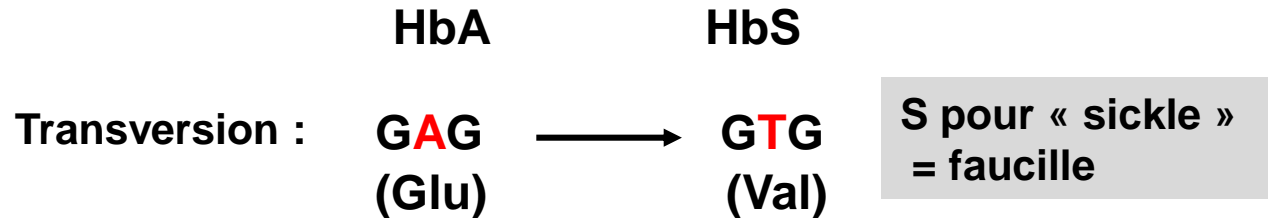


Témoin  
 (ADN non muté)

## Applications PCR-RFLP : Maladies monogéniques.

- **Exemple 1 (la Drépanocytose):** Une mutation A>T du gène engendre la **DISPARITION** d'un site de restriction *Bsu361*

Anémie hémolytique sévère due à une mutation **homozygote** d'une chaîne de  $\beta$  globine.



-La valine est un **AA hydrophobe** qui rend moins soluble la chaîne de  $\beta$  globine.

HbS cristallise (à l'état non oxygénée) et **forme des bâtonnets** qui empêchent l'érythrocyte de se déformer dans les capillaires qui se bouchent  
⇒ **défaut d'oxygénation des tissus : hypoxie.**

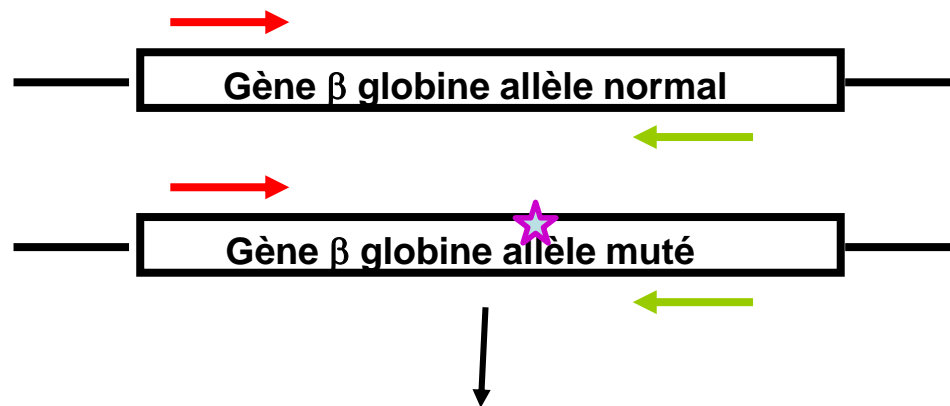
-**Remarque** : il existe 750 variantes de l'hémoglobine dues à des substitutions uniques d'acides aminés dans les chaînes de globine.

-L'enzyme de restriction Bsu 36I coupe le site **CCTNAGG**  
(N = l'une des bases A, C, G, ou T)

HbA	Leu	Thr	Pro	Glu	Glu	
	CTG	ACT	<b>CCT</b>	<b>GAG</b>	<b>GAG</b>	→ Séquence reconnue et coupée par Bsu36I
HbS	CTG	ACT	<b>CCT</b>	<b>GTG</b>	<b>GAG</b>	→ Séquence non coupée par Bsu36I

**Val**

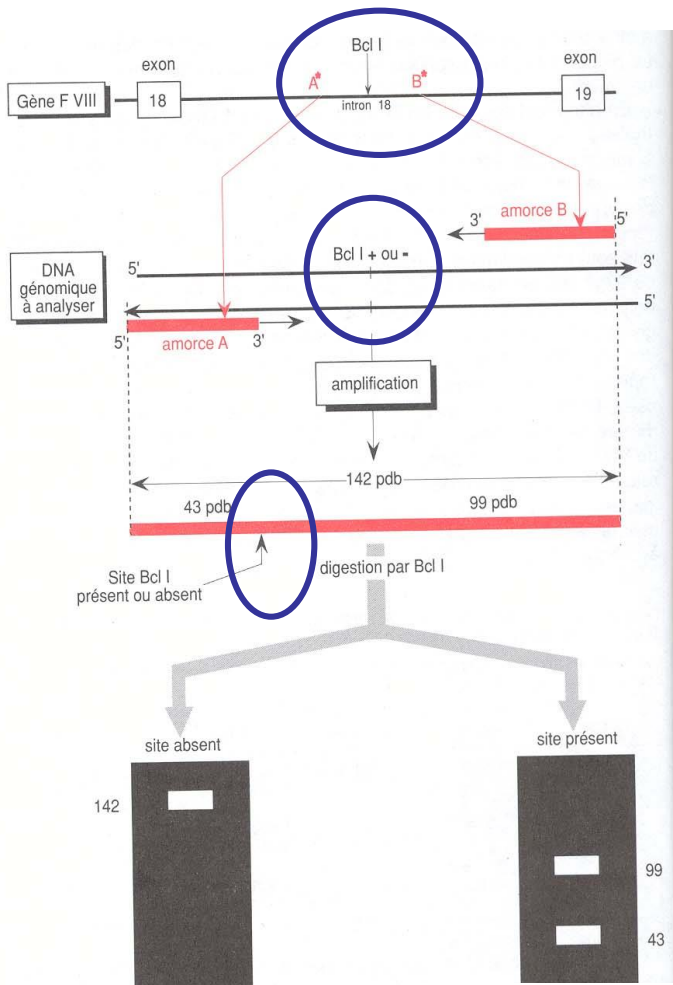
Une mutation A>T du gène engendre la **DISPARITION** d'un site de restriction **Bsu361**



- Amplification par PCR
- Coupure par Bsu 361
- Analyse sur gel des fragments obtenus (3 génotypes)



- **Exemple 2 (hémophilie):** Une mutation dans l'intron 18 du gène codant le FVIII engendre la **CREATION d'un site de restriction *Bcl*** (= allèle M ou +)

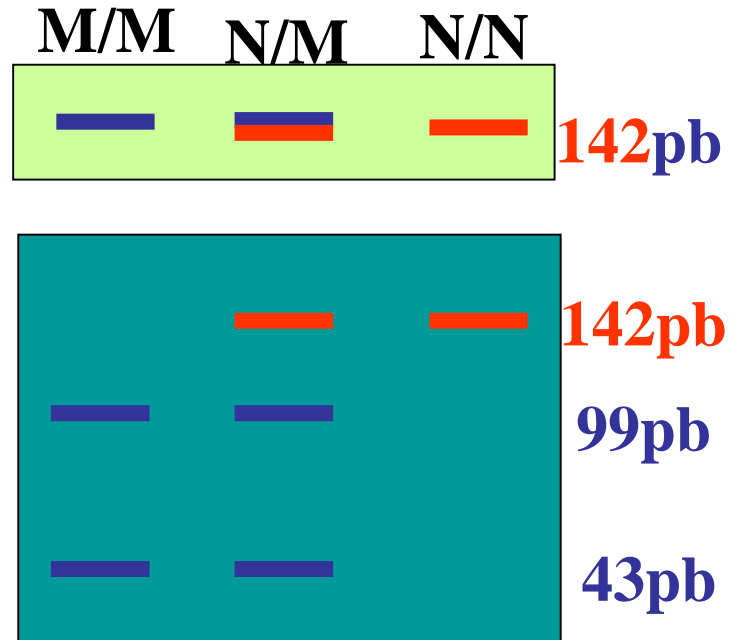


**Allèle N(-)**

**Allèle M(+)**

**PCR**

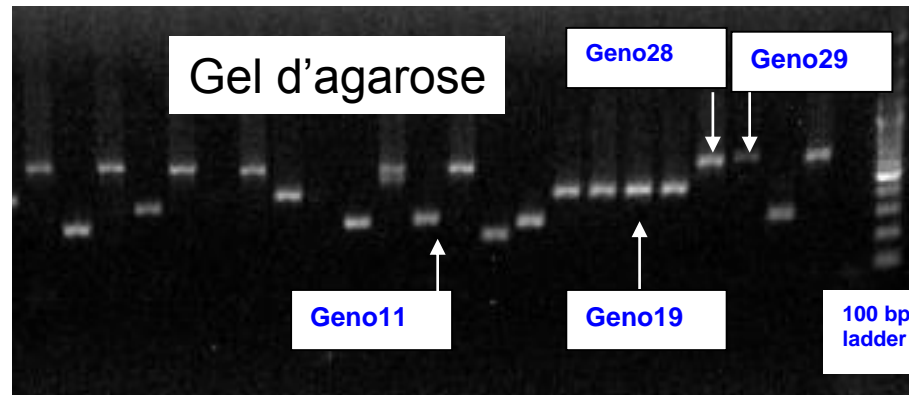
**Digestion *Bcl* des fragments PCR**



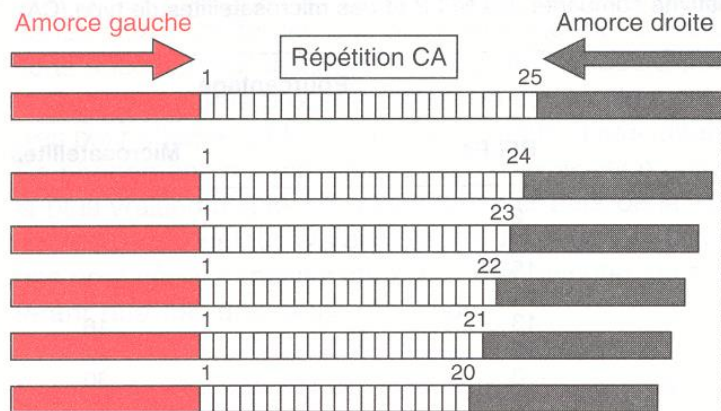
**Le profil PCR-RFLP indique si l'individu porte la mutation *Bcl* du gène codant le FVIII (homozygote ou hétérozygote)**

La PCR (classique) pour l'analyse d'une variabilité d'un nombre de répétition (Ex: STR ou VNTR)

Ex VNTR (rep de 18nt)



Gel d'acrylamide (+ résolutif)



Amplimère  
(taille en pdb)

100  
98  
96  
94  
92  
90

**Figure 9-10** Amplification par PCR d'un microsatellite de type  $(CA)_n$ . Les amorces de PCR sont des séquences uniques situées de part et d'autre du microsatellite. Dans l'exemple choisi, six allèles allant de 25 à 20 répétitions sont indiqués avec la taille des amplimères. Ceux-ci sont en général résolus par électrophorèse en gel de polyacrylamide dénaturant.

6 tailles de fragment PCR possibles (haplotypes)

Ex STR : 6 allèles

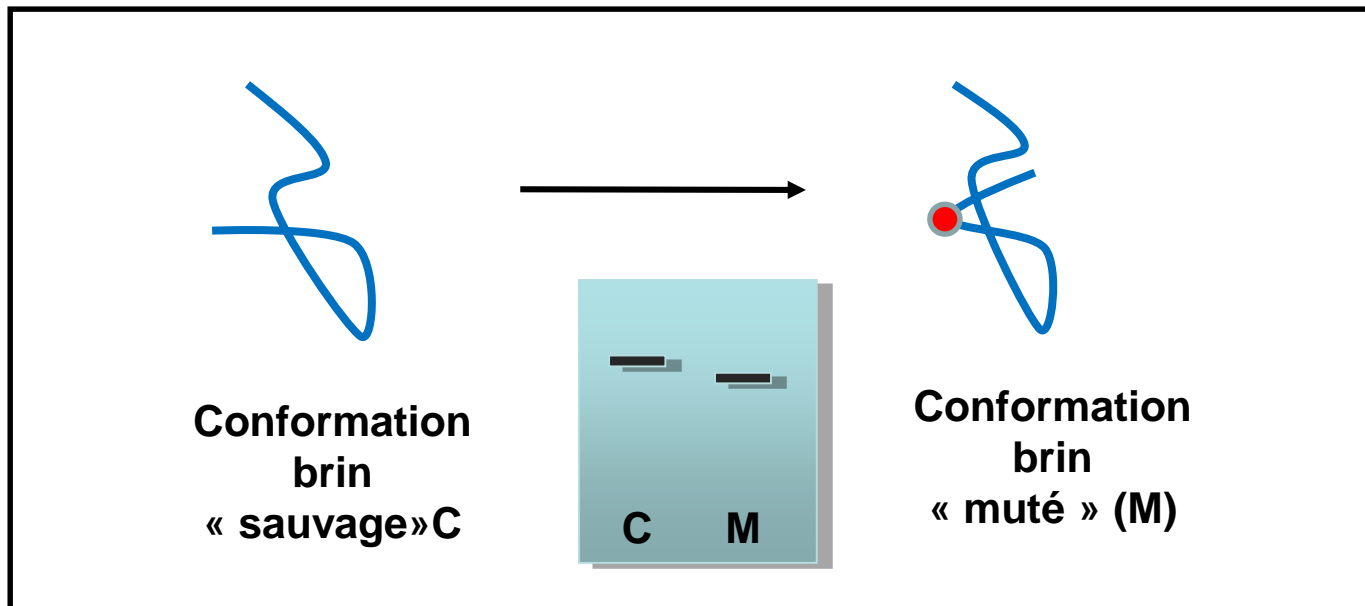
Nombreux génotypes possibles (diploïde)

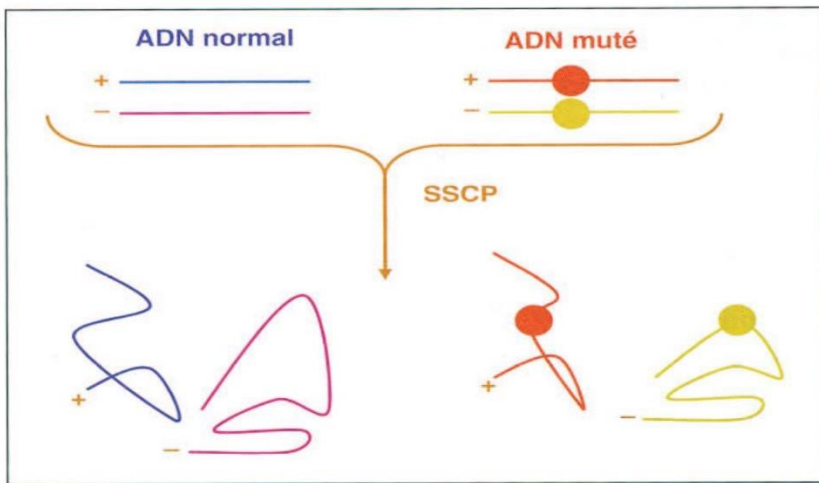


# Analyse de microlésions par la méthode SSCP (Single Strand Conformational Polymorphism)

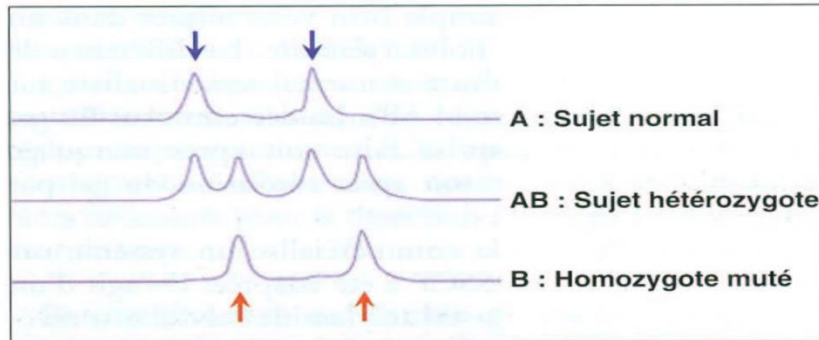
## PRINCIPE

-L'ADN simple brin présente des **conformations** dans l'espace qui vont se modifier **en fonction de la présence ou non d'une mutation** pathogène ou d'un polymorphisme.





**Figure 11.5A** Principe de la technique SSCP. Cette technique est différente à la fois de la D-HPLC et de la DGGE car sa réalisation nécessite une électrophorèse dans un milieu non dénaturant. Elle est réalisée sur des ADN simple brin. La migration de ces derniers va se faire en fonction de leur conformation secondotertiaire. En présence d'une mutation, la structure conformationnelle de l'ADN monobrin se modifie. Par conséquent, le profil de migration du brin correspondant sera modifié.



**Figure 11.5C** Principe de la technique SSCP : aspect de l'électrophorèse capillaire. Électrophorèse capillaire réalisée sur un automate de séquence (ABI, Applied Biosystems). Trois profils de migration sont représentés : sur la figure **A**, on distingue les deux simples brins normaux (« homoduplex ») (flèches bleues) représentés chez un sujet témoin (sans mutation) ; sur la figure **AB**, deux simples brins normaux et deux simples brins mutés « hétéroduplex », représentant l'ADN provenant d'un sujet ayant une mutation à l'état hétérozygote ; sur la figure **B**, un sujet homozygote malade avec deux simples brins correspondant aux simples brins mutés (« homoduplex ») (flèches rouges).

-1- Amplification par PCR de la région d'intérêt

2- Dénaturation des fragments d'ADN

3- Séparation électrophorétique des fragments d'ADN simple brin



- Cela va se traduire par **une différence de mobilité sur gel d'électrophorèse non dénaturant** (gel d'acrylamide)

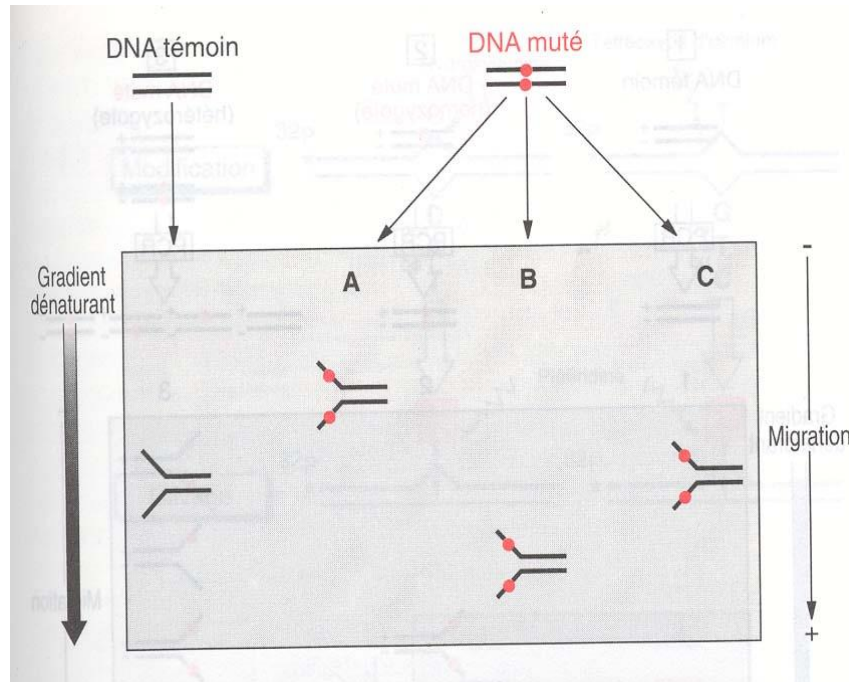
(environ 90% des mutations d'un fragment de taille inférieure à 300 pb peuvent être détectées par cette méthode).

- 4 bandes (simple brin) seront visibles chez un hétérozygote



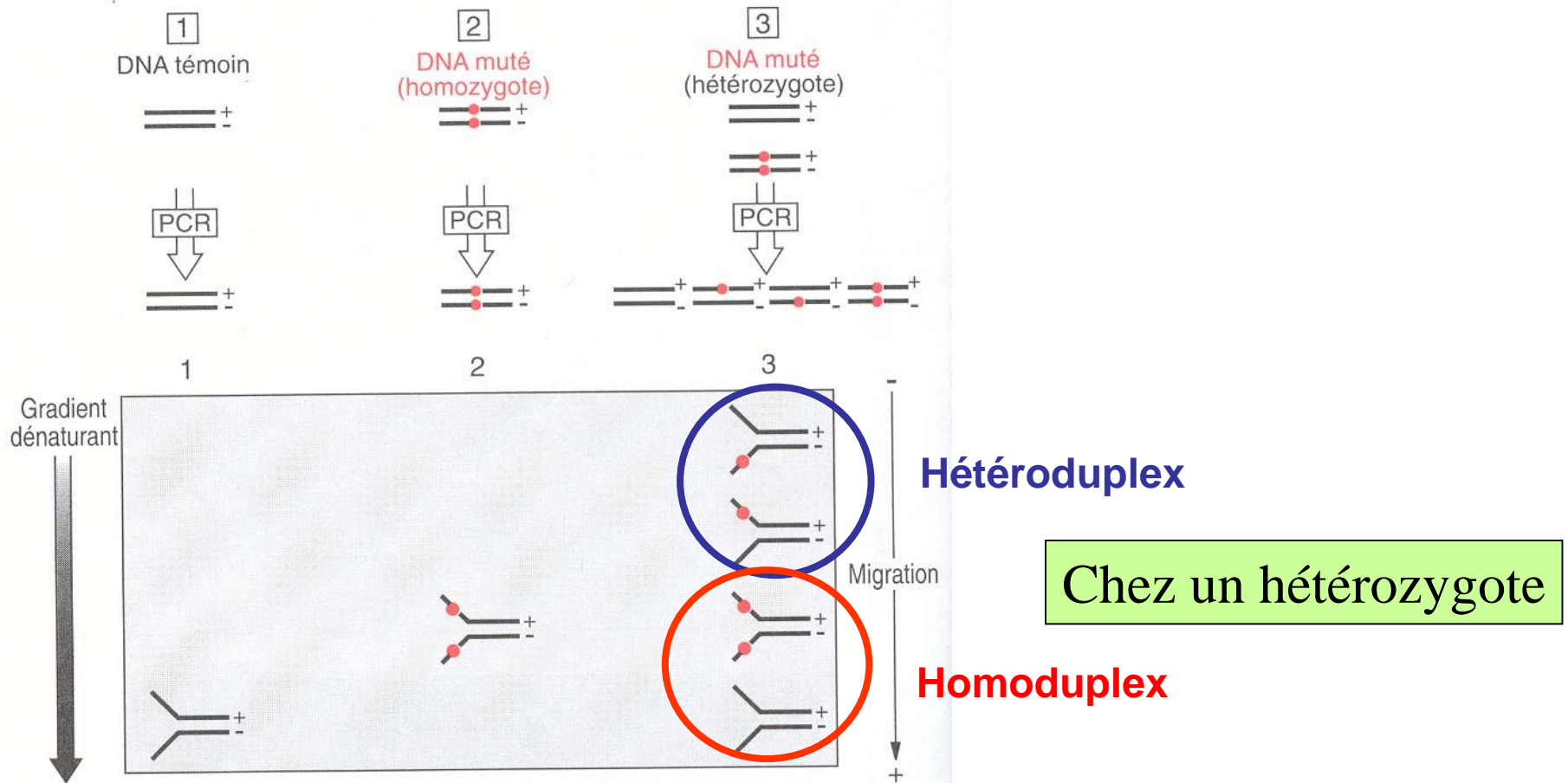
# Analyse de microlésions par la méthode DGGE (Denaturing Gradient Gel Electrophoresis)

## PRINCIPE



-Une mutation (Ex: A, B ou C) de l'ADN peut générer une variation de point de fusion ( $T_m$ ) entre un fragment d'ADN double brin (partiellement dénaturé) N ou M.

-Il en résulte une différence de migration des ADN double brin N ou M (partiellement dénaturé) dans un gel d'électrophorèse d'acrylamide dénaturant (gradient dénaturant généré par l'urée).



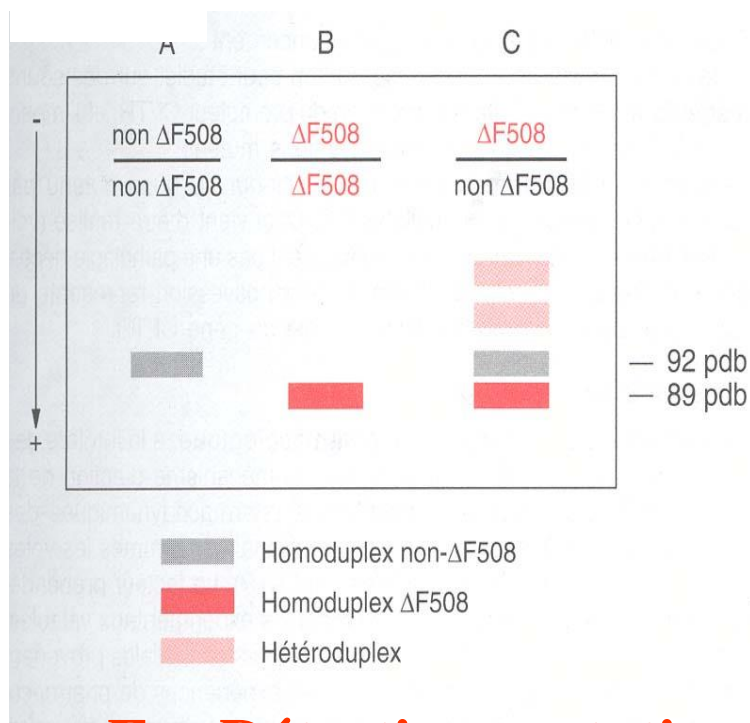
Cette technique (de part les successions d'étapes de dénaturation/hybridation/élongation de la PCR et le fait que l'ADN migre sous forme double brins partiellement dénaturés) génère des **hétéroduplexes** (associations brin+N/brin-M ou brin+M/brin-N).

Ces hétéroduplexes migrent différemment que les **homoduplexes** (associations brin+N/brin-N et Brin+M/brin-M)

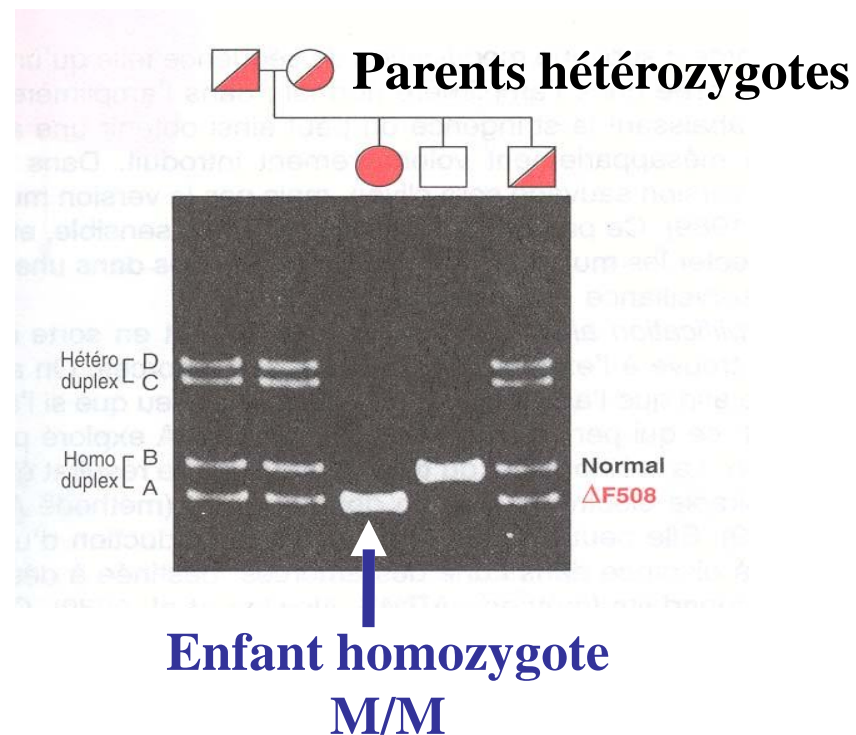
Position	504	505	506	507	508	509	510	511
<b>NORMAL</b>								
DNA	...GAA	AAT	ATC	ATC	TTT	GGT	GTT	TCC...
Protéine	Glu	Asn	Ile	Ile	Phe	Gly	Val	Ser
<b>CF <math>\Delta</math>F508</b>								
DNA	...GAA	AAT	ATC	AT	- - - T	GGT	GTT	TCC...
Protéine	Glu	Asn	Ile	Ile		Gly	Val	Ser

**Figure 14-44 L'anomalie  $\Delta$ F508 dans le gène CFTR**

La lésion est une délétion de 3 bases à cheval sur les codons 507 et 508 (exon 10). Les deux nucléotides provenant du codon 507 et le nucléotide provenant du codon 508 reconstituent un codon ATT (Ile) sans perturbation du cadre de lecture, ni changement de sens du codon 507. Il en résulte la perte du résidu Phe en 508.



**Ex: Détection mutation  $\Delta$ F508 (F508del) de la mucoviscidose par DGGE**



**(environ 95% des mutations d'un fragment de taille inférieur à 500 pb peuvent être détectées par cette méthode).**

**- 4 bandes (partiellement double brin) seront visibles chez un hétérozygote**

# Analyse de microlésions par la méthode D-HPLC (Denaturing High Performance Liquid Chromatography)

## PRINCIPE

-Le principe est le même que pour la DGGE à la différence qu'on analyse les différences de migration entre **hétéroduplexes** et **homoduplexes** (de part leur hydrophobicité différente) sur une colonne qui contient une phase mobile composée de **TEAA** (**triéthyl-amonium acétate, chargé +**) et d'acétonitrile, et une phase stationnaire électriquement neutre et hydrophobe (billes).

-Le TEAA est adsorbé à la surface des billes et de part sa charge + va lier **l'ADN qui est chargé négativement (PO4-)**.

-La dénaturation, provoquée à la fois par la chaleur et par un gradient d'acétonitrile, permet la séparation des homoduplexes et des hétéroduplexes.

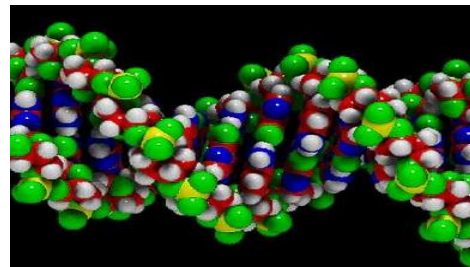
-A plus faible concentration d'acétonitrile, les hétéroduplexes, en raison des mésappariements qui les composent, vont se dénaturer plus vite que les homoduplexes. Leur temps de rétention sera donc plus réduit et ils seront donc élués les premiers.

=> Cette méthode a progressivement remplacé la DGGE. Cette technique est de même sensibilité que la DGGE (**environ 90% des mutations détectées**), plus rapide (7 à 20min contre 4h) et permet d'analyser des microlésions sur des fragments PCR >500pb.

Une molécule bifonctionnelle :

le **TEAA (triéthylamonium acétate)** :

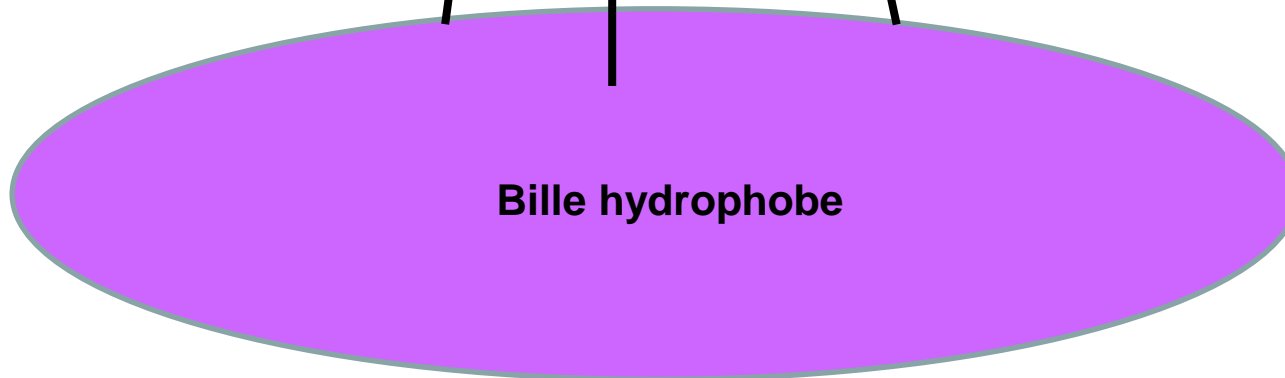
se lie par son **N<sup>+</sup>** sur les **phosphates négatif de l'ADN**,  
et par des **groupes éthyle** sur la **surface d'une bille hydrophobe**



**PO<sub>4</sub><sup>-</sup>**

**N<sup>+</sup>**

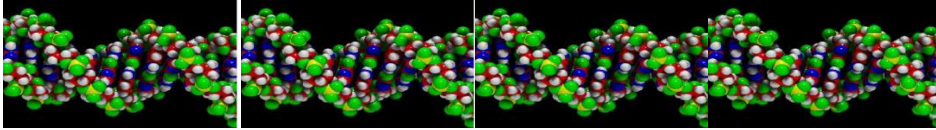
TEAA



(couplage ADN/Billes)



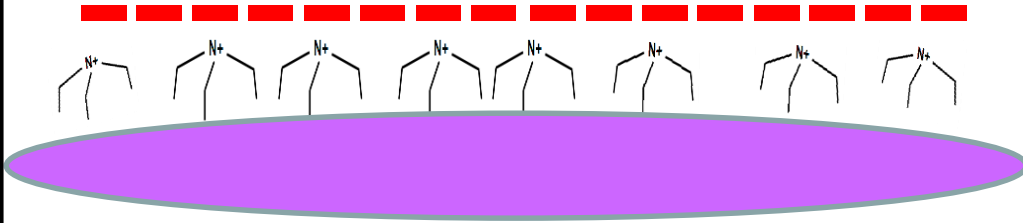
$\text{PO}_4^-$  Charges négatives = ■



Charges négatives serrées sur l'ADN double brin



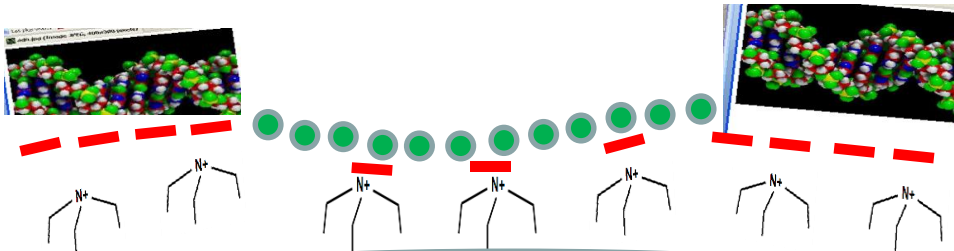
Forte liaison sur les billes  
(cas des **homoduplexes**)



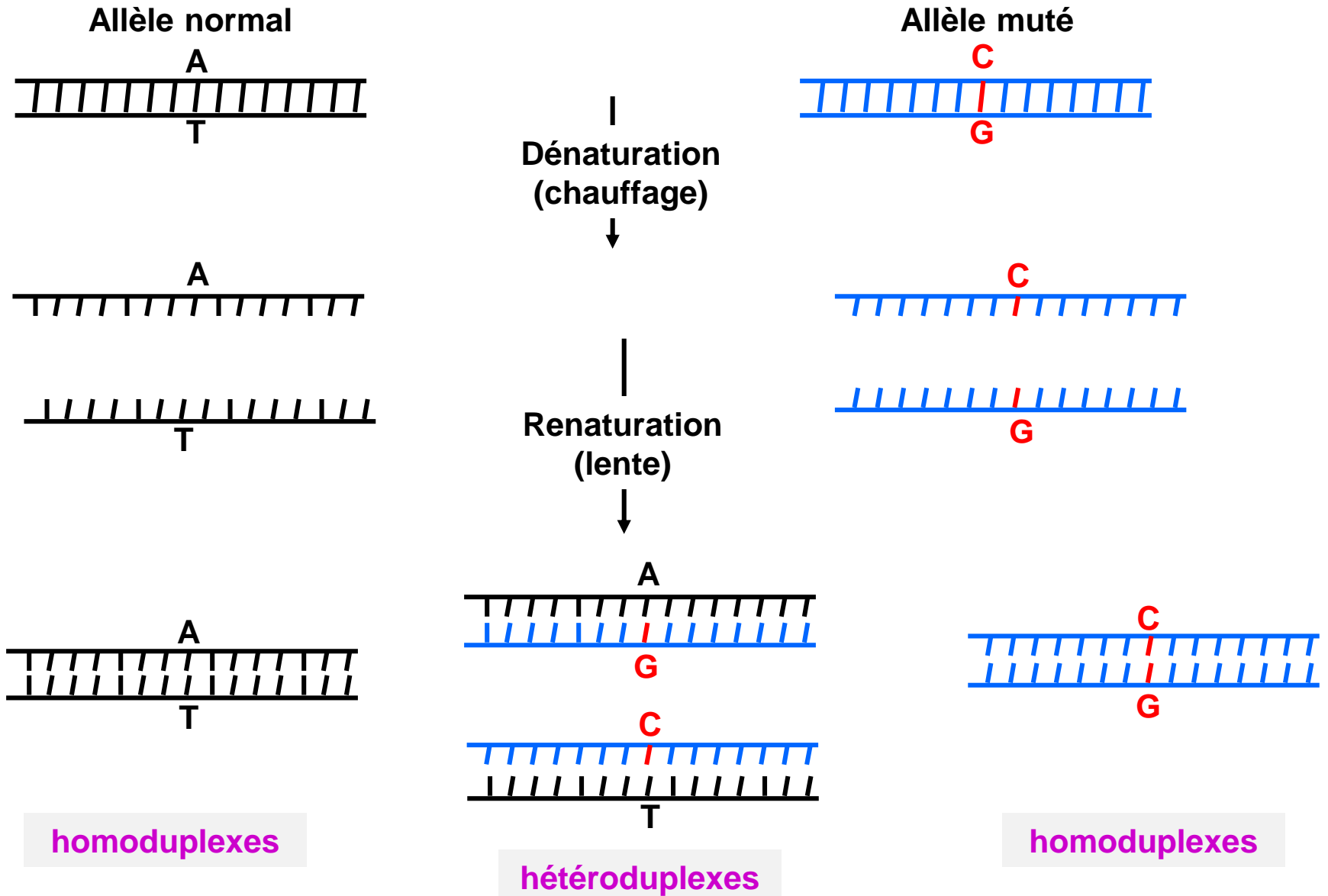
Moins de charges sur la partie du milieu  
(dénaturée par chauffage)



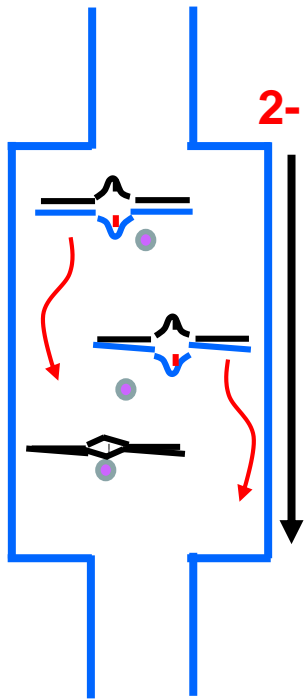
Liaison plus faible sur les billes  
(cas des **hétéroduplexes**)



# 1-Amplimères obtenus chez un sujet hétérozygote



## 2- Gradient décroissant d'acétonitrile



- La colonne est **chauffée** à une température très précise (1/10 de degré près) pour que les duplexes d'ADN se séparent **partiellement** (*plus ou moins suivant qu'il s'agit d'un homo ou d'un hétéroduplexe*).
- Le **gradient d'acétonitrile** interfère sur le temps de rétention des homoduplexes et des hétéroduplexes

Les hétéroduplexes s'éluent en premier car (voir avant) ils sont moins liés aux billes

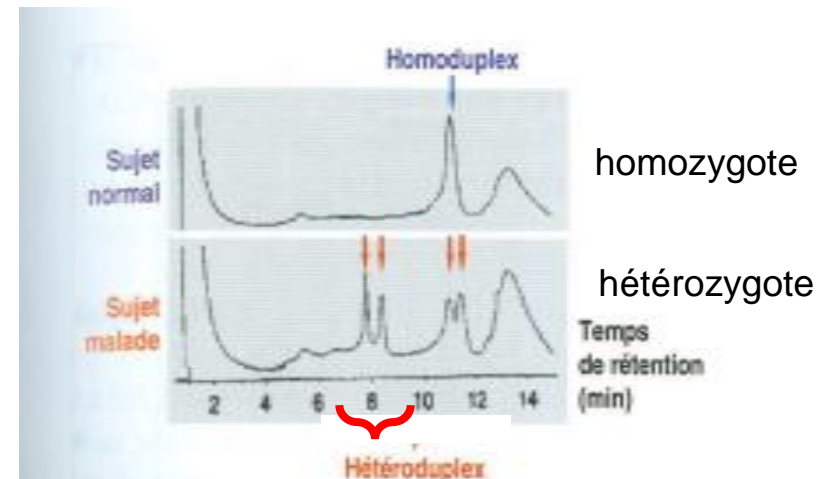
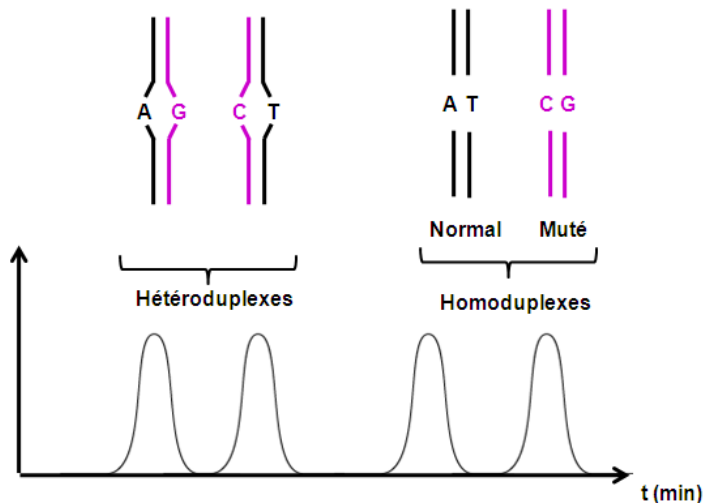


Figure 11.4 Représentation de la position des hétéroduplex et des homoduplex par D-HPLC. La D-HPLC est issue du même principe que la DGGE. Elle permet une séparation des homoduplex et hétéroduplex. Contrairement à la DGGE qui nécessite au moins 4 h pour sa réalisation, la D-HPLC est rapide, puisque seulement 7 à 20 min sont nécessaires pour traiter un échantillon.

# Analyse de microlésions par PCR en temps réel et méthode HRM (High Resolution Melting: analyse des courbes de fusion à haute résolution)

## PRINCIPE



-Technique actuellement en pleine expansion !

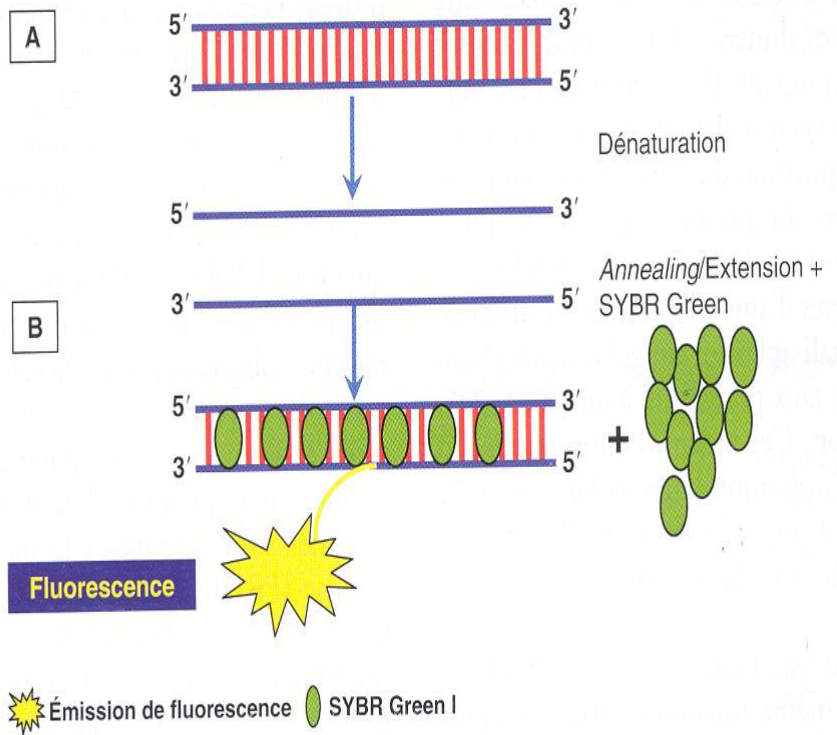
-De faibles quantité d'ADN génomique (muté et non muté) sont **amplifiées dans un appareil de PCR en temps réel.**

-A l'issue de l'amplification, les températures de fusion des duplexes porteurs ou non de la mutation sont analysées.

# La PCR en temps réel ou PCR quantitative (qPCR)

## Principe général :

**C'est le même principe que la PCR classique (3 étapes)  
Mais on ajoute un intercalant de l'ADN (ex; SYBR Green 1)**



### Avantages par rapport à la PCR classique:

- Quantification d'ADN (ou d'ARNm via leur passage en ADNc par transcriptase reverse = RT-qPCR) possible;
- Plus rapide car pas besoin de faire de gel pour l'analyse des résultats
- possibilité de mesure du Tm des brin d'ADN synthétisés par quantification de fluorescence. Détection spécifique et sensible de pathogènes d'intérêt vétérinaire - Quantification de la charge bactérienne, virale ou parasitaire.



Révélation de l'amplification:

qPCR

=> Par mesure de fluorescence en temps réel (sur écran)

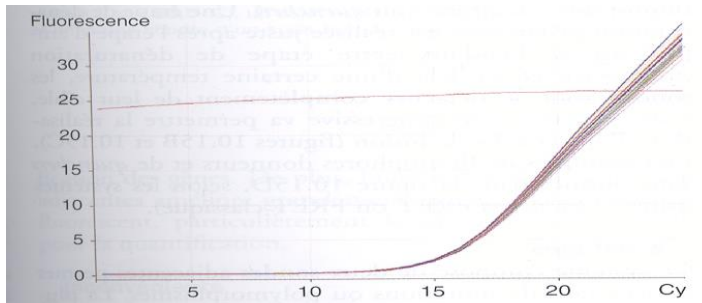
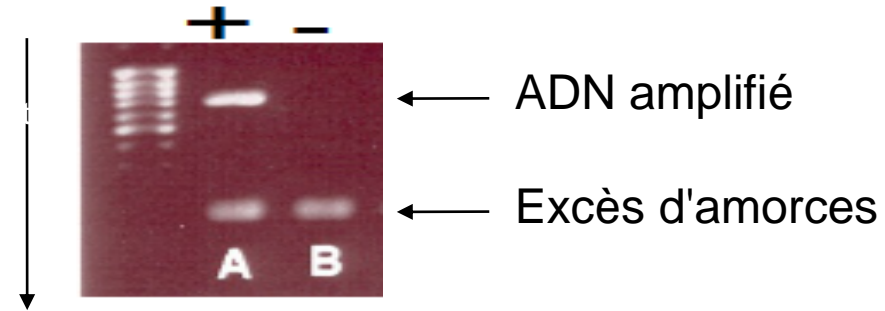


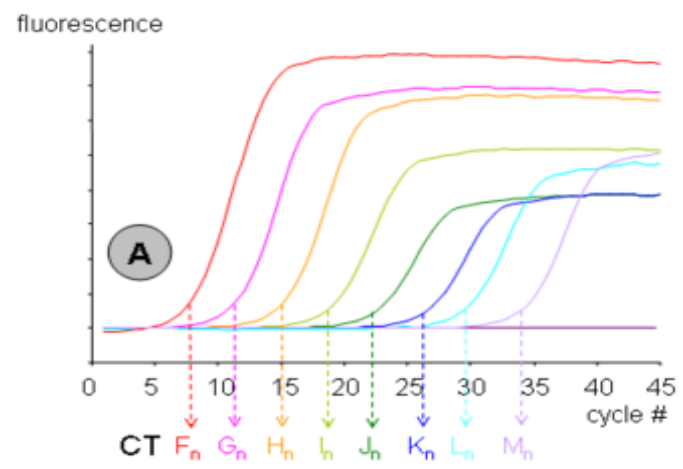
Figure 10.13 PCR en temps réel avec SYBR™ Green I. Aspect de la fluorescence en phase d'élongation. L'axe des abscisses correspond aux cycles de la PCR, l'axe des ordonnées correspond à l'intensité de la fluorescence qui est proportionnelle à la quantité des produit amplifié (échelle linéaire). (Image reproduite avec l'aimable autorisation du Dr Thomas Kaiser [Corbett Research].)

PCR

=> L'analyse de la PCR nécessite le dépôt des échantillons sur un gel d'électrophorèse (et leur coloration par le bromure d'éthidium, intercalant de l'ADN qui fluoresce sous les UV)



[ADN] F = 10G = 10<sup>2</sup>H = 10<sup>3</sup>I = 10<sup>4</sup>J = 10<sup>5</sup>K = 10<sup>6</sup>L = 10<sup>7</sup>M = 10<sup>8</sup>N

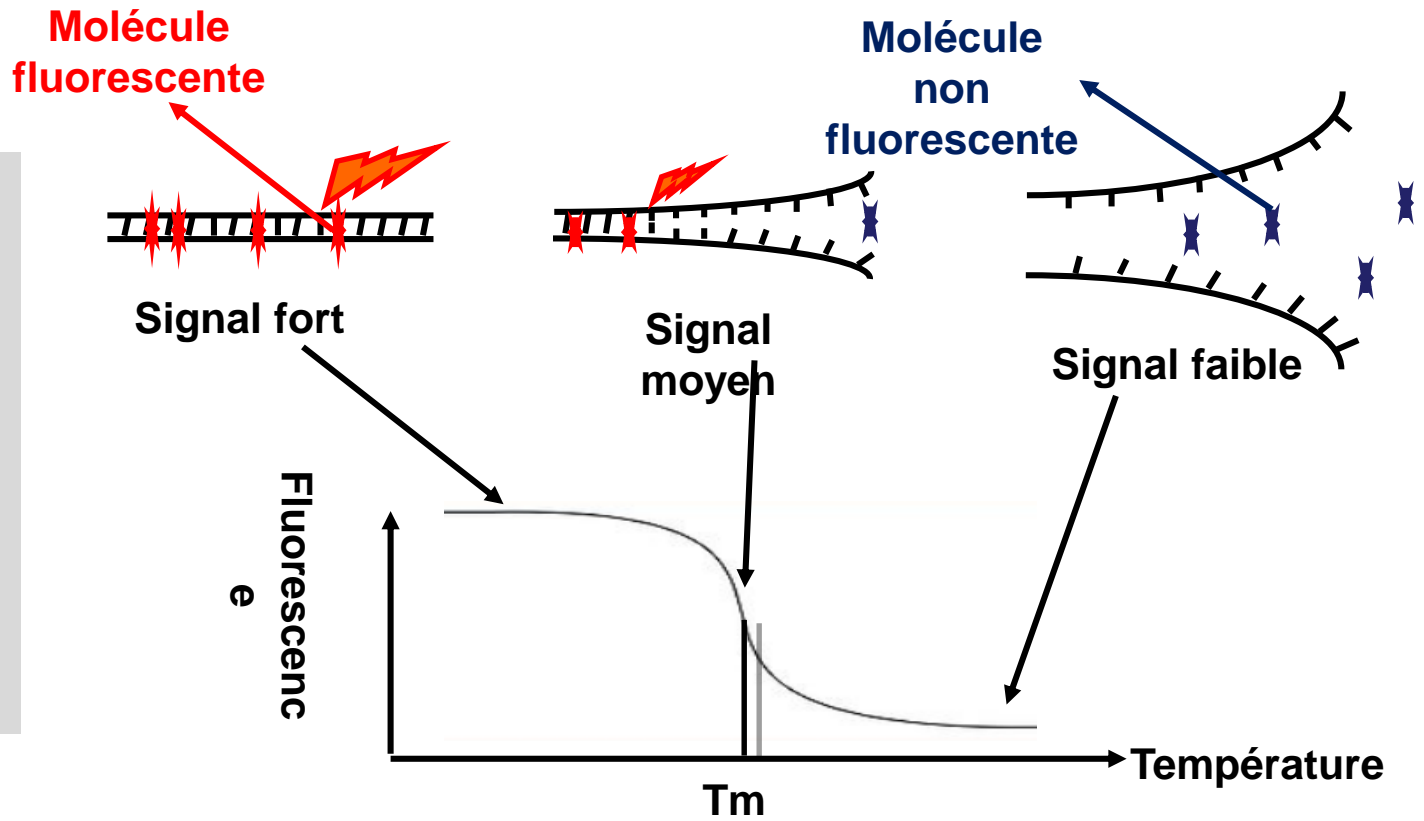


Exemple de quantification d'ADN (ou d'ADNc) sur une gamme de dilution par qPCR : => Plus la quantité d'ADN ciblé est élevée plus la détection est précoce (courbe rouge)

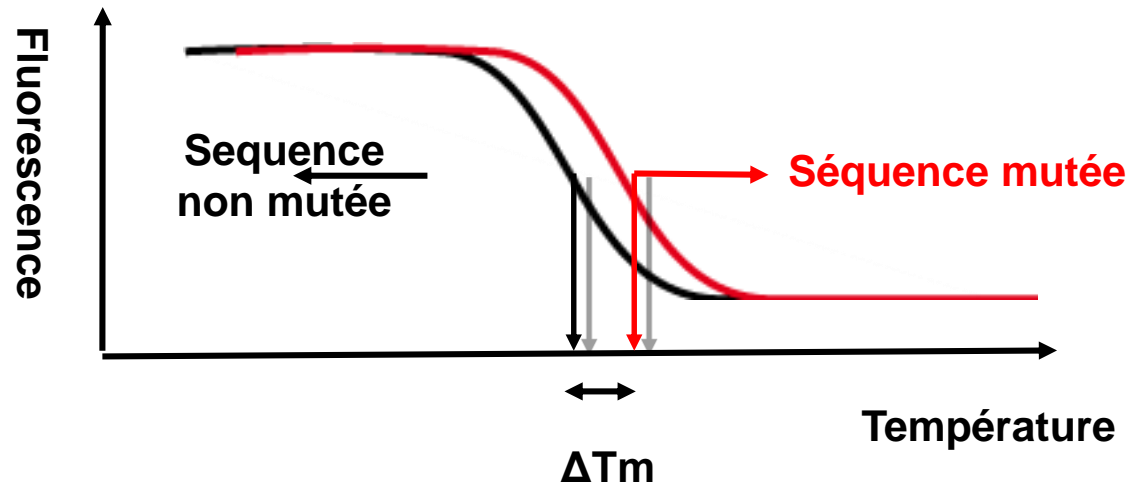
Ce qui constitue l'originalité de la méthode HRM, c'est de pouvoir mesurer **en temps réel** cette fusion des brins de façon **à déterminer précisément la température de fusion ( $T_m$ )**.

Ceci est réalisé par l'utilisation d'une molécule intercalante (à saturation) qui ne devient fluorescente que lorsqu'elle est insérée dans un ADN double brin

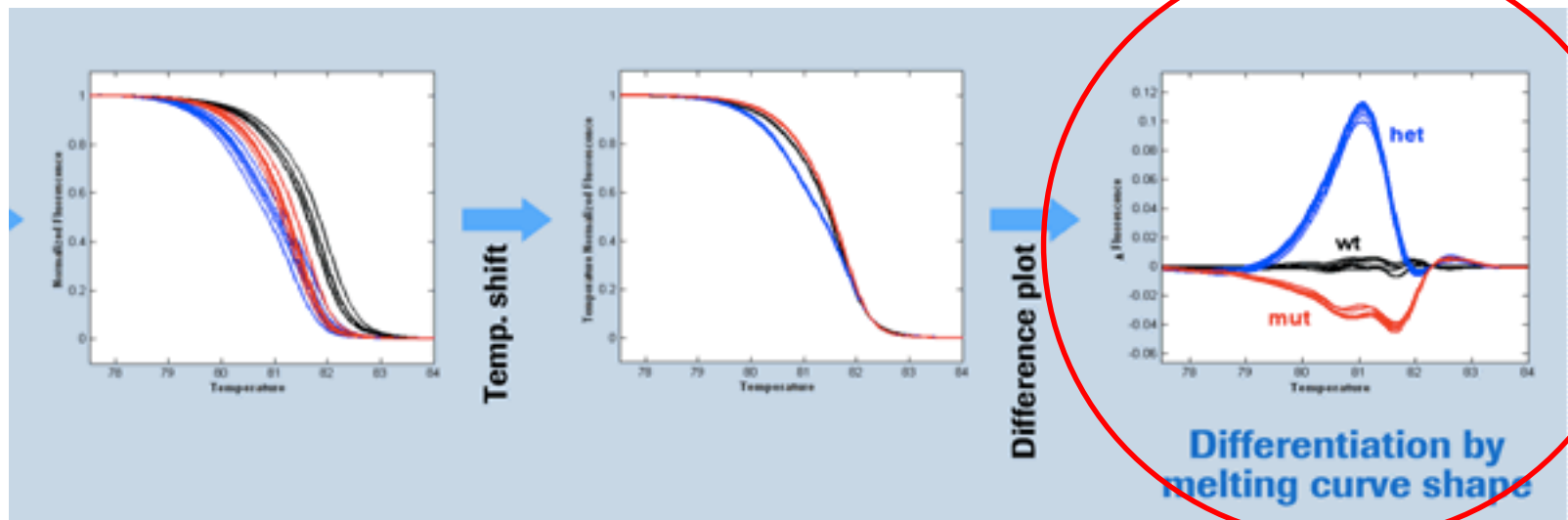
Après PCR, la température va être augmentée progressivement jusqu'à fusion complète des brins.



La séquence mutée n'aura pas tout à fait la même température de fusion ( $T_m$ )



Des logiciels vont transformer ces courbes afin d'analyser quantitativement les différences correspondant à plusieurs ADN



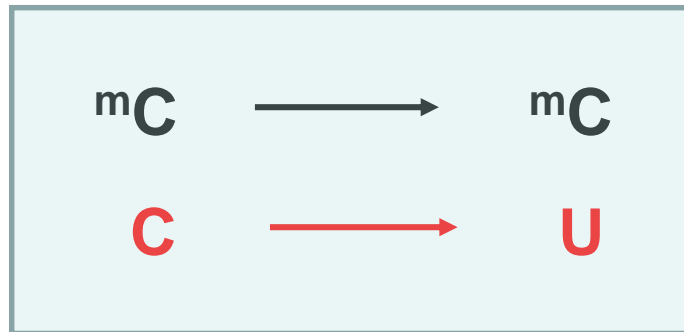
# Analyse des épimutations (méthylation) par méthodes DHPLC ou HRM

-Intérêt important dans l'analyse des **modifications épigénétiques**.

-RAPPEL : L'ADN peut être méthylé en position 5' des cytosines du motif 5'-CpG-3'. La présence de ces 5mC peut affecter le degré d'expression de certains gènes, par exemple ceux soumis à empreinte (cf Cours Bio Mol UE11 DFGSP2).

## PRINCIPE DE L'ANALYSE (Conversion au bisulfite)

-L'ADN est soumis à un traitement par le bisulfite de sodium qui transforme **les cytosines non méthylées en uracile (U)**.



# Principe



Dénaturation



Conversion  
au bisulfite



1 motif CpG méthylé  
sur ses cytosines

Les C **non méthylés**  
sont transformés en U

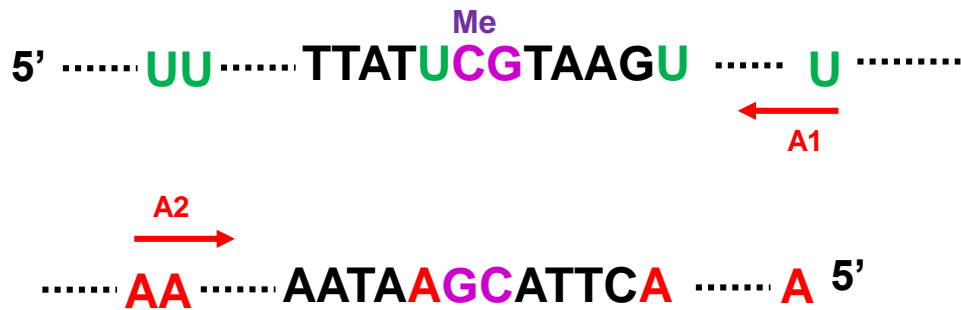
Les brins **ne sont plus complémentaires** :  
on peut choisir des amorces spécifiques  
d'un seul des brins pour l'amplifier.  
*Ici on choisira d'amplifier celui du  
haut*



Simple  
élongation



dénaturation



On choisit un  
couple d'amorces

(A1, A2) :

Pour l'instant seul A1 s'hybride sur  
le brin du haut, puis une simple  
polymérisation est réalisée à partir  
de cette amorce

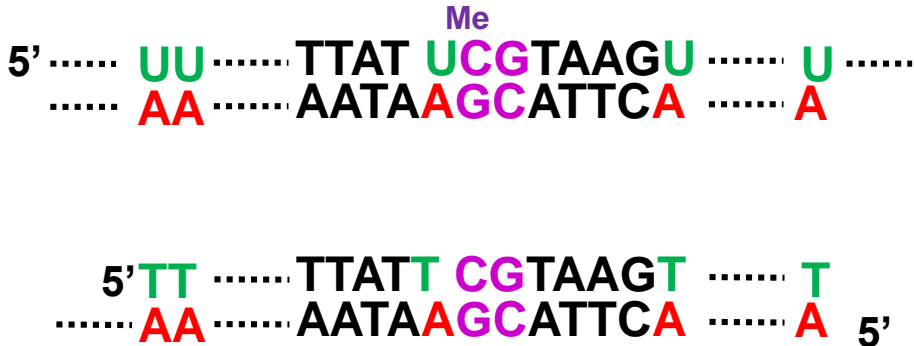
*(A se met en face des U)*

Le double brin généré (1/2 long) est  
alors dénaturé.

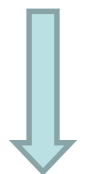
A1 et A2 vont à présent  
s'hybrider et des cycles de PCR  
« classique » vont être réalisés =  
amplification.

*(T se met en face des A)*





On voit déjà apparaître un brin qui contient des T à la place des U  
 Après quelques cycles, ces doubles brins vont devenir très majoritaires,  
 La quantité des brins 1/2 longs issus des brins matriciels et les brins matriciels eux-mêmes sera négligeable



Double brin majoritaire en fin de PCR

## Exemple : analyse d'un hétérozygote

### Cas d'un allèle méthylé



Bisulfite +  
PCR spécifique d'un brin



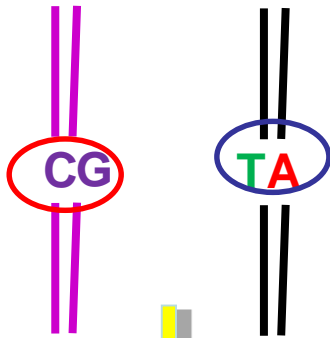
### Cas d'un allèle non méthylé



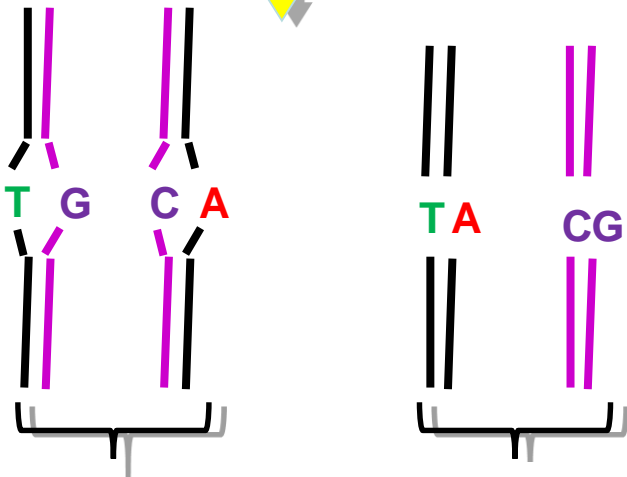
Bisulfite +  
PCR spécifique d'un brin



Ces deux séquences ne diffèrent que par les nucléotides qui sont placés aux endroits où se trouvaient une cytosine +/- méthylée.



Dénaturation et  
renaturation lente.



Hétéroduplexes    Homoduplexes.

- Plus le nombre de cytosines méthylées contenues dans les brins analysés est grand plus les différences en sortie de colonne seront importantes.
- Cette méthode permet d'avoir un aperçu global d'une régions méthylée (de préférence sur plusieurs sites) .
- L'analyse des amplicons obtenus peut aussi être réalisée **par HRM** (ou par **pyroséquençage**)

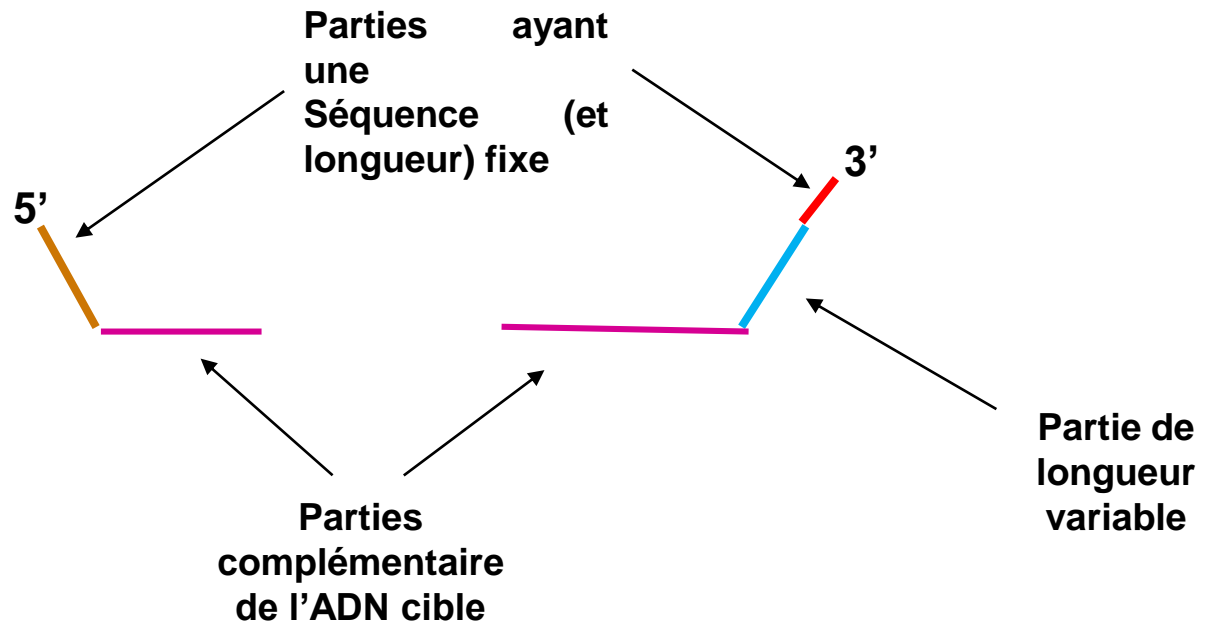
Ce mélange est passé sur  
colonne DHPLC  
comme dans le cas de la  
détection de mutations.

# Analyse de macrolésions (délétion ou insertion) par MLPA ( Multiplex Ligation Dependent Probe Amplification)

## PRINCIPE

On dispose de :

### a) 2 amorces très spéciales

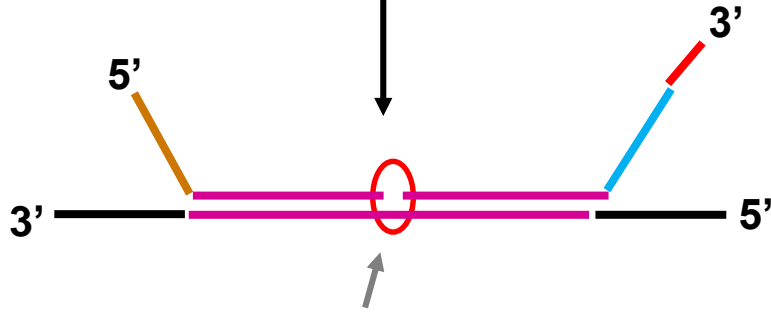


### b) Une séquence d'ADN cible



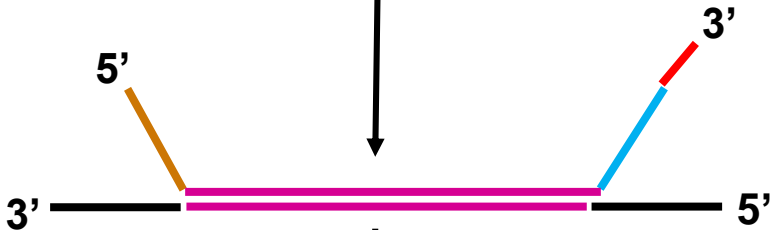
# La réaction de ligation-amplification

Hybridation  
des parties complémentaires  
des amorces

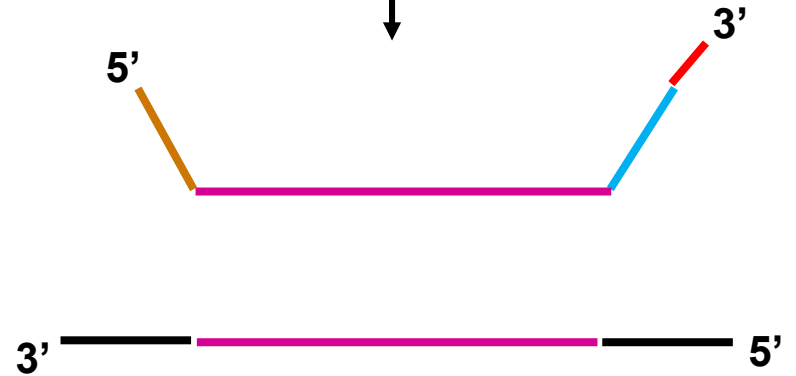


Il manque une liaison  
phosphodiester ici

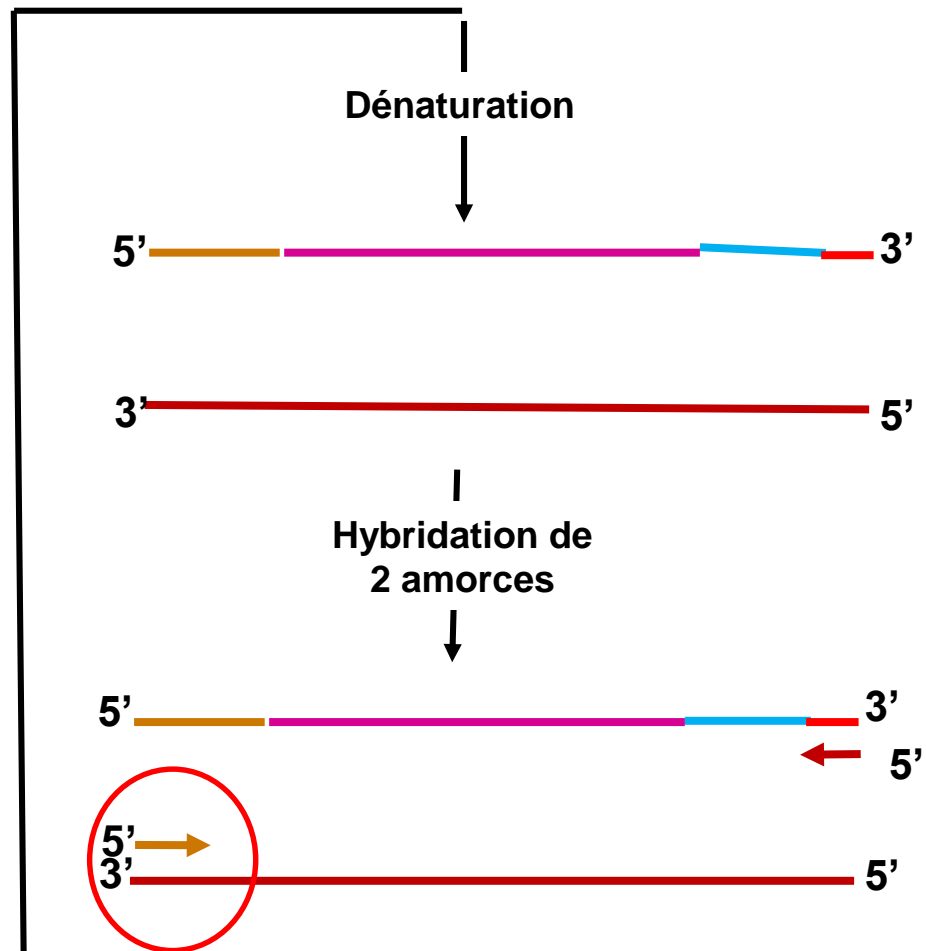
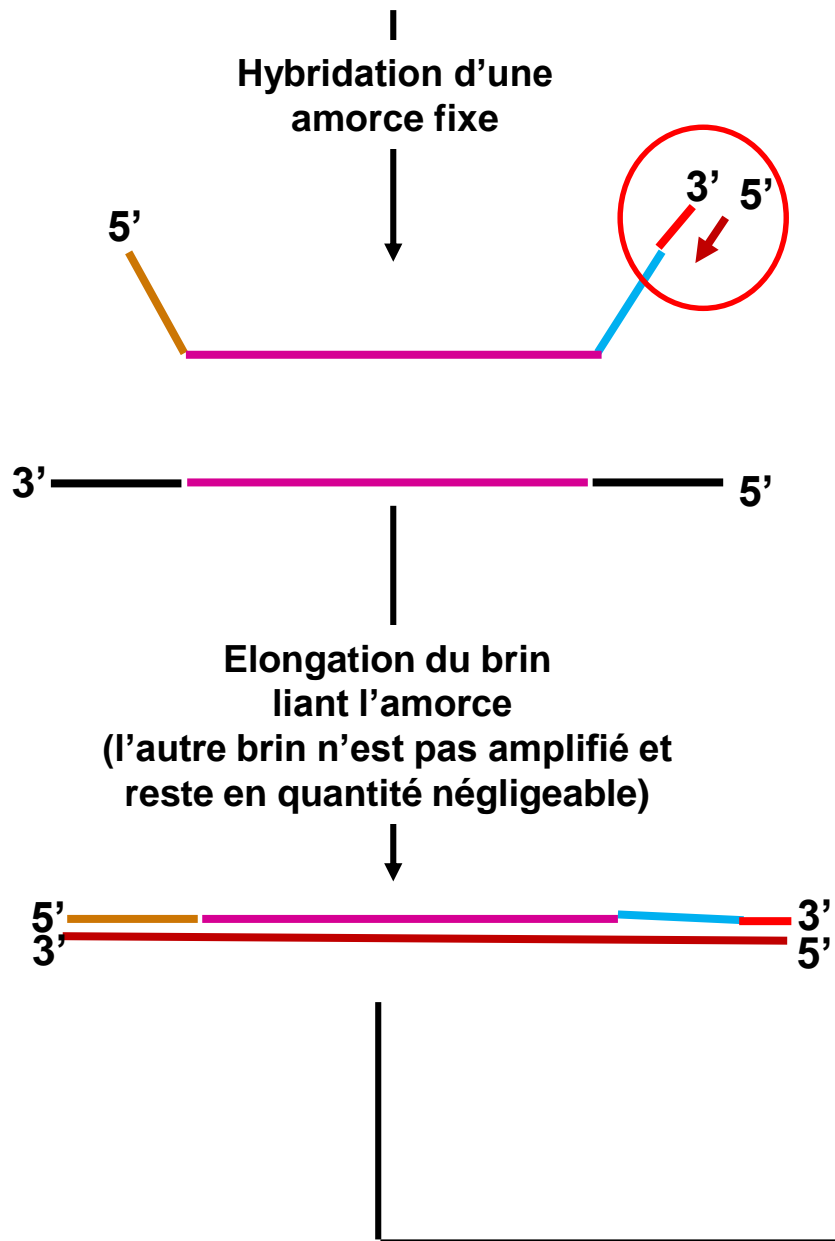
Ligation par une ligase

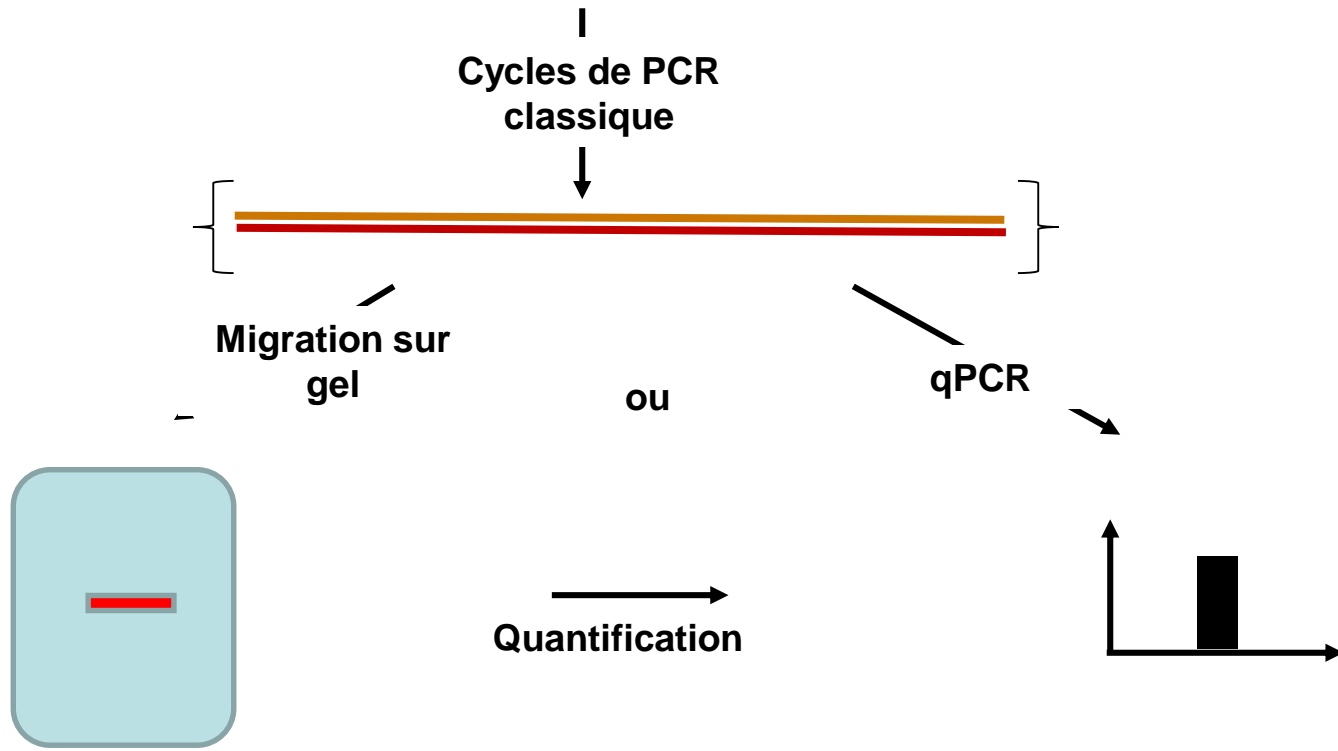


Dénaturation









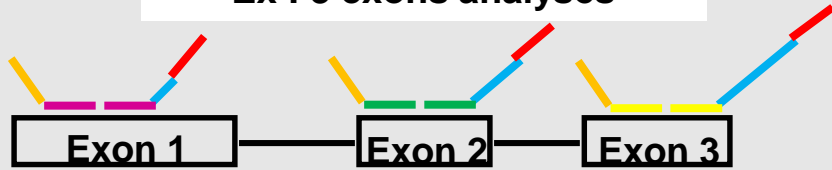
L'analyse par **MLPA** permet de mettre en évidence des **macrodéletions** (et aussi des duplications) **de larges fragments**

⇒ La détection d'un signal PCR signifie que la séquence recherchée est **PRESENTE** sur la matrice étudiée (la quantification indique la perte ou le gain de séquence ciblée)

⇒ A l'inverse, une absence de signal PCR signifie que la séquence recherchée est **ABSENTE** sur la matrice étudiée

## Exemple d'application MLPA : DETECTION MACRODELETION

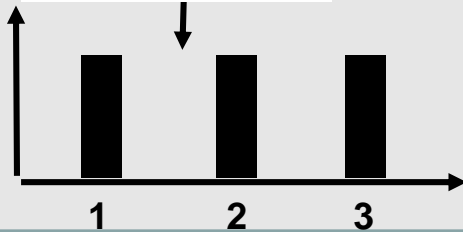
Ex : 3 exons analysés



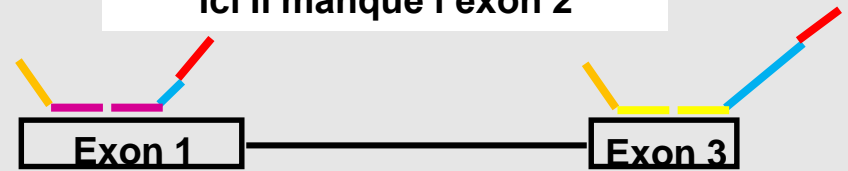
Ligation, PCR



Quantification  
PCR



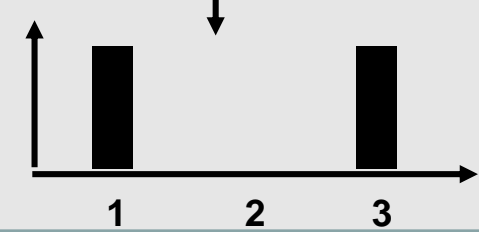
Ici il manque l'exon 2



Ligation, PCR



Quantification  
PCR



=> Les amorces ne peuvent pas se fixer sur l'exon manquant. La PCR ne développe alors aucun signal.

# B- OUTILS MOLECULAIRES D'IDENTIFICATION DES MUTATIONS

## 1- Basés sur l'amplification génique

- Rappels sur la méthode de base (PCR classique)
- La PCR en temps réel (Real Time PCR)
- la PCR-RFLP (détection mutation touchant un site de restriction)
- la SSCP (Single Strand Conformational Polymorphism)
- la DGGE (Denaturing Gel Gradient Electrophoresis)
- DHPLC (Denaturing High Performance Liquid Chromatography)
- la HRM (Analyse des courbes de fusion à haute résolution)
- la MLPA (Multiplex Ligation dependant Probe Amplification)

## -2- Basés sur le séquençage d'ADN

- Rappels sur la méthode de base de séquençage d'un ADN (Sanger & Coulson)
- Le pyroséquençage et la SMRT

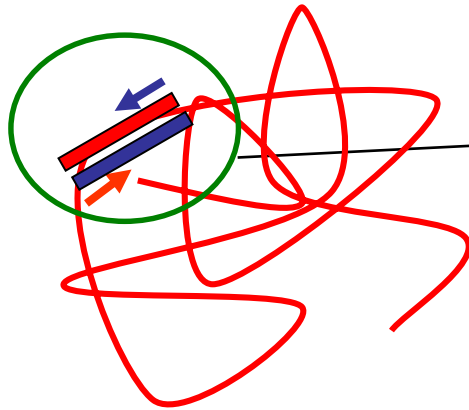
## -3- Basés sur l'hybridation moléculaire

- Southern blot
- puces à ADN

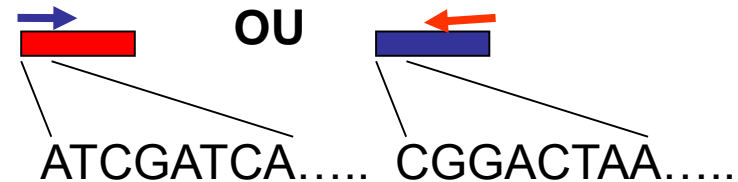
## 2- IDENTIFICATION DE MUTATIONS PAR SEQUENCAGE

- Rappels sur la méthode enzymatique de base de séquençage d'un ADN (Sanger & Coulson)
- Le pyroséquençage
- Le SMRT (Single Molecule Real Time)

**Séquence d'ADN d'intérêt du génome humain (connu)**



*Ciblage du gène par **1 amorce** ADN simple brin*



**Objectif = Détermination de l'agencement précis des bases A,C,G et T dans l'ADN d'intérêt ciblé**

**Applications multiples:**

- Étude des génomes
- Identification de mutations dans un génome
- etc



# LE SEQUENCAGE DE GENOMES

La connaissance de la structure d'un génome dans son entièreté peut passer par son séquençage.

Cependant, la taille des génomes étant de plusieurs millions (ou milliards) de bases, il est nécessaire de coupler les approches de biologie moléculaire avec celle de l'informatique pour pouvoir traiter un nombre aussi important de données.

## 1- Séquençage par la méthode de Sanger & Coulson (1980-) :

Deux grands principes de séquençage de génome entier sont utilisés. Dans les deux cas, l'ADN génomique est préalablement fragmenté par des méthodes enzymatiques (enzymes de restrictions) ou physiques (ultrasons):

- la **méthode de séquençage par ordonnancement hiérarchique** consiste à classer les fragments génomiques obtenus avant de les séquencer.

- la **méthode globale** (ou *whole-genome shotgun*) ne fait pas de classement des fragments génomiques obtenus mais les séquence dans un ordre aléatoire. Une analyse bioinformatique permet ensuite de réordonner les fragments génomiques par chevauchement des séquences communes. Il s'agit de la méthode de **séquençage enzymatique utilisant des ddNTP**, initialement imaginée dans le laboratoire de F. SANGER à Cambridge à la fin des années 1970 pour séquencer les premiers génomes de virus.

=> La principale différence entre ces deux principes est que l'ordonnancement hiérarchique essaie d'aligner un jeu de clones de grande taille (~ 100 kb) alors que dans la méthode globale le génome entier est réduit en fragments de petite taille qui sont séquencés puis alignés.

# LE SEQUENCAGE DE GENOMES

## 2- Séquençage par hybridation (via puces ADN)

Le séquençage par hybridation repose sur l'utilisation de puces à ADN contenant de plusieurs centaines (pour les puces de première génération) à plusieurs milliers d'oligonucléotides. L'ADN à analyser est coupé en de multiples fragments qui sont ensuite incubés sur la puce où ils vont s'hybrider avec les oligonucléotides dont ils sont complémentaires. La lecture de la puce (la détection des oligonucléotides hybridés), permet d'obtenir le spectre de la séquence d'ADN, c'est-à-dire sa composition en sous-séquences de  $n$  nucléotides, où  $n$  est la taille des sondes sur la puce utilisée. Le traitement informatique du spectre permet ensuite de reconstituer la séquence entière.

## 3- Séquençage haut débit (HTS) (2005-)

On désigne par séquençage haut débit (**HTS** pour *high-throughput sequencing*) aussi appelé **NGS** pour *next-generation sequencing* un ensemble de méthodes apparues à partir de 2005 produisant des millions de séquences en un *run* et à faibles coût.

Elles se caractérisent par l'utilisation d'approches massivement parallèles, permettant de séquencer des centaines de milliers de fragments simultanément. Elles **s'affranchissent des étapes de clonage et de constitution de banques génomiques**. Elles permettent de séquencer à partir de molécules uniques d'ADN.

⇒ **En gros, deux types de techniques :**

**a) celles qui nécessitent une amplification de la matrice par PCR.**

**ex : le pyroséquençage**

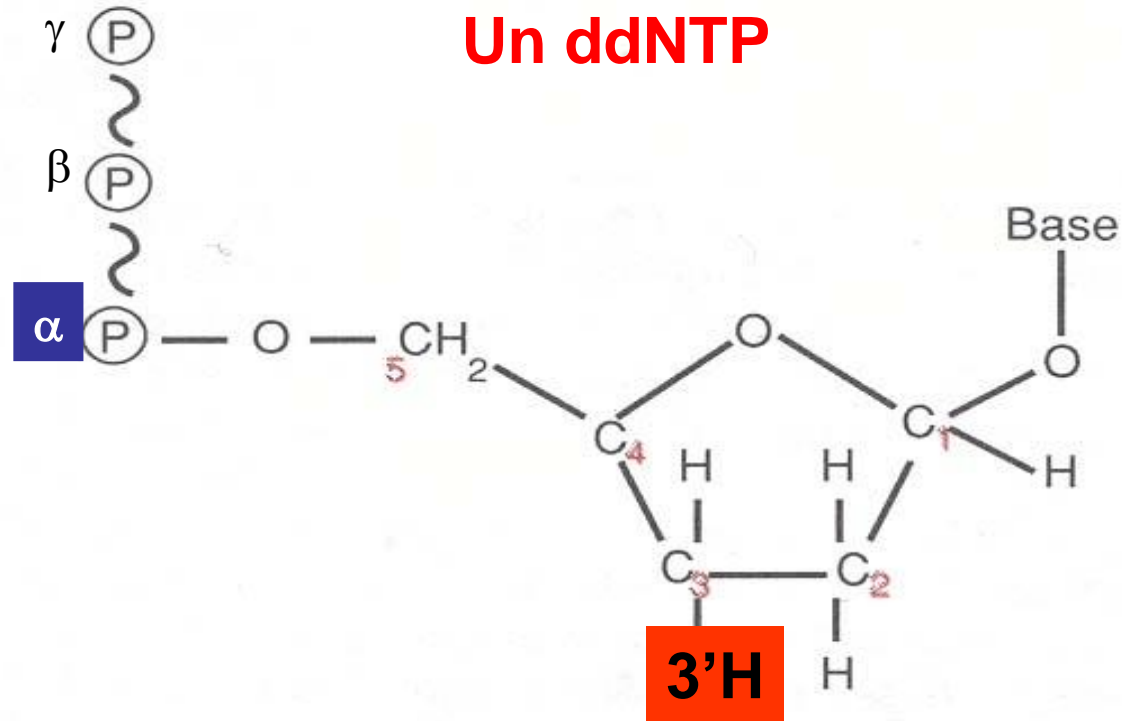
**b) celles qui ne nécessitent pas d'amplification.**

**ex : la séquence en temps réel d'une simple molécule (SMRT)**

Méthode	Longueur de la lecture	précision	Lecture par expérience	temps d'expérience	coût par 1 million de bases (en dollar américain \$)	Avantages	Inconvénients
<b>Séquençage d'une seule molécule en temps réel (SMRT) (Pacific Biosciences)</b>	10,000 pb à 15,000 pb en moyenne (14,000 pb N50); longueur de lecture maximale >40,000 bases	87%	50,000 par cellule, ou 500–1000 megabases	30 minutes à 4 heures	\$0.13–\$0.60	lectures longues. Rapide. Détecte 4mC, 5mC, 6mA	débit modéré, l'équipement peut être très coûteux
<b>Ion semiconductor (Séquençage Ion Torrent)</b>	jusqu'à 400 pb	98%	jusqu'à 80 millions	2 heures	\$1	l'équipement le moins cher, rapide	erreurs d'homopolymères
<b>Pyroséquençage (454)</b>	700 pb	99.9%	1 million	24 heures	\$10	lectures longues, rapide	l'expérience coûte cher, erreurs d'homopolymères
<b>Séquençage par synthèse (Illumina)</b>	50 à 300 pb	99.9%	jusqu'à 6 milliards	1 à 11 jours	\$0.05 à \$0.15	Potentiel de rendement élevé de séquence, selon le modèle de séquenceur et l'application souhaitée	L'équipement peut être très coûteux. Nécessite des concentrations élevées d'ADN.
<b>Séquençage par ligation (séquençage SOLiD)</b>	50+35 ou 50+50 bp	99.9%	NA	20 minutes à 3 heures	\$2400	lectures longues. Utile pour de nombreuses applications.	Plus cher et peu pratique pour les grands projets de séquençage. Cette méthode nécessite aussi du temps pour l'étape de clonage de plasmide ou PCR.

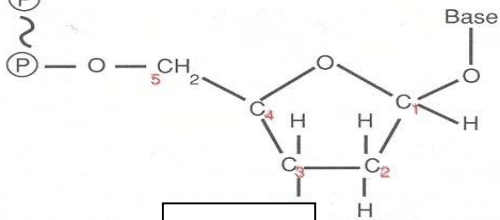
**Comparaison des méthodes de séquençage nouvelle-génération (avantages et inconvénients)**

# La méthode enzymatique de Sanger & Coulson



**Structure des di-désoxynucléotides utilisés dans la technique de Sanger**  
Le remplacement de l'hydroxyle en 3' par un hydrogène fait que ce nucléotide ne peut plus contracter de liaison phosphodiester, ce qui arrête la synthèse de DNA.

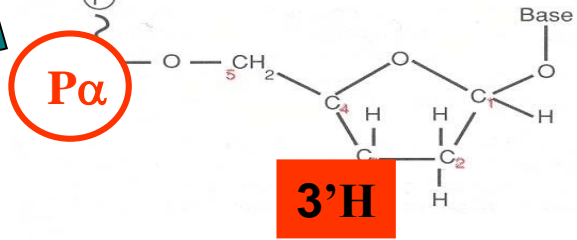
# 5'P- Chaîne d'ADN en élongation



5'P

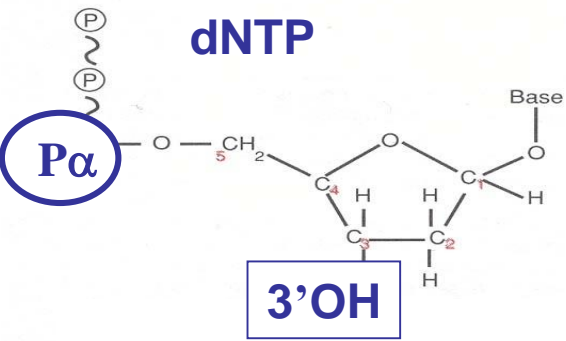
3'OH

+ ddNTP



3'H

dNTP



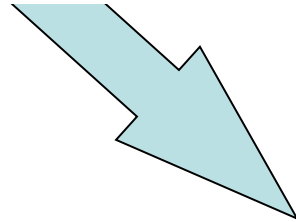
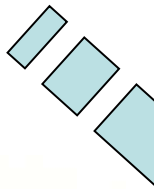
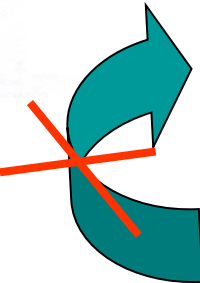
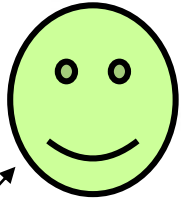
3'OH

Repose sur un principe de terminaison de chaines d'ADN en élongation par un ddNTP

3'H

Liaisons phosphodiester

STOP



- Le principe:

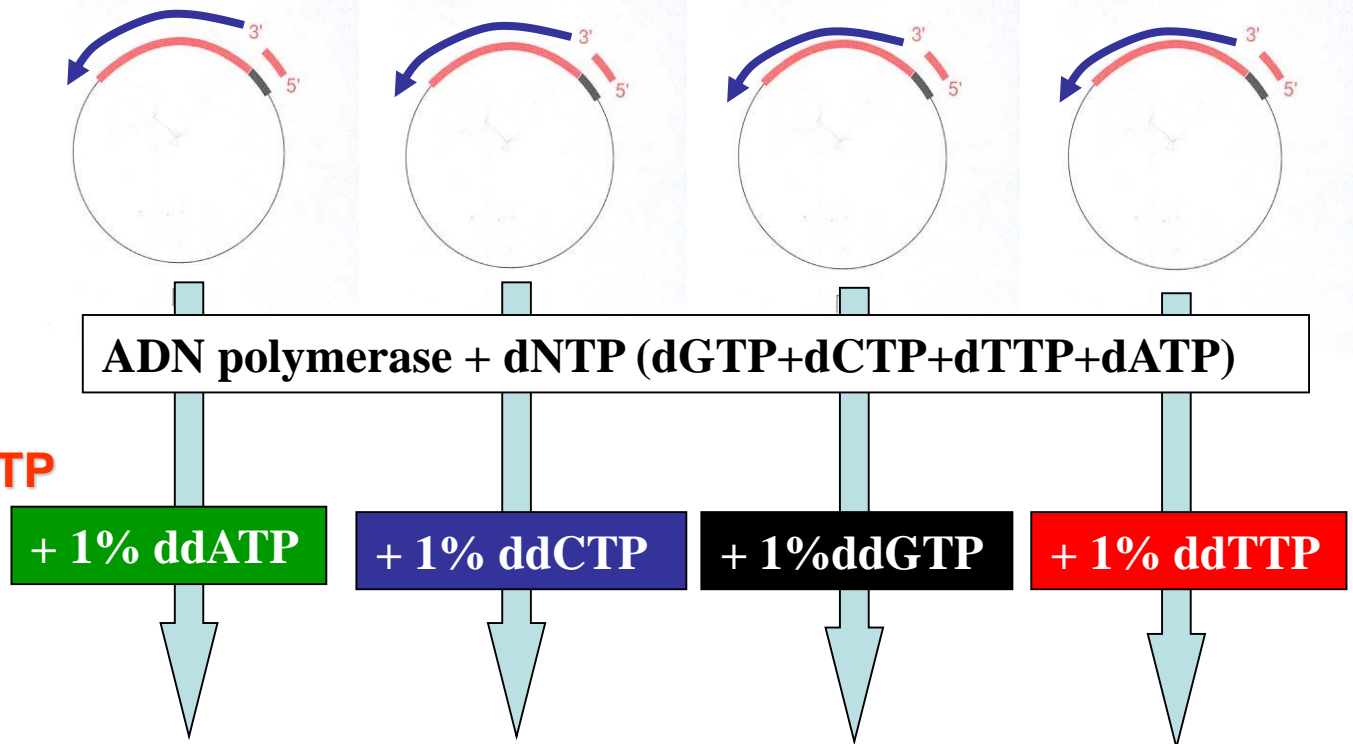
=> Polymérisation 5'-3'

-4 réactions (désignées ddA, ddC, ddG et ddT) sont effectuées en parallèle

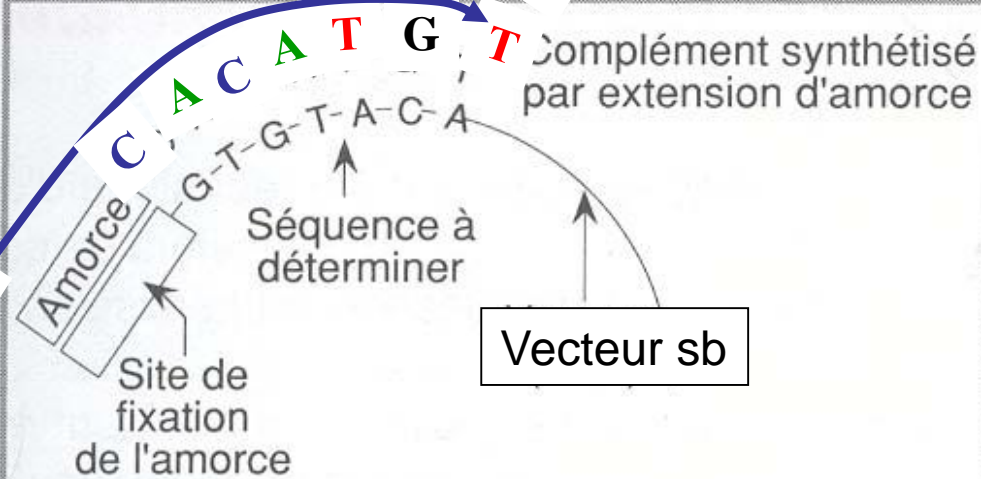
-Les ddNTP sont ajoutés à une quantité < à celles des dNTP

-Matrice ADNsb  
-ADN polymérase  
-Une amorce  
-dNTP  
→ Commun aux  
4 réactions

→ 1 seul des 4 ddNTP  
est ajouté aux  
4 réactions







### Réaction ddT



### Réaction ddA



### Réaction ddG



### Réaction ddC



=> Polymérisation 5'-3'

→ On génère autant de fragments (par extension d'amorce) qu'il y a de bases à séquencer

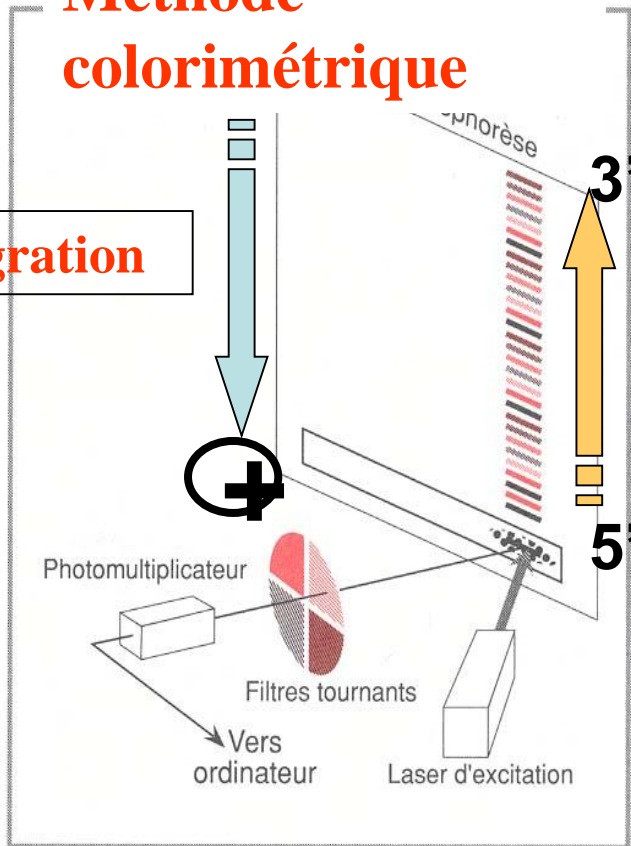
→ Tous ces fragments sont différents d'une base en taille et se terminent par un ddNTP

→ La taille du fragment indique la position de la base dans la séquence à déterminer

# La révélation des fragments d'ADN :

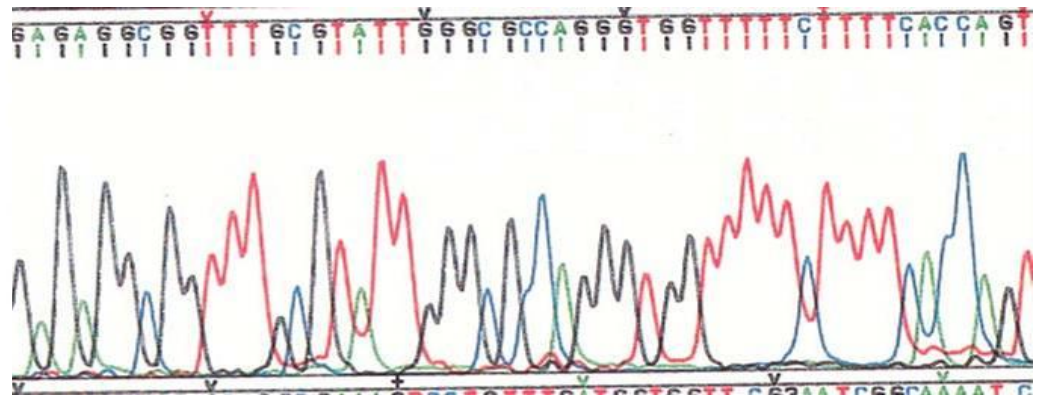
## Méthode colorimétrique

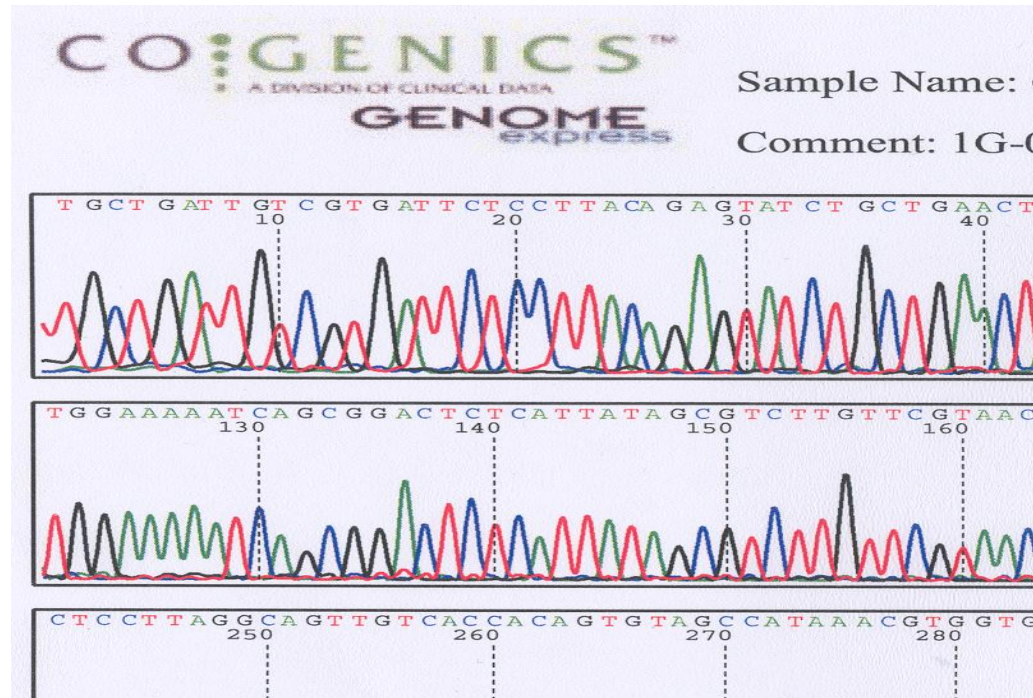
Migration



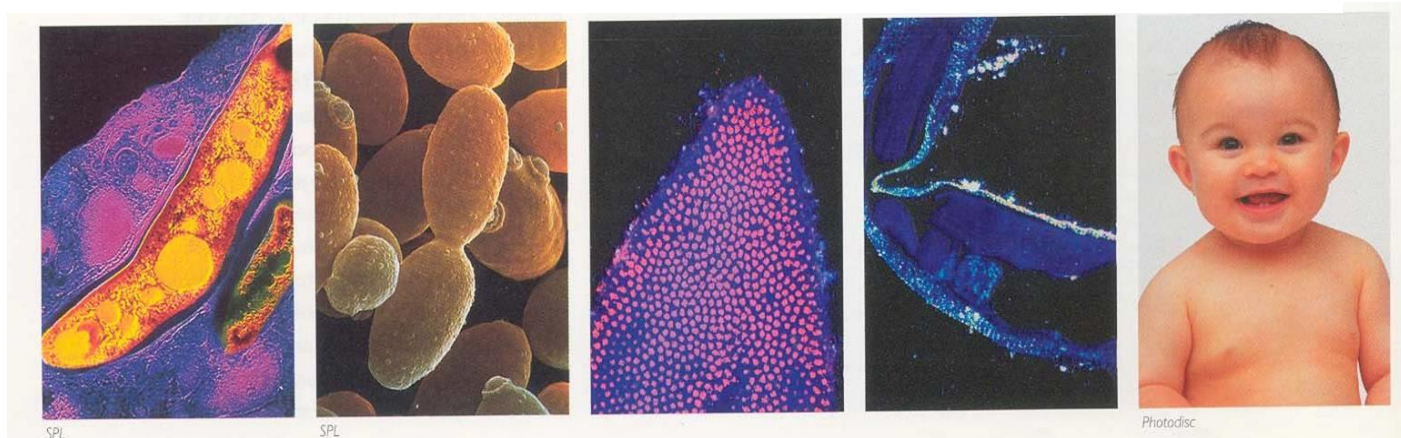
Lecture  
(séquence  
synthétisée)

5' → 3'





ORGANISME mycobactérie levure Drosophile ver nématode Homme



NBRE GENE 4000 6000 13000 18000 25000

# Le pyroséquençage

## PRINCIPE

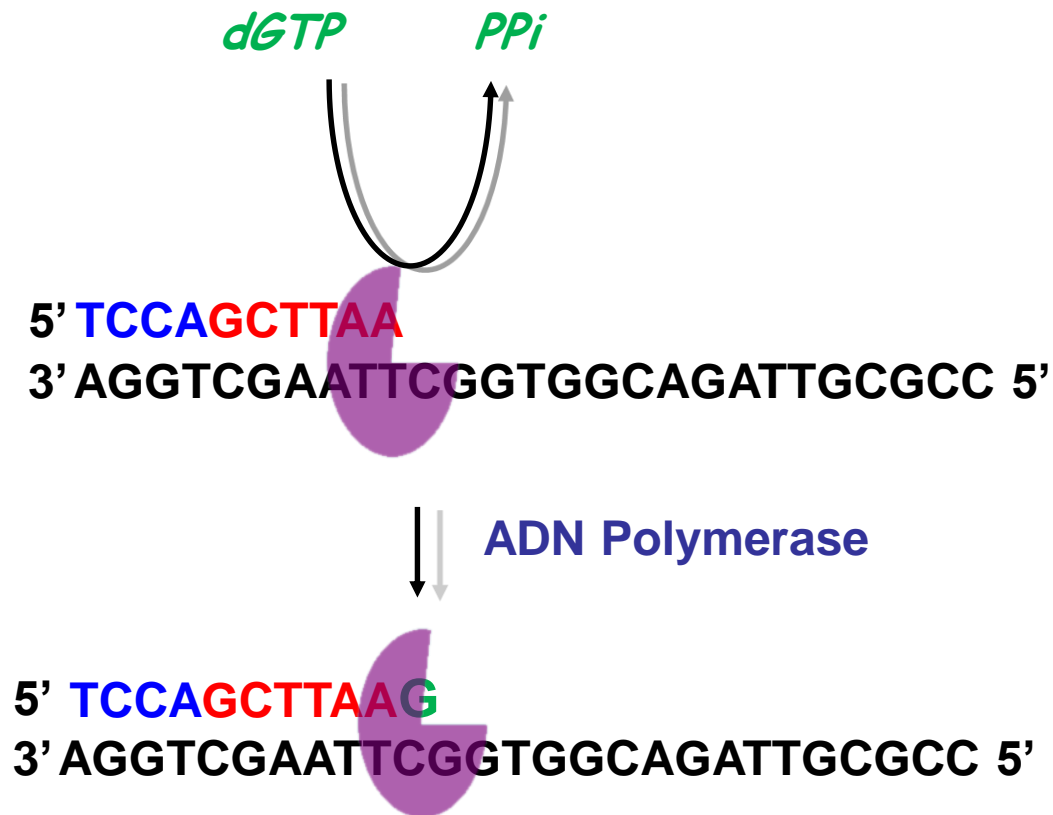
- **L'ADN est préalablement amplifié** (il faut disposer d'une quantité suffisante de matrice).
- Il est dénaturé, puis mis en présence d'une **amorce**, et d'une ADN polymérase.

5' **TCCA**  
3' AGG**T**CGAATTCGGTGGCAGATTGCGCC 5'

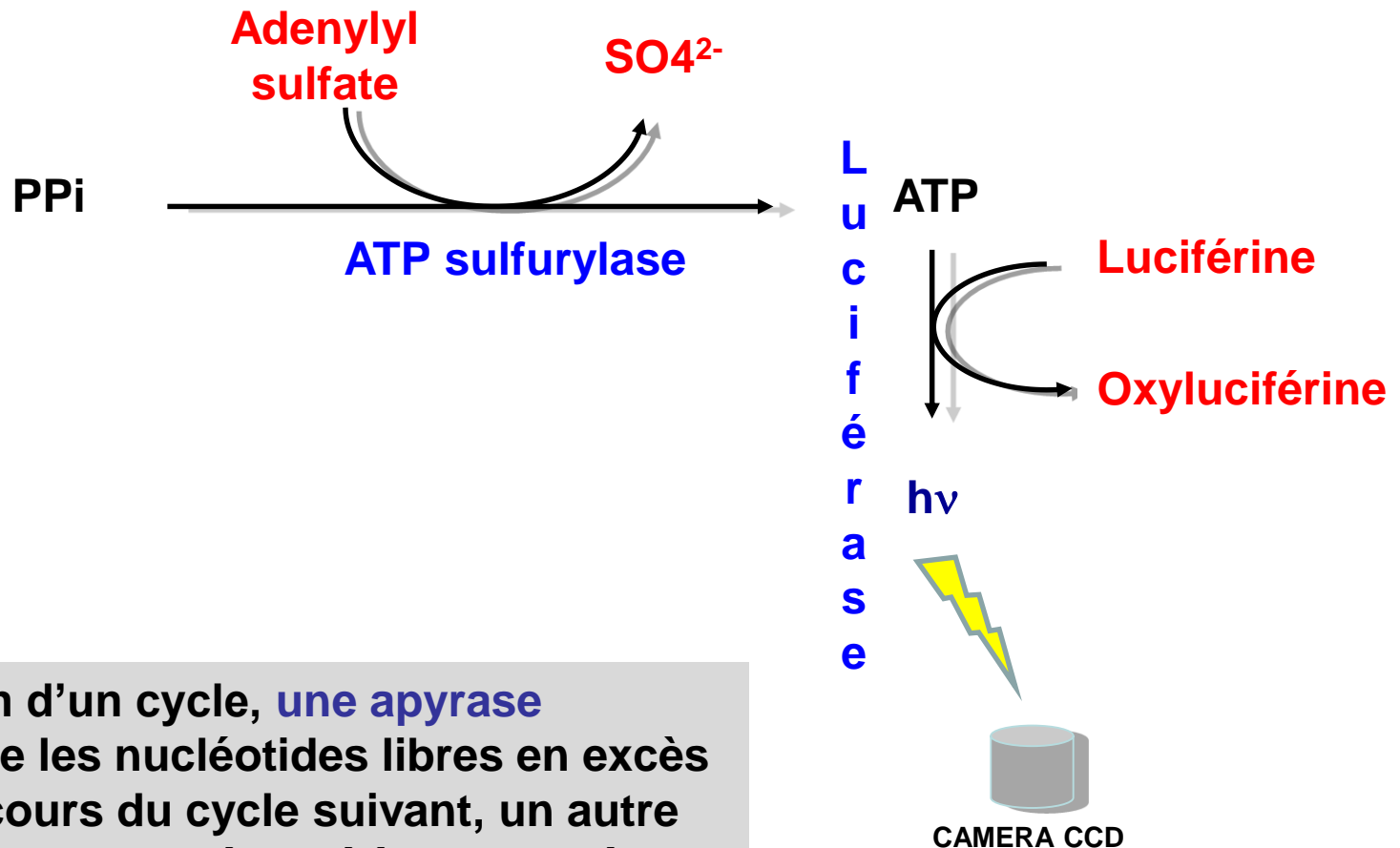
- Contrairement au séquençage classique de Sanger où tous les dNTP sont mélangés, ici ils seront ajoutés **séquentiellement**, **mais tout se passant dans même tube**.

Les cycles consistent à présenter à la polymérase séquentiellement les 4 nucléotides.

-A l'instant t du dessin ci-dessous, c'est seulement lorsque se présentera le dGTP (*qui est le nucléotide attendu par la polymérase*) que le PPI est libéré (*c'est lui qui sera à l'origine d'un flash lumineux*) et le G incorporé.



-La réaction qui provoque le flash à partir du PPI libéré est la suivante :



-A la fin d'un cycle, **une apyrase** dégrade les nucléotides libres en excès et au cours du cycle suivant, un autre nucléotide est présenté à la polymérase



## Résumé pyroséquençage:

Ici, les nucléotides ne sont pas ajoutés tous ensemble comme dans une réaction de séquençage normale (de type Sanger & Coulson) mais l'un après l'autre.

Si le nucléotide ajouté dans le milieu réactionnel correspond à celui attendu par l'ADN polymérase, il est incorporé dans le brin en cours de synthèse (d'élongation) et libère un pyrophosphate.

Une ATP sulfurylase vient alors transformer ce pyrophosphate (PPi) en ATP qui est alors utilisé, couplé à une luciférine, par une luciférase. On a alors production d'oxyluciférine et d'un signal lumineux (une apyrase dégrade les nucléotides en surplus).

C'est le séquenceur ou plus précisément un capteur CCD (Charge-Coupled Device) qui va capter ce signal lumineux et le reproduire sous forme d'un pic sur le *pyrogramme*.

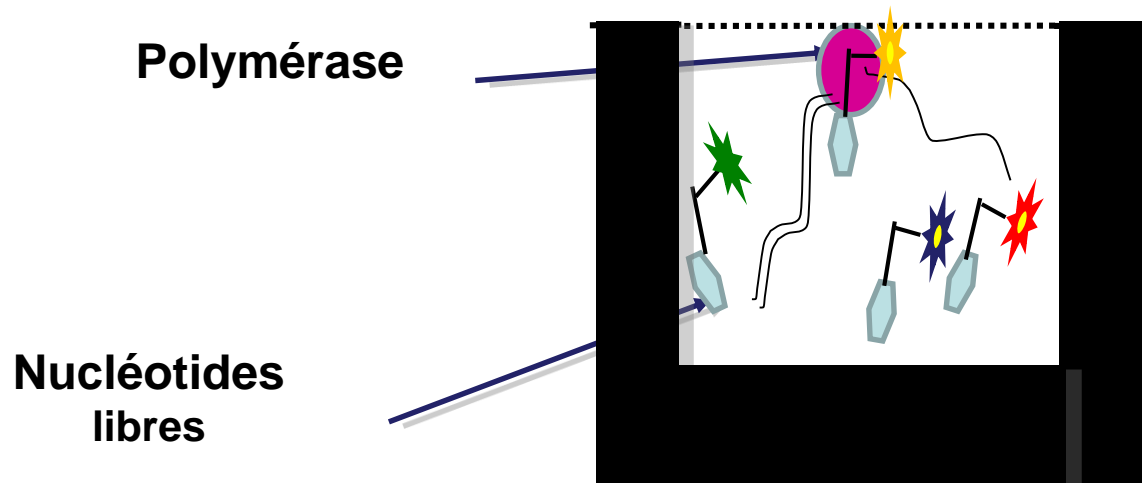
=> La hauteur de ce pic est fonction de l'intensité du signal lumineux, elle-même proportionnelle au nombre de nucléotides incorporés en même temps. On peut donc déduire la séquence à partir de la taille des pics obtenus.

# La SMRT (single molecule real time).

Ici, pas d'amplification de l'ADN, plusieurs molécules d'ADN de séquence différentes sont séquencées en même temps (séquençage en parallèle).

## PRINCIPE

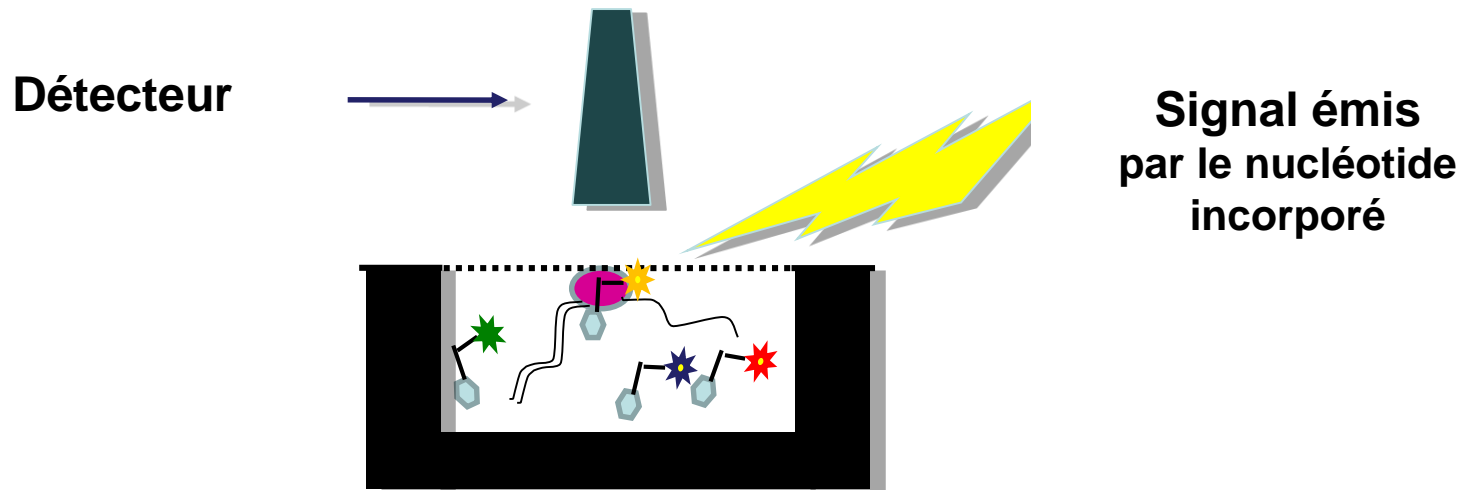
-une ADN polymérase est fixée sur une microcuvette de volume **extrêmement petit** ( $10^{-21}$ L).



-Exemple : Puce constituée d'une lame ayant 3000 micro trous. Chacun d'eux contient une molécule d'ADN différente

-Chacun des 4 dNTP est lié à un chromophore différent (qui émet une lumière fluorescente de longueur d'onde différente).

-La lumière excitatrice ne rentre pas profondément dans la cuve et n'excite que le chromophore lié à la polymérase  
*(donc n'excite pas les chromophores des nucléotides libres qui en moyenne sont plus bas)*

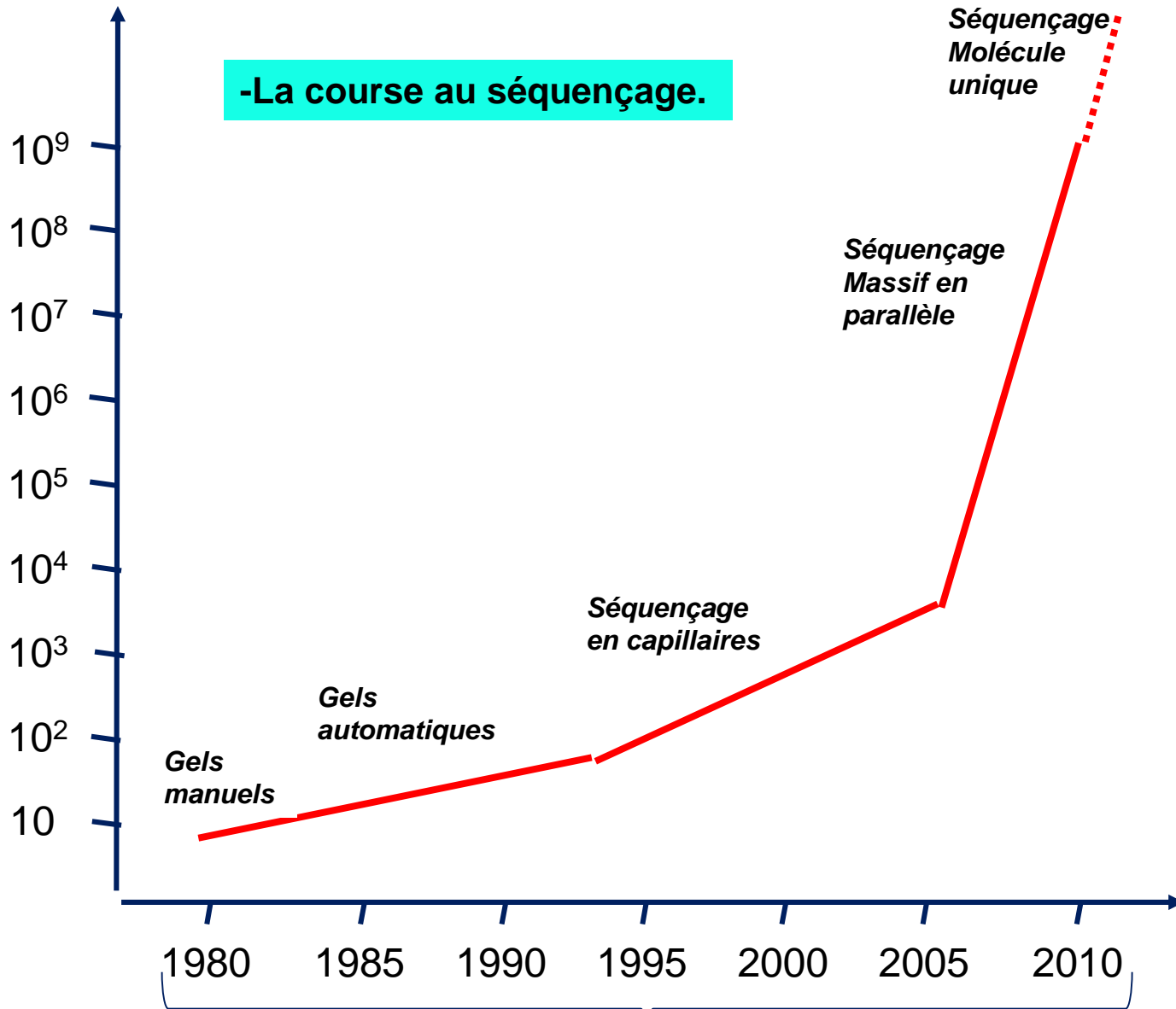


-Durant les 10ms où la polymérase « tient » le nucléotide marqué (juste avant de l'incorporer à la chaîne), la lecture se fait .

-La polymérase clive ensuite la liaison entre le nucléotide et le fluorophore et incorpore le nucléotide à la chaîne.

-La lecture se fait en même temps que la chaîne se forme  
*(à la vitesse de la polymérase (10nt/sec)), donc très rapide.*

**-La course au séquençage.**



Le coût du séquençage d'un génome humain pourrait être bientôt de 1000\$

CHUTE DU COÛT

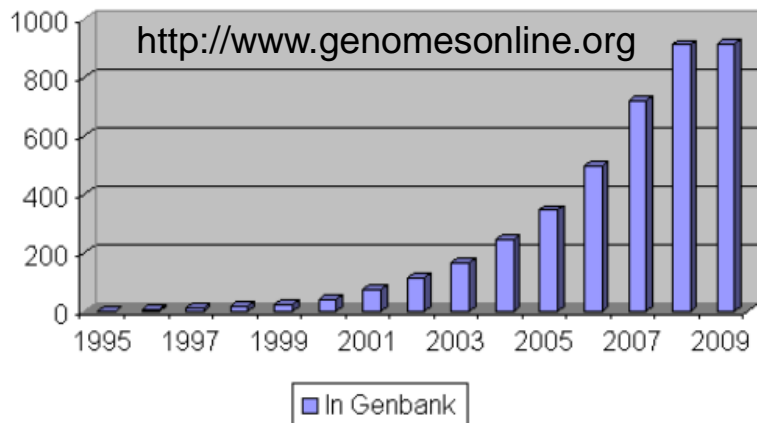
En trente ans le taux de séquençage est multiplié par un facteur  $10^8$

D'après M Stratton

## Applications des méthodes de séquençage : détections des mutations pathogènes et variants, méthylation de l'ADN, étude des génomes.

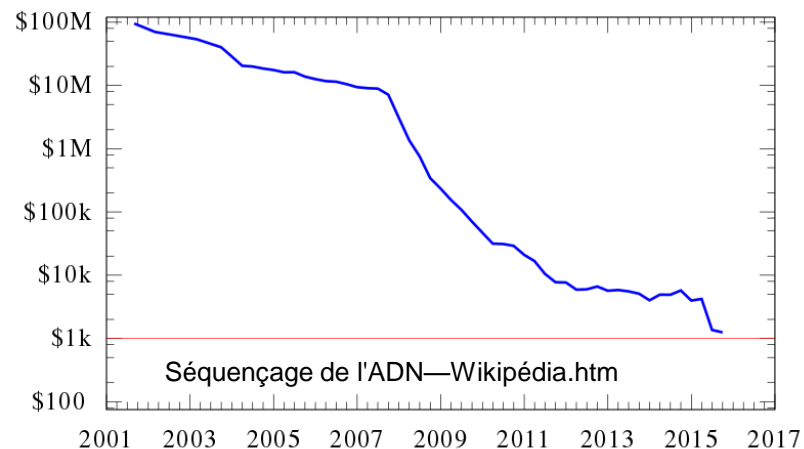
-La diminution du coût de séquençage pourrait permettre à cette technique de devenir rapidement dominante dans la caractérisation des génomes et la détection des mutations pathogènes et des variants (SNP, épimutations type méthylation, etc).

- Le séquençage de génomes d'organismes divers :
  - ⇒ plus de 1000 génomes séquencés 2009.



Le premier séquençage humain a été annoncé en 2003 et a coûté plus 2 milliards de dollars. L'évolution de la technique a permis une très forte baisse des prix.

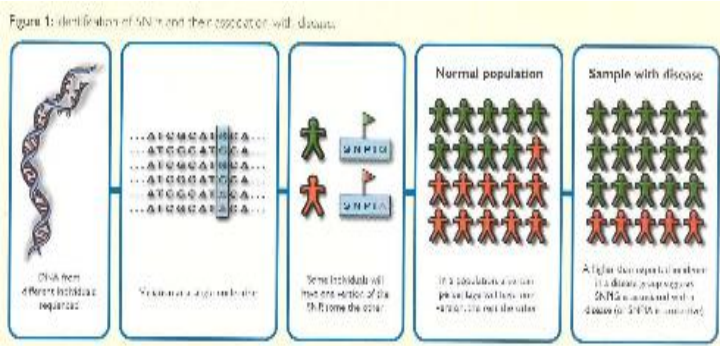
Cost to sequence a human genome (USD)



# Analyse de la diversité des génomes au sein d'individus d'une même espèce par séquençage

## Variation in the genome

Uncovering the variations in the human genome – in particular the single nucleotide polymorphisms – is generating almost as much excitement as the sequence itself. A new map of 1.42 million SNPs, the output of the International SNP Working Group, will prove invaluable in the search for genetic contributions to common diseases.



- identifier les mutations (ou épimutations) impliquées dans maladies (entre autre)

## Exemples d'objectifs du séquençage:



- étudier la relation facteurs environnementaux, gènes et maladies

Science & Vie d'aujourd'hui

KITS GÉNÉTIQUES

Que valent les analyses d'ADN grand public?

FAITS & CHIFFRES

Les tests génétiques utilisés en laboratoire, pour le dépistage de maladies génétiques par exemple, sont apparus il y a une vingtaine d'années; il en existe aujourd'hui plus de 1400, coûtant de 400 à 2400 €. Il y a quelques années, lire un seul SNP coûtait 1 \$, c'est moins de 1 centime aujourd'hui. 23AndMe, société leader du secteur, demande 315 € pour analyser 600000 SNPs du génome d'un client, deCODEme 778 € pour 1 million de SNPs. Navigenics est plus cher (1980 €) mais fournit un conseil génétique.

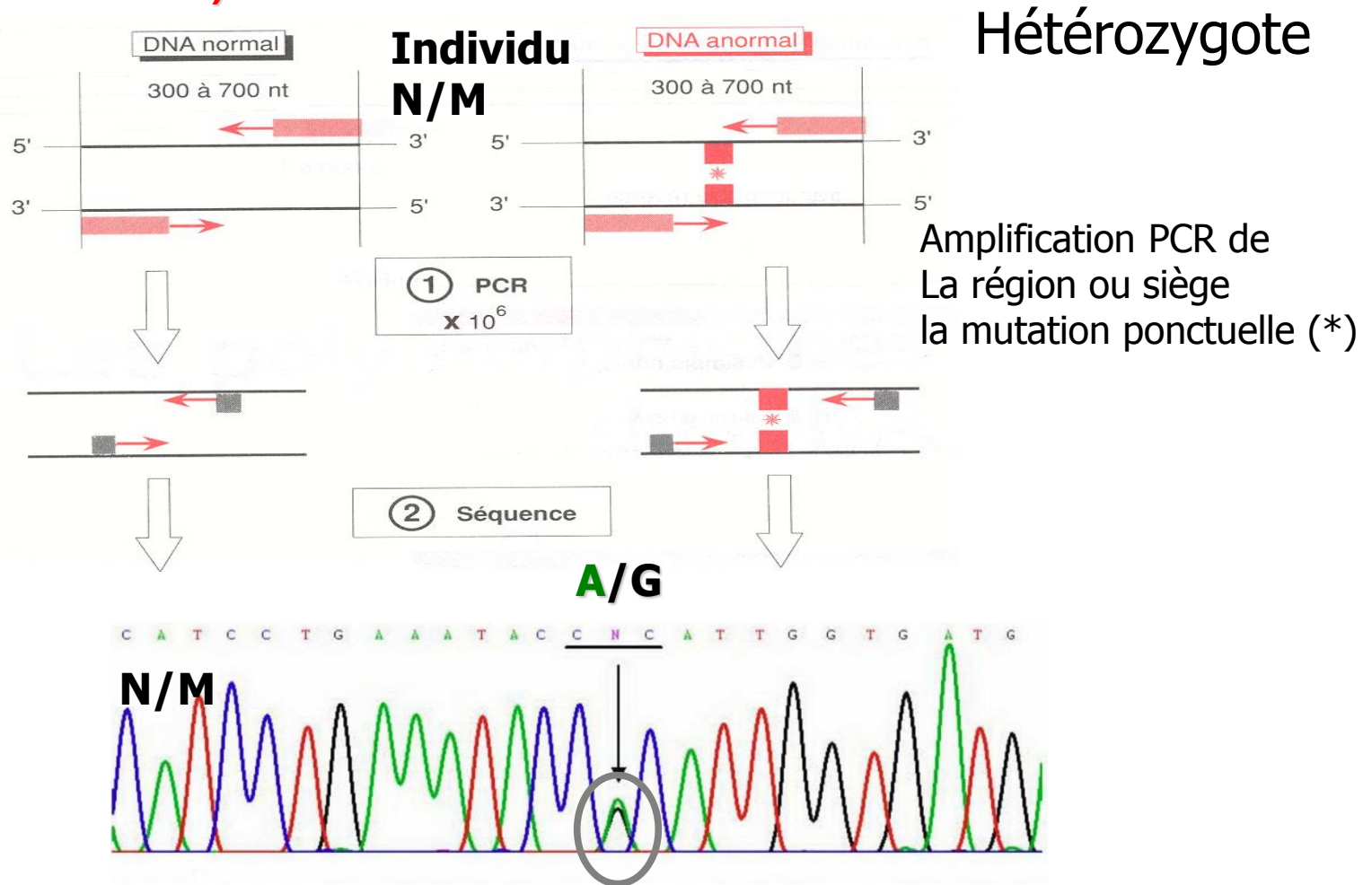
### Testing cancer patients

- New machines allow us to test cancer patients for mutations like B-Raf V600E from blood samples in 24 hours.
- Recent work at CR1 demonstrates the ability to monitor treatment response using mutation analysis.
- Information about these mutations can allow doctors to tailor the treatment for each patient.
- Patients and the NHS will benefit from "personalised cancer medicine".

- mise au point de kits génétiques pour le grand public (« médecine » prédictive ou personnalisée)



# - détection SNP par PCR et séquençage chez un hétérozygote (Sanger&Coulson)

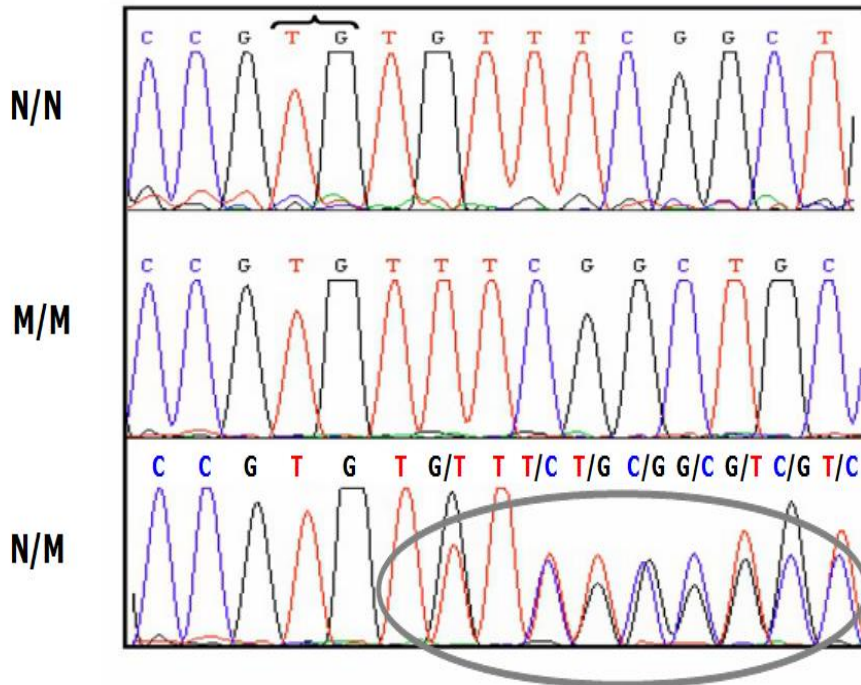


=> La superposition de 2 pics chez un hétérozygote indique la présence d'un SNP

- détection délétion (ou insertion) par séquençage  
(Sanger&Coulson)

Exemple de profils dans cas d'une délétion de 2 bases (TG) sur un gène

Exemple de délétion (F508del) impliquée dans maladie génétique monogénique (mucoviscidose)



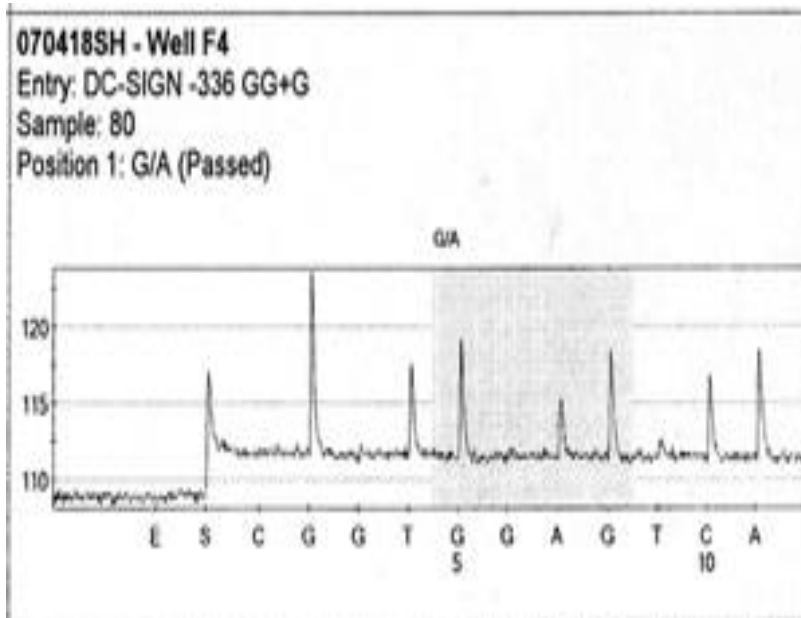
=> Superposition de nombreux pics chez hétérozygote

Position	504	505	506	507	508	509	510	511
<b>NORMAL</b>								
DNA	...GAA	AAT	ATC	ATC	TTT	GGT	GTT	TCC...
Protéine	Glu	Asn	Ile	Ile	Phe	Gly	Val	Ser
<b>CF ΔF508</b>								
DNA	...GAA	AAT	ATC	AT	- - -	T	GGT	GTT TCC...
Protéine	Glu	Asn	Ile	Ile			Gly	Val Ser

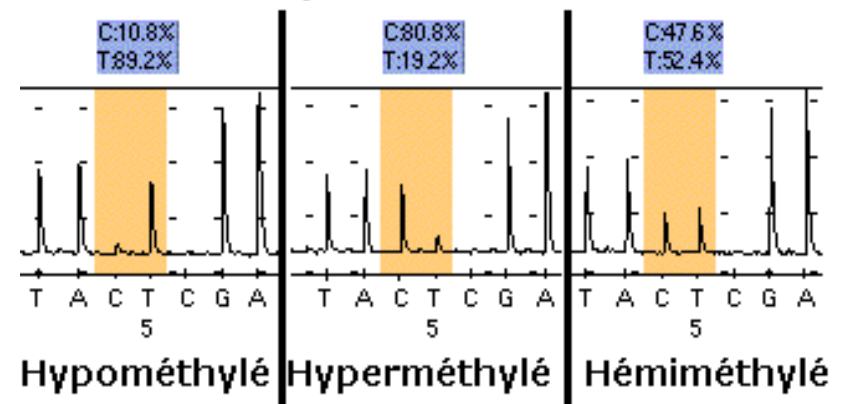
**Figure 14-44 L'anomalie ΔF508 dans le gène CFTR**  
La lésion est une délétion de 3 bases à cheval sur les codons 507 et 508 (exon 10). Les deux nucléotides provenant du codon 507 et le nucléotide provenant du codon 508 reconstituent un codon ATT (Ile) sans perturbation du cadre de lecture, ni changement de sens du codon 507. Il en résulte la perte du résidu Phe en 508.

## Détection de mutations, épimutations par pyroséquençage:

Par ailleurs, en cas de **mélange de nucléotides à une même position (polymorphisme de séquence)**, la taille des pics permet d'avoir une quantification de la proportion de brins porteurs de l'un ou l'autre des nucléotides.



**Application à l'étude de méthylation** : après traitement au Bisulfite de l'ADN, l'information épigénétique est transformée en un polymorphisme génétique C ou T au niveau de chaque CG de la région. L'ADN est dans un premier temps amplifié par PCR, puis passé au Pyroséquençeur. Les proportions respectives de C et de T pour chaque position peuvent donc être mesurées par pyroséquençage, et cette proportion reflète le pourcentage de brins méthylés ou déméthylés à ce niveau dans l'échantillon initial.



-Le pyroséquençage allié à la transformation de l'ADN par le bisulfite permet d'**analyser quantitativement la méthylation** d'une portion de séquence.

# B- OUTILS MOLECULAIRES D'IDENTIFICATION DES MUTATIONS

## 1- Basés sur l'amplification génique

- Rappels sur la méthode de base (PCR classique)
- La PCR en temps réel (Real Time PCR)
- la PCR-RFLP (détection mutation touchant un site de restriction)
- la SSCP (Single Strand Conformational Polymorphism)
- la DGGE (Denaturing Gel Gradient Electrophoresis)
- DHPLC (Denaturing High Performance Liquid Chromatography)
- la HRM (Analyse des courbes de fusion à haute résolution)
- la MLPA (Multiplex Ligation dependant Probe Amplification)

## -2- Basés sur le séquençage d'ADN

- Rappels sur la méthode de base de séquençage d'un ADN (Sanger & Coulson)
- Le pyroséquençage et le SMRT (Single Molecule Real Time)

## -3- Basés sur l'hybridation moléculaire

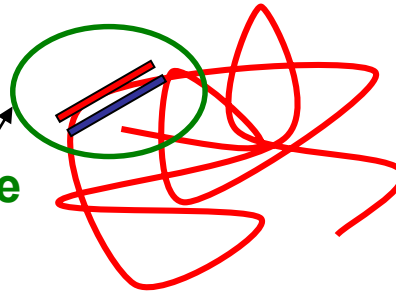
- Southern blot
- puces à ADN

# IDENTIFICATION DE MUTATION PAR Southern blot

Permet d'**identifier**, à l'aide d'une sonde spécifique, **un gène** (ou toute séquence d'ADN génomique) **dans un génome fragmenté**

*(dans le cas d'un génome non séquencé, cette méthode donne une idée de l'organisation génique d'une séquence isolée dans ce génome)*

Séquence d'ADN  
d'intérêt du génome  
humain  
(connu)



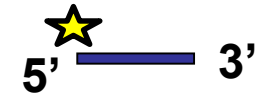
Sonde ADNdb

(Quelques centaines de pb)



Sondes ADNsb

(Quelques dizaines de base)



OU



## 1- Le Southern blot

*Ciblage du gène par 1 sonde  
ADN double brin ou simple brin  
(hybridation moléculaire)*

# Notions d'hybridation moléculaire (**les sondes ADN**)

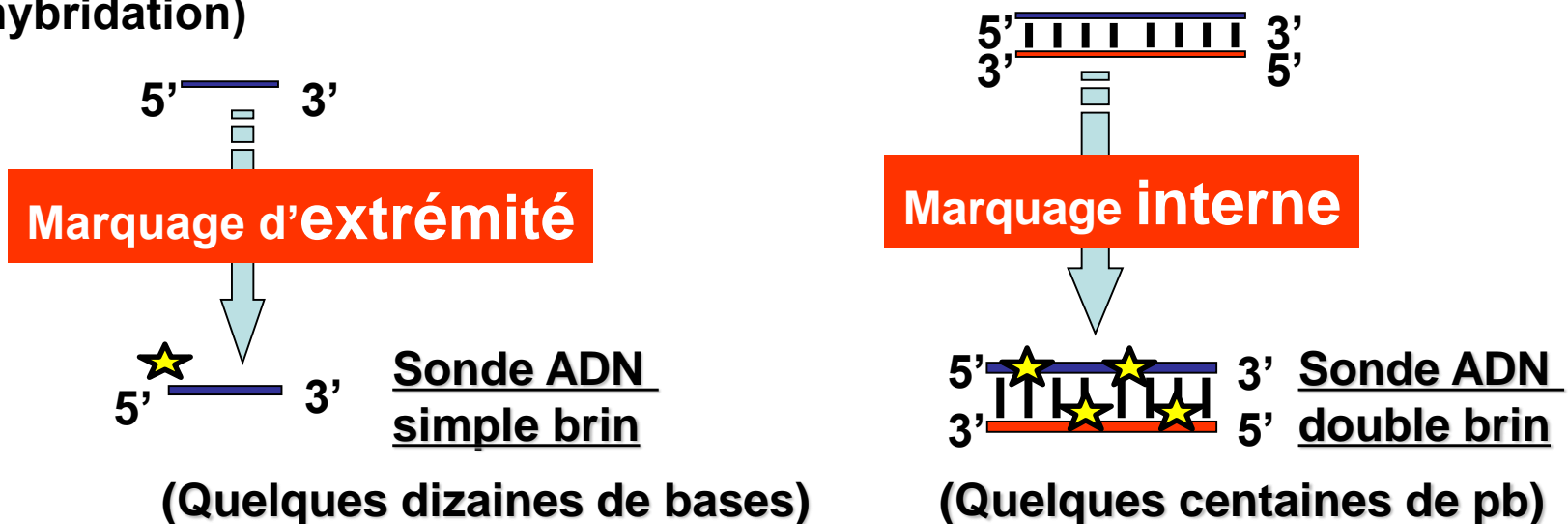
## Une **SONDE ADN**:

- correspond à une séquence (partielle) d'un gène (ou tout autre séquence d'ADN) que l'on cherche à détecter

- moyen de détection très spécifique d'un gène (par appariement de bases complémentaires A/T et G/C entre la sonde et le gène) dans un génome par la méthode du Southern blot.

*(Remarque: dans le cas d'ARNm, on parle de ribosonde et de Northern blot)*

– la séquence d'ADN de la sonde est simple (oligonucléotide) ou double brin et on peut la marquer (★) à volonté (en vue de la détection de l'hybridation)



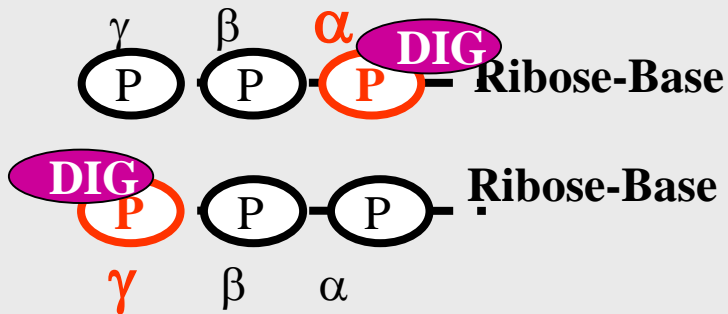


# Notions d'hybridation moléculaire (marquage d'une sonde)

Le marqueur de la sonde est un isotope radioactif ( $P^{32}$  ou  $P^{33}$ ) ou non (la digoxygénine par exemple) porté par le phosphate en position  $\alpha$  ou  $\gamma$  d'un désoxyribonucléoside triphosphate (dNTP)

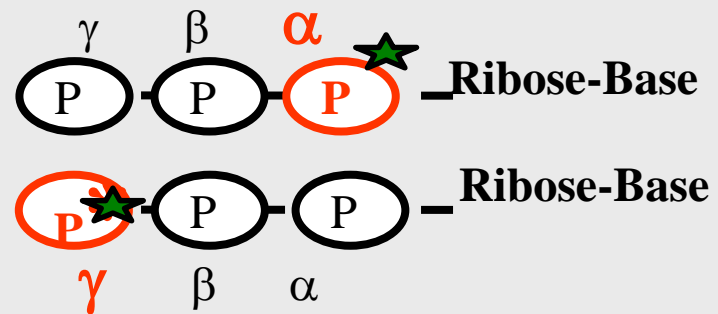
## Sondes non radioactive « FROIDES »

(Ex: digoxygénine) 



## Sondes radioactives « CHAUDES »

( $P^{32}$ ,  $P^{33}$ ) 



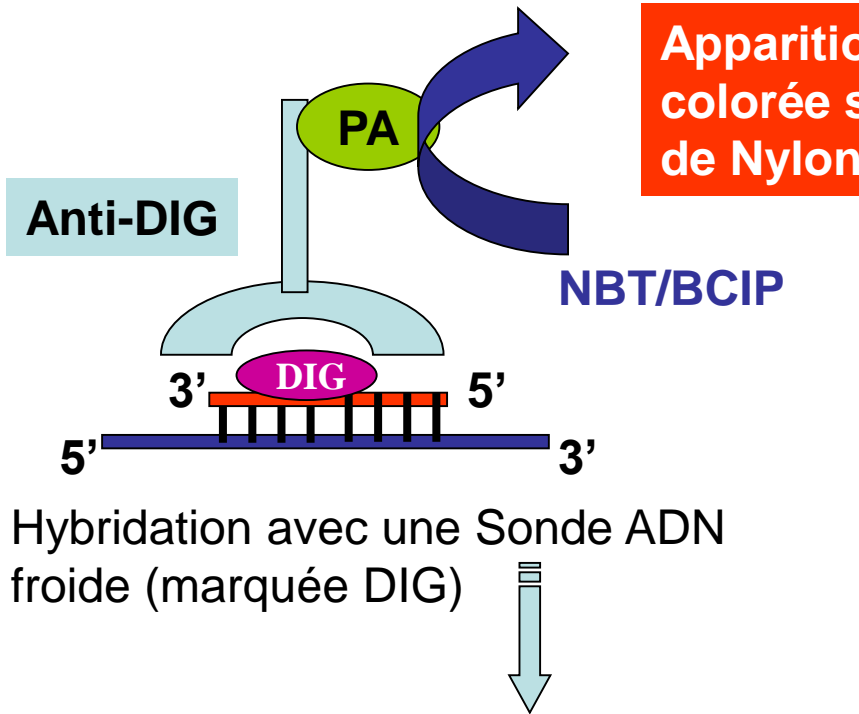
Ce dNTP marqué est :

- intégré dans une sonde ADNdb par polymérisation d'ADN (via ADN polymerase)

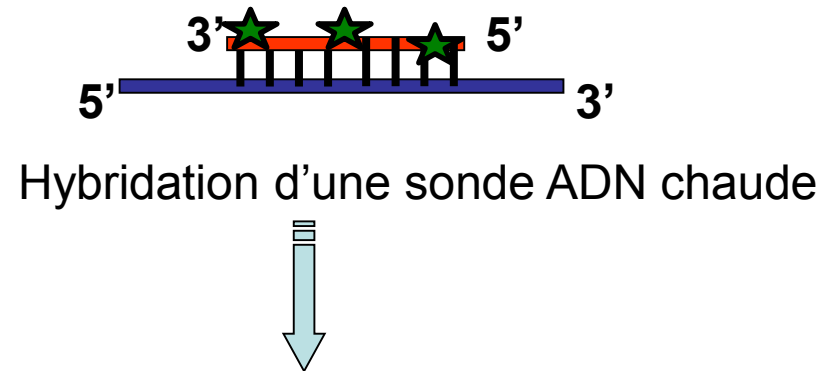
- transféré en extrémité de la sonde simple brin par une polynucléotide kinase



# Notions d'hybridation moléculaire (révélation de l'hybridation)



Révélation de l'hybridation par **colorimétrie** (via un anticorps anti-Digoxygénine (anti-DIG) couplé à la phosphatase alcaline (PA) et ajout de substrats (NBT/BCIP)).



Révélation de l'hybridation par **autoradiographie** (exposition sur un film à l'obscurité puis développement du film;

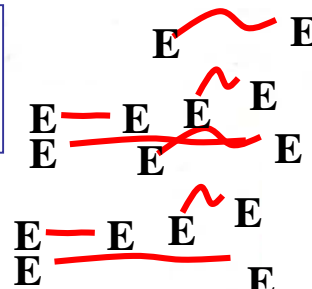
Apparition d'une bande radioactive sur un film

# Le Southern blot (principe)

la méthode nécessite **beaucoup plus d'ADN** (quelques  $\mu\text{g}$ )  
**que la PCR** (quelques ng). De plus, les délais d'obtention  
des résultats sont **beaucoup plus longs** (24h contre 2-4h en  
PCR)

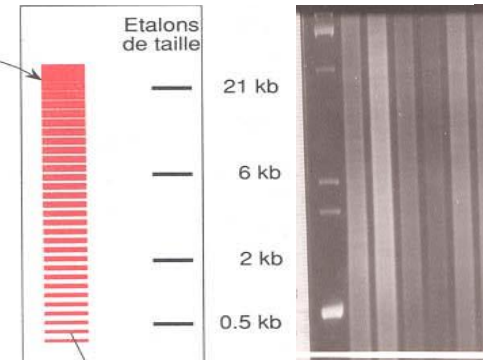
**ADN génomique  
purifié (quelques  $\mu\text{g}$ )**

**Digestion  
Par ER** 1



**Séparation des  
fragments selon leur  
taille par électrophorèse** 2

**Coloration BET**

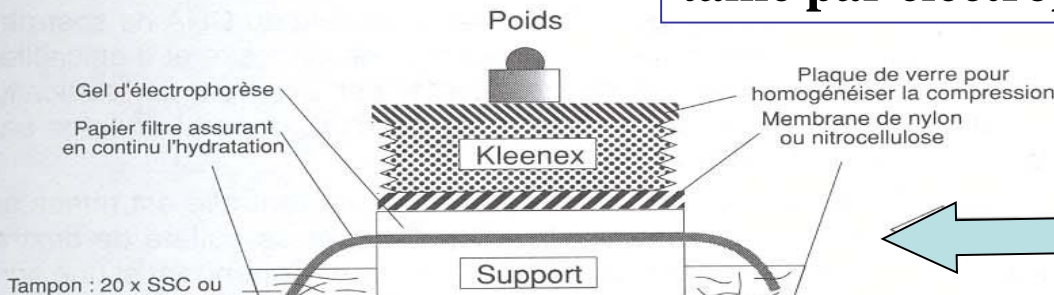


Visualisation des  
fragments de DNA  
séparés en fonction  
de leur taille

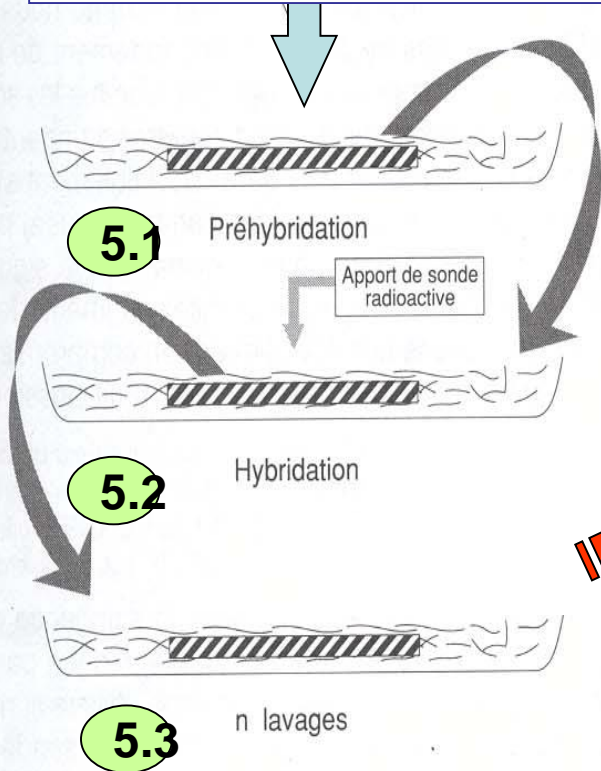
**Dénaturation des  
fragments d'ADN** 3

Traitement des gels par la soude  
⇒ cassure + passage en simple brin

**Transfert (BLOT) des fragments  
d'ADN simple brin sur mb NYLON** 4



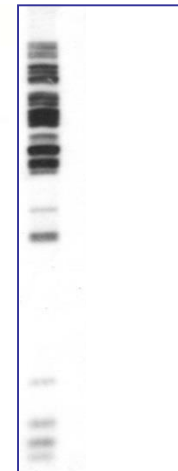
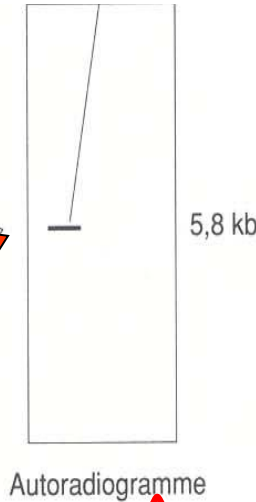
# Hybridation de la membrane de Nylon avec une sonde ADN correspondant au gène 5



Profil simple

OU

Profil complexe



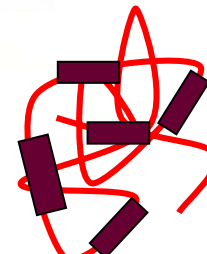
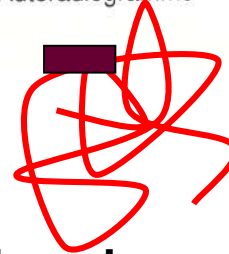
Révélation de l'hybridation et interprétation 6

interprétation de l'hybridation 7

Organisation génomique de la séquence hybridée:

Copie unique

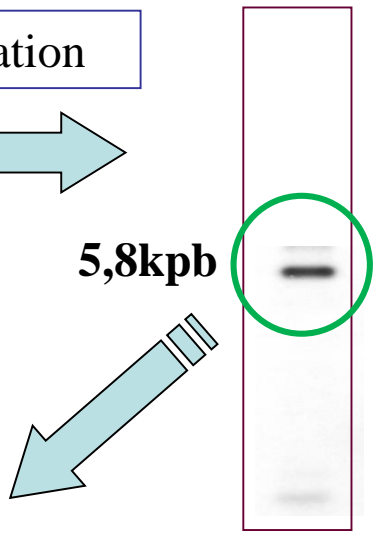
multicopie



# Le Southern blot (détection d'un gène dans un génome)

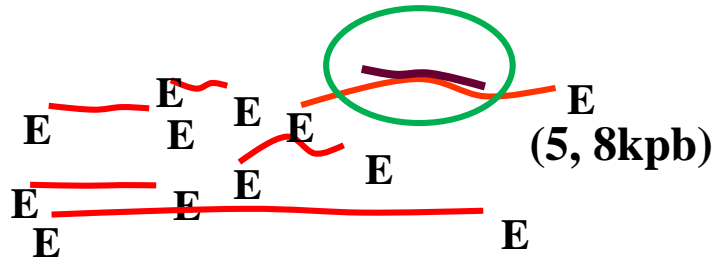
Sonde 1  
(500pb)

Hybridation

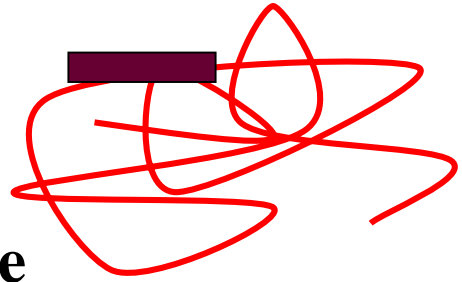


Cas d'un profil simple

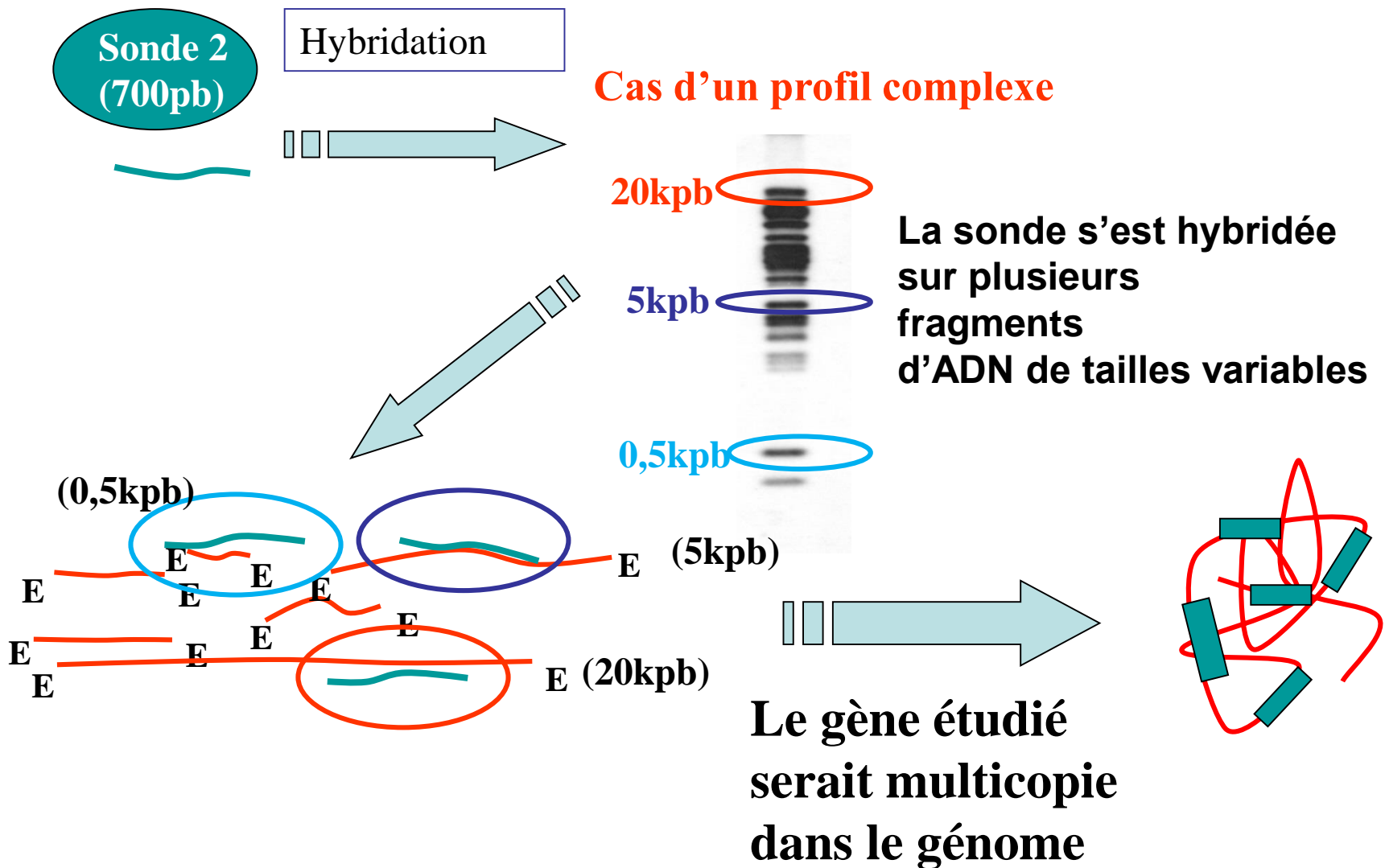
La sonde s'est hybridée sur un unique fragment d'ADN (ne signifie pas que le gène fait 5,8kpb mais qu'il est localisé dans un fragment génomique de 5,8 kpb)



Le gène étudié serait en copie unique dans le génome



# Le Southern blot (détection d'un gène dans un génome)

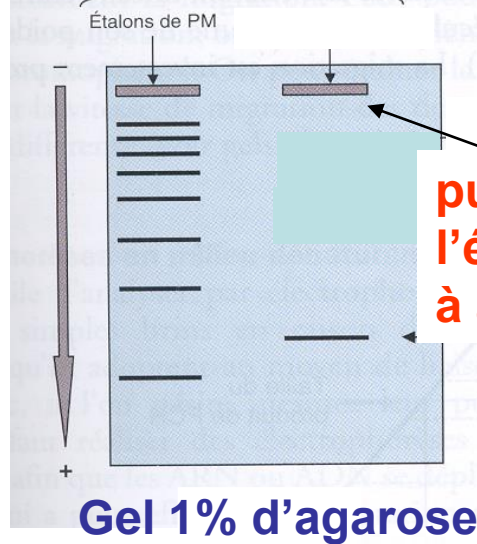




# Séparation électrophorétique (classique ou en champ pulsé) de l'ADN et visualisation

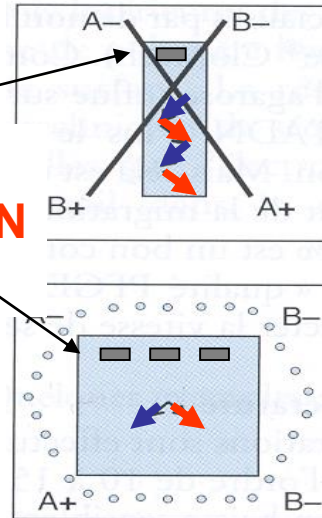
## 1 orientation de migration

### Electrophorèse classique



## 2 (ou plus) orientations de migration

### Electrophorèse en Champ pulsé



puits de dépôt de l'échantillon d'ADN à analyser

Fragment d'ADN

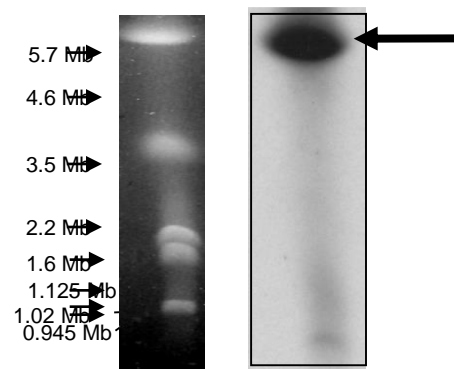
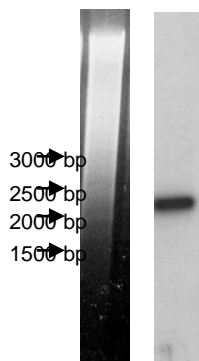
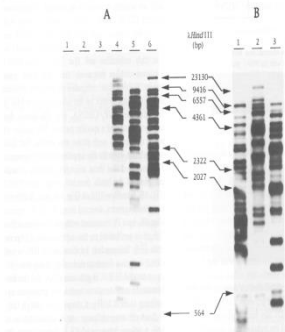
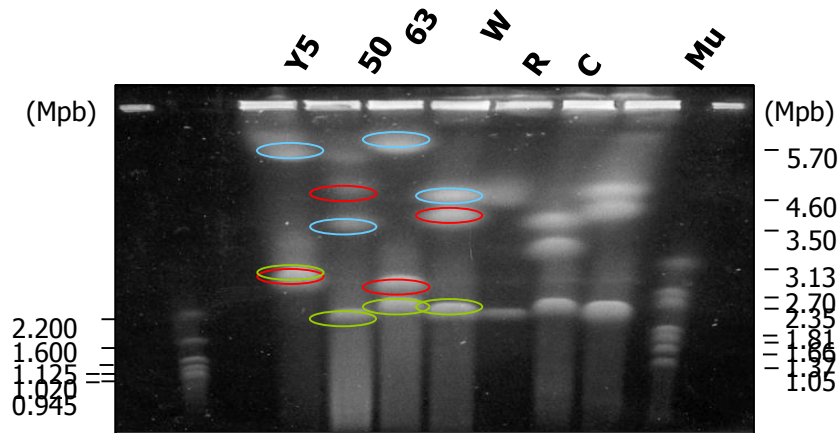


Fig. 4. The P903.1 clone belongs to a multigene family and is distinct from other described multicopy sequences of *P. falciparum*. Southern blot analysis of a PUP/CB genomic DNA (A) digested with *AccI* (lane 4), *AclI* (lane 5) and *HinfI* (lane 6), and probed under stringent conditions with the P903.1 locus, together with *SalI*-digested human DNA (lane 1), *P. chabaudi* DNA (lane 2) and *B. abortus* DNA (lane 3). (B) PUP/CB genomic DNA digested with *TaqI* and hybridized successively with the P903.1 (lane 1), T98.9 (lane 2) and P903.1 (lane 3) probes. Position and size in bp of the *HinfI* fragments of the background *ADNA* are marked by arrows.

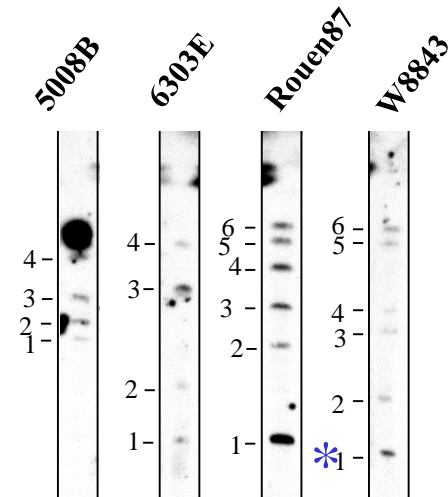
# Application Southern à la détection de macrolésions (Ex: Mise en évidence de macrodélétion/macroinsertion)

observée chez des individus différents d'une même espèce (autre que l'Homme)

Analyse PFGE de différent individus (la sonde hybride sur le chromosome rouge)



Analyse PFGE-RFLP de différents individus (la sonde indique la présence d'au moins 4 à 6 copies génique sur le chromosome rouge)

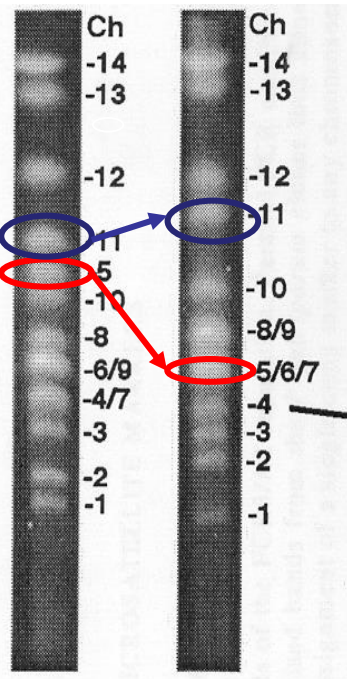


4 copies

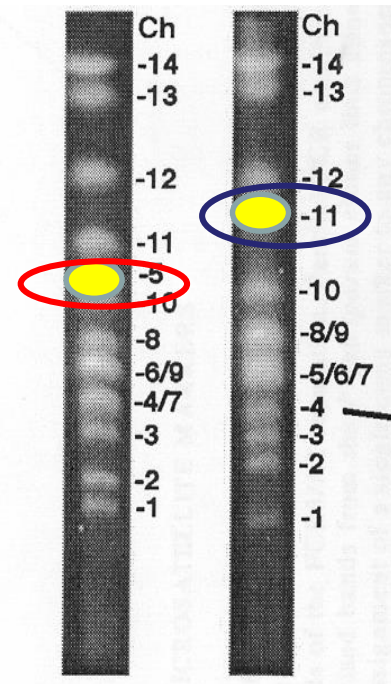
6 copies

# Application Southern à la détection de macrolésions (Ex: Mise en évidence de polymorphisme chromosomique et d'une translocation chromosomique)

observée chez des individus différents d'une même espèce (autre que l'Homme)



**polymorphisme chromosomique**  
(chr5 diminue de taille alors que le chr11 augmente de taille)



**translocation chromosomique**  
(suivi d'un gène (hybridation en jaune) situé normalement sur le chr5 et qui se retrouve sur chr11)

# Les puces à ADN

## Généralités

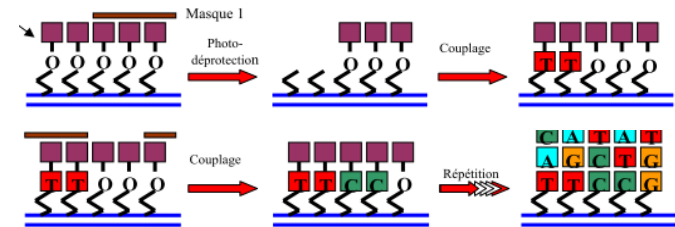
Des séquences d'ADN plus ou moins longues (qui seront l'équivalent de sondes) qui représentent des morceaux de gènes sont disposées sur des supports à l'aide de différentes techniques :



Robot qui utilise une aiguille fine pour déposer des polynucléotides



Simplex dépôts de polynucléotides (sondes) plus ou moins longs



ou plus récemment, synthèse *in situ* de sondes (système de protection avec des masques, et réaction induite par la lumière).

Une biopuce, **puce à ADN**, ou micromatrice d'ADN, est un ensemble de molécules d'ADN fixées (« sondes») en rangées ordonnées sur une petite surface qui peut être du verre, du silicium ou du plastique.

Cette biotechnologie récente permet d'analyser le niveau d'expression des gènes (transcrits) dans une cellule, un tissu, un organe, un organisme ou encore un mélange complexe, à un moment donné et dans un état donné par rapport à un échantillon de référence.

Les puces à ADN sont aussi appelées *puces à gènes*, *biopuces*, ou par les termes anglais « *DNA chip*, *DNA-microarray*, *biochip* ».

=> Le principe de la puce à ADN repose sur la propriété que possède l'ADN dénaturé de reformer spontanément sa double hélice lorsqu'il est porté face à un brin complémentaire (**réaction d'hybridation**). Les quatre bases azotées de l'ADN (A,G,C et T) ont en effet la particularité de s'unir deux à deux par des liaisons hydrogènes (A = T et T = A ; G ≡ C et C ≡ G).

## Principe d'utilisation des puces à ADN :

Concrètement, les ARN totaux sont extraits de cellules (A ou B), dont on veut comparer l'expression des gènes avec un étalon, et subissent une amplification qui va permettre d'obtenir une quantité de matériel génétique suffisante pour l'expérience.

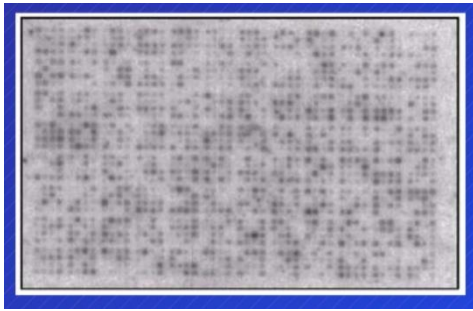
Ensuite ces ARNm sont transformés en (ADNc) par la technique de rétrotranscription et marqués par un colorant : soit la Cyanine 3 (**fluorochrome vert ; ADNcA**) soit la Cyanine 5 (**fluorochrome rouge ; ADNcB**).

On met ensuite les ADNc obtenus dans une puce contenant des fragments d'ADN (sondes), en même temps que l'ADNc étalon. Les 2 types d'ADNc (A et B) s'hybrident sur les mêmes séquences complémentaires présentes sur la puce.

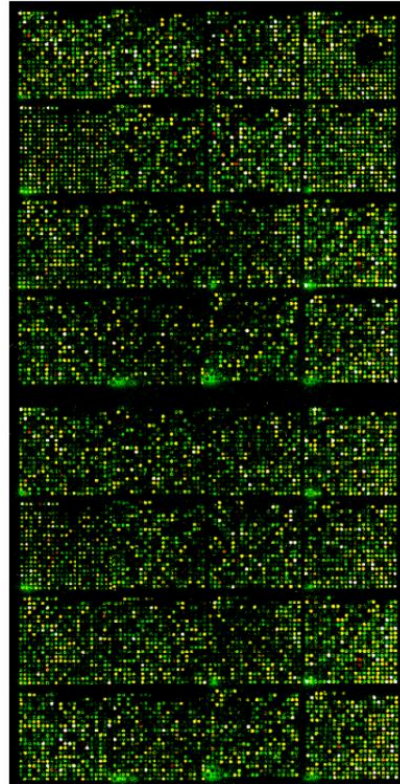
Chaque point (ou *spot*) de la puce va être analysé individuellement par un scanner à très haute résolution, et ce à la longueur d'onde d'excitation de la Cyanine 3 puis de la Cyanine 5. L'image scannée va être traduite en niveaux de gris. On va ensuite comparer l'intensité du signal entre le vert et le rouge. En fonction de l'intensité du signal il y aura plus ou moins de pixels pour chaque point de la puce. À chaque point (ou *spot*) est attribué une valeur d'intensité normalisée par rapport à l'ADN "étalon" : on parle de *spike*. Chacune des valeurs peut être analysée par des techniques de bio-informatique, ce qui permet d'estimer avec plus ou moins de précision l'intensité d'expression d'un gène.

Par exemple on peut marquer **l'ADNc du malade en vert** et **du traité en rouge**, ou bien, du **témoin en rouge** et du **traité en vert**. Ce marquage se fait habituellement grâce à une enzyme : la polymérase T7 qui amplifie l'ARNm et incorpore les cyanines pour un marquage optimal. Une fois marqués ces ADNc sont déposés sur la lame de verre qui, elle-même, possède fixée à sa surface, des fragments de génome humain recouvrant tous les gènes présents dans une cellule. Les molécules d'ADN fixées sur la lame sont appelées des *sondes* même si la nomenclature peut varier. Des dizaines de milliers de sondes peuvent être fixées sur une même puce.

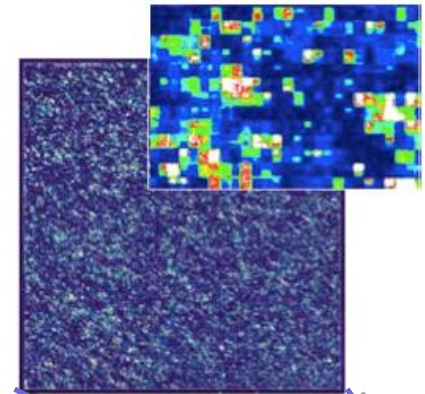




**100-200  
spots/cm<sup>2</sup>  
sur nylon**



**1000 à 10000  
spots/cm<sup>2</sup>  
sur verre**



**250000  
spots/cm<sup>2</sup>  
sur verre**



**La densité des dépôts n'a pas cessée d'augmenter ces dernières années**

## Applications en Biologie médicale

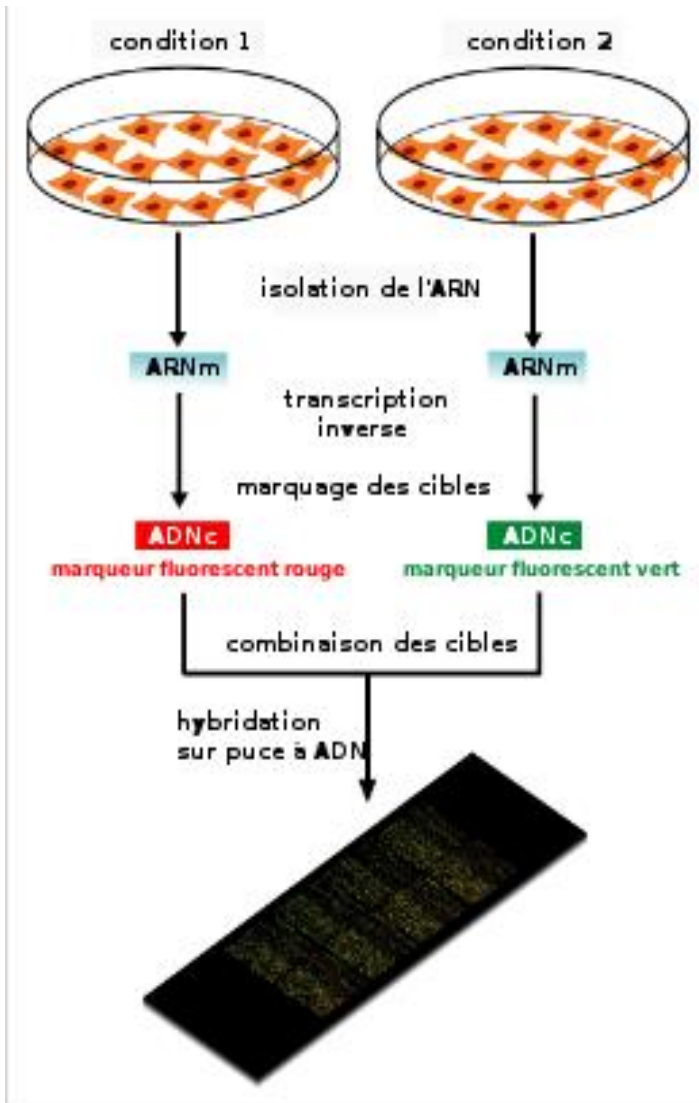
L'utilisation des puces à ADN connaît un essor croissant notamment dans le domaine de la **cancérologie** pour le typage tumoral d'après leur profil génétique.

L'utilisation **des puces à ADN comme outil de diagnostic** présente l'avantage de faire appel à de nombreux marqueurs : plusieurs milliers de gènes peuvent être criblés simultanément pour fournir une signature du type cellulaire étudié. Si l'on considère que chaque type de tumeur présente une signature génétique unique, ce système permet virtuellement de distinguer et classer tous les types de tumeurs.

=> Si un patient est porteur d'une maladie, les brins extraits de l'ARN d'un patient (et rétrotranscrits en ADNc), vont s'hybrider avec les brins d'ADN synthétiques représentatifs de la maladie.

Les puces à ADN permettent donc de comparer l'expression des gènes de deux types cellulaires différents, de faire de l'étude des gènes exprimés sur un grand nombre de patients pour observer l'effet d'un médicament (anti-cancéreux par exemple), de regarder l'effet d'un traitement sur l'expression des gènes, de comparer tissus sains contre tissus malades, traités contre non-traités etc...

L'approche Puce à ADN permet en une seule expérience qui dure environ 2 jours d'avoir une estimation sur l'expression de plus de 30000 gènes.



## \*Puces d'expression ( analyse des ARNm)

=> connaitre les différences entre les ARN produits par les cellules A et ceux produits par les cellules B.

## \*Puces CGH Array (perte et gain d'ADN génomique).

=> Permet d'évaluer les pertes ou les gains de séquences d'ADN génomique entre un individu sain et un individu malade (profilage des tumeurs).

## \*Puces de génotypage (détection de milliers de SNP; READIT™SNP).

-Il existe des puces (**dites de génotypage**) capables de détecter en une seule expérience des milliers de SNP (polymorphisme du simple nucléotide) existant dans l'ADN génomique :

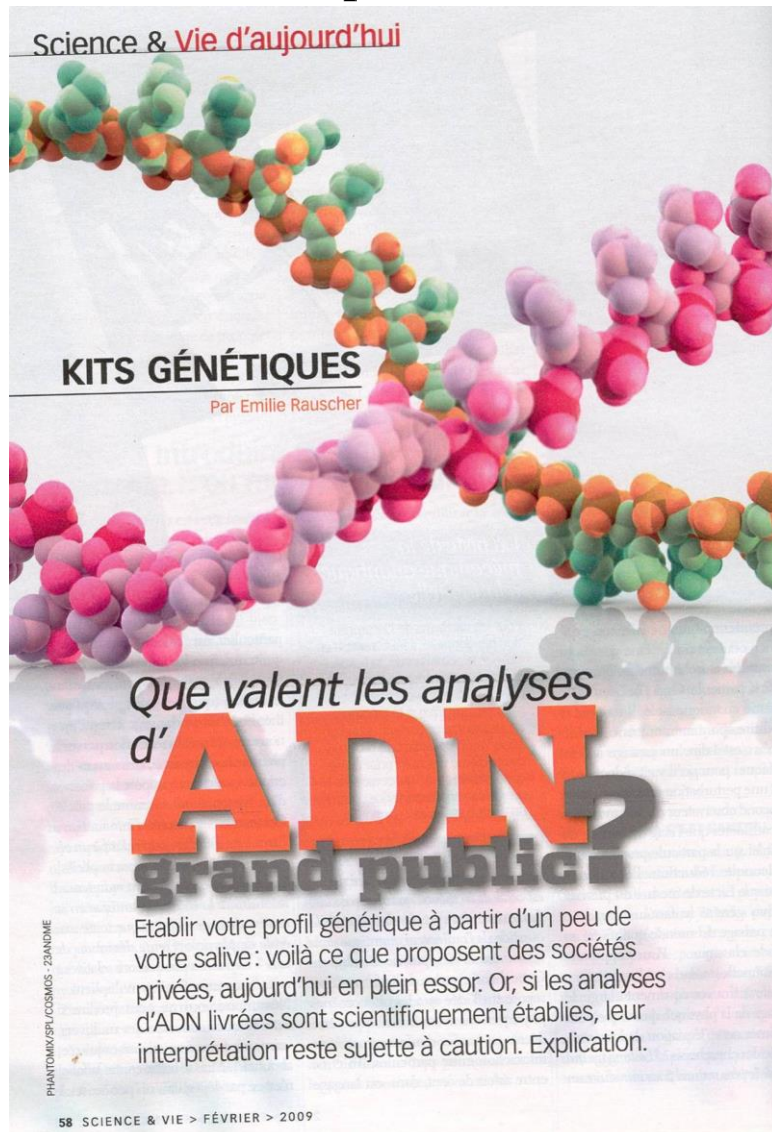
-Il a été postulé que ces polymorphismes pouvaient être en partie à la base de **la susceptibilité individuelle** à certaines maladies multifactorielles (ex diabète, dépression....) ou au traitements.



**-Mais les résultats sont encore peu clairs (pas de prise en compte de l'environnement!)**



*mise au point de kits génétiques pour le grand public*  
(« médecine » prédictive)



> FAITS & CHIFFRES

Les **tests génétiques** utilisés en laboratoire, pour le **dépistage** de maladies génétiques par exemple, sont apparus il y a une vingtaine d'années; il en existe aujourd'hui plus de **1 400**, coûtant de **400 à 2 400 €**. Il y a quelques années, lire un seul **SNP** coûtait **1 \$**, c'est moins de **1 centime** aujourd'hui. 23AndMe, société leader du secteur, demande **315 €** pour analyser **600 000 SNPs** du **génom**e d'un client, deCODEme **778 €** pour **1 million de SNPs**, Navigenics est plus cher (**1 980 €**) mais fournit un conseil génétique.

# Health Learn for the present.

## Take a more active role in managing your health

Knowing how your genes may impact your health can help you plan for the future and personalize your healthcare with your doctor.



**Find out if your child will be at risk for 49 inherited conditions and learn about steps you can take.**



Knowing your health risks will help you and your doctor figure out health areas to keep an eye on about disease risks

about carrying

**Exemple: <https://www.23andme.com/>**

23andMe estimates your genetic chances of getting  
**Type 2 Diabetes**



AS LOW AS

**8 %**



AS HIGH AS

**52 %**



23andMe will tell you:  
Your genetic risk  
What you can do

\*This result based on our calculations for males with European ancestry



Personalize your healthcare by knowing in advance how you will respond to certain medications like Warfarin.

about drug response

# Why should you know your genetic risk?



## Make better lifestyle choices.

On average, **one person in five** develops diabetes by age 79. Variations in your DNA tested for by 23andMe might raise your risk to **one in three**, making your lifestyle choices on factors like exercise and weight control even more critical.



## Be on the lookout for common conditions.

Age-related macular degeneration is the most common cause of irreversible vision loss in the Western world among people over 60. By learning if your genes put you at increased risk for this condition, you can choose to go in for more frequent eye exams and take other measures to protect your sight.



## Prepare for serious diseases.

Some inherited mutations greatly increase your likelihood of developing certain diseases later in life. For example, one rare mutation tested by 23andMe is associated with a nearly 60% lifetime risk of Parkinson's disease.

»See our health reports on 247 diseases and conditions.



23andMe isn't just about you. Our research arm, 23andWe, gives customers the opportunity to leverage their data by contributing it to studies of genetics. With enough data, we believe 23andWe can produce revolutionary findings that will benefit us all.

## Get involved in a new way of doing research.

Direct research by participating in studies of conditions and traits you care about.

Join an effort to translate basic research into improved health care for everyone.

Support 23andMe's efforts to discover new genetic associations that could shed more light on your data.

## Participate in research while exploring your own genetics.

Take surveys that collect important data for scientific research.

Learn new things about yourself—and what your genes may have to do with them.

Find out which traits make you stand out from the crowd.

## Join our research communities.

**Parkinson's Disease:** Recent discoveries suggest that genetics plays a greater role in Parkinson's disease than was previously thought. Help us understand how both genes and environment affect risk for PD by joining over 5,000 people that are already part of this movement.

**Roots into the Future:** Learn about your **roots** through genetic ancestry testing while contributing to research that will make a difference for the African American community in the **future**. Help us make discoveries about how DNA contributes to the development of disease in African Americans.

**Sarcoma:** There are clues that factors in our DNA contribute to the development of sarcomas of bone and soft tissues, but we still don't understand them very well. Join this study to help us predict how likely someone is to develop sarcoma or how well a particular treatment will work for a specific patient.

**Myeloproliferative Neoplasms (MPN):** Some genetic associations with MPNs have been discovered, but a great deal about this group of rare blood and bone marrow cancers still remains unknown. Join the 23andMe MPN Research Initiative to help accelerate the search for answers.

