

UNIVERSITE SIDI MOHAMMED BEN ABDELLAH  
FACULTE DE MEDECINE ET DE PHARMACIE  
FES



Année 2015

Thèse N°067/15

# LE SÉQUENÇAGE D'ACIDE DÉSOXYRIBONUCLÉIQUE: PRINCIPE TECHNIQUE, INDICATIONS MÉDICALES ET EXPÉRIENCE DU CHU HASSANE II DE FÉS

THESE

PRESENTEE ET SOUTENUE PUBLIQUEMENT LE 13/07/2015

PAR

Mr. AHAKOUD MOHAMED

Né le 06 SEPTEMBRE 1985 à BOULMANE

POUR L'OBTENTION DU DOCTORAT EN MEDECINE

MOTS-CLES :

ADN – ARN – SEQUENCAGE – PCR – ELECTROPHORESE – NUCLEOTIDE

JURY

M. SQALLI HOUSSAINI TARIK ..... PRESIDENT  
Professeur Agrégé de Néphrologie  
M. OULDIM KARIM ..... RAPPORTEUR  
Professeur agrégé de Génétique  
M. EL AZAMI EL IDRISI MOHAMMED..... } JUGES  
Professeur Agrégé d'Immunologie  
Mme. BENNIS SANAE ..... }  
Professeuse agrégée de Biologie Cellulaire

# Plan

Liste des Abréviations .....	6
Introduction .....	8
<b>HISTORIQUE</b> .....	10
I. L'expérience de Griffith : .....	11
II. La difficulté à accepter l'ADN comme support de l'hérédité: .....	14
III. La structure de l'ADN : .....	18
1- L'élaboration du modèle par Watson et Crick : .....	18
2- La molécule sémantique : .....	20
IV. Séquençage d'ADN : .....	21
<b>BASSE FONDAMENTALES</b> .....	26
I. Définition : .....	27
II. Les nucléotides : .....	28
1- Les bases azotées .....	28
1-1. Les bases pyrimidiques et les bases puriques .....	28
1-2. Des dérivés : molécules d'intérêt biologique .....	29
1-3. Des analogues synthétiques .....	30
1-4. Des propriétés importantes des bases azotées .....	30
2- Les nucléosides .....	33
2-1. Le pentose des nucléosides .....	33
2-2. La liaison osidique .....	33
2-3. Nomenclature .....	34
2-4. Les nucléosides naturels .....	35
2-5. Des nucléosides naturels et singuliers .....	35
2-6. Des analogues synthétiques .....	35
3- Les nucléotides .....	36

3-1.	Nomenclature .....	36
3-2.	Les nucléotides d'intérêt biologique .....	37
III.	les acides nucléiques : .....	39
1-	La structure primaire des polymères .....	39
1-1.	La taille des polymères nucléiques .....	39
1-2.	La liaison phosphodiester : .....	40
2-	Les différents groupes ionisables des acides nucléiques .....	40
2-1.	Les groupes phosphates .....	41
2-2.	Les azotes des cycles des bases pyrimidiques et puriques .....	41
2-3.	Les groupes amino liés aux cycles des bases .....	42
2-4.	Les formes tautomères (lactame et lactime) .....	42
2-5.	En guise de conclusion.....	42
3-	L'hydrolyse des acides nucléiques .....	44
3-1.	L'hydrolyse chimique .....	44
3-2.	L'hydrolyse enzymatique .....	44
4-	Structure spatiale des acides désoxyribonucléiques.....	46
4-1.	Quelques résultats expérimentaux sur les acides désoxyribonucléiques .....	46
4-2.	Les doubles hélices .....	49
	<b>SEQUENÇAGE 1<sup>ère</sup> GENERATION .....</b>	<b>53</b>
I.	Méthode de Maxam et Gilbert : .....	54
II.	Méthode de Sanger : .....	55
	<b>SEQUENÇAGE NOUVELLE GENERATION.....</b>	<b>59</b>
I.	Le séquençage de seconde génération : l'explosion de nouvelles technologies .....	60
1-	Les trois grands du séquençage : .....	62

1-1	454/Roche .....	62
1-2	Solexa/Illumina.....	64
1-3	SOLiD .....	67
2-	Les cadets de la seconde génération .....	68
2-1	Ion Torrent .....	68
2-2	Polonator G007.....	69
2-3	Complete Genomics.....	69
3-	La « miniaturisation » des séquenceurs : .....	71
II.	Séquençage de 3eme génération .....	72
1-	HeliScope .....	72
2-	Technologies des nanopores.....	72
3-	Pacific Biosciences .....	74
4-	Starlight .....	75
III.	La génomique personnelle : .....	75
IV.	Les applications du séquençage à haut débit.....	76
1-	Le séquençage de novo.....	76
2-	Le reséquençage.....	77
3-	Les analyses fonctionnelles.....	78
<b>MATERIELS ET METHODES</b>	.....	<b>80</b>
I.	Prélèvement : .....	81
II.	Techniques de préparation et de purification des acides nucléiques : .....	81
1-	Par sel : .....	81
1-1	Principe : .....	81
1-2	Réactifs nécessaires : .....	81
1-3	Protocole expérimental : .....	82
2-	Par Kit : .....	83

2-1	Principe : .....	83
2-2	Protocole expérimental : .....	83
3-	Dosage de l'ADN : .....	87
III.	Amplification des séquences nucléotidiques par PCR : .....	87
1-	Principe : .....	87
2-	Condition de PCR : .....	88
3-	Témoins de la réaction PCR : .....	88
4-	Contrôle des produits PCR : .....	89
IV.	Séquençages d'ADN : .....	89
1-	Purification par ExoSAP® : .....	89
2-	Reaction de sequence par BigDye® Terminator v 3. 1 Cycle Sequencing : .	90
3-	Purification de la réaction de séquence avec le Kit BigDye-XTerminator : .	91
4-	Détermination de la séquence : .....	92
V.	Les outils de bio-informatique : .....	93
1-	Logiciel BLAST (nucléotide BLAST) : .....	93
2-	Logiciel NCBI (CCDS database) : .....	93
3-	Logiciel NCBI (CCDS variation viewer) : .....	93
<b>RESULTATS</b>	.....	<b>94</b>
<b>DISCUSSION</b>	.....	<b>119</b>
I.	Champs d'application des NGS : .....	120
1-	Maladies génétiques mendéliennes .....	120
1-1	Tests préconceptionnels et dépistage des hétérozygotes .....	120
1-2	Diagnostic prénatal et préimplantatoire .....	122
1-3	Diagnostic des maladies mendéliennes .....	124
2-	Maladies dites multifactorielles .....	127
3-	Oncologie .....	128

3-1 Oncogénétique constitutionnelle .....	129
3-2 Hémopathies malignes.....	129
3-3 Tumeur malignes .....	132
II. Limitations actuelles des méthodes de séquençage à haut débit :.....	135
III. Au Maroc :.....	137
<b>RESUMES</b> .....	141
<b>Liste des figures</b> .....	145
<b>Liste des tableaux</b> .....	148
<b>Références</b> .....	149

## Liste des Abréviations

<b>A</b>	Adénine
<b>ADN</b>	Acide Désoxyribonucléique
<b>ADP</b>	Adénosine Diphosphate
<b>AMPc</b>	Adénosine Monophosphate Cyclique
<b>ARN</b>	Acide Ribonucléique
<b>ARNm</b>	Acide Ribonucléique Messenger
<b>ARNr</b>	Acide Ribonucléique Ribosomique
<b>ARNt</b>	Acide Ribonucléique De Transfert
<b>ATP</b>	Adénosine Triphosphate
<b>BDT</b>	BigDye Terminator
<b>C</b>	Cytosine
<b>CCD</b>	Charge Coupled Device
<b>CpG</b>	Segment d'ADN double brin riche en paires de bases CG
<b>Da</b>	Dalton
<b>dAMP</b>	Désoxyadénosine Monophosphate
<b>dATP</b>	Désoxyadénosine Triphosphate
<b>dCTP</b>	Désoxycytidine Triphosphate
<b>ddATP</b>	Didésoxyadénosine Triphosphate
<b>ddNTP</b>	Didésoxyribonucléotide Triphosphate
<b>DFNA (gène)</b>	Gène Deafness A
<b>DFNB (gène)</b>	Gène Deafness B
<b>DFNX</b>	Gène Deafness X
<b>dGTP</b>	Désoxyguanosine Triphosphate
<b>DNase</b>	Désoxyribonucléase
<b>DS</b>	Déviation Standard
<b>dTTP</b>	Désoxythymidine Triphosphate
<b>EDTA</b>	Éthylène Diamine Tétra-Acétique
<b>FAD</b>	Flavin Adénine Dinucleotide
<b>FRET</b>	Förster Resonance Energy Transfer
<b>G</b>	Guanine
<b>Gb</b>	Giga-base
<b>GMPc</b>	Guanosine Monophosphate Cyclique
<b>HTS</b>	High Throughput Sequencing
<b>HUGO</b>	Human Genome Organisation
<b>IDH1, IDH2</b>	Gène Isocitrate Dehydrogenase 1 Et 2
<b>IPA</b>	Isopropanol
<b>Mb</b>	Méga-base

---

<b>NAD</b>	Nicotinamide Adenine Dinucleotide
<b>NGS</b>	Next-Generation Sequencing
<b>NIH</b>	National Institutes Of Health
<b>PAPS</b>	Adénosine 3'-(mono)Phosphate 5'-PhosphoSulfate
<b>Pb</b>	Paire De Base
<b>PCR</b>	Polymerase Chain Reaction
<b>Pur</b>	Purine
<b>Pyr</b>	Pyrimidine
<b>RNAse</b>	Ribonucléase
<b>SLB</b>	Solution de Lyse de globules Blancs
<b>T</b>	Thymine
<b>TE</b>	Tris-EDTA
<b>Tm</b>	Temperature of Melting
<b>UGMO</b>	Unité de Génétique Médicale et Oncogénétique
<b>USA</b>	United States of America
<b>USD</b>	United States Dollar

## Introduction

Depuis la description de la structure de l'ADN en 1955 jusqu'à nos jours, la biologie a connu une technologies dont le séquençage constitue l'un des évènements clés. En ce début de troisième millénaire déjà riche en nouvelles technologies, nous assistons à une nouvelle révolution dans le domaine du séquençage [1].

Allan Maxam, Walter Gilbert (USA) et Frederick Sanger (Royaume-Uni), les pionniers de la conquête de la séquence, ont mis au point deux méthodes très différentes permettant d'accéder à la lecture de la séquence. Maxam et Gilbert ont exploité des stratégies de dégradation chimique sélective quand Sanger choisit une stratégie de synthèse enzymatique sélective. L'histoire ne retient effectivement que Sanger, à juste titre, puisque cette stratégie bénéficiant de l'invention de la PCR et du développement de l'électrophorèse capillaire, permettant de simplifier la partie séparative et analytique.

C'est en 2007 que la deuxième génération des appareils NGS a été développée. Les termes « deuxième » et « troisième génération » sont de séquençage direct de Sanger dite de « première génération » [2].

Le séquençage de l'ADN constitue une méthode dont le but est de déterminer la succession linéaire des bases A, C, G et T prenant part à la structure de l'ADN. La lecture de cette séquence permet d'étudier l'information biologique contenue par celle-ci. Étant donné l'unicité et la spécificité de la structure de l'ADN chez chaque individu, la séquence de l'ADN permet de nombreuses applications dans le domaine de la médecine, comme, par exemple, le diagnostic, les études génétiques, l'étude de paternité, la criminologie, la compréhension de mécanismes

physiopathologiques, la synthèse de médicaments, les enquêtes épidémiologiques [1].

Dans de nombreuses publications, le terme séquençage peut se retrouver sous deux dénominations différentes qu'il est important de connaître. Dans les études de génomes, le terme de reséquençage (expression pouvant prêter à confusion) est utilisé à la place de séquençage. Cette dénomination, essentiellement utilisée en génétique, désigne le séquençage d'un segment d'ADN suivi de la comparaison du résultat obtenu avec celui d'une séquence de référence connue. Un autre terme est également employé : le séquençage de novo. Dans ce cas, il s'agit du séquençage d'un génome pour lequel il n'existe pas de séquence référence. Il s'agit donc de la détermination d'une séquence inconnue [1].

A travers cette thèse, nous avons abordé la technique du séquençage d'ADN sur tous les plans, théorique et pratique :

- ✓ Théorique : en abordant les différentes méthodes de la technique, classique ou moderne, avec la mise au point de l'expérience de notre service de génétique médicale et oncogénétique dans ce domaine.
- ✓ Pratique : un CD-ROM intitulé « séquençage d'ADN », destiner aussi bien aux étudiant en médecine qu'aux résidents en génétique médicale et doctorants en science, qui aborde de manière simplifié et pratique les différents étapes de la technique de séquençage de l'ADN et aussi donne une petite aperçu sur les nouvelles générations de séquenceurs.

# HISTORIQUE

## I. L'expérience de Griffith :

Le premier phénomène qui allait permettre de progresser dans l'identification du support de l'hérédité est celui de la transformation bactérienne, rapporté en 1928 par l'anglais Fred Griffith (1877 – 1941). Celui-ci travaille alors au laboratoire de pathologie du ministère de la santé du Royaume Uni.



Un médecin et bactériologue anglais, devenu célèbre par l'expérience scientifique qui porte son nom. Griffith mis en évidence l'absorption d'un facteur génétique par une souche de pneumocoques (1928).

Figure 1 :

Fred Griffith (1877 – 1941)

Griffith décrit deux souches de pneumocoques *Diplococcus pneumoniae* : la souche R et la souche S. La souche S doit son aspect à une capsule polysaccharidique qu'elle synthétise autour d'elle. Cette souche est mortelle pour la souris lorsqu'elle lui est injectée. A l'inverse, la souche R ne synthétise pas une telle capsule, et elle n'est pas nocive lorsqu'elle est injectée à une souris. On sait aujourd'hui que cette différence entre les deux souches est due à une mutation, chez la bactérie R, du gène codant l'enzyme responsable de la synthèse de la capsule.

Griffith observe tout d'abord que l'injection de bactéries S, si elles ont été préalablement tuées par la chaleur, n'est plus létale pour la souris. Pour une raison qui nous est toujours inconnue, Griffith décide alors d'injecter conjointement des bactéries S chauffées mélangées à des bactéries R vivantes. Cette fois, les souris meurent de septicémie. Les bactéries R, au contact des bactéries S tuées, ont donc acquis un caractère pathogène qu'elles ne possédaient pas précédemment. Ce phénomène a été appelé transformation bactérienne, et il a été par la suite reproduit chez plusieurs autres espèces bactériennes.

Il existe en fait plusieurs souches de pneumocoques (types I, II, ou III), discernables grâce à des tests immunologiques. Lorsque qu'une souche R de type III est injectée avec une souche S de type II inactivée, la bactérie virulente qui est ré-isolée de la souris tuée est toujours du type II. Ce changement est stable et définitif.

Ceci suggère donc qu'il existe chez les cellules un "facteur transformant", probablement libéré par la chaleur, susceptible d'être intégré par d'autres bactéries, et qui leur confère de façon héréditaire de nouvelles propriétés génétiques.

Ce phénomène représentait un test d'activité biologique, grâce auquel on pouvait envisager de déterminer la nature du matériel génétique responsable de la transformation de bactérie du type R en bactérie de type S. Griffith ne sut pas en tirer lui-même avantage, et la nature du "facteur transformant" sera élucidée plus de 10 ans plus tard par Avery et ses collègues.

La nature biochimique du matériel génétique mis en évidence par Griffith est élucidée en 1944 par les travaux d'Oswald Avery, et de ses collègues (Colin McLeod, et McLyn McCarthy), qu'ils mènent à l'Institut Rockefeller de New-York. Ils reprennent

les expériences de Griffith sur la transformation bactérienne, et cherchent à purifier le facteur transformant du pneumocoque.

Cette caractérisation prendra 10 ans, elle les conduit à montrer que ce facteur n'est autre que l'ADN : en effet, l'ADN extrait d'une souche S suffit à lui seul pour transformer une souche non virulente en souche virulente. La transformation des bactéries R s'effectue par incorporation de fragments d'ADN provenant des bactéries



Un médecin américain d'origine canadienne. Une expérience sur les pneumocoques qu'il fit en 1944 avec ses collaborateurs Colin MacLeod et Maclyn McCarty fut décisive pour établir le rôle de l'acide désoxyribonucléique comme support de l'hérédité.

Figure 2 :

Oswald T. Avery (1877–1955)

S tuées. On sait aujourd'hui que de tels fragments sont capables de rentrer dans une bactérie vivante, et de s'intégrer au chromosome de celle-ci en lieu et place de la région homologue.

L'identification de l'ADN comme principe transformant est à l'époque suffisamment extraordinaire pour nécessiter qu'une telle découverte soit étayée par des arguments indiscutables. Aussi Avery et ses collègues effectuent-ils leurs analyses avec un soin particulièrement méticuleux. Tous les contrôles alors disponibles sont testés : l'absence de protéine dans les préparations est testée par

divers réactifs chimiques, leur composition chimique est analysée par des moyens chimiques ou spectrophotométrique.

Enfin, l'utilisation d'enzymes montre que le pouvoir transformant réside bien dans l'ADN, puisque la DNase anéantit ce pouvoir, alors que la RNase ou des protéinases le laisse intact. [3]



Figure 3 :

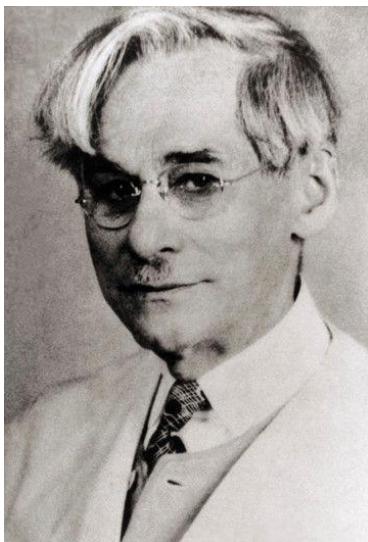
Maclyn McCarty avec Francis Crick et James D. Watson

## **II. La difficulté à accepter l'ADN comme support de l'hérédité:**

Malgré une accumulation croissante de preuves jusqu'au début des années 50, la communauté scientifique n'a pas accepté facilement que l'ADN puisse être le support de l'hérédité. Selon les thèses alors les plus largement acceptées, l'ADN n'est qu'une molécule simple, et donc incapable de véhiculer une information complexe. La théorie tétra-nucléotidique proposée par Phoebus Aaron Levene (1869 – 1940), stipulait que la structure de l'ADN est régulière et monotone, comprenant un enchaînement répétitif des 4 bases azotées.

Levene était alors un des plus grands spécialistes des acides nucléiques, c'est lui qui avait identifié le désoxyribose comme un des constituants de l'ADN. Les protéines, dont on avait perçu l'immense diversité, semblaient de bien meilleurs candidats pour véhiculer une information génétique.

La découverte d'Avery a été accueillie avec beaucoup de scepticisme, ils furent interprétés par beaucoup comme le résultat d'une contamination des préparations d'ADN par de faible quantité de protéines ou d'une autre substance. Et même si ces préparations d'ADN étaient absolument pures, il était possible d'envisager que cette molécule ne soit pas elle-même porteuse d'information, mais qu'elle joue



PhoebusLevene est né à Sagor en Russie. Il a grandi à St-Petersbourg et a étudié la médecine à l'Académie Impériale Militaire de Médecine.

Figure 4 :

Phoebus Aaron Levene (1869 – 1940)  
simplement un rôle de commutateur : toutes les informations nécessaire à la transition de la forme R vers la forme S auraient déjà été présentes dans chacune des souches, et l'ADN n'aurait alors contribué qu'à la transition d'un type vers l'autre.

Avery lui-même n'a pas véritablement cherché à imposer ses conclusions. Ainsi, l'importance fondamentale de ces travaux ne sera reconnue que tardivement, et le comité Nobel ne le retiendra pas pour l'attribution d'un prix. Lorsqu'en 1953 Watson et Crick décriront la structure de l'ADN, ils ne prendront même pas la peine de citer ces travaux dans leur article.

Certains scientifiques ont cependant saisis immédiatement la portée immense des travaux d'Avery. Ce fut en particulier le cas d'E. Chargaff, J. Lederberg, G. Beadle, ou de A. Lwoff.

Chargaff est un biochimiste d'origine autrichienne ayant émigré aux USA en 1934. Il fait partie des scientifiques qui saisissent immédiatement la portée immense des travaux d'Avery sur la transformation bactérienne. Lorsque Avery publie ses résultats, Chargaff dirige un laboratoire de biochimie à l'université de Columbia. Le thème central en est alors la biochimie des lipoprotéines. Quand il prend connaissance des travaux d'Avery, Chargaff comprend aussitôt que l'ADN occupe une place centrale dans les mécanismes héréditaires, et il décide de consacrer désormais les activités de son laboratoire à l'étude des acides nucléiques.

En 1950, il publie ses travaux sur le contenu en bases azotées de l'ADN chez diverses espèces, réalisés grâce aux progrès de la chromatographie sur papier. Il montre alors que le rapport  $A+T/C+G$  est variable selon les espèces, mais constant pour tous les membres d'une espèce donnée. L'ADN est donc porteur d'une certaine spécificité, cette molécule n'a pas une structure polymérique monotone, elle est donc susceptible de contenir une information. Ces travaux contribuent à répandre l'idée que l'ADN puisse être une molécule porteuse de l'information génétique.

Chargaff montre par ailleurs que le rapport C/G ou A/T est à l'inverse constant et quasiment égal à un chez toutes les espèces étudiées. Cette dernière observation sera déterminante pour l'élaboration du modèle de la structure de l'ADN par Watson et Crick quelques années plus tard.

A la suite du succès remporté par ce modèle, Chargaff est très réticent à en reconnaître l'entière paternité à ses auteurs, pour lesquels il n'a jamais eu grande estime (à la suite de sa première rencontre avec Watson et Crick, il les avait comparés à deux clowns). Il continuera à revendiquer pour son propre compte le modèle d'appariement des bases, alors qu'il n'y avait lui-même jamais pensé, et n'ayant mis en évidence que les rapports égaux d'adénine et de thymine, de cytosine et de guanine. [4]

Al Hershey est, avec Delbrück et Luria, la troisième personnalité marquante du Groupe du phage. Sa première contribution importante est la mise en évidence d'une génétique des bactériophages : Il montre en effet en 1946 que peuvent apparaître des mutations chez ces organismes, caractérisés par différentes formes de plages de lyse. Ces résultats confirment donc ceux obtenus par Luria l'année précédente.

Hershey parvient aussi à détecter des phénomènes de recombinaison chez les phages : en infectant simultanément une souche bactérienne avec deux phages portant des mutations distinctes, il observe l'apparition de phages recombinants, portant soit simultanément les deux mutations, soit aucune d'entre elles. Des observations similaires seront rapportées par Delbrück.

Mais l'expérience la plus célèbre de Hershey concerne l'étude du rôle que jouent chacun des constituants du phage (ADN et protéines) dans la transmission de l'information génétique. Pour cela, il utilise – avec Martha Chase – le bactériophage

T2. Leur expérience est une des premières expériences de biologie moléculaire où des isotopes radioactifs permettent de tracer des molécules. Ils utilisent un marquage isotopique différentiel de chacun des constituants du phage : du phosphore radioactif, incorporé à l'ADN, et du soufre radioactif, incorporé aux protéines de la capsid. Ces phages sont utilisés pour infecter des bactéries. Immédiatement après l'infection, il est possible de séparer par agitation mécanique les bactéries des phages qui les ont infectés, et de séparer par centrifugation les particules phagiques des bactéries infectées. Lorsqu'elles sont remises en culture, ces bactéries produisent des phages. Or l'analyse de la répartition de la radioactivité montre que ces bactéries ne contiennent que du phosphore radioactif : seul l'ADN a donc pénétré dans la cellule, et cette fraction est responsable à elle seule de la reproduction du phage. C'est donc l'ADN qui détient l'information génétique.

Bien que cette expérience fût dans sa conception plus grossière que celle d'Avery, son impact fut au moins aussi important. Ceci était dû en partie au fait qu'elle avait été effectuée par des membres du Groupe du phage, qui représentait alors les scientifiques les plus influents dans le domaine de la biologie moléculaire. D'autre part, ces résultats arrivaient pratiquement simultanément avec l'élucidation de la structure de l'ADN par Watson et Crick. [5][6][7]

### **III. La structure de l'ADN :**

#### **1 – L'élaboration du modèle par Watson et Crick :**

La structure en double hélice de l'ADN est élucidée par Watson et Crick en 1953. Watson a décrit dans son livre passionnant *La double hélice* (1968) le récit de la formidable découverte réalisée avec Crick. Les deux chercheurs disposent alors des éléments suivants :

(i) la composition chimique de l'ADN (désoxyribose, bases azotées, et groupements phosphate).

(ii) les clichés de diffraction aux rayons X d'ADN cristallisé, clichés dus principalement à Rosalind Franklin et Maurice Wilkins du King's College. Ces clichés montrent une figure en croix, caractéristique des structures en hélice.

(iii) les travaux de Erwin Chargaff, qui avaient montré que pour toute molécule d'ADN, le nombre de molécules d'adénine est égal au nombre de molécules de thymine, et que celui de cytosine est égal à celui de guanine.

(iv) les analyses en microscopie électronique, qui avaient montré que le diamètre de la molécule d'ADN est de 20 Å, ce qui suggérait que cette molécule comportait deux chaînes de désoxyribose-phosphate.

C'est en élaborant successivement plusieurs modèles moléculaires que Watson et Crick réussissent à proposer une structure qui satisfasse à l'ensemble des données cristallographiques et biochimiques alors disponibles. Cette structure est aujourd'hui connue de tous, elle est devenue l'emblème de la biologie moléculaire : deux brins constitués des groupements phosphates et des sucres forment une double hélice où les orientations de chacun des brins sont opposées. Sur les sucres de chacun des deux brins sont liées les bases azotées, chaque base d'un brin étant maintenue en vis-à-vis d'une base de l'autre brin par des liaisons hydrogène. Une cytosine fait toujours face à une guanine, et une adénine à une thymine. Les deux brins d'une molécule d'ADN sont dits complémentaires.

Crick, Watson, et Wilkins reçurent en 1962 le prix Nobel pour ces travaux, qui a été qualifiée par Peter Medawar de "la plus grande réussite scientifique de notre siècle". Rosalind Franklin aurait vraisemblablement été associée à ce prix si la maladie ne l'avait prématurément emportée.



Figure 5 :

Francis Crick, James D. Watson,  
Rosalind Franklin

## 2- La molécule sémantique :

L'ordre des bases le long de la molécule définit l'information génétique portée par l'ADN. Les possibilités offertes par l'agencement des quatre bases de l'ADN sont immenses ( $4^n$  pour une séquence de  $n$  bases, soit par exemple plus d'un milliard de combinaisons pour une séquence de 10 bases).

L'ordre des bases nucléotidiques détermine l'enchaînement des acides aminés des protéines (par l'intermédiaire des processus de transcription et de traduction).

Une structure à une dimension détermine donc des structures à 3 dimensions. Ce paradoxal "gain d'information" est réalisé spontanément, par le repliement des protéines au fur et à mesure de leur synthèse, en fonction des contraintes thermodynamiques imposées par leur séquence et par le milieu dans lequel elles sont synthétisées.

Les êtres vivants disposent donc d'un système informatif à une dimension, entièrement compris dans une séquence linéaire, qui peut être facilement copié ou dupliqué. De cette séquence linéaire découle celle des protéines, qui adoptent spontanément la structure à trois dimensions nécessaire à leur fonction.

[8][9][10][11][12]

#### **IV. Séquençage d'ADN :**

Un bref historique des évolutions du séquençage de l'ADN permet de comprendre ce que certains nomment révolution technologique... certainement concernant l'univers de la biologie moléculaire une innovation comparable à celle la PCR.

Allan Maxam, Walter Gilbert (USA) et Frederick Sanger (Royaume-Uni), les pionniers de la conquête de la séquence, ont mis au point deux méthodes très différentes permettant d'accéder à la lecture de la séquence. Maxam et Gilbert ont exploité des stratégies de dégradation chimique sélective quand Sanger choisit une stratégie de synthèse enzymatique sélective. Pour ces découvertes, Gilbert et Sanger ont été récompensés par le prix Nobel de chimie en 1980. L'histoire ne retient effectivement que Sanger, à juste titre, puisque cette stratégie bénéficiant de l'invention de la PCR (invention de KaryMullis publiée en 1986, lui aussi nobélisé par le prix de chimie en 1993) et du développement de l'électrophorèse capillaire, permettant de simplifier la partie séparative et analytique. C'est ainsi que des

sociétés comme Perkin Elmer, Beckman Coulter et AppliedBiosystem ont investi le marché sur le principe même du séquençage Sanger sans le révolutionner mais en permettant son automatisation et l'augmentation du nombre de réactions de séquences analysables par jour. Une fois amorti le coût des séquenceurs, le coût de revient d'une séquence n'a cessé de baisser permettant de banaliser l'accès à la séquence.

Bien que se montrant de plus en plus performant, des applications comme le séquençage complet de génomes eucaryotes supérieurs, les approches métagénomiques (nécessité de cloner), les études de modulation de transcrits (la méthode SAGE les permettait sur la base du séquençage Sanger, cette méthode est lourde et nécessite toujours une phase de clonage) ont connu des limites quasi infranchissables (méthodes nécessitant trop de temps et de capitaux). Prenons à titre d'exemple les projets de séquençage du génome humain. Ces projets auront nécessité plus de 10 ans de travail et de 300 millions de USD pour le projet privé de Celera et un peu moins de 3 milliards de USD pour le projet HUGO (HUMANGenome Organisation). Le match entre Craig Venter (fondateur de Celera avec le soutien de Perkin Elmer) et James Watson (premier directeur du NIH coordinateur du consortium international) se solde par un match nul entaché de polémiques avec les publications en 2001 des premières séquences du génome humain.

L'année 2005 a connu à grands bruits, l'arrivée de nouvelles méthodes de séquençage à haut débit. Ces méthodes de séquençage massif font appel aux techniques de clonage et d'amplification moléculaire, leur spécificité relevait de leur stratégie de lecture.

En effet, la société 454 (rachetée par Roche en 2007), utilise des méthodes de pyroséquençage (luminescence par libération de pyrophosphate) de fragments d'ADN

isolés dans des micro-gouttes comme micro-réacteur de PCR isolés au sein d'une émulsion, la société Solexa (rachetée par Illumina en 2007), utilise des méthodes d'amplification sur support solide permettant l'incorporation de bases terminateurs de chaîne réversibles marqués par des fluorochromes. La société Agencourt (rachetée par AppliedBiosystem en 2006), a quant à elle basé son système de détection sur le principe de l'amplification par émulsion et hybridation-ligation chimique.

Depuis 2007, une certaine frénésie s'empare du monde de la biologie moléculaire, un très grand nombre de machines que l'on qualifie tour à tour de NGS (NextGenerationSequencing), HTS (High ThroughputSequencing), ou encore de manière plus appropriée séquençage multi-parallélisé se trouvent disponibles et évoluent vers le plus de profondeur, ou vers le moindre coût, rendant obsolètes les versions précédentes.

Cette révolution technologique est sortie des laboratoires pour attirer l'attention d'investisseurs et autres banquiers conscients qu'une ruée vers l'or accompagnant la ruée vers la séquence pouvait s'engager. Dans ce sens nous mettons à disposition un document datant de 2007 réalisé pour le compte de la Deutsche Bank (ce document très complet a été co-réalisé par nombre de consultants en biotechnologies, il permet de décortiquer assez finement les 3 technologies présentes sur le marché à partir de 2005 et de montrer les attentes des mondes scientifique et financier)

Afin de terminer ce bref tour d'horizon il semble nécessaire de hiérarchiser ces nouvelles technologies. Il est possible de distinguer tout d'abord deux grands groupes : les technologies sans amplification (permettant de séquencer une seule molécule d'ADN matriciel, citons exemple de la technologie SMRT de Pacific

Biosciences, la technologie tSMS de Helicos Biosciences ou encore celle toujours en développement mais très attendue développée par Oxford Nanopore) et avec amplification (toutes les technologies amplifiant une matrice clonale d'adn issu d'une librairie)

Parmi cette seconde classe, il est possible de distinguer trois grandes classes de matériel :

-les usines à produire de la séquence (type Hiseq 2000 de Illumina permettant de séquencer jusqu'à 200 Gb par run de 8 jours.

-les séquenceurs de pailleasse (type PGM de IonTorrent de taille et prix beaucoup plus modeste permettant de réaliser des run autour de 100 Mb avec une perspective de 1 Gb pour début 2012).

-les séquenceurs haut débit en voie d'obsolescence (454 de Roche, les versions des séquenceurs Illumina antérieures au HiSeq 1000 etc.)

Cet historique est bien entendu un prétexte pour planter le décor et présenter les forces en présence. Après l'âge des pionniers (Maxam et Sanger), celui de la douce exploitation (séquenceur capillaire) vient celui de l'accélération que certains nomment révolution... ces technologies marquent quoi qu'il en soit un tournant pour la génomique et pour les applications actuelles et en devenir qui y sont liées. Enfin la diffusion de ces technologies va de paire avec un changement de physionomie des laboratoires : la biotechnologie est une science composite où bio-informatique, biologie cellulaire et moléculaire mais aussi physique et statistique deviennent de plus en plus interdépendantes.

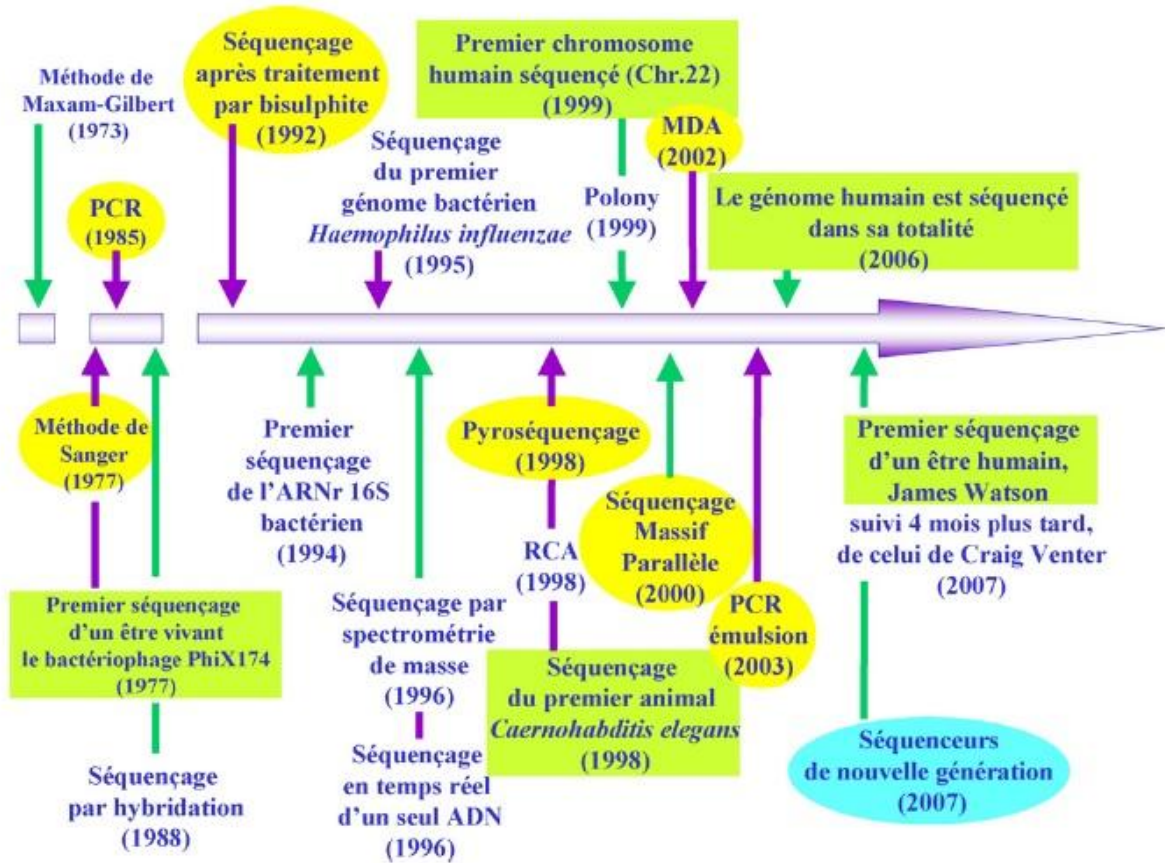


Figure 6 :

Quelques étapes importantes démontrant l'évolution des progrès du séquençage.

# **BASSE FONDAMENTALES**

## **I. Définition :**

Le médecin suisse MIESCHER isola en 1869, à partir de leucocytes de pus de plaies infectées, une substance organique remarquablement riche en phosphore qu'il baptisa nucléine. Sa nature chimique fut élucidée par les travaux initiateurs de KOSSEL à partir de 1882 jusqu'à ceux de LEVENE en 1927 : il s'agissait de l'acide désoxyribonucléique dont l'abréviation est ADN (DNA : anglo-saxon), polymère d'unités dénommées nucléotides.

Les polynucléotides biologiques sont :

- le support moléculaire de l'information génétique : l'ADN (et ARN pour certains virus) est le support de l'hérédité et du codage des composés biologiques (les ARN, les protéines)

- des effecteurs de l'expression de l'ADN en peptides et protéines : acide ribonucléique dont l'abréviation est ARN (RNA : anglo-saxon) regroupés en trois classes :

- les ARN messagers (ARNm)

- les ARN de transfert (ARNt)

- les ARN ribosomiaux (ARNr)

Les nucléotides ont des fonctions variées et importantes, ce sont :

- des composés à "haut potentiel énergétique" dont dépend l'activation de molécules entrant dans des réactions endergoniques dans les réactions cellulaires,

- des composés structuraux de coenzymes,

- des seconds messagers intracellulaires de signaux et des médiateurs extracellulaires,

- des régulateurs d'activité de protéines lors de modifications covalentes.

## II. Les nucléotides :

Un nucléotide résulte de :

1) la condensation d'un ose (pentose) avec une base nucléique (hétérocycle azoté) qu'on appelle nucléoside.

2) l'estérification de l'ose d'un nucléoside par l'acide phosphorique produit un nucléotide.

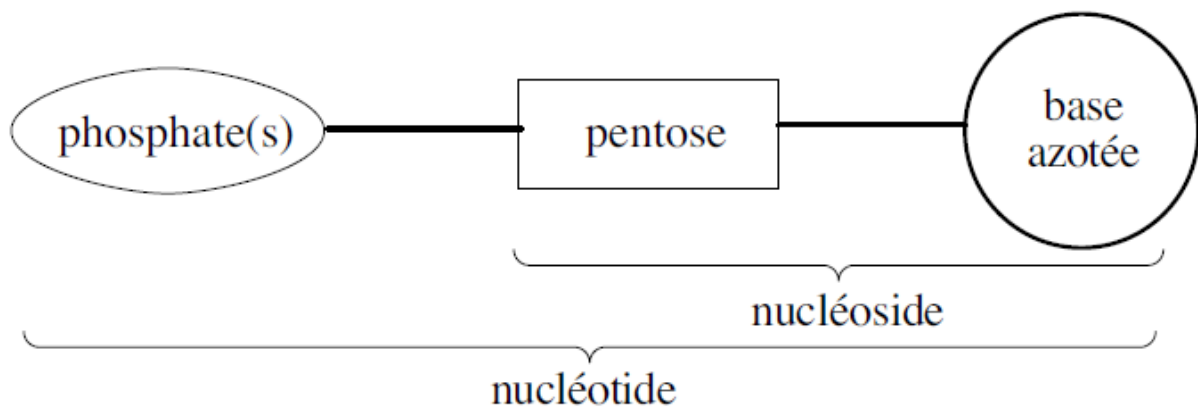


Figure 7:

Schémas de la composition d'un nucléotide.

### 1- Les bases azotées

Cinq bases majeures, partagées en deux séries, entrent dans la composition des nucléotides et leurs polymères. Elles vont conférer aux composés biologiques dont elles font partie des propriétés capitales.

#### 1-1. Les bases pyrimidiques et les bases puriques

Les dérivés oxy ou/et amino de la pyrimidine et de la purine forment les deux familles de base des nucléotides naturels.

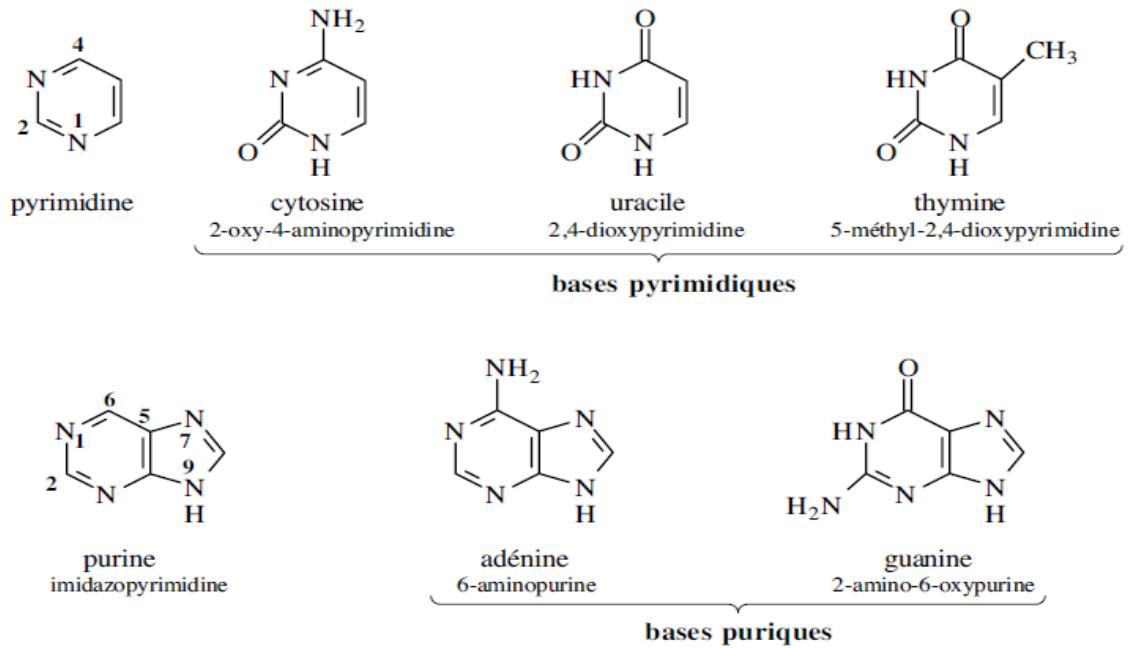


Figure 8:

Les différentes bases azotes.

Les noms courants des différentes bases n'ont aucun lien avec la nomenclature classique de la chimie organique, certains font référence à leurs conditions de découverte (thymine : thymus de veau).

Les nucléotides de l'ADN, comme ceux de l'ARN ne comportent que quatre de ces bases azotées :

- deux puriques communes aux deux types d'acides nucléiques
- une pyrimidique commune : la cytosine
- une pyrimidique spécifique : l'uracile pour l'ARN et son dérivé méthylé, la thymine pour l'ADN.

### 1-2. Des dérivés : molécules d'intérêt biologique

- Lorsqu'elles ne sont pas recyclées, les bases puriques sont dégradées en acide urique par passage par des formes désaminées hypoxanthine et xanthine. Celui-ci, très peu soluble, est excrété par les primates.

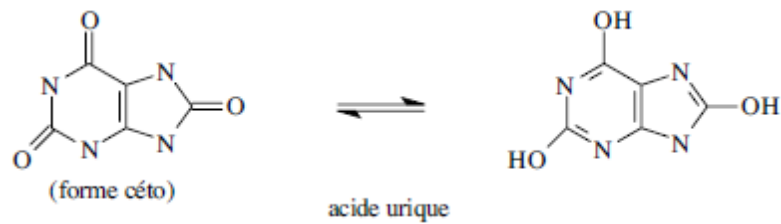


Figure 9 :

### Dégradation des bases azotées

#### 1-3. Des analogues synthétiques

Des analogues des bases nucléiques sont utilisés :

- comme molécules de marquage en biologie moléculaire : 5-bromouracile sur lequel peuvent être greffés des molécules marqueurs
- comme agents thérapeutiques, agissant en compétition avec les bases naturelles, ils bloquent la multiplication des bactéries, la mitose :
  - antitumoraux : ce sont les dérivés fluorés (5-fluorouracile), thiols (6-thiopurine) ou encore aza (N remplace un C : 8-azaguanine) qui sont utilisés
  - un antiviral classique (acyclovir) est un dérivé de la guanine : (2-hydroxyéthoxy)9-méthylguanine

#### 1-4. Des propriétés importantes des bases azotées

Leurs formules chimiques indiquent :

- les hétérocycles azotés sont susceptibles d'ionisation
- les doubles liaisons créent des systèmes conjugués pour lesquels certaines propriétés physiques sont remarquables (spectre, hydrophobicité, empilement).

La conjugaison des doubles liaisons

La résonance entre de nombreux atomes délocalise les électrons  $\pi$  des doubles liaisons avec les conséquences suivantes :

- la molécule est fortement stabilisée dans une configuration plane
- la molécule existe sous différentes formes tautomères.

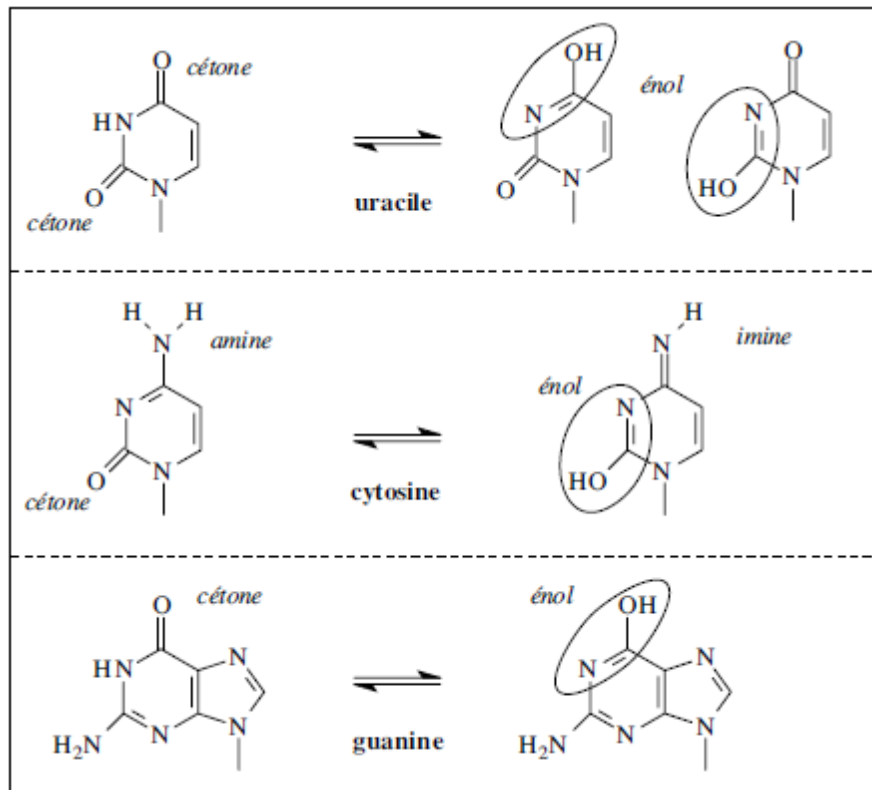


Figure 10 :

## Différentes forme des bases azotées

Equilibres tautomériques à pH 7 des différentes bases liées à un ose (nucléoside ou nucléotide) :

- forme lactame à gauche (cétone)
- forme lactime à droite (énol)

Les formes prépondérantes à pH7 sont les formes lactame (cétone) et amino.

Les propriétés spectrales

Les hétérocycles des différentes bases ainsi que leurs dérivés, nucléosides ou nucléotides, présentent des spectres d'absorption caractéristiques dans l'ultraviolet, spectres dépendant du pH. L'aire de ces spectres dans cette région est plus élevée pour les purines (à deux cycles) : leurs absorptions sont donc plus importantes. Ces

propriétés optiques sont communément utilisées pour la détection, le dosage et le contrôle de pureté d'acides nucléiques.

La fluorescence de ces bases est par contre inutilisable : l'émission se situe dans la région UV 300–320 nm et elle est très faible (400 fois plus faible que celle du tryptophane pour les purines et 2500 fois pour les pyrimidines).

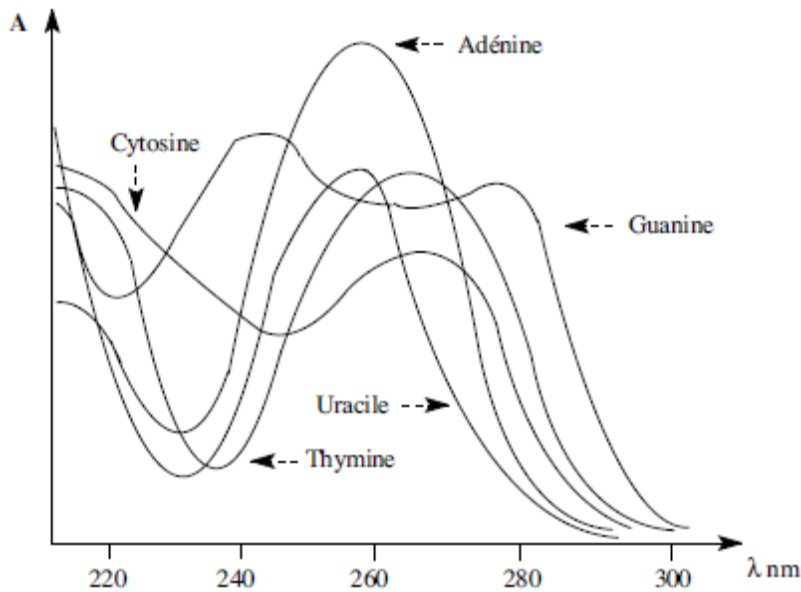


Figure 11 : Spectre d'absorption de différentes bases azotes

Les transformations chimiques des bases

1) la lente désamination est spontanée dans les cellules (100 fois moins importante pour les purines par rapport aux pyrimidines) :



2) les radiations altèrent les bases :

- l'irradiation dans l'ultraviolet ouvre les liaisons de deux bases superposées et les pontes par des liaisons covalentes
- les radiations ionisantes (rayons X ou gamma) ouvrent les cycles et les cassent.

3) de nombreux agents chimiques réagissent avec les bases :

- l'acide nitreux ( $\text{HNO}_2$ ) et l'hydrogénosulfite de sodium  $\text{HNO}_3$  ont une action désaminante. Ils font partie des conservateurs dans l'industrie alimentaire.

- les espèces réactives de l'oxygène (peroxydes, radicaux libres) font subir des dommages oxydatifs

## 2- Les nucléosides

Une liaison covalente (liaison N-osidique) fixe les bases à un pentose.

### 2-1. Le pentose des nucléosides

Il s'agit du ribose dans les acides ribonucléiques (ARN) et de son dérivé 2-désoxyribose dans les acides désoxyribonucléiques (ADN) avec les caractéristiques suivantes :

- l'énantiomère est de la série D
- l'ose est sous forme hémiacétalique (furanose)
- l'anomère est en conformation

### 2-2. La liaison osidique

La liaison avec la base est de type N-osidique entre le carbone 1' (carbone anomérique) du furanose en conformation  $\beta$  et l'azote N1 des pyrimidines et N9 des purines.

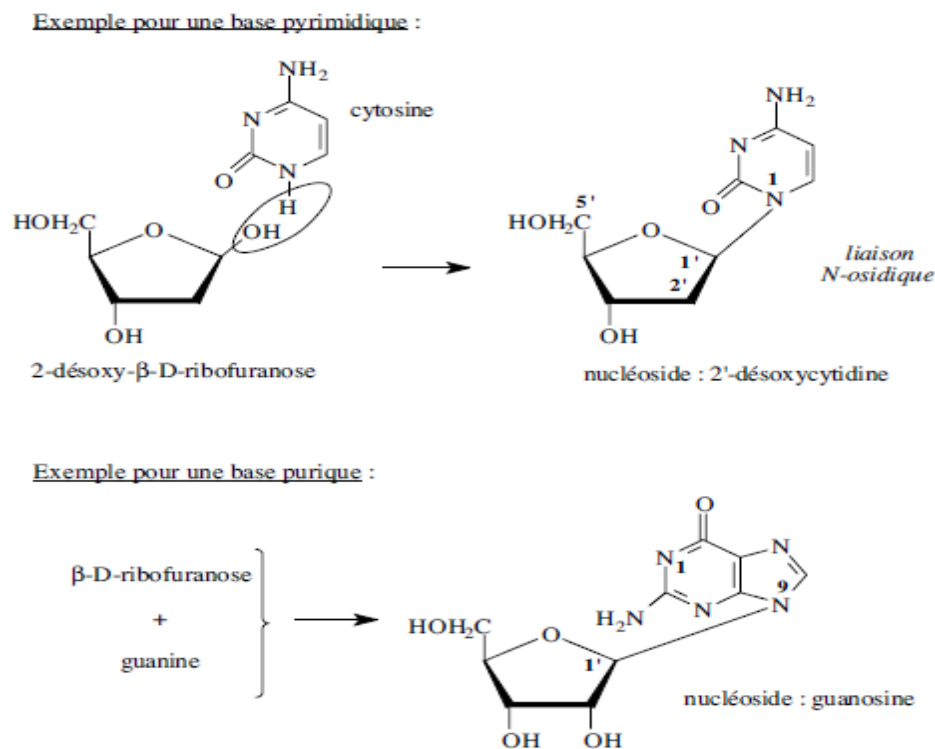


Figure 12 :

Liaison N-osidique

### 2-3. Nomenclature

Les noms des nucléosides ont comme suffixe :

- "osine" pour les nucléosides puriques
- "idine" pour les nucléosides pyrimidiques

Bases azotées		Ribonucléosides		Désoxyribonucléosides	
Nom	Symbole	Nom	Symbole	Nom	Symbole
cytosine	Cyt	cytidine	Cyd	2'-désoxycytidine	dCyd
uracile	Ura	uridine	Urd	2'-désoxyuridine	dUrd
thymine	Thy	thymidine*	Thd	(2'-désoxy)thymidine*	
adénine	Ade	adénosine	Ado	2'-désoxyadénosine	dAdo
guanine	Gua	guanosine	Guo	2'-désoxyguanosine	dGuo
xanthine	Xan	xanthosine	Xao	2'-désoxyxanthosine	dXao
hypoxanthine	Hyp	inosine	Ino	2'-désoxyinosine	dIno

Tableau 1: Nomenclatures des différents nucléotides

\* le nom de thymidine a été consacré pour la 2'-désoxythymidine : pour différencier ce ribonucléoside, on le désigne sous le nom de ribosylthymidine.

#### **2-4. Les nucléosides naturels**

Les nucléosides apparaissent en tant qu'intermédiaires dans le métabolisme des nucléotides, mais aussi comme cofacteurs ou coenzymes :

- la S-adénylméthionine est un cofacteur de transfert de groupement méthyle. Le carbone 5' de l'adénosine est lié de manière covalente avec le groupement sulfhydryl de l'acide aminé méthionine.

- le coenzyme B12 (ou vitamine B12), dont la fonction permet de déplacer des groupes de manière intramoléculaire, est la 5'-désoxyadénosine liée au cobalt du noyau corrine.

#### **2-5. Des nucléosides naturels et singuliers**

- On trouve dans les ARN de transfert (ARNt) de l'uracile qui est lié au ribose non par l'azote mais par le carbone C5 : c'est une liaison C-osidique. On trouve aussi des dérivés 2'-Ométhylés du ribose : la présence du méthyle donne une résistance à l'hydrolyse enzymatique ou même alcaline à l'ARNt.

- Certains microorganismes produisent des nucléosides libres dans lesquels l'ose est l'arabinose.

L'arabinosylcytosine est utilisé comme médicament dans le traitement de l'herpès, l'arabinosyladénosine dans les leucémies.

- La puromycine est un antibiotique sécrété par Streptomyces. La base est l'adénosine Nméthylée, le pentose la ribosamine dans lequel le carbone C3' est lié par une liaison amide à une O-méthyltyrosine.

#### **2-6. Des analogues synthétiques**

- les formes désoxy, halogénées ou non, des nucléosides puriques ou pyrimidiques sont des bactériostatiques, par exemple la 3'-désoxyadénosine.

- la 6-azauridine (carbone 6 remplacé par un azote) sont des médicaments utilisés dans les maladies prolifératives de l'épiderme.

### 3- Les nucléotides

Ce sont des esters-phosphates de nucléosides (condensation alcool-acide).

Exemple de nucléotides :

- la phosphorylation peut concerner un ou plusieurs hydroxyles de l'ose.
- le groupe ester-phosphate peut être engagé soit :
  - avec d'autres molécules d'acide phosphorique (ADP, ATP)
  - avec un autre acide (adénosine 3'-phosphate 5'-phosphosulfate)
  - avec un autre hydroxyle par une deuxième liaison ester dans les nucléotides

Cycliques.

Exemple de nucléotides :

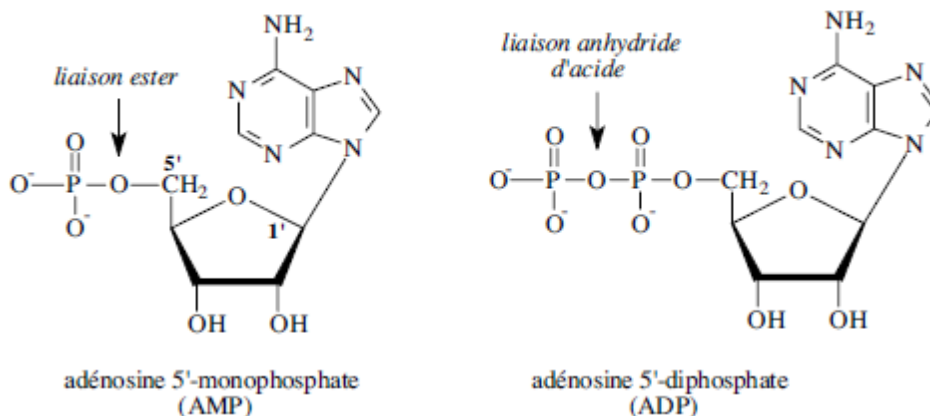


Figure 13 :

Les liaisons esters

#### 3-1. Nomenclature

Les noms des nucléotides obéissent à la règle suivante :

nucléoside n'[n"] - nb phosphate : (n' numéro du carbone portant le phosphate)

nucléoside n':[n"] - nb phosphate, cyclique pour les nucléotides cycliques

Les symboles ou abréviations classiques sont pour les nucléotides à une seule liaison ester :

– symbole nucléoside n' P[P]

– symbole nucléoside (une lettre) N P (N nombre de phosphate : M 1, D 2, T 3)

pour le symbole à 3 lettres qui suppose que la liaison ester est porté par le carbone 5'.

Lorsque l'ose est le désoxyribose, le symbole est précédé de d (exemple dAMP)

Pour leur caractère acide, dû au groupement phosphate, les nucléotides monophosphates ont aussi la dénomination suivante : acide n' [désoxy] (radical de la base) idylique : (n' numéro du carbone portant le phosphate)

Par exemple :

– adénosine 3', 5'-biphosphate : l'adénosine porte 2 groupements phosphate, l'un sur la carbone C3 et l'autre sur le C5.

– adénosine 3'-(mono)phosphate 5'-phosphosulfate (PAPS) : l'adénosine porte un groupement phosphate en C3' et un groupement phosphate en C5' qui est lui-même lié à un sulfate par une liaison anhydride d'acide.

### **3-2. Les nucléotides d'intérêt biologique**

Les nucléosides monophosphates

La phosphorylation sur le carbone C5' est la plus courante. L'AMP est l'un des composés que l'on trouve dans de nombreuses réactions du métabolisme.

Les nucléosides 5'-phosphates sont les éléments de bases de polymères tels que l'ADN et l'ARN que l'on trouve dans tous les organismes vivants.

Les nucléosides diphosphates

Le PAPS (adénosine 3'-(mono)phosphate 5'-phosphosulfate) est impliqué dans des réactions de sulfatation des glycanes et des lipides.

L'ADP (et les autres nucléosidesPP classiques) est un composé qui a de nombreuses différentes fonctions, citons parmi elles :

- molécule à haut potentiel énergétique (hydrolyse de la liaison anhydride d'acide)

- molécule intermédiaire dans la production d'ATP

- activateur d'enzyme allostérique comme la L-glutamate-déshydrogénase

Les nucléosides triphosphates

- L'une des molécules les plus "universelles" est l'ATP : ses deux liaisons anhydride d'acide (enthalpie libre d'hydrolyse d'une liaison de l'ordre de  $30\text{kJ}\cdot\text{mol}^{-1}$ ) sont une source d'énergie utilisable :

- par des réactions de catalyse enzymatique impossibles sans couplage

- pour le transport transmembranaire actif

- pour la contraction musculaire

- etc...

D'autres nucléosides 5' triphosphates jouent un rôle identique à celui de l'ATP mais à un degré beaucoup plus minoritaire.

Les nucléosides 5' triphosphates sont aussi des donneurs de phosphates.

- Les nucléosides 5' triphosphates sont les précurseurs de base dans la biosynthèse des acides nucléiques.

Des constituants de coenzymes

Citons :

- l'adénosine 3'-phosphate 5'-diphosphate est une partie de la molécule d'un coenzyme d'acétylation, le coenzyme A.

- l'adénosine 5'-diphosphate fait partie de la structure de deux coenzymes d'oxydo-réduction : le nicotinamide adénine dinucléotide (NAD) et la flavine adénine dinucléotide (FAD)

- le PAPS est impliqué dans des réactions de sulfatation des glycanes et des lipides

## Les seconds messagers

Un second messenger est une molécule qui assure le relais intracellulaire d'un signal extracellulaire. L'AMPc fut identifié comme le second messenger universel et dans quelques cas, ce fût le GMPc. Il est produit sur la face intracellulaire de la membrane plasmique après fixation externe d'une molécule signal.

### III. les acides nucléiques :

Les acides nucléiques sont des enchaînements de nucléosides 5'-phosphates dont l'assemblage est réalisé par une liaison phosphodiester. Les deux types d'acides nucléiques sont :

– ADN (acide désoxyribonucléique) composé de :

– un ose qui est le 2'-désoxyribose

– la base est soit : adénine ou guanine (purine), soit cytosine ou thymine (pyrimidine)

– ARN (acide ribonucléique) composé de :

– un ose qui est le ribose

– la base est soit : adénine ou guanine (purine), soit cytosine ou uracile (pyrimidine)

Remarquons que l'ADN peut contenir des bases méthylées et que l'ARN contient de nombreuses différentes bases modifiées (voir le paragraphe des bases).

La différence entre les 2 oses a des conséquences très importantes entre ces deux polymères : la présence de l'hydroxyle en 2' du ribose interdit tout appariement pour former des duplex de chaînes.

#### 1– La structure primaire des polymères

##### 1-1. La taille des polymères nucléiques

La taille des acides nucléiques est exprimée en trois unités selon l'usage :

– la longueur

- la masse moléculaire en Dalton (Da)
- le nombre de nucléotides (ou bases), noté b, pour les molécules simple brin et le nombre de paires de base, noté pb, pour les molécules double brin.

### 1-2. La liaison phosphodiester :

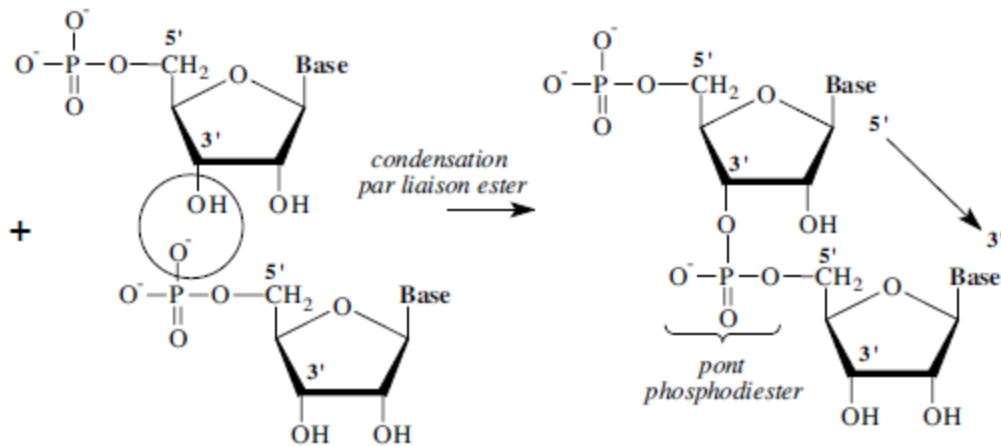


Figure 14 :

### Liaison phosphodiester

Les acides nucléiques sont des enchaînements de nucléosides 5'-phosphate dont l'assemblage est réalisé par une liaison phosphodiester :

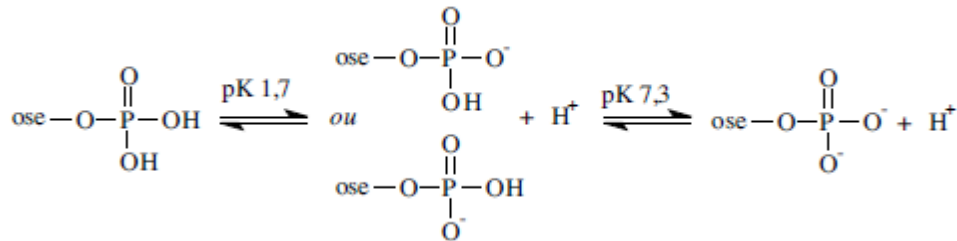
La chaîne est vectorisée : elle est écrite de gauche à droite et dans le sens, extrémité phosphate 5'→3'. C'est le sens dans lequel les séquences d'acides nucléiques sont utilisées comme molécules informationnelles (transcription, traduction).

## 2- Les différents groupes ionisables des acides nucléiques

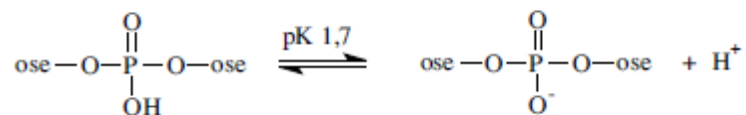
Les différents groupes ionisables des acides nucléiques sont de trois types et d'un type particulier correspondant aux formes tautomères de l'équilibre forme céto et forme énol(lactame et lactime). La nature de l'ose n'a aucun effet pour ces phénomènes d'ionisation.

## 2-1. Les groupes phosphates

- le groupe phosphomonoester possède deux fonctions acides dont les valeurs des pK sont respectivement 1,5 à 2 et 7,2 à 7,5.



- le groupe phosphodiester, participant aux ponts phosphodiester, possède une seule fonction acide de pK égal à 1,5 à 2



Chacun de ces groupes possède donc au pH physiologique (6,5) un seul groupement ionisé : l'acide nucléique portera un nombre de charges négatives (contribution des phosphates) égal au nombre de nucléotides.

## 2-2. Les azotes des cycles des bases pyrimidiques et puriques

Lorsque les bases puriques ou pyrimidiques sont liées à un ose (liaison N-osidique), la protonation d'un azote du cycle correspond au cas plus simple des bases de type pyrrole ou pyridine. Il faut noter que la conjugaison des liaisons du doublet libre de l'azote avec les doublets des électrons  $\pi$  aboutit à la règle simple suivante :

- la protonation d'un azote d'un cycle empêche la protonation des autres azotes.

- l'uracile et la thymine n'ont aucun azote protoné.

### **2-3. Les groupes amino liés aux cycles des bases**

Les groupements amino liés au cycle peuvent être protonés et ce plus facilement qu'un azote du cycle. Le doublet de l'azote de ce groupement amino peut se conjuguer partiellement avec les électrons du cycle, empêchant la protonation des autres azotes.

### **2-4. Les formes tautomères (lactame et lactime)**

Nous avons vu (page 5) que la présence d'un groupement céto lié à un cycle entraînait l'existence de formes tautomères par le passage de la forme céto en forme énol. Cette dernière fonction peut libérer un proton à pH basique (acide faible de pK d'environ 9,2 à 9,5), entraînant l'apparition d'une charge négative.

### **2-5. En guise de conclusion**

Les bases puriques et pyrimidiques, liées à un ose (liaison N-osidique), portent des groupes ionisables qui peuvent libérer des protons :

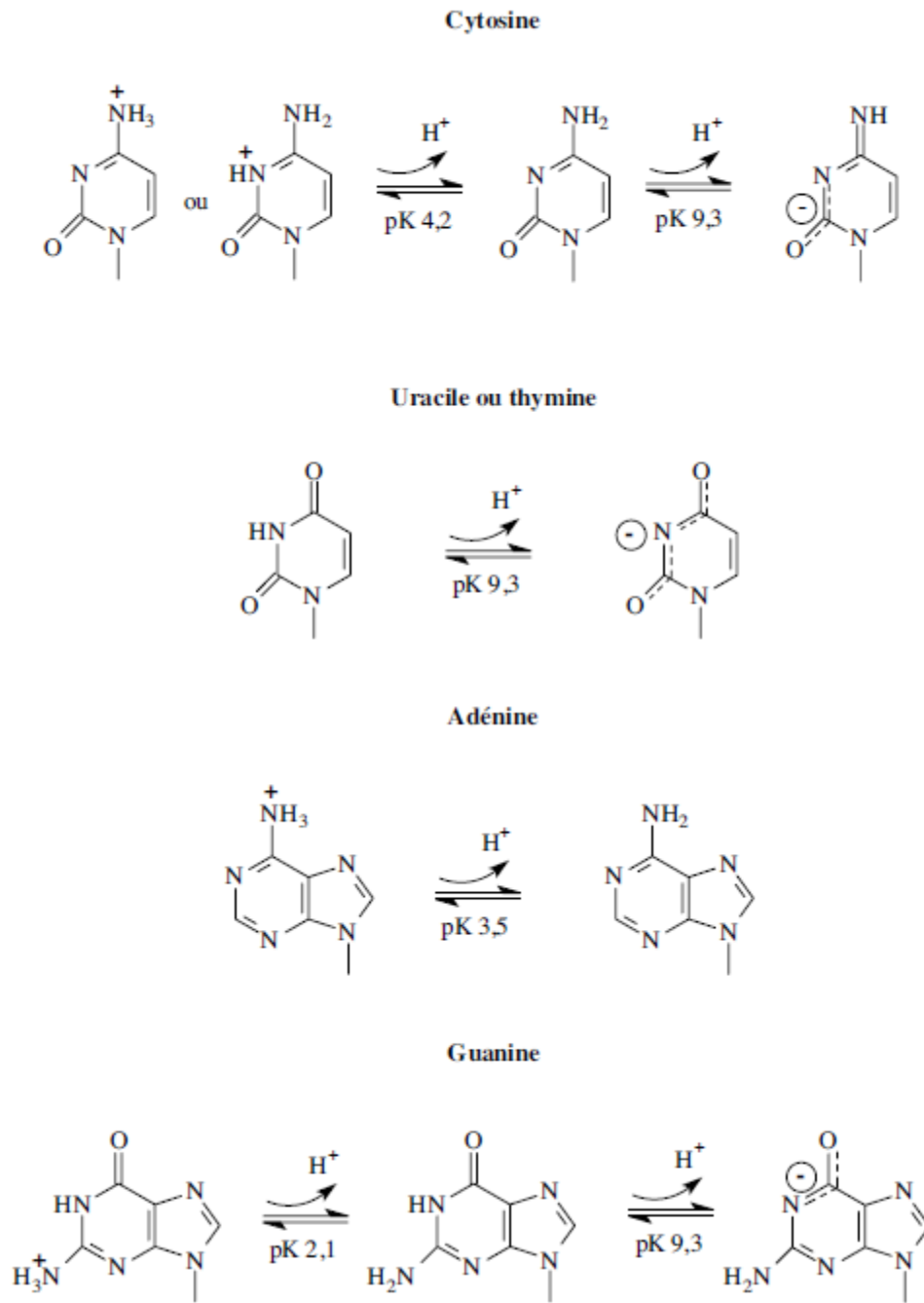


Figure 15 :

Charge des bases azoteseufonction du pH

A un pH voisin du pH physiologique (6,5), les bases puriques ou pyrimidiques ne portent aucune charge.

Les acides nucléiques dans une solution dont le pH est voisin du pH physiologique sont des polymères chargés négativement, charges dont la contribution est uniquement due aux groupements phosphates.

### 3– L'hydrolyse des acides nucléiques

La dégradation d'un polynucléotide peut être chimique ou enzymatique, elle concerne :

- l'enchaînement phosphodiester
- les unités nucléotidiques : composants et liaison osidique

#### 3-1. L'hydrolyse chimique

Le traitement acide affecte de la même façon les ADN et les ARN

– la dégradation du squelette phosphodiester est obtenue dans des conditions drastiques (acide concentré et chauffage) auxquelles ne résistent pas les autres liaisons, cette dégradation conduit à la libération d'un mélange de phosphates, oses et bases.

– dans des conditions douces (pH 4), seules les liaisons N-osidique avec les purines sont hydrolysées.

Les ARN et les ADN réagissent différemment à l'hydrolyse alcaline

– les ADN résistent aux pH basiques : par exemple, à pH 13 et à 37°C on a une dizaine de coupures de ponts phosphodiester par million de ponts

– les ARN sont totalement hydrolysés en leurs ribonucléotides en quelques minutes à 37°C et à pH 11. C'est la présence de l'hydroxyle libre en 2' qui permet cette hydrolyse qui donne un intermédiaire cyclique 2':3' P, aboutissant à des nucléotides 2' P ou 3' P.

#### 3-2. L'hydrolyse enzymatique

Les enzymes qui catalysent l'hydrolyse de la liaison phosphodiester des acides nucléiques, présents dans la plupart de toutes les cellules, sont

desphosphodiesterases spécifiques appelées nucléases. Des endonucléases de très haute spécificité sont présentes dans les bactéries, ce sont des désoxyribonucléases désignées sous le nom d'enzyme de restriction.

### Les nucléases

Elles présentent des niveaux de spécificité et sont classées par :

- leur mode d'attaque de la chaîne : extrémité (exo) ou intérieur (endo)
- leurs spécificité vis-à-vis du substrat : ADN, ARN ou les deux et de la structure, simple ou double brin
- leur spécificité de reconnaissance des sites : bases ou leur enchaînement
- le type de coupure de la liaison phosphodiester.

nucléases	substrats	type coupure	spécificité coupure
<b>exonucléases</b>			
phosphodiesterase de venin	ARN, ADN (s)	3'	extrémité 3'
phosphodiesterase de la rate	ARN, ADN (s)	5'	extrémité 5'
exonucléase I d' <i>E.Coli</i>	ADN (s)	3'	extrémité 3'
exonucléase III d' <i>E.Coli</i>	ADN (d)	3'	extrémité 3'
<b>endonucléases</b>			
endonucléase S1 d' <i>Aspergillus</i>	ARN, ADN (s, d)	3'	aléatoire
ARNase T1 d' <i>Aspergillus</i>	ARN (s)	5'	-G # N-
ARNase de pancréas	ARN (s)	5'	-Pyr # N-
ADNase II de thymus	ADN (s)	5'	-dPyr # dPur-

Tableau 2: Exemple de quelques nucléases avec leurs spécificités

s: simple brin, d : double brin

N : représente n'importe lequel des nucléotides

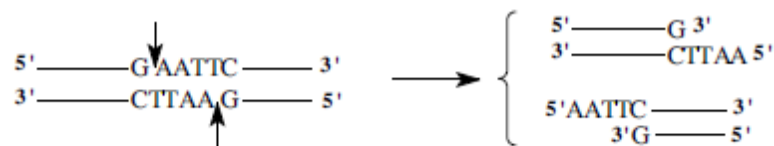
### Les enzymes de restriction

Chaque espèce de bactéries produit une collection d'endo-désoxyribonucléases, dont le site est spécifique d'une séquence de 4 à 10 nucléotides, qui leur permet de s'opposer à l'infection de certains virus en hydrolysant leur ADN, sans hydrolyser leur propre ADN protégé par le biais d'une

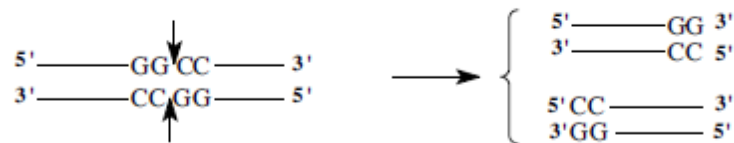
méthylation des séquences partielles spécifiques. On parle de restriction virale et on a appelé ces nucléases des enzymes de restriction.

Ces désoxyribonucléases hydrolysent un pont phosphodiester de chacune des 2 chaînes d'acide désoxyribonucléique dont la structure est un double brin. L'ADN double brin est formé de deux chaînes complémentaires de sens opposé (voir plus loin). La plupart du temps la séquence nucléotidique reconnue (lue dans le même sens) est identique sur les deux brins, on dit que le site est palindromique :

Voici deux exemples d'enzymes de restriction :



Hydrolyse par l'enzyme de restriction : EcoRI  
site : GAATTC (coupure débordante)



Hydrolyse par l'enzyme de restriction : HaeIII  
site : GGCC (coupure franche)

Figure 16 :

Exemples d'enzymes de restriction

Ces enzymes sont un outil de choix pour toutes les techniques de biologie moléculaire d'autant que le nombre d'enzymes de restriction purifiés est relativement grand : environ 500.

#### 4- Structure spatiale des acides désoxyribonucléiques

##### 4-1. Quelques résultats expérimentaux sur les acides désoxyribonucléiques

La composition en bases

Les molécules d'ADN présentent la caractéristique suivante :

- quelle que soit l'origine de l'ADN, le nombre de purines est toujours égal au nombre de pyrimidines :  $[Pur] = [Pyr]$  ou encore  $[A] + [G] = [T] + [C]$

- de plus, les fractions molaires des bases sont telles que :  $A = T$  et  $G = C$

Cette caractéristique est désignée sous le nom de règle de Chargaff qu'il observa en 1940. Les bases A et T sont dites complémentaires, il en est de même pour G et C. Bien sûr les proportions ( $[A] + [T]$ ) et ( $[G] + [C]$ ) ne sont pas égales et varient de 35 à 75% selon l'ADN étudié.

#### Le titrage des groupes basiques

Si on titre les bases de l'ADN (charge moyenne de la molécule en fonction du pH) en présence d'un agent dénaturant, comme l'urée, le pK moyen obtenu est de l'ordre de 4,5.

Si on titre les bases de l'ADN dans des conditions non dénaturantes, on trouve un pK moyen augmenté d'une unité, environ 5,5

#### L'effet hyperchrome

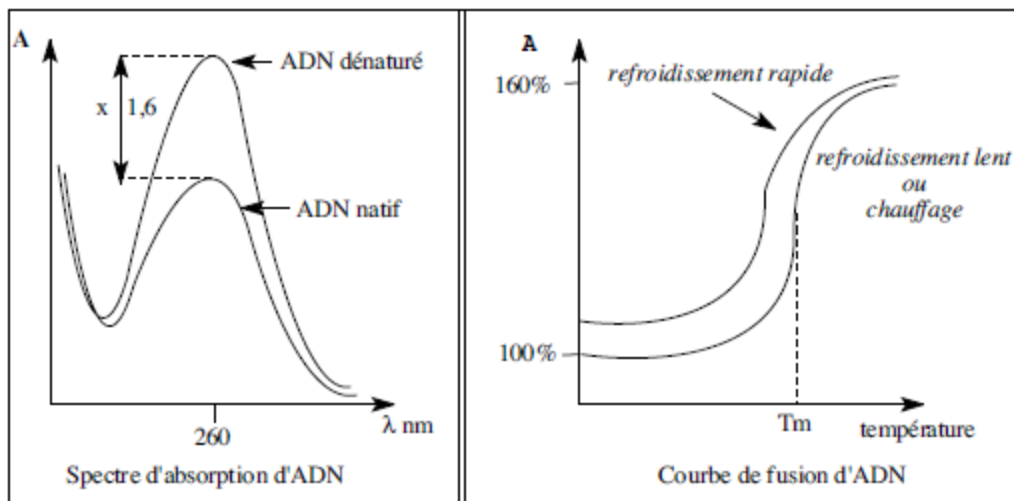


Figure 17 :

#### Spectre d'absorption de l'ADN

- Le spectre d'absorption de l'ADN natif n'est pas identique à celui du même ADN dénaturé la chaleur (chauffage à 100°C) ou par l'urée ou encore à pH très

alcalin. L'ADN dénaturé a une absorption à 260 nm plus élevée que l'ADN natif, d'un facteur 1,6. Cette propriété est appelée l'effet hyperchrome ou hyperchromicité.

- L'absorption de l'ADN natif, à 260 nm, en fonction de la température (courbe de fusion), présente l'allure d'une sigmoïde : le point d'inflexion de cette courbe, qui correspond à la demi-variation d'absorbance, est la température de fusion de la molécule, notée  $T_m$ .

- Lors d'un refroidissement lent, l'absorption suit la courbe de fusion en sens inverse. Lors d'un refroidissement rapide, l'absorption ne suit pas la courbe de fusion en sens inverse mais une autre courbe qui n'aboutit à la même valeur originale de l'absorption, mais à une valeur plus élevée : c'est le phénomène d'hystérésis.

- Il faut noter que la température de fusion  $T_m$  est dépendante de la force ionique du milieu et qu'elle diminue lorsque cette dernière augmente (dans des milieux où  $[NaCl] > 1M$ ).

- La valeur de la température de fusion  $T_m$  est d'autant plus élevée que le pourcentage de bases G + C est grand :

- 65 °C pour l'ADN de E. Coli où G+C est égal à 50%,

- 76 °C pour l'ADN de P. aeruginosa où G+C est égal à 68%

La structure révélée par la diffraction aux rayons X

L'analyse aux rayons X de cristaux de molécules d'ADN, réalisée en 1953 par Watson et Crick, indique une structure ayant la configuration suivante :

- une double hélice formée de deux chaînes d'ADN dont les paires de bases sont complémentaires

- les deux chaînes sont antiparallèles

- les enchaînements oses-phosphates forment les squelettes hélicoidaux parallèles extérieurs, les plans des oses étant presque perpendiculaires au plan des bases
- les plans des bases sont perpendiculaires à l'axe de l'hélice et chacune des bases d'une chaîne est appariée à celle de l'autre chaîne par des liaisons hydrogène, les deux bases étant situées dans un même plan
- les bases appariées sont complémentaires (A / T) et (G / C) et sont à l'intérieur du cylindre central de l'hélice.

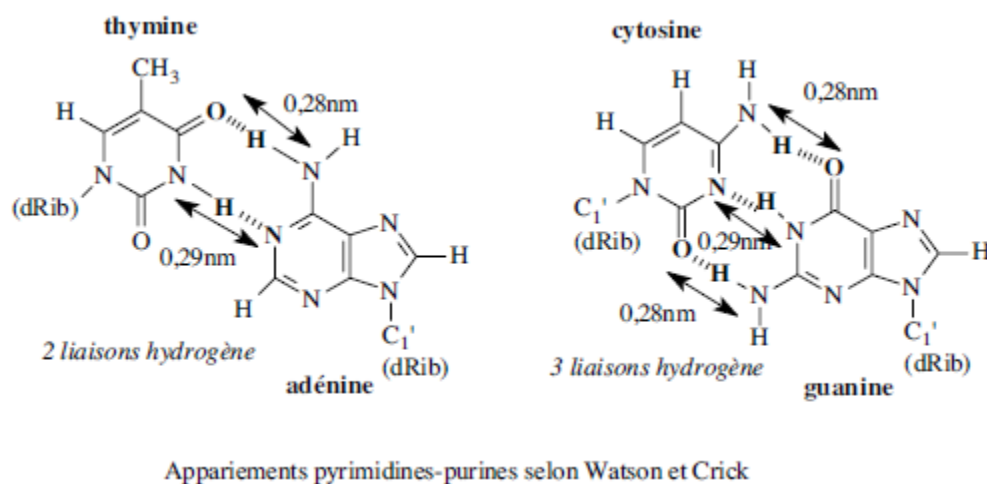


Figure 18:

Appariement pyrimidines-purines selon Watson et Crick

#### 4-2. Les doubles hélices

Une double hélice est formée de deux chaînes antiparallèles d'ADN qui s'enroulent soit à droite (sens des aiguilles d'une montre) soit à gauche.

Cette double hélice est caractérisée par :

- son axe principal,
- le sens d'enroulement,
- le pas.

Du fait de la position de l'axe au centre de chaque paire de base complémentaire et de leur attachement dissymétrique sur le squelette ose-phosphate, deux sillons d'ouverture et de profondeurs différentes vont alterner sur les flancs de la double hélice : La stabilité de la structure secondaire de ces différentes double hélices d'ADN est essentiellement due aux :

- liaisons hydrogène entre les bases complémentaires de chacun des brins
- interactions hydrophobes et électrostatiques des bases successives empilées dans la structure de l'hélice, dont la distance des plans varie de 0,26 à 0,37 nm.

Cette stabilité n'entraîne pas une rigidité de la molécule d'ADN et celle-ci peut adopter des conformations différentes selon les régions.

Plusieurs conformations correspondant à des sens d'enroulement différents ou des pas différents ont été trouvées, les principales étant :

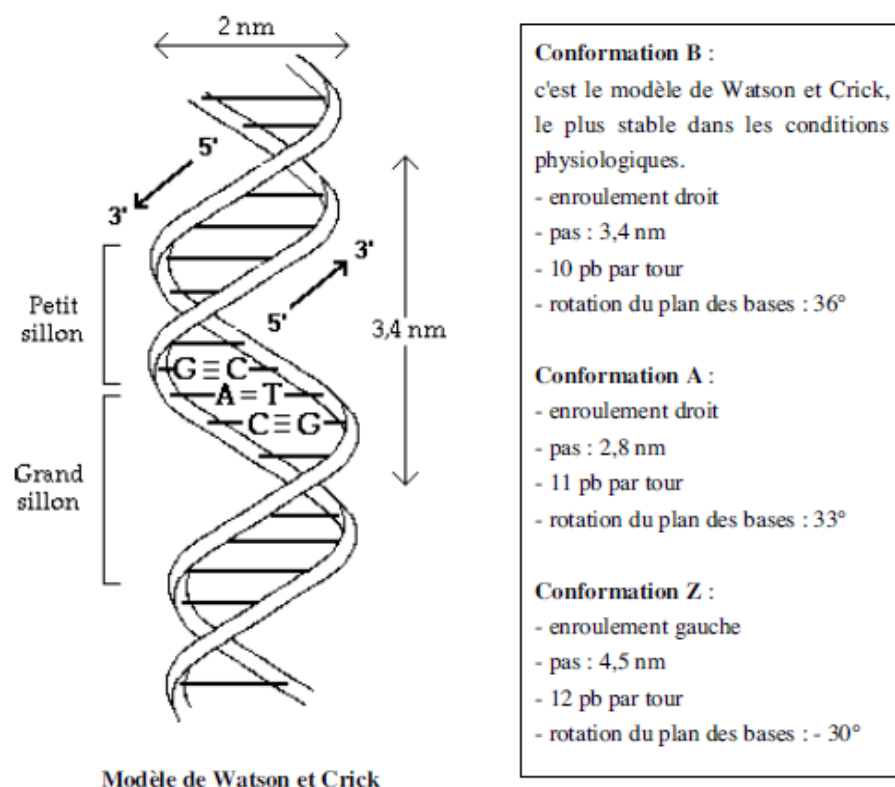


Figure 19:  
Modèle de Watson et Crick de l'ADN (conformation B)

La stabilité de la structure secondaire de ces différentes double hélices d'ADN est essentiellement due aux :

- liaisons hydrogène entre les bases complémentaires de chacun des brins
- interactions hydrophobes et électrostatiques des bases successives empilées

dans la structure de l'hélice, dont la distance des plans varie de 0,26 à 0,37 nm.

Cette stabilité n'entraîne pas une rigidité de la molécule d'ADN et celle-ci peut adopter des conformations différentes pour différentes régions.

C'est une hélice droite de pas égal à 3,4 nm (10 pb par tour).

- L'inclinaison des bases par rapport à leur axe de rotation perpendiculaire à l'axe principal de l'hélice est de  $1^\circ$ .

- La distance entre deux plans de bases successifs est de 0,34 nm.

#### Conformation A

Lorsque la teneur en eau d'une solution contenant une molécule d'ADN est diminuée, par exemple lors de la cristallisation, la molécule change de conformation et adopte une conformation notée A. Ce changement est réversible.

La conformation A a les spécificités suivantes :

- hélice droite (idem conformation B)
- hélice plus compacte
- pas de 2,8 nm (11 pb par tour)
- distance entre les plans de bases successives de 0,26 nm
- les plans des bases sont tournés de  $180^\circ$  par rapport à la conformation B
- l'inclinaison des bases par rapport à leur axe de rotation perpendiculaire à

l'axe principal de l'hélice est de  $20^\circ$ .

Cette conformation est trouvée in vivo dans :

- l'ADN de certaines spores bactériennes, formées en réponse à la dessiccation du milieu

- les hybrides ADN-ARN qui se forment transitoirement à l'amorce de la réplication, et pendant la transcription.

### Conformation Z

Cette conformation est présente dans des régions courtes de l'ADN dans une conformation générale de type B (native). Ces régions spécifiques sont en général des segments de séquence alternée Pur/Pyr (G-C-G-C).

La conformation Z a les spécificités suivantes :

- hélice gauche
- pas de 4,5 nm (12 pb par tour) : la molécule est plus étirée dans cette conformation
- distance entre les plans de bases successives de 0,37 nm
- l'inclinaison des bases par rapport à leur axe de rotation perpendiculaire à l'axe principal de l'hélice est de 9°.

Cette conformation est trouvée in vivo pour des segments de la molécule d'ADN, avec des enchaînements alternés Pur/Pyr dont les bases sont souvent méthylées. Celle-ci aurait un rôle dans l'expression des gènes.

Ces différentes conformations décrivent la structure secondaire des molécules d'ADN double brin. Ces conformations sont très stables, mais possèdent une flexibilité assez grande qui peut définir des structures tertiaires (structure dans l'espace d'une double hélice) différentes : celles-ci vous seront décrites dans des cours ultérieurs et vous verrez, par exemple, les phénomènes de surenroulement.

# SEQUENÇAGE 1<sup>ère</sup> GENERATION

## I. Méthode de Maxam et Gilbert :

Cette méthode n'a plus guère qu'un intérêt historique, bien qu'elle ait permis le séquençage, parmi d'autres, de l'opérateur/promoteur de l'opéron :

Des fragments de restriction (<200pb) sont purifiés à partir d'un gel et marqués à leurs extrémités 5' par kination avec l'enzyme poly nucléotide kinase et du  $\gamma$ -[32P]-ATP.

Les deux brins sont dénaturés à la chaleur, puis séparés en gel d'acrylamide.

Chaque brin est soumis à quatre réactions d'hydrolyse partielle qui vont cliver la molécule avant les A et les G, ou avant les G, ou avant les C, ou avant les C et T. On va générer pour chaque réaction une famille de molécules clivées à une position variable en 3' de l'extrémité marquée.

Les produits sont séparés par l'électrophorèse en gel d'acrylamide contenant de l'urée. Ces gels sont très résolutifs (deux fragments de longueur différant par 1 nucléotide sur 500 ou plus seront séparés).

Le gel est autoradiographié : on verra que les molécules ayant l'extrémité 5' marquée. La séquence est lue directement sur l'autoradiographie en comparant le résultat des 4 réactions.

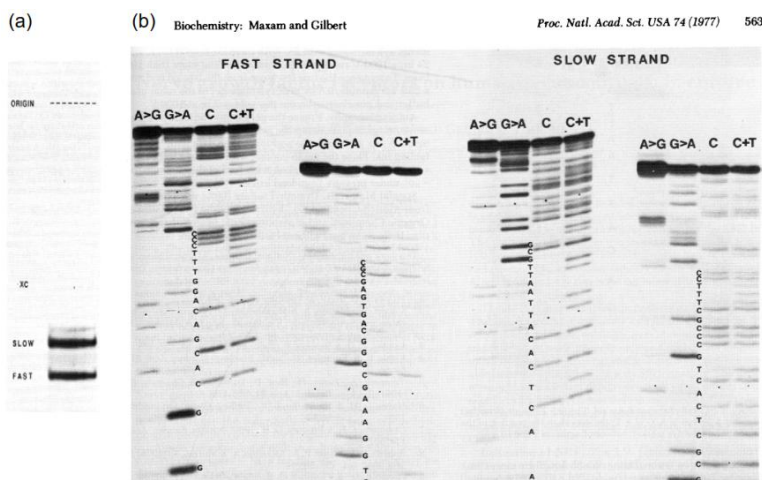


Figure 20 :

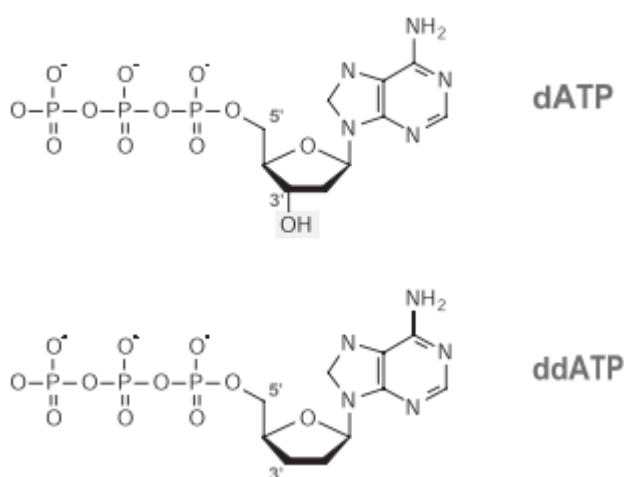
Principe de la méthode Maxam et Gilbert

La compilation des séquences chevauchantes sur les deux brins permet d'obtenir sans ambiguïté la séquence du gène entier. Utilisant les agents chimiques qui lèsent l'ADN, assez lourde à mettre en pratique, cette méthode n'est plus utilisée que dans quelques cas particuliers.

## II. Méthode de Sanger :

La méthode des didésoxyribonucléotides, inventée il y a une vingtaine d'années dans le laboratoire de Fred Sanger à Cambridge en Grande-Bretagne, est aujourd'hui universellement employée pour séquencer l'ADN. Elle repose sur l'allongement par l'ADN polymérase d'un brin à partir d'une amorce, en utilisant un autre brin d'ADN comme matrice.

Cet allongement est réalisé en présence des quatre désoxyribonucléotides triphosphate (dATP, dTTP, dGTP, dCTP), monomères utilisés par la polymérase, et d'un analogue didésoxyribonucléotide (ddNTP, figure 21) qui joue le rôle de terminateur de chaîne.



Dans le didésoxyribonucléotide (ddNTP), le remplacement du groupement 3'-OH par un 3'-H empêche la formation d'une liaison phosphodiester du côté 3'.

Ces nucléotides modifiés peuvent toutefois être incorporés par l'ADN polymérase car ils possèdent un côté 5'-triphosphate normal.

Les règles d'appariement A-T et G-C sont respectées lors de l'incorporation des ddNTP. Ainsi le ddATP sera incorporé lorsqu'on trouvera en regard un T sur le brin matrice.

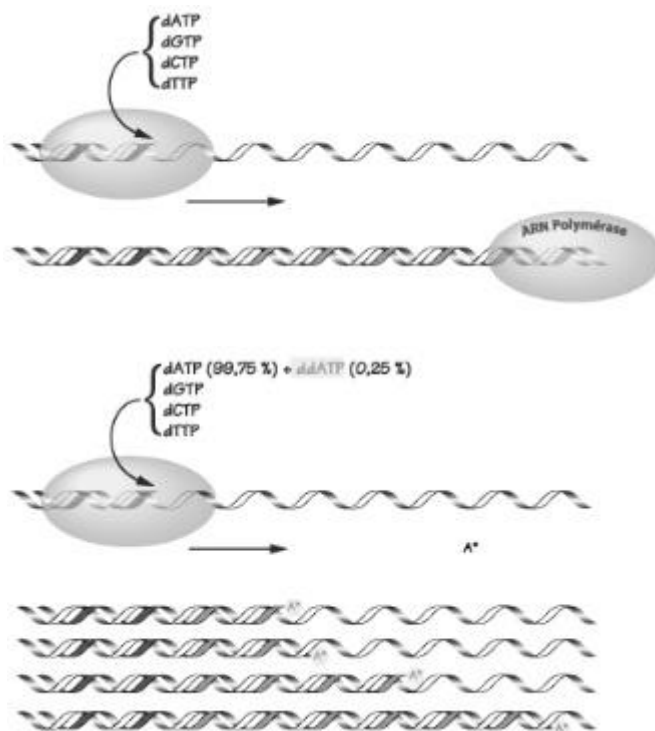
Figure 21 :

Structure des didésoxyribonucléotides.

Du fait de l'incorporation spécifique de l'analogue par la polymérase, on obtient un mélange de fragments qui se terminent sélectivement aux positions correspondant au nucléotide choisi (dans la figure 22, les A).

Le principe de la méthode de séquençage est illustré sur la figure 22 :

On effectue ainsi quatre réactions en parallèle, chacune avec l'un des quatre ddNTP, et l'on sépare les fragments obtenus par électrophorèse. Afin de pouvoir identifier les fragments d'ADN synthétisés par la polymérase et en particulier pour pouvoir les distinguer de l'ADN matrice, on les marque avec un traceur fluorescent. Celui-ci est accroché à l'une de ses deux extrémités, soit en 5', sur l'amorce de séquençage, soit en 3' sur le didésoxyribonucléotide terminateur.



En présence d'un brin d'ADN matrice et des quatre dNTP, l'ADN polymérase est capable d'allonger un brin d'ADN complémentaire, à partir d'un oligonucléotide amorce hybridé au brin matrice.

Lorsqu'un didésoxyribonucléotide est incorporé par la polymérase, celui-ci agit comme un terminateur de chaîne, bloquant tout allongement ultérieure. Cette incorporation se produit de manière aléatoire, avec une fréquence dépendant du rapport de la concentration du didésoxyribonucléotide sur celle du désoxyribonucléotide correspondant.

Figure 22 :

Principe du séquençage par la méthode de Sanger.

Les séquenceurs automatiques modernes utilisent un système de détection in situ pendant l'électrophorèse. Le faisceau d'un laser émettant dans la bande d'absorption du fluorophore traverse le gel (figure 23).

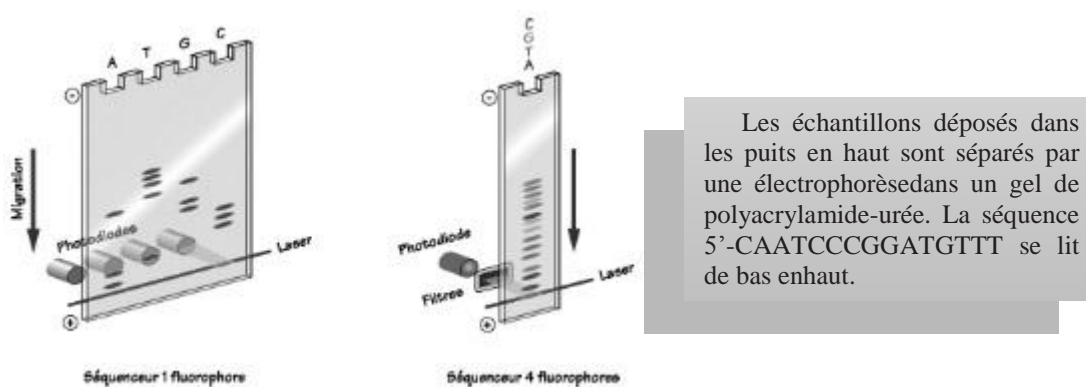


Figure 23 :

Séquençage automatique sur des séquenceurs 1 et 4 fluorophores.

Pendant la migration, lorsqu'une bande d'ADN passe devant le faisceau, un signal de fluorescence est émis. Celui-ci est capté par une photodiode située en regard du gel. Le signal est amplifié puis transmis à l'ordinateur de contrôle et analysé par un logiciel spécialisé.

Dans des conditions favorables, cette technique permet de lire jusqu'à 1000 nucléotides par fragment séquencé. En routine, la moyenne est de l'ordre 500 à 800 nucléotides par expérience.

Deux méthodologies cohabitent actuellement, reposant sur l'utilisation d'un fluorophore unique ou bien sur celle de quatre traceurs fluorescents possédant des spectres d'émission distincts. Dans le premier cas, les quatre mélanges correspondant à chacun des quatre ddNTP sont déposés sur des puits distincts du gel. L'analyse se fait sur la migration comparée des fragments dans les quatre pistes

résultantes. Dans l'autre système, on utilise un fluorophore différent dans chacune des quatre réactions de séquençage. Une solution consiste à utiliser des ddNTP modifiés chacun par un traceur spécifique. Après avoir effectué les quatre réactions de polymérisation, on les mélange et on dépose dans le même puit sur le gel. La reconnaissance des nucléotides se fait alors sur la base des propriétés d'émission du traceur qui passe devant le faisceau laser, au moyen de filtres colorés sélectifs. L'analyse est alors effectuée sur une seule piste du gel.

La technique à quatre fluorophores est un peu plus onéreuse, car elle nécessite une chimie un peu plus diversifiée. En revanche, elle présente l'avantage d'être bien adaptée aux plus hauts débits, car plus d'échantillons peuvent être analysés sur le même gel. Dans les séquenceurs de dernière génération, le gel de polyacrylamide rectangulaire classique est remplacé par un capillaire réutilisable (le principe de la séparation et de la détection reste le même).

Cette technique permet de réduire la durée des expériences de quelques heures à quelques dizaines de minutes et de minimiser le temps passé par l'opérateur pour sa préparation. Les modèles multi-capillaires les plus performants peuvent, en principe, traiter jusqu'à 1000 échantillons par jour, soit un débit théorique de 0,5 Mbases de séquence brute par jour et par machine.

Les centres de séquençage massif possèdent aujourd'hui plusieurs dizaines de ces machines.

Les réactions de séquençage peuvent également être réalisées par des robots qui réalisent automatiquement pipetages, mélanges et incubations et minimisent les risques d'erreur humaine. La préparation des matrices d'ADN reste l'étape la plus lourde à automatiser, même si de nombreux progrès ont été accomplis.

# SEQUENÇAGE NOUVELLE GENERATION

## **I. Le séquençage de seconde génération : l'explosion de nouvelles technologies**

Depuis 2005, l'émergence de nouvelles technologies considérées comme la nouvelle ou seconde génération de séquençage (NGS, next-generation sequencing) a permis de séquencer avec des débits qui évoluent encore aujourd'hui de façon phénoménale. On a pour habitude de comparer cette évolution à l'augmentation exponentielle énoncée par la loi de Moore en 1965 selon laquelle la puissance des processeurs allait doubler tous les ans à prix constant : les technologies de séquençage ont évolué beaucoup plus vite (voir Figure 24).

Après celui de Craig Venter en 2007, plusieurs génomes personnels ont été publiés comme preuve de principe de ces nouvelles technologies. En 2008 c'est celui de James D. Watson, biochimiste célèbre pour sa découverte de la double hélice d'ADN [13], qui est généré sur un séquenceur FLX de Roche [14]. En seulement deux mois et pour moins d'1 million de \$, 24,5 millions de bases sont générées à une profondeur de 7,4x et elles couvrent 95% du génome. En 2009, un des co-fondateurs de la société Helicos Biosciences, Stephen R. Quake, séquence son génome [15] avec une profondeur de 28x et une couverture du génome de 90% pour un coût de 48000 \$. En 2009, 4 autres génomes humains ont été décrits : ceux d'un homme yoruba du Nigeria [16] séquencé à une profondeur de 30x, de 2 coréens [17][18] à une profondeur de 28 et 29x et d'un chinois Han [19] à une profondeur de 36x. Le séquençage du génome a également commencé à servir les intérêts de la médecine et c'est ainsi que la même année ont été publiés les génomes de deux patients atteints de leucémie [20] [21] à une profondeur de 33 et 23x. Fin 2010, le nombre de génomes séquencés approche les 3000 et fin 2011 il faut désormais le multiplier par 10 [22].

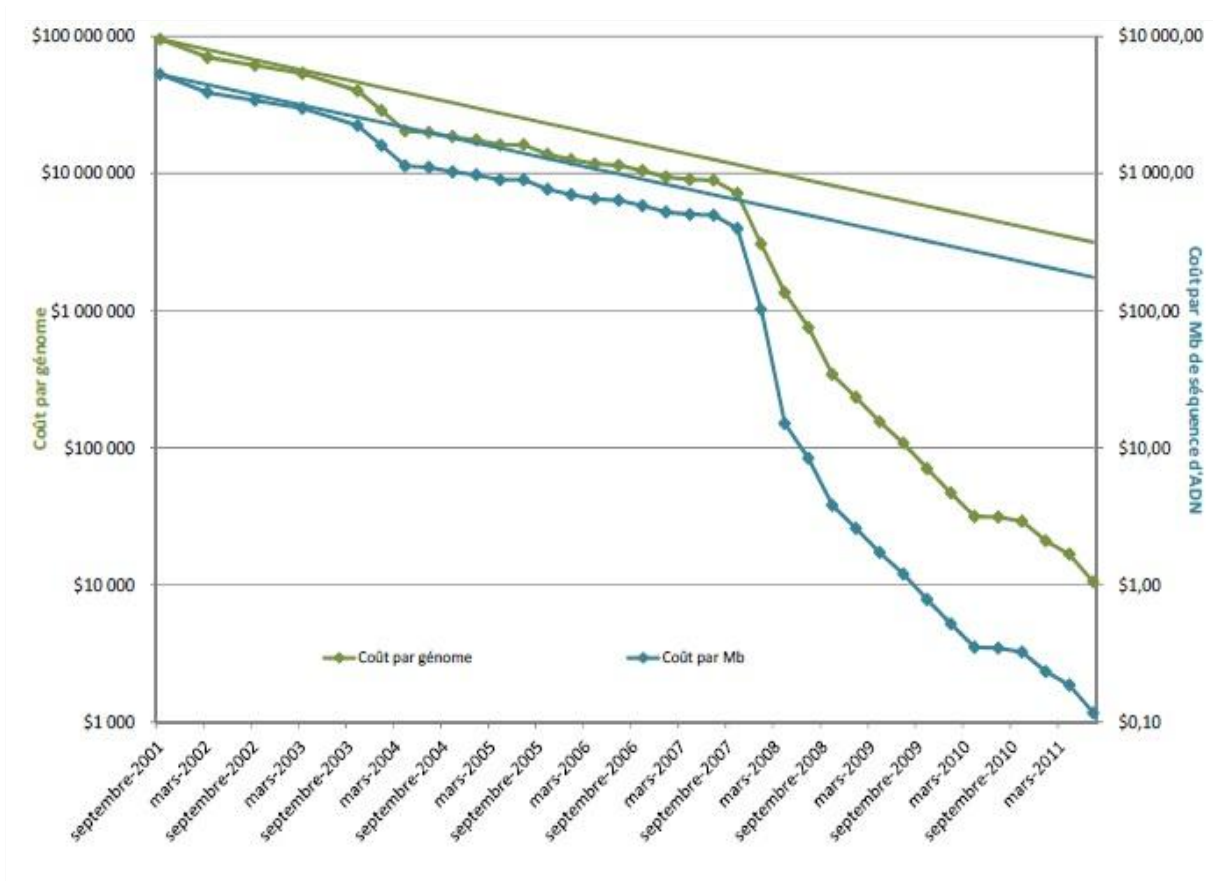


Figure 24 :

### Evolution des coûts du séquençage

Chiffres d'après Wetterstrand KA. DNA Sequencing Costs : Données du NHGRI Large-Scale Genome Sequencing Program

Cette augmentation est corrélée avec une diminution drastique des coûts : on approche de l'objectif du génome à 1000 \$ [23] fixé quelques années auparavant et les plus optimistes prévoient que ce seuil sera atteint d'ici deux ans même si certains s'accordent à dire que ce coût supportera seulement les réactifs et n'englobera ni l'amortissement des instruments, ni la main-d'œuvre ou encore le stockage et le management des données [24].

Quatre plateformes de séquençage de seconde génération sont actuellement disponibles sur le marché, proposant différentes versions de machines (voir Tableau

3). Elles utilisent des chimies qui leur sont propres ; le choix d'investir dans l'une ou l'autre de ces technologies reposera en partie sur l'application finale et devra prendre en compte les caractéristiques intrinsèques à chacune [25].

### 1- Les trois grands du séquençage :

Les trois technologies de séquençage qui dominent actuellement le marché se démarquent les unes des autres par les chimies qui les constituent. Elles se décomposent cependant toutes en 4 grandes étapes principales : la préparation des bibliothèques qui contient une étape d'amplification par PCR, les cycles de réactions de séquençage, la prise d'image après chacun de ces cycles pour déterminer le nucléotide correspondant, puis l'analyse des données. Ces nouvelles générations de machines ont pour avantage leur capacité à analyser de grands génomes à haute résolution grâce à la parallélisation des réactions.

#### 1-1 454/Roche

Le premier des séquenceurs de nouvelle génération a été commercialisé par 454 Life Sciences en 2005, depuis racheté par Roche, mais la plateforme est toujours connue sous le nom générique de 454. Jonathan M. Rothberg en a élaboré la technologie et a démontré sa robustesse avec le séquençage du génome de *Mycoplasmagénitalium*[26].

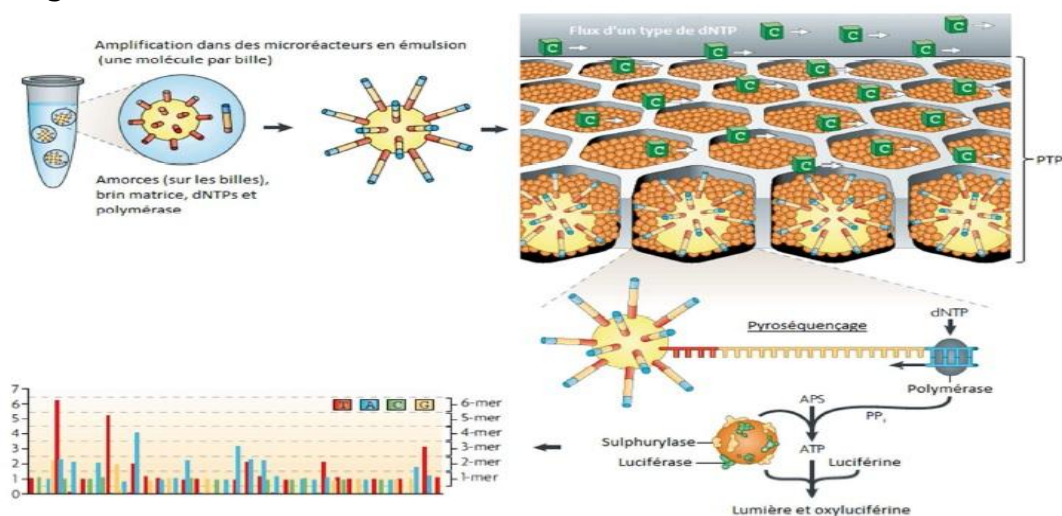


Figure 25 :

Aperçu de la technologie de séquençage 454

La spécificité de cette technologie repose sur une PCR en émulsion pour l'amplification des fragments à séquencer : la PCR a lieu dans une microgoutte renfermant une microbille d'agarose en phase aqueuse, séparée des autres billes (plusieurs millions) par de l'huile (voir Figure 25). On obtient ainsi des copies d'un seul fragment d'ADN par bille. Chacune des billes est ensuite déposée dans un des 1,6 millions de puits d'un support solide appelé PTP (PicoTiter Plate) [27].

Des réactions de pyroséquençage ont alors lieu à l'échelle du picolitre dans chaque puits : un flux de nucléotides (chaque nucléotide l'un après l'autre) traverse la PTP et lorsque l'un d'entre eux est incorporé par la polymérase, un pyrophosphate (PPi) est libéré. Deux enzymes, ici contenues dans un autre type de billes beaucoup plus petites, permettent alors une cascade enzymatique : en présence du PPi libéré, une sulphurylase convertit de l'adénosine phosphosulphate (APS) en adénosine triphosphate (ATP). L'ATP permet ensuite la conversion de luciférine en oxyluciférine par une luciférase. Cette dernière réaction produit de la lumière et l'intensité de la réponse lumineuse est directement proportionnelle à la quantité de nucléotides incorporés.

On obtient jusqu'à 900 Mb de données en une dizaine d'heures, soit 15000 fois plus qu'avec la première génération de séquenceurs. Cette technologie présente la plus grande taille de lecture (jusqu'à 700 pb par rapport à 100 pb à ses débuts) et sa grande précision la rend appropriée au séquençage de novo. Cependant, l'un de ses défauts, intrinsèque au pyroséquençage, reste la quantification des homopolymères.

Rothberg annonça en 2005 que cette technologie permettrait de séquencer le génome de James D. Watson pour seulement 1 million de \$. Le projet s'acheva en 2007 en respectant ce budget et Rothberg et ses collaborateurs publièrent l'un des premiers génomes humains complets [14] après avoir publié celui de Néanderthal en

2006 [28] [29] pour lequel on montrera cependant avoir majoritairement séquencé de l'ADN humain moderne contaminant [30].

## **1-2 Solexa/Illumina**

L'entreprise Solexa a commercialisé son premier séquenceur avec succès en 2006 : le Genome Analyzer (GA). La spécificité de cette technologie repose sur une amplification en pont (bridge PCR) des fragments à séquencer. Elle a lieu sur une surface de verre appelée flow cell (FC), similaire à une lame de microscope, divisée en huit lignes (à l'origine, une ligne par échantillon). Les fragments de la librairie à séquencer possèdent des adaptateurs à leurs extrémités. Ceux-ci vont leur permettre de se fixer de façon aléatoire sur la FC, par hybridation sur les amorces qui en couvrent la surface (voir Figure 26). Un nouveau brin est alors synthétisé par une polymérase (Figure 26a) : il est fixé de façon covalente à la FC. Le brin d'origine est alors éliminé par dénaturation (b) et l'extrémité libre du brin restant s'hybride à une amorce adjacente pour former un pont (c). La polymérase synthétise à nouveau le brin complémentaire pour former un pont d'ADN double brin (d) puis les deux copies sont libérées par dénaturation (e). Le cycle d'amplification en pont (étapes c à e) recommence pour former à terme un regroupement d'ADN clonai en une zone appelée cluster (f). Les brins anti-sens (correspondant aux amorces vertes) sont ensuite clivés (g) : c'est la linéarisation.

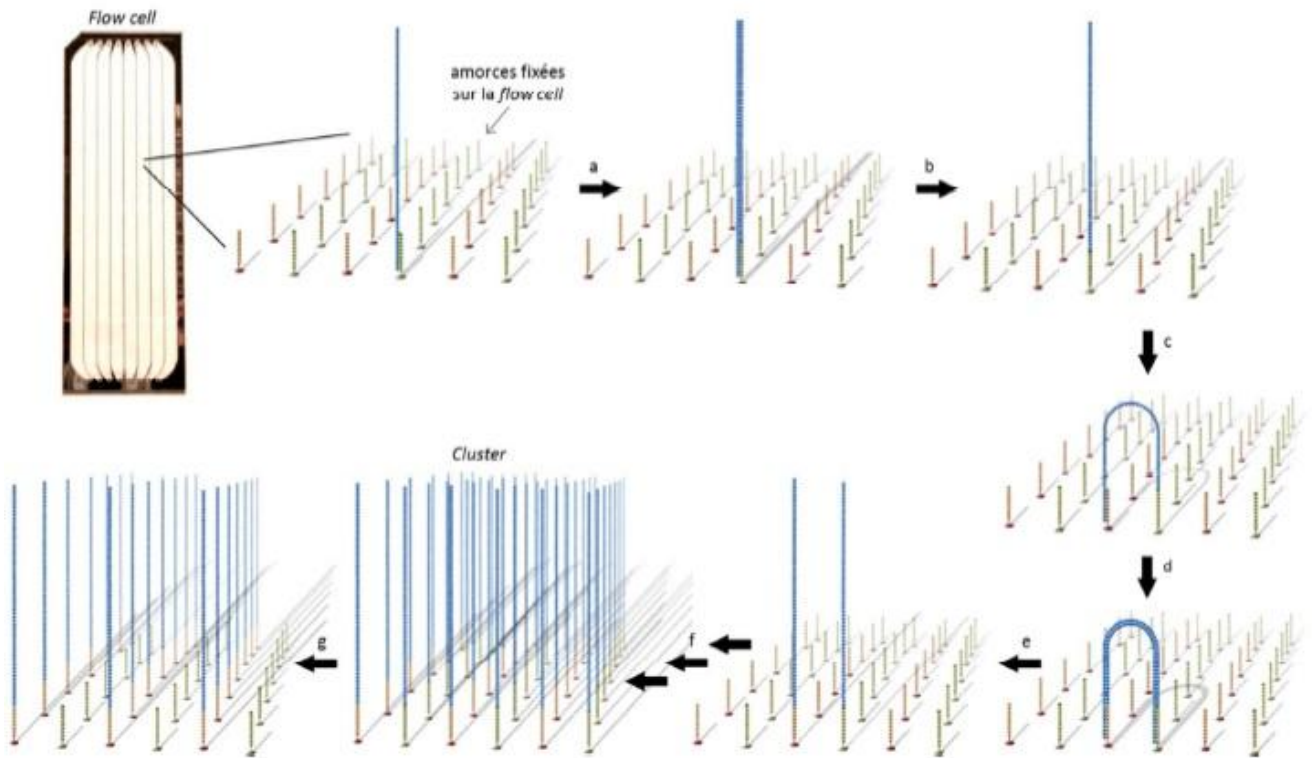


Figure 26 :

Amplification en pont de la technologie Illumina

L'extrémité 3' libre des fragments d'ADN est bloquée et l'amorce de séquençage s'y hybride (voir Figure 27). Le séquençage s'effectue sur des centaines de millions de clusters simultanément, grâce à une chimie de terminateurs réversibles : des nucléotides bloqués marqués par fluorescence sont ajoutés, l'un

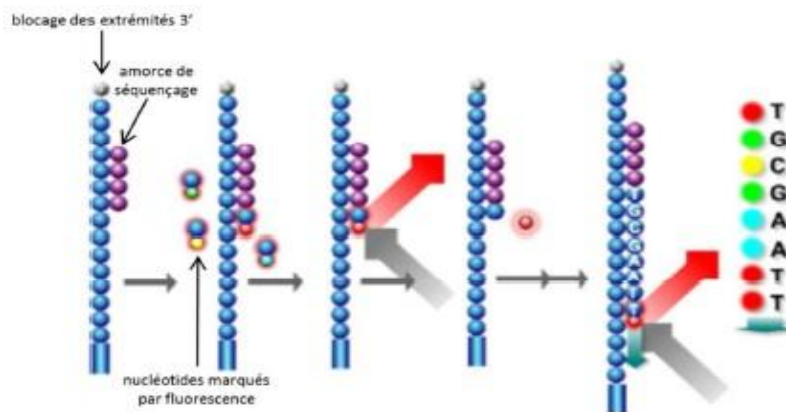


Figure 27 :

Aperçu de la technologie de séquençage Illumina

d'entre eux est incorporé, la fluorescence émise est relevée puis le fluorophore et le bloqueur sont clivés permettant l'ajout d'un nouveau nucléotide. A chaque cycle d'incorporation, une base peut être déterminée.

Cette chimie a pour avantage de séquencer correctement les homopolymères. Le Genome Analyzer a cependant pour inconvénient de lire peu de bases, (36 à ses débuts, jusqu'à 150 bases aujourd'hui), ce qui le rend toutefois approprié à l'analyse de génomes dont on a une bonne annotation. En novembre 2006, Illumina a acquis Solexa et est maintenant le leader sur le marché du séquençage, notamment avec le lancement en février 2010 de son HiSeq2000, à un prix de 690000 \$, qui peut produire jusqu'à 600 Gb en 10 jours, soit presque 700 fois plus que le dernier séquenceur de Roche, grâce à une méthode optique qui permet de travailler sur les deux surfaces de la FC et sur deux FCs en parallèle. Elle planifie désormais de l'optimiser pour aboutir mi-2012 au HiSeq2500 qui permettrait le séquençage d'un génome en un seul jour. Illumina possède la plateforme de séquençage la plus largement utilisée à travers le monde, en proposant un coût de séquençage du génome humain de seulement 10000 \$.

### 1-3 SOLiD

Le SOLiD (Sequencing by Oligonucleotide Ligation and Detection) a été la troisième plateforme de séquençage de nouvelle génération, commercialisée par Applied Biosystems (aujourd'hui Life Technologies) depuis 2007. La technologie

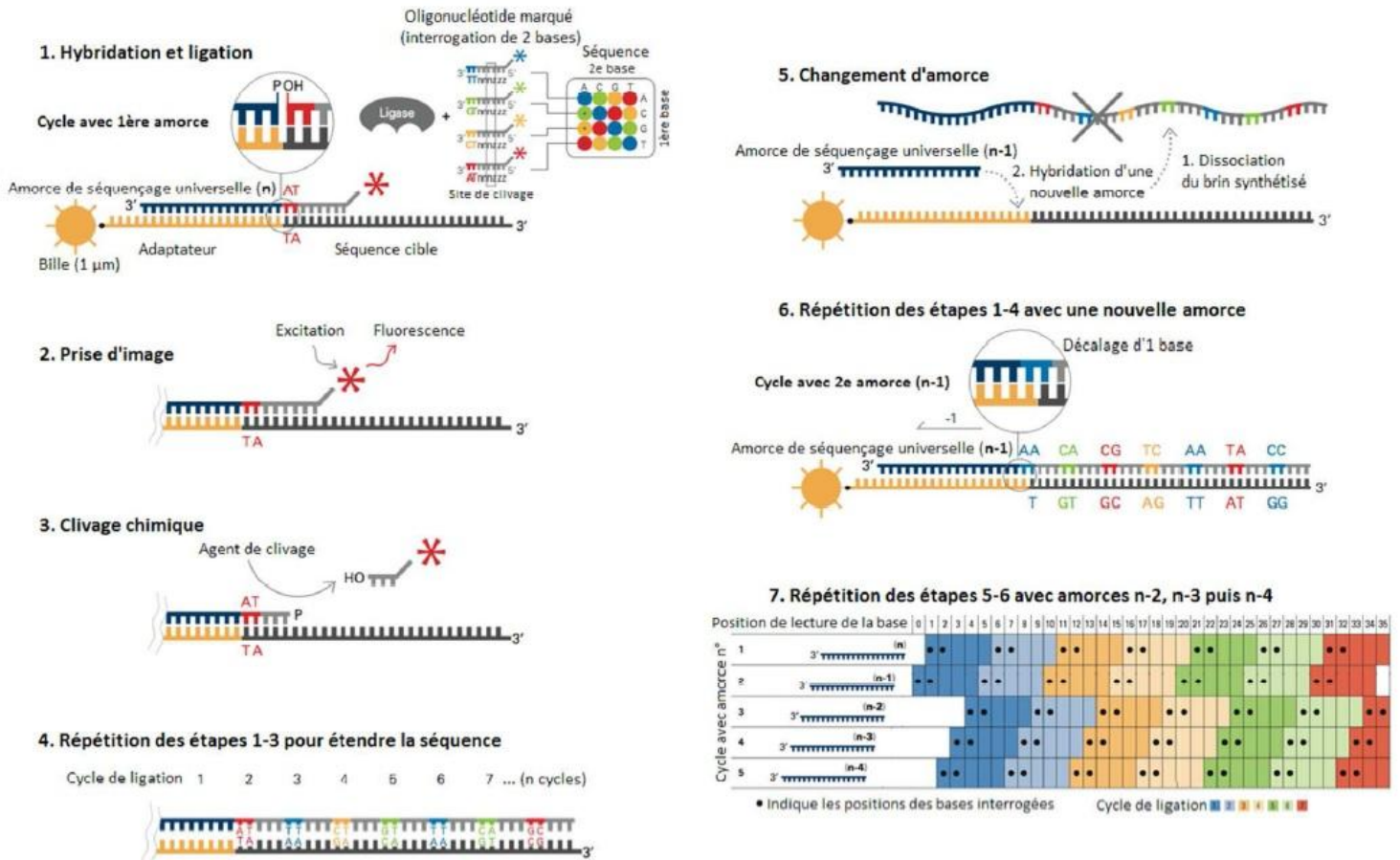


Figure 28 :

#### Aperçu de la technologie de séquençage SOLiD

repose également sur une PCR en émulsion sur billes. Le séquençage ne s'effectue pas par synthèse comme sur les plateformes précédentes mais par ligation. Une amorce de séquençage universelle se fixe sur l'adaptateur puis des oligonucléotides dégénérés de 8 bases, marqués par fluorescence, sont ajoutés (voir Figure 28).

Dès que l'un d'entre eux correspond à la séquence adjacente à l'amorce, la ligase le fixe et de la fluorescence est émise, permettant d'identifier l'oligonucléotide fixé et d'interroger ainsi ses deux premières bases. Un clivage chimique retire les

bases 6 à 8 ainsi que le fluorophore puis les oligonucléotides sont alors ajoutés à nouveau : on identifie les bases 6 et 7 de notre fragment à séquencer, puis dans un troisième temps les bases 11 et 12, et ainsi de suite jusqu'aux bases 31 et 32. Un deuxième cycle de ligation est alors entamé avec une amorce universelle se fixant en position  $n-1$  et on identifie les bases en position 0 et 1, 5 et 6, et ainsi de suite jusqu'aux bases 30 et 31. 3 nouveaux cycles de ligation sont effectués grâce à des amorces se fixant en  $n-2$ ,  $n-3$  puis  $n-4$ . Le nombre de cycles de ligation, détection et clivage détermine ainsi la longueur de lecture, de 35 bases dans le cas présenté ici à 75 bases. Chaque base est lue deux fois avec cette technologie, ce qui explique sa grande précision et qui la rend adaptée au reséquençage ou à l'analyse de polymorphismes. Néanmoins, la complexité de fonctionnement de cette technologie en est un inconvénient puisqu'elle implique un lourd travail d'analyse.

## **2- Les cadets de la seconde génération**

### **2-1 Ion Torrent**

En 2007, Jonathan M. Rothberg fonde la compagnie Ion Torrent, rachetée en 2010 par Life Technologies, qui commercialisera en décembre 2010 son premier séquenceur, le PGM (Personal Genome Machine) pour 50000 \$. Sa technologie repose sur la libération naturelle d'un ion  $H^+$  après incorporation d'un nucléotide par une polymérase. Ce phénomène entraîne une modification du pH pouvant être détectée par une puce de silicium semi-conductrice composée de plusieurs millions de transistors. Cette technologie est qualifiée de PostLight car aucun intermédiaire de lumière n'est utilisé contrairement aux méthodes citées précédemment : c'est une modification chimique qui entraîne la création du signal. L'absence de marquage fluorescent et de système de détection optique ou encore l'utilisation de micropuces standards explique le faible coût de cette machine qui pourrait avoir le même impact

sur les technologies que l'ordinateur personnel il y a une trentaine d'années, à condition d'en améliorer encore les taux d'erreurs existants.

La dernière version de ce séquenceur (puce 318) lui permet désormais de lire 1 Gb avec une grande précision en seulement deux heures et la longueur de lecture atteindra les 400 pb en 2012. Le PGM a ainsi servi à décoder le génome de la souche d'*Escherichia coli* qui a fait des ravages en Europe et provoqué un scandale sanitaire en mai 2011 [31]. L'équipe de Rothberg a ensuite utilisé cette technologie pour séquencer le génome de Gordon Moore, co-fondateur d'Intel, à une profondeur de 10,6x [32]. Life Technologies envisage le lancement d'un nouvel instrument pour mi-2012, le Ion Proton, qui permettrait le séquençage d'un génome en 1 jour pour 1000 \$.

## **2-2 Polonator G007**

Le Polonator G007 est distribué par Danaher Motions, une filiale du groupe Dover. C'est une plateforme libre d'accès grâce à laquelle les utilisateurs préparent eux-mêmes leurs réactifs en se basant sur les travaux publiés par le laboratoire de George Church [33]. L'accès aux protocoles et aux logiciels est gratuit. Cette technologie est basée sur une amplification en émulsion et sur un séquençage via une synthèse par ligation qui fonctionne de la même manière que sur le SOLiD mais seule une base est interrogée à l'aide d'oligonucléotides dégénérés.

## **2-3 Complete Genomics**

Complete Genomics a développé une technologie permettant de séquencer de faibles quantités d'ADN à faible coût en réactifs sur une lame de silicium. La librairie est préparée par RCA (Rolling Circle Amplification) qui permet de créer des nanoballs d'ADN [34]. Le séquençage s'effectue ensuite par ligation. Cette technologie a permis de séquencer 3 génomes humains à une profondeur de 45 à

87x pour 4400 \$ par génome. Cependant, Complete Genomics ne propose désormais que des services de séquençage haut-débit.

### 3- La « miniaturisation » des séquenceurs :

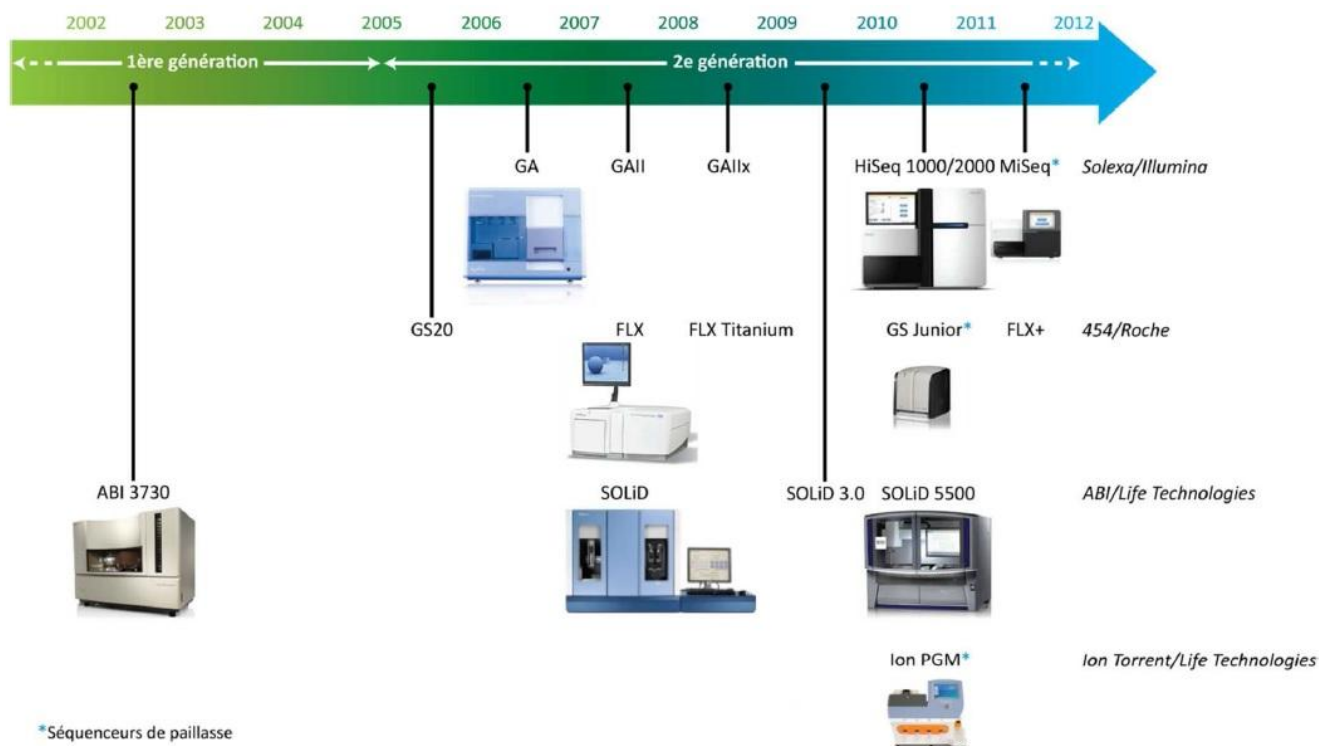


Figure 29 :

#### Chronologie de la commercialisation des séquenceurs haut-débit

En parallèle des différentes versions ou modèles existant pour ces séquenceurs (voir Tableau 3 et Figure 29) les entreprises développent des versions plus petites de leurs machines appelées séquenceurs de paillasse qui seront à la portée de plus petits laboratoires.

Roche a lancé en 2010 le GS Junior, commercialisé à 100000 \$ et capable de générer 35 Mb en seulement 10h. Illumina a présenté son MiSeq en janvier 2011 pour 125000 \$, capable de générer 1 Gb en 26h et dont les performances concurrencent directement le séquenceur de paillasse PGM de Life Technologies. Ce dernier est vendu plus de la moitié du prix d'un MiSeq mais les utilisateurs, la

plupart ayant une plus grande expérience de la technologie Illumina, préféreront opter pour un séquenceur Illumina.

## **II. Séquençage de 3eme génération**

Grâce aux avancées en microfluidique et en technologie des nanopores, la troisième génération est bientôt prête à émerger. La différence majeure entre la deuxième et la troisième génération de séquenceurs réside dans leur capacité à séquencer directement des molécules d'ADN de façon individuelle sans aucune amplification préalable. On peut donc également séquencer de l'ARN sans devoir le convertir au préalable en ADNc. Deux plateformes de séquençage de troisième génération ont été ou sont actuellement disponibles sur le marché et une dernière est en stade de développement avancé.

### **1 – HeliScope**

Helicos Biosciences a développé le premier séquenceur d'une molécule unique : le HeliScope Single Molecule Sequencer. A la différence de la chimie utilisée sur la plateforme Illumina, ici les nucléotides sont marqués avec le même fluorophore. Helicos ne vend désormais plus d'appareil et propose un service de séquençage.

### **2 – Technologies des nanopores**

Le séquençage par nanopores est une technologie prometteuse puisqu'elle permet de déterminer une séquence ADN à la résolution du nucléotide sans aucune amplification, par lecture directe. Les différentes bases et leur statut de méthylation peuvent être déterminés en temps réel grâce au courant qui traverse le pore, avec une grande précision (99,8%) et à faible coût puisqu'aucun système optique n'est nécessaire ainsi que peu de traitement des données [35]. La longueur de lecture est aussi plus grande que celle des autres technologies.

Une molécule unique d'ADN traverse un pore formé par une protéine ancrée dans une bicouche lipidique par application d'un potentiel (voir Figure 30). Une

exonucléase clive chaque nucléotide à l'entrée du nanopore et celui-ci est alors détecté de façon électronique via une cyclodextrine [36]. Aucun traitement ni marquage préalables ne sont nécessaires et de faibles quantités d'ADN suffisent.

Le deuxième point fort de cette technologie est sa capacité à identifier les modifications épigénétiques puisqu'une 5-méthylcytosine peut être différenciée d'une cytosine.

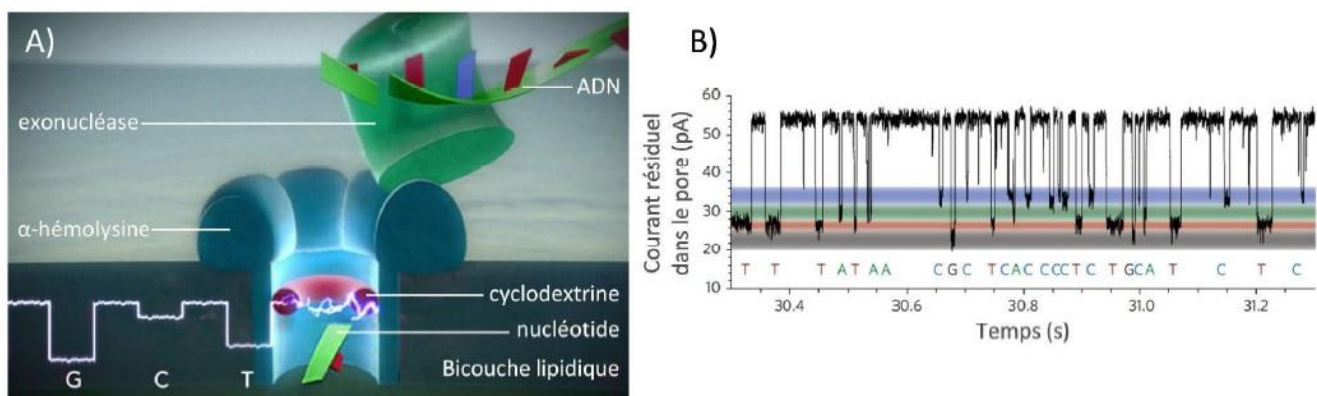


Figure 30 :

### Détection de nucléotides par l'utilisation d'un nanopore

A) Structure utilisé pour le séquençage d'une molécule d'ADN unique

B) Courant résiduel mesuré à travers le nanopore permettant de discriminer chaque nucléotide

En 2010 certains groupes ont mis en œuvre 2 principaux types de pores protéiques : le laboratoire de HaganBayley a d'abord optimisé un pore à partir d'une protéine mutée d' $\alpha$ -hémolysine A de staphylocoque. Un groupe de l'université de Washington a ensuite développé un nanopore à partir de la porine A de *Mycobacteriumsmegmatis* ou MspA supposée avoir de meilleures propriétés grâce à sa plus petite taille [37]. Oxford Nanopore Technologies, dont les travaux se basent sur ceux de HaganBayley, optimise également différents nanopores. Certaines équipes mettent aussi au point des nanopores à base de silicium ou encore de graphène qui est un matériau conducteur et qui a pour avantage de servir de pore et d'électrodes à la fois [38]. Enfin on a vu récemment le développement de nanopores hybrides grâce à des protéines intégrées dans une membrane à l'état solide [39]. Les

défis actuels de ces technologies restent la gestion d'un brin d'ADN qui se meut trop rapidement à l'entrée ou à l'intérieur du pore, et de prouver qu'elles fonctionnent sur de longs fragments de 1 à 5 kb. Aucune machine n'est commercialisée à l'heure actuelle, mais Oxford Nanopore Technologies vient d'annoncer le lancement de son instrument GridION pour 2012 ainsi que d'un système de la taille d'une clé USB, le MinION dont le coût ne devrait dépasser les 900 \$.

### **3- Pacific Biosciences**

Pacific Biosciences a développé le premier instrument capable de séquencer une molécule unique en temps réel, là encore par lecture directe et sans amplifier le matériel de départ : il s'agit du PacBio RS. Il utilise une structure composée de cellules SMRT (Single Molecule Real Time) [40] [41]. Chacune de ces cellules contient 75000 nanostructures appelées détecteurs ZMW (Zero-Mode Waveguide) de 100 nm de diamètre, c'est-à-dire plus petits que les longueurs d'onde utilisées sur la plateforme (532 et 643 nm). La lumière ne peut donc pas s'y propager, d'où le terme de mode-zéro.

Chacun de ces ZMW contient une polymérase qui y est immobilisée et qui incorpore des nucléotides liés à un fluorophore, libérant ainsi celui-ci à l'extérieur du ZMW et mettant fin au signal lumineux. L'intervalle de temps entre chaque pic ainsi que la durée de chaque pic de fluorescence sont propres à chaque nucléotide et permettent ainsi leur identification. Ces paramètres sont également différents pour des nucléotides méthylés et il est donc possible, sans traitement préalable, de différencier une cytosine d'une méthylcytosine ou encore d'une hydroxyméthylcytosine[42] [43].

La longueur de lecture moyenne est de plus de 1000 pb, la préparation de l'échantillon dure environ 30 minutes et les données sont produites en quelques

minutes seulement, ce qui en fait une des technologies les plus prometteuses des années à venir.

#### **4- Starlight**

Life Technologies développe un séquenceur de molécule unique appelé Starlight qui utilise des quantum dots, des particules nanométriques à base de semi-conducteur et dont les propriétés de fluorescence peuvent être contrôlées par leur taille. Excitées par un laser, ces particules transfèrent leur énergie aux nucléotides marqués par fluorescence lorsque ceux-ci sont incorporés par la polymérase [44]. De plus, il est possible de remplacer la polymérase lorsqu'elle s'essouffle. On peut ainsi théoriquement séquencer la longueur désirée de fragment. En l'absence de communication de la part de Life Technologies, le doute subsiste quant à une éventuelle commercialisation de cet appareil.

### **III. La génomique personnelle :**

Plusieurs compagnies offrent désormais des services de génomique personnelle. Elles s'adressent directement au grand public et proposent de décoder son génome à des prix relativement abordables. Il s'agit en réalité d'étudier des polymorphismes impliqués dans un panel de maladies. C'est le cas de 23andMe, deCODEme, Navigenics et bien d'autres qui proposent pour parfois moins de 100 \$ de prédire votre avenir génétique. D'autres grands du séquençage proposent cette fois-ci de séquencer votre génome complet. C'est le cas de Complete Genomics ou encore de Knome qui propose de vous révéler les secrets de votre ADN pour 5000 \$.

Tous les séquenceurs déjà commercialisés ont été certifiés pour une utilisation dans des laboratoires de recherche mais les scientifiques tentent de plus en plus de les orienter vers une utilisation clinique grâce à laquelle les médecins pourraient traiter leurs patients de façon personnalisée en leur fournissant des traitements adaptés. Il faut cependant désormais donner un sens biologique à ces quantités

pharaoniques de données produites depuis plusieurs années avant d'espérer entrer dans cette nouvelle ère de la biologie. Le Personal Genome Project lancé en 2009 recrute des volontaires prêts à donner leur génome à la science. Il a pour but de séquencer 100000 génomes afin de construire une base de données qui permettrait de mieux les appréhender. De nouveaux challenges informatiques sont alors lancés [45] afin de gérer ces quantités de données, les interpréter de façon correcte mais également les protéger pour assurer une confidentialité aux patients. De nombreux problèmes éthiques sont désormais soulevés [46] [47] et la communauté scientifique devra rapidement prendre ce sujet en main pour encadrer au mieux l'utilisation des données de génomique personnelle.

#### **IV. Les applications du séquençage à haut débit**

Le champ d'application de ces nouvelles méthodes de séquençage est très vaste. En effet à partir du moment où il est possible d'obtenir une molécule d'ADN, le séquençage à haut débit peut être utilisé. Grâce à la capacité de ces machines à fournir de grandes quantités de séquences ou de travailler sur un grand nombre d'échantillons en parallèle, ces outils permettent de couvrir plusieurs technologies différentes employées jusqu'à maintenant comme les puces à ADN, le séquençage classique de Sanger ou la PCR quantitative à haut débit. Les applications disponibles avec le séquençage à haut débit peuvent globalement se regrouper en 3 grandes catégories : le séquençage de novo, le reséquençage et les analyses fonctionnelles.

##### **1 – Le séquençage de novo**

La première application qui a été mise en avant grâce à ces nouvelles technologies de séquençage, ce sont leurs capacités à remplacer le séquençage Sanger traditionnel et donc à fournir la séquence de génomes inconnus. Pour réussir à obtenir des versions de génome de bonne qualité il est souvent nécessaire de combiner plusieurs méthodes. Le pyroséquençage permet grâce à des lectures

longues de construire une première version du squelette du génome quand les méthodes par terminateurs réversibles vont corriger les erreurs présentes dans cette première reconstruction pour produire un brouillon du génome de qualité. Dans le domaine médical ces outils sont utilisés pour la découverte de génomes d'agents pathogènes inconnus ou de nouveaux virus [48].

## **2- Le reséquençage**

Toujours dans le domaine où la connaissance de la séquence est importante, le reséquençage est utilisé quand la séquence du génome de référence est déjà connue. Le séquençage à haut débit est alors employé pour connaître quelles sont les variations génomiques de l'échantillon qui est étudié en comparaison avec celui pris comme référence. Ces approches sont certainement parmi les plus employées actuellement dans le domaine médical. Ces outils ont typiquement pour vocation de remplacer les méthodes traditionnelles d'hybridation génomique comparative (CGH) et de permettre de préciser les diagnostics soit de façon préventive soit pour caractériser une pathologie déjà déclarée. De nombreux articles ont été publiés dans ce domaine et le foisonnement des études d'associations en cours démontre de l'engouement réel du monde médical pour ces approches. Il est ainsi possible par exemple d'effectuer un diagnostic prénatal non invasif pour certaines maladies génétiques [49], de détecter les variants associés à des maladies génétiques [50] [51] [52] [53] [54], de typer ou de suivre l'évolution des tumeurs cancéreuses chez les patients [55] [56] [57]. Dans le domaine de la toxicologie, on peut imaginer que le séquençage permettra non seulement de suivre les effets mutagènes potentiels des molécules thérapeutiques sur les tissus qui ne sont pas les tissus cibles, mais on sera également capable de décrire précisément les modifications génomiques induites.

Le reséquençage peut également être utilisé pour caractériser les différentes souches d'agents pathogènes. Dans le cas d'études épidémiologiques, le séquençage à haut débit va permettre des gains de temps très importants par rapport aux méthodes traditionnelles [58], ce qui peut s'avérer crucial en situation d'urgence. L'exemple de la découverte de l'origine sud asiatique de la souche de choléra qui a infecté Haïti en 2010 grâce au séquençage à haut débit, a ainsi mis en évidence que la contamination provenait d'une région géographique éloignée probablement due aux activités humaines [59].

Enfin le dernier domaine d'application du séquençage à haut débit, à la frontière entre reséquençage et séquençage de novo, concerne la métagénomique. Le but est de découvrir dans un mélange complexe l'ensemble des organismes qui le composent comme par exemple la flore intestinale [60]. On peut ainsi imaginer qu'en séquençant le sérum de patients, il sera possible de déterminer quels sont les agents pathogènes qui y sont présents et qui sont à l'origine de la pathologie observée. Il pourrait aussi être possible d'envisager en toxicologie le suivi de l'incorporation de vecteurs viraux dans différents tissus d'un organisme en dehors de ceux ciblés par le vecteur de thérapie génique.

### **3- Les analyses fonctionnelles**

Le dernier champ d'application des méthodes de séquençage à haut débit concerne la génomique fonctionnelle. Dans ce domaine, l'important n'est pas de connaître la séquence d'ADN des échantillons, mais plutôt de quantifier le nombre et le type d'éléments biologiques présents.

Il est par exemple possible de :

Déterminer quelles quantités et quels types d'ARN s'expriment dans une cellule (RNA-Seq) [61], voire sélectivement les ARNs messagers fixés sur les ribosomes et donc en cours de traduction (traductome) [62].

Détecter les phénomènes de transcription qui sont utilisés pour la régulation à travers les longs ARN non codants et toute la famille des petits ARN (Small RNA-Seq).

Connaître les régions de l'ADN où se fixent les facteurs de transcription (ChIP-Seq) [63] et de déterminer les modifications épigénétiques d'un génome en cartographiant ses sites de méthylation (MeDIP-Seq) [64].

Ces applications vont concurrencer de plus en plus les méthodes traditionnelles basées sur les puces à ADN.

Le séquençage à haut débit permet en effet de s'affranchir des biais de l'hybridation, il dispose d'une gamme dynamique plus grande et la résolution des résultats obtenus est disponible à la base près. Ces outils sont donc intéressants pour déterminer les effets des molécules thérapeutiques lors de test in vitro sur des cultures de cellules, in vivo sur des modèles animaux ou lors des essais cliniques. Le fait de pouvoir soit être exhaustif sur l'ensemble de toutes les réponses génomiques possibles, soit de cibler une région précise et de passer un grand nombre d'échantillons en parallèle est un atout important pour la caractérisation et le criblage de biomarqueurs [65].

# **MATERIELS ET METHODES**

## **I. Prélèvement :**

Avant tout prélèvement, une fiche qui porte les renseignements, l'accord et le consentement de l'intéressé est obligatoire. On prélève deux tubes EDTA de sang, pour l'extraction d'ADN par kit ou par sel. Les tubes peuvent être conservés à +4°C ou -20°C pour utilisation ultérieure.

## **II. Techniques de préparation et de purification des acides nucléiques :**

L'extraction de l'ADN (acide désoxyribonucléique) est une technique qui isole de l'ADN à partir d'une cellule en quantité et en qualité suffisante pour permettre son analyse. Au niveau du laboratoire de Génétique médicale et d'oncogénétique, l'extraction se fait selon deux méthodes : soit par SEL ou par KIT commercialisé (invitrogèn). Nous avons utilisé l'extraction par Kit pour les sangs anciens pour obtenir une meilleure qualité de l'ADN.

### **1- Par sel :**

#### **1-1 Principe :**

Le sang doit être initialement et vigoureusement mélangé à une solution hypotonique pour faire éclater les globules rouges. Le lysat est centrifugé et après élimination du surnageant, le culot cellulaire, contenant les leucocytes, est traité par une solution de lyse des Globules blancs (SLB) contenant la protéinase K. (enzyme qui digère les protéines cellulaires). Les protéines seront par la suite éliminées par l'intermédiaire d'une force ionique du NaCl (5 M) et la précipitation de l'ADN génomique est effectuée en utilisant une solution d'éthanol absolu à froid (-20 °C).

#### **1-2 Réactifs nécessaires :**

- ✓ Tris-EDTA 20 mM / 5 mM

- ✓ Tris-EDTA 10 mM / 1 mM
- ✓ SLB à +42°C
- ✓ Protéinase K 10 mg/ml
- ✓ Na Cl 5 M
- ✓ Ethanol 75 % à +4°C
- ✓ Eau Distillé stérile

### **1-3 Protocole expérimental :**

L'extraction d'ADN est réalisée à partir du sang total, conservé dans un tube EDTA et stocké à -20°C. Une décongélation de l'échantillon est donc nécessaire.

- ✓ Lyse des globules rouges

Une fois décongelé, le sang est récupéré dans un tube de 15 ml, traité avec deux volumes du tampon d'extraction TE 20/5 (V/V), puis incubé dans la glace pendant 20 mn. Ces étapes permettent la lyse des globules rouges. Après incubation, une centrifugation à 2500 tr/mn pendant 5 mn est réalisée, et le surnagent est ensuite éliminé.

Les étapes précédentes sont répétées (ajout de deux volumes de TE 20/5, incubation....) jusqu'à l'obtention d'un culot blanc ne renfermant que les globules blancs.

- ✓ Lyse des globules blancs

Le culot est suspendu dans 3 ml de SLB (Solution de lyse des globules blancs), puis 100 µl de protéinase K (10mg/ml) sont additionnés. L'incubation dure une nuit sous agitation douce à 42°C.

- ✓ Dénaturation et précipitation des protéines et impuretés

Après l'incubation, 4 ml d'eau distillée stérile et 4 ml de NaCl 5 M sont ajoutés, le tube est par la suite homogénéisé, puis centrifugé à 3000 tr/mn pendant 30 mn.

✓ Précipitation et lavage de l'ADN

Le surnageant est récupéré dans un nouveau tube de 50 ml, auquel deux volumes d'éthanol absolu froid sont ajoutés. Après homogénéisation, "la méduse" d'ADN se forme dans le tube. Cette dernière sera récupérée dans un tube eppendorff de 1.5 ml, puis lavée à l'éthanol 75 %.

Après évaporation pour éliminer toute trace de l'éthanol, l'ADN est dilué dans 200 µl de TE (Tris-EDTA) puis stocké à +4°C pour toute utilisation ultérieure (pour un stockage prolongé il est préférable de conserver l'ADN à -20°C).

## 2- Par Kit :

### 2-1 Principe :

Après digestion enzymatique et solubilisation des lipides dans une solution spécifique, l'ADN chargé négativement est mis en contact avec des billes magnétiques chargées positivement afin qu'il s'accroche à celles-ci. La solution est éliminée alors que les billes sont maintenues à l'aide d'un aimant.

Une autre solution de composition différente est ensuite ajoutée aux billes afin de neutraliser leur charge. L'ADN est alors libéré dans cette seconde solution.

### 2-2 Protocole expérimental :

Notez : Pour 2 ml des échantillons de sang, employez les volumes de réactif indiqués pour 3 ml des échantillons de sang.

1. Vortexer le tube contenant les billes Magnétiques GeneCatcher™ pour resuspendre et distribuer les billes dans le tampon de stockage.

2. Ajouter les billes Magnétiques resuspendu et le tampon de Lysis (L13) à un tube 50 ml stérile. Mélangez les billes en tourbillonnant doucement le tube jusqu'à ce que les billes soient distribuées.

3. Ajouter 2–10 ml du sang bien mélangé au tube contenant les billes et inversez doucement le tube étonné 3 fois pour mélanger les billes.

4. Incuber à la température ambiante pendant 5 minutes pour permettre à l'ADN de se lier aux billes. Pendant l'incubation, agiter doucement le tube de temps en temps.

5. Placer le tube de 50 ml sur le Support de Séparateur Magnétique pendant 3 minutes.

6. Sans enlever le tube du Séparateur Magnétique, éliminer soigneusement le surnageant en utilisation une pipette de 5 ml sans toucher les billes.

7. Enlever le tube du Séparateur Magnétique.

8. Ajouter 5 ml du tampon de Lysis (L13) au tube sans tenir compte du volume du sang et inversez doucement le tube 3 fois pour mélanger les billes et enlever les contaminations résiduelles par lavage.

9. Incuber à la température ambiante pendant 30 secondes.

10. Placer le tube sur le Séparateur Magnétique pendant 20 secondes.

11. Sans enlever le tube du Séparateur Magnétique, éliminer le surnageant en utilisation une pipette de 5 ml sans toucher les billes.

12. Passer immédiatement à la Digestion Protéase.

### **Protéase Digestion**

1. Mettre le bain marie à 65°C.

2. Enlever le tube contenant les billes magnétiques du Séparateur Magnétique

3. Ajouter le tampon de Protéase et la Protéase au tube et vortexer le tube pour disperser les billes (30 secondes).

Notez : Pour 2 ml des échantillons de sang, employez les volumes de réactif indiqués pour 3 ml des échantillons de sang.

4. Incuber le tube à 65°C pendant 10 minutes et incuber ensuite à température ambiante (10–20 minutes).

5. Inversez doucement le tube deux fois pour resuspendre billes.

6. Passer immédiatement au lavage d'ADN.

### **Lavage d'ADN**

1. Ajouter l'isopropanol (IPA) 100 % au tube et mélanger en inversant doucement le tube jusqu'à l'apparition d'une forme d'agrégation visible.

Note : L'absence d'agrégation après 2–3 minutes indique un très faible taux d'ADN dans l'échantillon.

2. Placer l'échantillon sur le Séparateur Magnétique pendant 30 secondes à 1 minute.

3. Sans enlever le tube du Séparateur Magnétique, éliminer le surnageant en utilisant une pipette de 5 ml sans toucher les billes.

4. Enlever le tube du Séparateur Magnétique.

5. Ajouter 3 ml 50 % (v/v) d'isopropanol sans tenir compte du volume de sang au tube et le mélanger en inversant doucement le tube 5 fois.

6. Placer l'échantillon sur le Séparateur Magnétique pendant 30 secondes.

7. Sans enlever le tube du Séparateur Magnétique, éliminer le surnageant en utilisant une pipette de 5 ml sans toucher les billes.

8. Garder le tube sur le Séparateur Magnétique pendant une minute pour permettre au reste du liquide d'arranger au fond du tube. Eliminer le surnageant en utilisant une pipette de 1 ml.

9. Sans enlever le tube du Séparateur Magnétique, ajoutez 250 µl du tampon Lavant (W12, sans tenir compte du volume de sang) en face de la boulette de billes.

10. Incuber pendant 1 minute à température ambiante.

11. Sans enlever le tube du Séparateur Magnétique, éliminer le surnageant en utilisant une pipette de 1 ml sans toucher les billes.

12. Répéter des étapes 9-11.

13. Passer immédiatement à l'élution d'ADN.

### **L'élution d'ADN**

1. Mettre le bain marie à 65°C.

2. Enlever le tube contenant les billes magnétiques du Séparateur Magnétique (étapes 12).

3. Ajouter 1 ml du tampon d'Elution (E5), sans tenir compte du volume de sang, au tube et agitent doucement l'échantillon en tourbillonnant le tube pour sortir la boulette de bille.

Pour les échantillons avec haut contenu d'ADN

4. Incuber à 65°C pendant 1 heure.

5. Enlever le tube du bain marie et mixer par une pipette de 800 µ l, pour disperser la boulette complètement.

Notez : Pour le haut contenu d'ADN, incubez les échantillons pendant 20 minutes complémentaires et répéter l'étape 5.

6. Placer le tube sur le Séparateur Magnétique avant que le surnageant ne soit clair et sans couleur (d'habitude 15 minutes à 1 heure, les échantillons visqueux exigent plus de temps).

7. Sans enlever le tube du Séparateur Magnétique, enlever soigneusement le surnageant contenant l'ADN en employant la pipette 1 ml sans toucher la boulette de billes et transférer le surnageant à un tube stérile.

8. éliminer les billes magnétiques employées.

### **Le stockage de l'ADN**

9. Stocker l'ADN épuré à  $-20^{\circ}\text{C}$  ou employer l'ADN pour les applications désirable en aval.

10. Pour éviter la congélation répétée d'ADN, stockez l'ADN épuré à  $4^{\circ}\text{C}$  pour l'utilisation immédiate ou l'aliquoter et le stocker à  $-20^{\circ}\text{C}$ .

### Dosage de l'ADN :

La concentration en ADN de l'échantillon est estimée par spectrophotométrie.

En effet, les acides nucléiques présentent un pic d'absorption dans l'ultraviolet. Le maximum de cette absorption se situe à 260nm. L'interférence par des contaminants se reconnaît par calcul d'un « ratio ». Les protéines absorbant à 280 nm, le ratio  $A_{260}/A_{280}$  est utilisé pour estimer la pureté de l'acide nucléique. L'ADN pur devrait avoir un ratio d'environ 1,8, tandis que l'ARN pur devrait avoir une valeur d'environ 2,0. L'absorption à 230 nm reflète la contamination de l'échantillon par des substances telles que les hydrates de carbone, les peptides, les phénols ou les composés aromatiques. Dans le cas d'échantillons purs, le ratio  $A_{260}/A_{230}$  devrait être d'environ 2,2.

Au laboratoire, le « NANODROP » (figure 11) est utilisé pour réaliser le dosage. Il suffit de déposer dans l'appareil  $2\mu\text{l}$  d'ADN extrait pour déterminer sa concentration et obtenir la courbe qui renseigne sur sa pureté.

## III. Amplification des séquences nucléotidiques par PCR :

### 1- Principe :

La réaction PCR (Polymerase Chain Reaction) permet d'amplifier in vitro une région spécifique d'un acide nucléique donné afin d'en obtenir une quantité suffisante pour le détecter et l'étudier. Pour se faire, une série de réactions permettant la réplique d'une matrice d'ADN double brin est répétée en boucle.

Ainsi, au cours de la réaction PCR, les produits obtenus à la fin de chaque cycle servent de matrice pour le cycle suivant, l'amplification est donc exponentielle.

Pour avoir la réplication d'un ADN double brin, il faut agir en trois étapes :

(1) dénaturation de l'ADN pour obtenir des matrices simple brin.

(2) hybridation des amorces spécifiques.

(3) réaliser la réaction de polymérisation du brin complémentaire par l'enzyme polymérase. A la fin de chaque cycle, les produits sont sous forme d'ADN double brin.

## **2- Condition de PCR :**

Pour procéder à l'amplification d'une séquence d'ADN, Les PCR ont été effectuées et optimisées sur l'ADN génomique préalablement dilué à 100 ng/ $\mu$ l en utilisant des amorces spécifiques pour chaque région étudiée.

Les réactions PCR ont été effectuées et optimisées pour chaque région cible dans un volume final de 20 $\mu$ l pour le mélange réactionnel. Et enfin, le mélange est déposé au niveau d'un thermocycleur (AppliedBiosystem) selon un programme déterminé.

Pour chaque région cible, les amorces sens et antisens sont indiquées, ainsi que la taille attendue pour chaque produit amplifié.

## **3- Témoins de la réaction PCR :**

Les témoins inclus pour chaque réaction PCR comprennent :

- Témoins positif qui a pour fonction de s'assurer des bonnes conditions de la PCR, ce témoin est représenté par l'ADN d'un sujet normal sain.

- Témoin négatif qui permet de s'assurer de l'absence de contamination, auquel il n'y a pas d'ADN.

#### 4- Contrôle des produits PCR :

Les produits PCR sont contrôlés sur gel d'agarose 2 % en présence d'un marqueur de taille, qui permet de vérifier la taille et la spécificité du produit amplifié.

### IV. Séquençages d'ADN :

Avant de séquencer nos produits PCR, il faudra les purifier pour éliminer l'excès d'amorces, dNTP et d'ADN polymérase, pour cela il 3 réactions essentielles comme suit :

#### 1- Purification par ExoSAP® :

ExoSAP-IT est conçu pour une purification rapide et efficace des produits de PCR. Il comporte deux enzymes hydrolytiques, l'exonucléase (Exo) qui dégrade les ADN simples brins et la phosphatase alcaline de crevette (SAP) qui hydrolyse les dNTPs libres, et en excès sans interférence avec les applications en aval. L'enzyme est active à 37°C et inactive à 80°C. Les fragments simples brins inférieurs à 100 pb sont ainsi dégradés.

- On utilise 8 µl de l'enzyme ExoSAP avec 10 µl du produit PCR.
- On le place au niveau du thermocycleur suivant le cycle suivant :

Température	Durée
37°C	3 mn
80 °C	15 min
10°C	∞

Tableau 4 : Cycle utilisé au niveau du thermocycleur

## **2- Reaction de sequence par BigDye® Terminator v 3. 1 Cycle**

### **Sequencing :**

Selon la réaction de Sanger qui repose sur l'incorporation alléatoire de didéoxynucléotides interrupteurs de chaîne (ddNTP) eux aussi présents dans le milieu réactionnel dont chacun est marqué par un fluorophore dont le spectre d'émission est spécifique. Une analyse spectrale va différencier les différents fluorochromes, associer la base correspondante et donc définir la séquence nucléotidique du brin d'ADN initial. Les fragments d'ADN synthétisés portent ce fluorophore terminal. On les appelle des terminateurs d'élongation ou "BigDyeTerminators" ou "Dye-labeledterminator".

Nous utilisons la technologie BigDyeTerminator\_version 2 (AppliedBiosystems : ABI).

La technologie BigDye Terminator (BDT) utilise un système de transfert d'énergie par résonance (FRET) entre deux fluorochromes fixés sur le même ddNTP et reliés entre eux par un linker. Le premier est une fluorescéine (6 carboxyfluorescéine) appelé fluorochrome donneur, commun aux quatre ddNTP. Le second est une dichlororhodamine (dRhodamine) qui joue le rôle de fluorochrome accepteur.

Le fluorochrome donneur est excité par un rayon laser à argon émettant à 488 nm et 514,5 nm. Son énergie de fluorescence émise (515–520 nm) est captée intégralement par le fluorochrome accepteur qui est excité à son tour. Le fluorochrome accepteur ou dichloroRhodamine est différent pour chaque type de ddNTP.

Le spectre de la fluorescence réémise sera ainsi spécifique de chaque type de ddNTP. Le transfert du signal de la fluorescéine vers la dRhodamine permet une

amplification du signal et par conséquent, une augmentation de la sensibilité de la technique (figure 31).

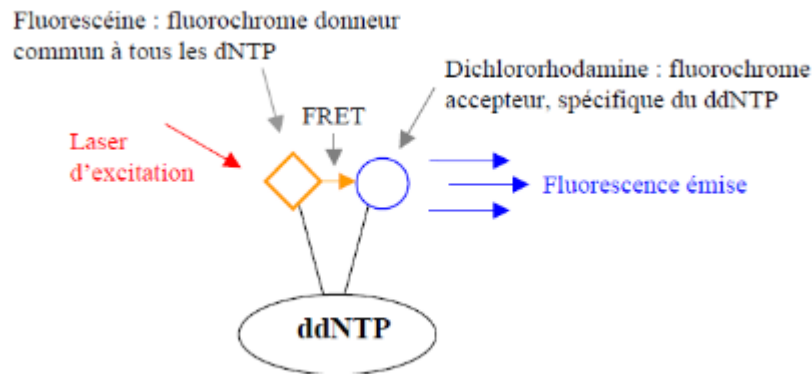


Figure 31 :

Mécanisme de la fluorescence par la technique du transfert d'énergie par résonance.

On effectue 30 cycles d'amplification sur le thermocycleur dans des conditions spécifiques à la réaction de séquence. Un cycle correspond à :

Une étape de dénaturation de l'ADN à 95°C pendant 10 secondes pour obtenir l'ADN sous forme simple brin.

Une étape d'hybridation à 50°C pendant 5 secondes (température moyenne d'hybridation pour la majorité des amorces utilisées).

Une étape d'élongation de l'ADN par la Taq polymérase à 60°C pendant 4 minutes. Cette température faible ralentit la Taq et va déplacer l'équilibre pour permettre une meilleure incorporation des ddNTP. Une fois l'étape de thermocyclage effectuée, il est nécessaire de purifier la réaction de séquence.

### 3- Purification de la réaction de séquence avec le Kit BigDye-

#### XTerminator :

Le kit BigDye® Xterminator™ permet la purification des produits de réaction de séquence en capturant les dyes non incorporés dans la réaction, les sels et autres

molécules chargées qui pourraient interférer lors de la détection des bases par électrophorèse capillaire.

Le kit contient : – un flacon “SAM Solution” (stockage à T° ambiante)

– un flacon : « XTerminator Solution (stockage à 4°C, ne pas congeler)

1. Déposer 10µl de produits de réaction de séquence dans la plaque Applied.

2. Centrifuger la plaque.

3. Préparer un mix contenant :

45 µl de solution SAM par puits

10 µl de XTerminator Solution par puits à l'aide des cônes spécifiques.

– Vortexer 10 sec le flacon XTerminator Solution et prélever immédiatement la quantité nécessaire.

4. Vortexer le mélange et ajouter 55µl dans chaque puits à l'aide de la pipette distritips ou de la pipette monocanal.

5. Sceller la plaque correctement avec un film adhésif

6. Mettre la plaque sur l'agitateur Eppendorf Mix Mate.

7. Lancer l'agitateur sur la position 1800 rpm pendant 30secondes et vérifiez que tout se mélange bien et que la plaque reste bien attachée.

8. Agiter à 1800 rpm pendant 30 minutes.

9. Centrifuger la plaque 2 minutes à 1000 g (Centrifugeuse HeraeusMultifuge 3S-R).

Les plaques peuvent être conservées 7 jours à 4°C (pas congelées) avant d'être analysées par le séquenceur.

#### **4– Détermination de la séquence :**

La plaque est placée dans le séquenceur (figure 14) qui est un automate d'électrophorèse capillaire. Ce dernier lance un flux électrique d'ions à travers un capillaire, ce qui entraine la migration des fragments d'ADN. Une fois arrivés au site

de détection, les quatre fluorochromes des ddNTP terminaux seront excités. Suite à cette excitation, chaque fluorochrome émettra une lumière à une longueur d'onde différente qui sera détectée puis convertie en séquence par le logiciel d'analyse des séquences.

Les séquences d'ADN sont déterminées par séquençage automatique (séquenceur à 8 capillaires appliedbiosystem 3500 DX)

## **V. Les outils de bio-informatique :**

### **1- Logiciel BLAST (nucléotide BLAST) :**

Le logiciel blast permet de comparer une séquence nucléique dite requête à une banque de séquences nucléiques sur les 2 brins, c'est-à-dire la séquence étudiée (brin +) et son complémentaire inversé (brin -).

### **2- Logiciel NCBI (CCDS database) :**

Ce logiciel permet de localiser la mutation au niveau de la partie codante du gène étudié et éventuellement de connaître l'acide aminé touché par la mutation sur la séquence protéique correspondante.

### **3- Logiciel NCBI (CCDS variation viewer) :**

Dès qu'on arrive à localiser la mutation, ce logiciel nous permet d'évaluer l'impact clinique de celle-ci (pathogénique ou pas), son origine (germinale ou somatique), son positionnement sur le génome entier et si elle a déjà été rapportée par autres auteurs.

# RESULTATS

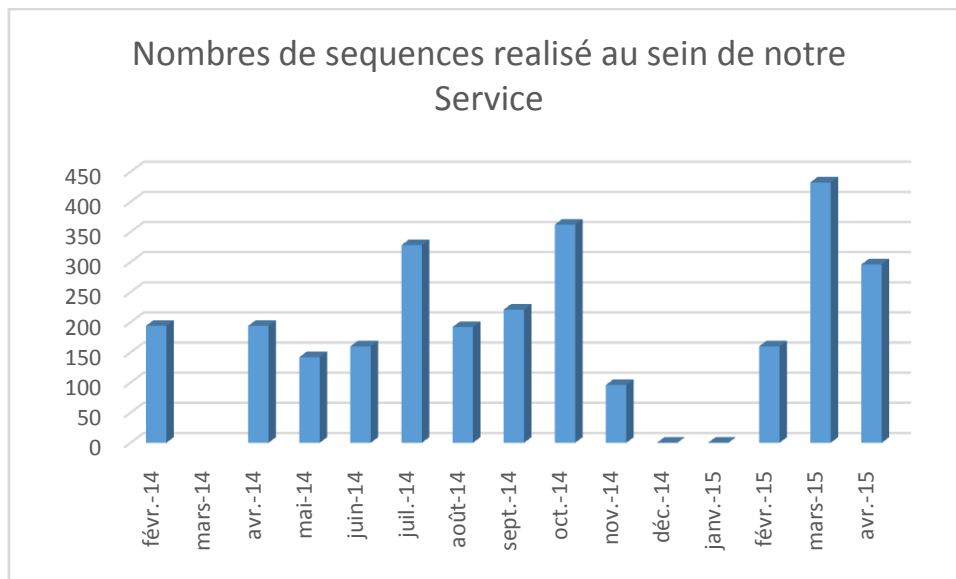
## I. Introduction:

Dans cette partie de notre travail, on va représenter de façon statistique le nombre de séquences étudié au sein de notre service de génétique médicale et oncogénétique durant la période entre février 2014 et avril 2015.

Durant cette période, on a pu séquencer 2777 séquences avec un pic au cours du mois de mars 2015.

Mois	Nombre de séquences
<b>févr-14</b>	194
<b>avr-14</b>	194
<b>mai-14</b>	142
<b>juin-14</b>	160
<b>juil-14</b>	328
<b>août-14</b>	192
<b>sept-14</b>	221
<b>oct-14</b>	362
<b>nov-14</b>	96
<b>févr-15</b>	160
<b>mars-15</b>	432
<b>avr-15</b>	296
<b>totale</b>	<b>2777</b>

Tableau 5 : Nombre de séquence séquencé dans la période entre 02/2014 et 04/2015



L'arrêt du séquenceur au cours des périodes : Mars 2014 ; Décembre 2014 ; Janvier 2015 est due à des périodes de réparation et de maintenance du séquenceur.

**Ces explorations concernent les pathologies suivantes :**

<b>Maladies auto inflammatoires héréditaires</b>	
Mutations majoritaires du gène MEFV	Fièvre méditerranéenne familiale
<b>Dysmorphologie</b>	
Mutations du gène PTPN11	Syndrome de Noonan
<b>Maladies neurosensorielles</b>	
Mutation 35delG de la connexine 26	Surdit� autosomiques r�cessives
<b>Oncologie/h�matologie</b>	

Mutation V617F du gène JAK2	Syndrome myeloprolifératifs non LMC
Recherche du facteur V de Leiden	Maladies thromboemboliques / maladie abortive
<b>Maladies neuromusculaires et anomalies du développement</b>	
Délétion de l'exon 7 du gène SMN	Amyotrophie spinale et hypotonies congénitales
Mutation 525delT du gène SGCG	Dystrophies musculaires des ceintures autosomiques récessives type 2C (LGMD2C)
<b>Anomalies de reproduction</b>	
Délétions AZF du chromosome Y	Infertilité masculine
Mutation dans HMG-box du gène SRY	Syndrome de dysgénésie testiculaire
<b>Prédisposition héréditaire aux cancers</b>	
Mutations des gènes : MLH1 et MLH2 (tous les exons)	Syndrome de Lynch
Mutations des gènes BRCA1 et BRCA2	Diagnostic pré-symptomatique du cancer du sein et de l'ovaire

Mutations des exons 12 et 13 du gène RET	Néoplasies endocriniennes multiples type 2A
<b>Oncogénétique tissulaire</b>	
Mutations du gène KRAS (exons 2, 3, 4)	Facteur pronostique et prédictif de réponse au Cetuximab dans les cancers colorectaux métastatiques
Mutation du gène BRAF (V600E)	cancers colorectaux mélanome
Mutations du gène EGFR	Lung cancer
Mutations des gènes KIT, PDGFRA	GISTs

Tableau 6 : Les différentes pathologies explorées au sein de notre service

## II. Exemples de pathologies étudiées au sein de service :

### 1-Pathologie constitutionnelle :

#### 1-1 La surdité d'origine génétique :

La surdité constitue le déficit sensoriel le plus fréquent. Elle affecte environ 1 enfant sur 1000 et sa prévalence augmente avec l'âge [1]. Actuellement, l'origine génétique est incriminée dans 60-80 % des cas [2].

10 % des surdités génétiques sont syndromiques c'est-à-dire associées à d'autres signes et 90 % sont isolées.

Le mode de transmission autosomique récessif est prédominant (près de 80 %), les formes autosomiques dominantes (20 %), liées à l'X (1 %) et mitochondriales (<1 %) ont également été décrites.

Les surdités génétiques sont, dans la grande majorité des cas, des maladies monogéniques et la déficience auditive est le plus souvent due à une atteinte cochléaire [3].

Les différentes formes de surdité non syndromiques ont, en général, été caractérisées par une analyse de liaison génétique dans de grandes familles, permettant de définir un locus. Le gène a souvent été identifié dans un second temps.

Une codification internationale a été établie pour nommer chaque locus de surdité non syndromique. Par convention et à mesure de la découverte des loci, le code commence soit par DFNA (pour deafness, autosomique dominant) soit par DFNB (pour deafness, autosomique récessif), soit par DFNX (pour deafness, liée à l'X). On donne ensuite un numéro par ordre de découverte : DFNB1 à 98, DFNA1 à 64. [3]

Ainsi plus de 67 gènes incriminés dans les surdités non syndromiques ont été identifiés à ce jour en particulier le gène GJB2 localisé en 13q11, qui code pour la Connexine 26 ; protéine transmembranaire intervenant dans le transport intercellulaire au niveau de l'oreille interne. Ce gène est muté dans environ 10 % de toutes les surdités de l'enfant et dans plus de 50 % des surdités récessives non syndromiques. [5]

Plus de 100 différentes mutations et de nombreux polymorphismes de GJB2 ont été décrits. La mutation, 35delG, représente 65 % des anomalies de ce gène dans 12 les populations caucasiennes. D'autres mutations fréquentes existent dans d'autres ethnies. Le taux de porteurs hétérozygotes sains dans la population caucasienne est estimé entre 1 et 3 %.

Ces mutations sont responsables d'une surdité neurosensorielle de transmission autosomique récessive (DFNB1), d'une surdité neurosensorielle beaucoup plus rare de transmission autosomique dominante (DFNA3), et de différentes surdités associées à des signes cutanés : syndrome de Vohwinkel (surdité associée à une hyperkératose palmo-plantaire), syndrome de KID, ...

*Expérience de notre service :*

Selon une étude faite au sein de notre service de génétique médicale et oncogénétique, à propos de 60 sujets présentant une surdité isolée durant les années 2013 et 2014. La majorité des patients ont bénéficié d'un interrogatoire détaillé, d'un examen ORL et général systématiques associés à une exploration fonctionnelle de l'audition et une imagerie du rocher, ainsi qu'une consultation génétique.

On a exclu de l'étude les patients présentant une surdité syndromique ou de cause environnementale, telle un ictère néonatal, une méningite bactérienne, une prématurité ou encore une infection virale prénatale.

*Résultats de l'étude :*

*Génotype et fréquence de la mutation 35delG :*

Parmi les 34 cas de surdité familiale, 7 patients étaient porteurs de la mutation 35delG à l'état homozygote avec une fréquence de 20,58 %. Cette même mutation a été retrouvée à l'état hétérozygote chez 2 patients soit avec une fréquence de 5,88 %.

La fréquence dans les cas sporadiques était de 15,4 % autant à l'état homozygote qu'à l'état hétérozygote.

Génotype	Cas familiauxn (%)	Cas sporadiquesn (%)
Homozygote 35delG+/35delG+	7 (20,58 %)	2 (5,88 %)
Hétérozygote 35delG+/35delG	2 (7,7 %)	2 (7,7 %)

Tableau 7 : Fréquence de la mutation 35delG dans la population étudiée

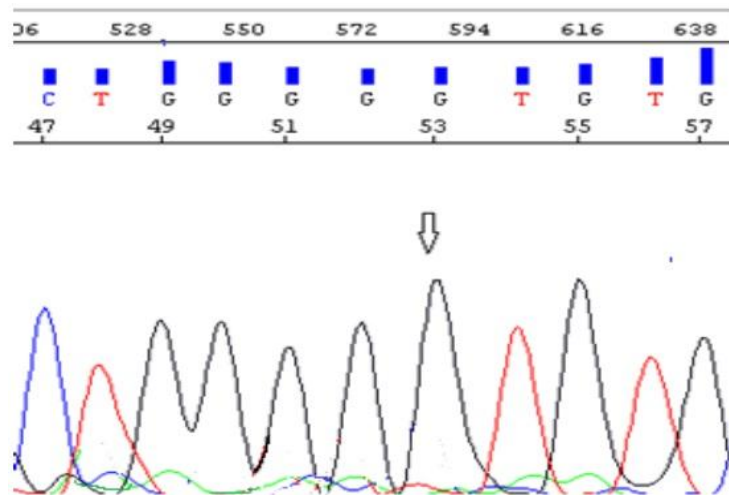


Figure 33:

Electrophéogramme de l'exon 2 du gène GJB2  
Mutation c.**35delG** à l'état homozygote (LGMO)

Autres mutations identifiées du gène de la connexine 26 :

L'étude du gène GJB2 a pu identifier, en plus de la mutation 35delG, deux autres mutations (tableau n°7) : une mutation c. 598G>A associée à la mutation 35delG, et une mutation c. 175G>C q l'état hétérozygote.

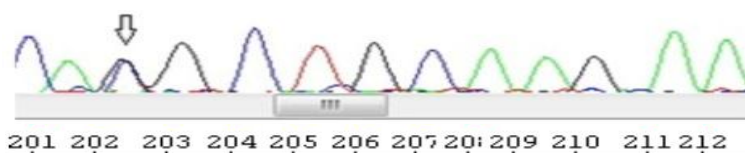
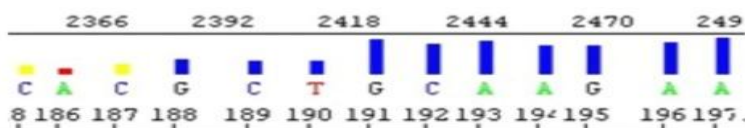


Figure 35 :

Electrophérogramme de l'exon 2 du gène GJB2  
Mutation c.175G>C à l'état hétérozygote

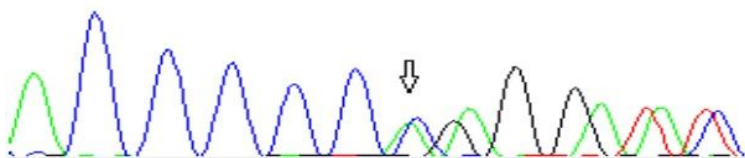


Figure 34:

Electrophérogramme de l'exon 2 du gène GJB2  
Mutation c.35delG à l'état hétérozygote

Génotype	Cas familiaux n (%)	Cas sporadiques n (%)
598G>A+/+	-	1
175G>C+/-	1	-

Tableau 8 : Fréquence des mutations c.598G>A et c.175G>C dans la population étudiée

Chez 45 patients nous n'avons pas pu retrouver une mutation du gène GJB2.  
Aucune mutation n'a été retrouvée au niveau de l'exon 1.

## **1-2 Syndrome de Noonan :**

Le syndrome de Noonan est caractérisé par une petite taille, une dysmorphie faciale caractéristique et des anomalies cardiaques congénitales. La prévalence à la naissance varie entre 1/1 000 et 1/2 500 naissances vivantes.

Les principaux traits du visage associés au syndrome de Noonan sont un hypertélorisme avec des fentes palpébrales antimongoloïdes, un ptosis, des oreilles bas implantées en rotation postérieure avec un hélix épais. Les anomalies cardiovasculaires les plus souvent associées sont une sténose pulmonaire et une cardiomyopathie hypertrophique. Les patients présentent aussi un cou palmé court, une déformation de la cage thoracique, une déficience intellectuelle légère, une cryptorchidie, des difficultés d'alimentation durant la petite enfance, une tendance aux saignements et des dysplasies lymphatiques.

Le mode d'hérédité est autosomique dominant. Chez la moitié des malades environ, la maladie est provoquée par des mutations faux-sens du gène PTPN11 localisé sur le chromosome 12, résultant en un gain de fonction de la phosphotyrosine phosphatase SHP-2 (protein tyrosine phosphatase, non-receptor type 11). Récemment, des mutations d'autres gènes de la voie RAS MAPK (gènes KRAS, SOS1, et RAF1) ont été identifiées chez une petite proportion de patients atteints du syndrome de Noonan.

Une analyse mutationnelle peut être réalisée à partir d'échantillons de sang et est recommandée devant toute suspicion de syndrome de Noonan. Cependant, le diagnostic ne peut pas être exclu en cas de résultat négatif, car la sensibilité des tests permettant de rechercher l'ensemble des gènes connus ne permet de confirmer le diagnostic que chez moins de 75% des patients.

Le diagnostic différentiel inclut le syndrome de Turner, le syndrome cardio-facio-cutané, le syndrome de Costello, la neurofibromatose de type 1 (NF1) et le syndrome LEOPARD.

Le diagnostic prénatal peut être réalisé par biopsie du trophoblaste ou par amniocentèse, si la mutation a été identifiée au préalable chez un membre de la famille. Cela étant, la complexité technique de cette analyse rend le diagnostic difficile sur un cas index dans le court laps de temps disponible au cours de la grossesse.

Le diagnostic préimplantatoire est également réalisable. Le diagnostic de syndrome de Noonan doit être envisagé chez tous les foetus présentant un polyhydramnios, des épanchements pleuraux, des oedèmes et une augmentation de la clarté nucale avec un caryotype normal.

Le traitement doit inclure la prise en charge des problèmes alimentaires rencontrés dans la petite enfance, l'évaluation de la fonction cardiaque et le suivi de la croissance et du développement moteur. La kinésithérapie et/ou l'orthophonie doivent être proposées si nécessaire.

Un examen oculaire et auditif complet doit être pratiqué durant les premières années de scolarisation. Des études de coagulation doivent être menées avant toute intervention chirurgicale. En bénéficiant de soins adaptés, la majorité des enfants atteints de syndrome de Noonan mènent une vie normale à l'âge adulte. Les signes et les symptômes s'estompent avec l'âge et la plupart des adultes atteints du syndrome de Noonan ne nécessitent aucun traitement médical particulier.

*Expérience de notre service :*

Méthodologie :

L'étude a porté sur les 16 patients NOONAN recrutés au CHU HASSAN II de Fès durant la période entre janvier 2009 et juin 2013. Le protocole PCR-Séquençage, des principaux exons touchés par les mutations, a été mis au point, et l'ADN extrait des différents patients a servi pour réaliser la PCR puis le séquençage des différents exons. Les séquences obtenues ont subi une analyse bio-informatique.

Résultats :

Sur les 16 patients étudiés, trois mutations décrites (Asp61Gly, Tyr63Cys et Asn308Ser) ont été confirmées chez quatre patients (25%), dont une dans l'exon 8 et deux dans l'exon 3. Ces quatre patients partagent les traits faciaux caractéristiques du syndrome de NOONAN. Trois parmi eux souffrent d'une sténose pulmonaire. Une autre substitution (g.63899C>T) dans l'intron 7 a été également observée chez trois autres patients. Cette substitution a été signalée dans la littérature comme étant un polymorphisme.

Les observations des patients :

**Observation 1 :**

Patient de 11ans, de sexe masculin, de petite taille, qui présente une dysmorphie faciale faite d'un hypertélorisme, avec des fentes palpébrales antimongoloïdes, un épicanthus. Des anomalies cardiovasculaires associées à type desténose congénitale de l'artère pulmonaire et une communication interauriculaire. Le patient présente aussi une déformation de la cage thoracique, une déficience intellectuelle légère, une cryptorchidie.

Chez ce patient on a réalisé une PCR- séquençage du gène PTPN11 : Une mutation 182A>G retrouvé chez lui.

Mutation localisé sur l'exon 3 du domaine N-SH2.

Elle la comme conséquence une mutation faux sens Asp61Gly.

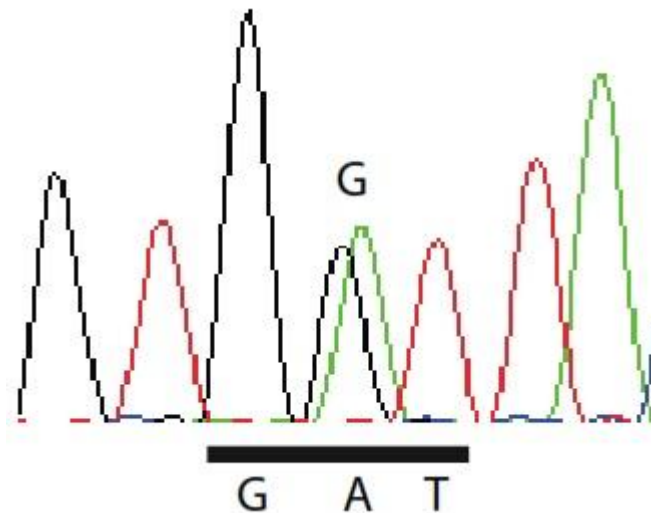


Figure 36:

Electrophérogramme de l'exon 3 du gène *PTPN11*  
Mutation c. **182A>G** à l'état hétérozygote

**Observation 2 :**

Patient de 11 ans et 3 mois, de sexe masculin, de petite taille (117 cm -4 DS), poids à 19Kg (-3DS) qui présente une dysmorphie faciale faite d'un facies triangulaire, micrognathisme, hypertélorisme, avec des fentes palpébrales antimongoloïdes, des oreilles bas implantées en rotation postérieure avec un hélix épais, un épicanthus, un palais ogival, des cheveux bas implantés. Des anomalies cardiovasculaires associées à type de sténose congénitale de l'artère pulmonaire. Le patient présente aussi un cou palmé court, une déformation de la cage thoracique, une déficience intellectuelle légère, une cryptorchidie.

Des anomalies squelettiques à type de cubitus valgus, brachydactylie et cyphoscoliose. Des anomalies hématologiques telles qu'une tendance aux saignements.

Chez ce patient on a réalisé une PCR- séquençage du gène *PTPN11* : Une mutation 182A>G retrouvé chez lui.

Mutation localisé sur l'exon 3 du domaine N-SH2.

Elle la comme conséquence une mutation faux sens Asp61Gly

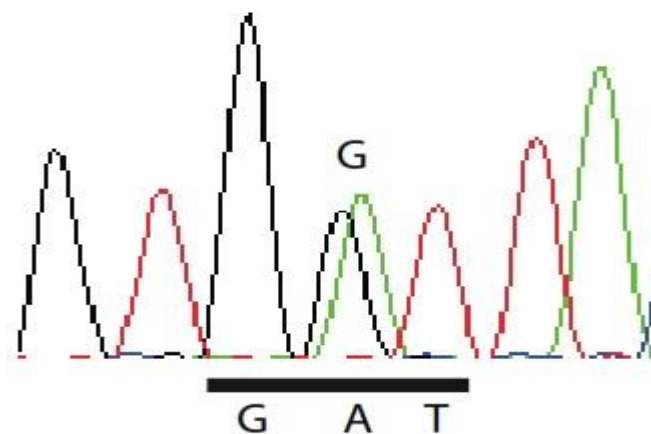


Figure 37 :

Electrophérogramme de l'exon 3 du gène *PTPN11*  
Mutation c.182A>G à l'état hétérozygote

**Observation 3 :**

Patient de 10ans et 7mois, de sexe féminin, de petite taille (127 cm  $-2.5$  DS), poids à 27Kg ( $-1$ DS) qui présente une dysmorphie faciale faite d'un facies triangulaire, hypertélorisme, ptosis, avec des fentes palpébrales antimongoloïdes, des oreilles bas implantées en rotation postérieure avec un hélix épais, un palais ogival, des cheveux bas implantés. Le patient présente aussi un cou palmé court, une déformation de la cage thoracique type thorax en entonnoir, une déficience intellectuelle légère, une cryptorchidie.

Des anomalies squelettiques à type de cubitus valgus. Des anomalies hématologiques telles qu'une tendance aux saignements.

Sans anomalies cardiovasculaires décelés.

Chez ce patient on a réalisé une PCR- séquençage du gène *PTPN11* : Une mutation 188A>G retrouvé chez lui.

Mutation localisé sur l'exon 3 du domaine N-SH2. Elle la comme conséquence une mutation faux sens Tyr63Cys.

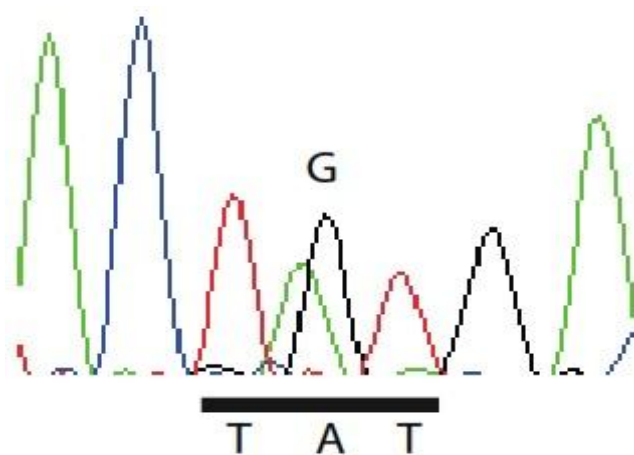


Figure 38 :

Electrophérogramme de l'exon 3 du gène *PTN11*  
Mutation c.188A>G à l'état hétérozygote

**Observation 4 :**

Patient de 4 ans et 8 mois, de sexe féminin, de taille normale (105 cm), poids à 17Kg qui présente une dysmorphie faciale faite d'un facies triangulaire, micrognathisme, hypertélorisme, avec des fentes palpébrales antimongoloïdes, des oreilles bas implantées en rotation postérieure avec un hélix épais, un épicanthus, un palais ogival, des cheveux bas implantés. Des anomalies cardiovasculaires associées à type de sténose congénitale de l'artère pulmonaire et une cardiomyopathie hypertrophique. Le patient présente aussi un cou palmé court, une déformation de la cage thoracique, une déficience intellectuelle légère.

Des anomalies squelettiques à type de cubitus valgus, brachydactylie et clinodactylie.

Chez ce patient on a réalisé une PCR- séquençage du gène *PTPN11* : Une mutation 923A>G retrouvé chez lui.

Mutation localisé sur l'exon 8 du domaine PTP.

Elle la comme conséquence une mutation faux sens Asn308Ser.

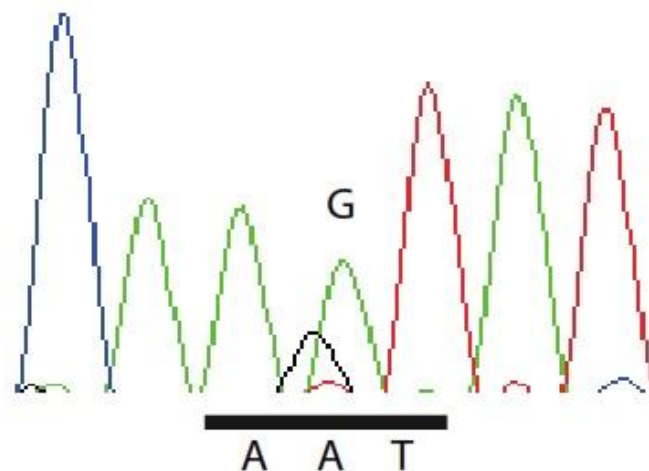


Figure 39 :

Electrophèrogramme de l'exon 8 du gène *PTPN11*  
Mutation c.903A>G à l'état hétérozygote

**Conclusion:** L'étude présente a permis de confirmer le diagnostic du Syndrome de NOONAN chez nos patients concernés, et de leur accorder une prise en charge multidisciplinaire adaptée et un conseil génétique adéquat. Cette étude nous a permis également d'avoir une idée sur le profil génétique de la population marocaine atteinte du syndrome de NOONAN.

### **1-3 La maladie périodique :**

Introduction. La fièvre méditerranéenne familiale (FMF), appelée aussi maladie périodique, est une maladie inflammatoire héréditaire. Elle présente une répartition géographique et ethnique particulière. En effet, elle est essentiellement observée dans les populations originaires du pourtour méditerranéen (Arabes, Juifs Sépharades, Turcs et Arméniens). C'est une affection génétique à transmission autosomique récessive, due à des mutations dans le gène MEFV porté par le chromosome 16. Plus de deux cents variants ont été décrits, mais seules cinq mutations (p.E148Q, p.M680I, p.M694I, p.M694 V et p.V726A) sont fréquemment associées à la majorité des cas de FMF.

Nous rapportons les résultats de trois familles marocaines avec récurrence de cette pathologie.

Patients et méthodes. 16 patients ont été recrutés à l'unité de génétique médicale du CHU Hassan II de FES, neuf femmes et sept hommes. L'ADN génomique a été extrait à partir de sang périphérique selon le protocole standard de « salting out », puis le gène MEFV a été exploré par PCR-séquençage à la recherche spécifique de la mutation la plus fréquente M694I dans l'exon 10.

Résultats. L'analyse moléculaire conduite chez les patients a montré que : six portaient la mutation M694I à l'état homozygote. Six portaient la mutation à l'état hétérozygote, suggérant qu'ils pourraient être hétérozygotes composites avec une autre mutation non encore recherchée. Quatre patients ne présentaient pas la mutation recherchée, avec une clinique très évocatrice de la maladie et une bonne réponse au traitement test, ce qui n'exclut pas la présence d'autres mutations non encore recherchées.

#### **1-4 Syndrome de RETT :**

Le syndrome de Rett est une maladie d'origine génétique causée dans près de 80% des cas typiques et 30% des formes atypiques par des mutations dans le gène MECP2 localisé en Xq28 qui est un répresseur de la transcription. La perte de la fonction de la protéine MeCP2 pourrait entraîner une surexpression de certains gènes, qui serait dommageable pour le développement du système nerveux central.

Cliniquement on note un développement quasi normal lors de la 1ère année suivi d'une régression rapide avec perte de l'utilisation volontaire des mains et retrait social. Une décélération de la croissance du périmètre crânien s'observe parallèlement et correspond à une atrophie cérébrale diffuse, affectant surtout la substance grise. Les patients évoluent ensuite vers un tableau de polyhandicap souvent compliqué d'épilepsie puis de l'apparition d'une scoliose.

Il n'y a pas de traitement à visée étiologique. Il est néanmoins important de proposer un traitement symptomatique ainsi qu'une prise en charge éducative adaptée.

Nous rapportons une étude réalisée chez 8 patientes âgées entre 2 et 11 ans, adressées pour régression psychomotrice, microcéphalie et mouvements stéréotypés des mains. L'analyse moléculaire a mis en évidence des mutations délétères du gène MeCP2 chez 3 d'entre elles.

Dans cette étude notre travail de recherche s'est focalisé sur la validation d'un nouvel examen moléculaire au sein de l'unité de génétique médicale et d'oncogénétique du CHU HASSAN II de Fès et la vérification de son efficacité dans l'amélioration de la qualité et de la spécificité du diagnostic.

## **2-Pathologie acquise :**

### **2-1 Gliomes :**

À côté de ces altérations génétiques et génomiques rencontrées dans les différents sous-types gliome, il existe une altération « transversale », présente à des fréquences variables, dans différents sous-types de gliomes diffus.

En effet, lors d'un séquençage haut débit portant sur plus 20 000 gènes dans une cohorte de 22 glioblastomes, des mutations ponctuelles récurrentes sur le codon 132 du gène Isocitrate dehydrogenase 1 (IDH1) ont été mises en évidence dans 12 % des tumeurs [25]. IDH1 est un gène localisé sur le chromosome 2q34 et code pour un enzyme du cycle de Krebs, l'isocitrate deshydrogénase. La forme sauvage de la protéine, cytosolique, catalyse de manière NADP-dépendante la décarboxylation oxydative l'isocitrate en alpha-cétoglutarate. La forme mutée, quant à elle, catalyse la réaction de réduction de l'alpha-cétoglutarate en 2 hydroxyglutarate aux propriétés pro-oncogéniques [26].

Des travaux complémentaires ont permis de préciser les mutations du gène IDH1 dans les gliomes. Les tumeurs gliales non mutées pour IDH1 peuvent être mutées sur le gène codant pour la forme mitochondriale d'IDH1, Isocitrate dehydrogenase 2 (IDH2). Ce dernier est porté par le chromosome 15q26.

Les mutations de IDH1 ou IDH2, sont mutuellement exclusives et sont observées dans : 90 % des astrocytomes de grade II, 73 % des astrocytomes de grade III, 5 % des glioblastomes primaires ou de novo, 85 % des glioblastomes secondaires, 84 % des oligodendrogliomes de grade II, 94 % des oligodendrogliomes de grade III et 100 % des oligoastrocytomes de grade II et III [27].

Les mutations des gènes IDH1 et IDH2 ont une valeur clinique diagnostique et pronostique importante. En effet, ces mutations sont présentes dans d'autres types tumoraux mais dans moins de 10 % des cas :

Les hémopathies (i.e. les leucémies aiguës myéloïdes, les syndromes myélodysplasiques, les syndromes myéloprolifératifs, les leucémies lymphoblastiques aiguës B) ;

Les cancers prostatiques ;

Les paragangliomes ;

Les cancers coliques ;

Les cancers thyroïdiens.

Par ailleurs, les mutations IDH1 et IDH2 ne sont pas mises en évidence dans les astrocytomes pilocytiques. Ces éléments font des mutations IDH1 et IDH2 un biomarqueur diagnostique pertinent dans les gliomes diffus et les tumeurs cérébrales en général. Enfin, ces mutations ont également une valeur pronostique dans les gliomes diffus indépendamment du grade de malignité de la tumeur [28,29].

Exemples de profils électrophérogramme réalisé au niveau de notre service :

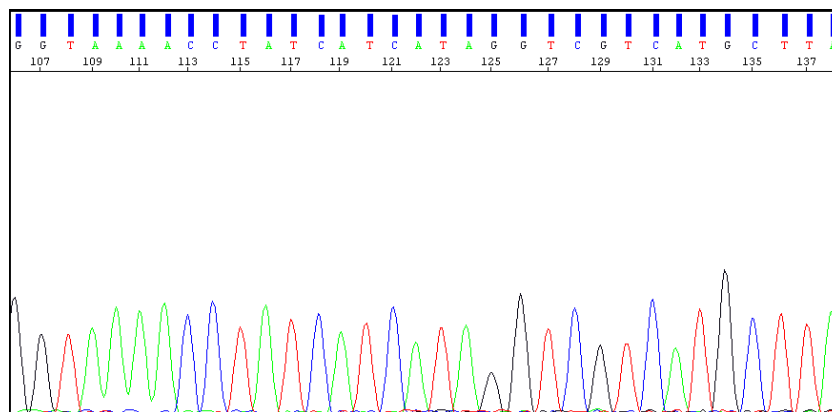


Figure 40 :

Electrophérogramme d'un patient qui présente un gliome avec un gène IDH1 **normal**

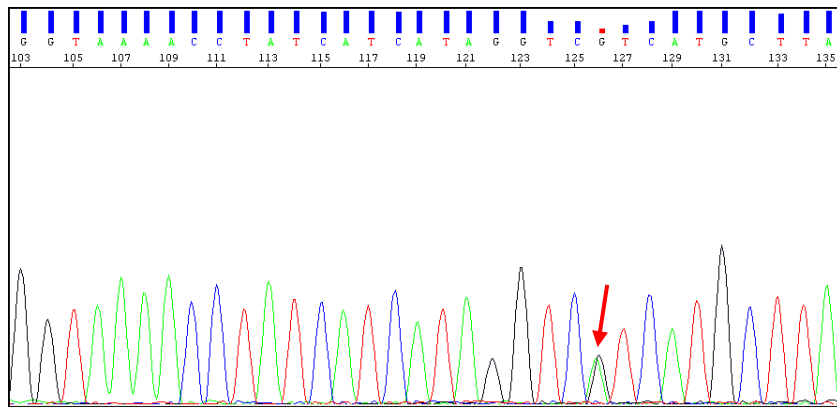


Figure 41 :

Electrophérogramme d'un patient qui présente un gliome avec une mutation R132H du gène IDH1

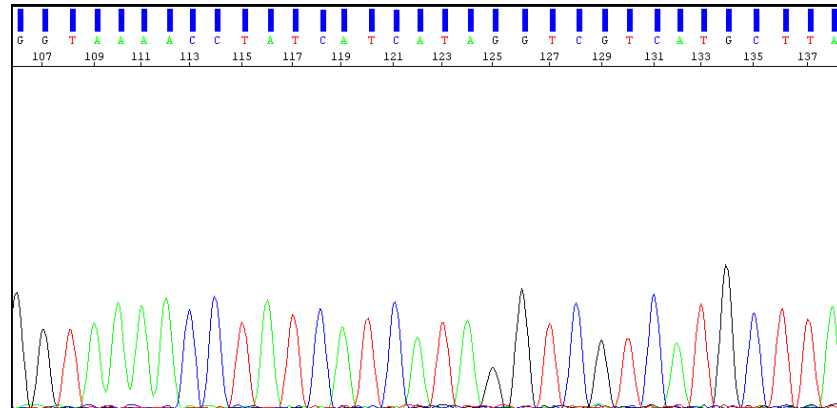


Figure 42 :

Electrophérogramme d'un patient qui présente un gliome avec un gène IDH2 normal

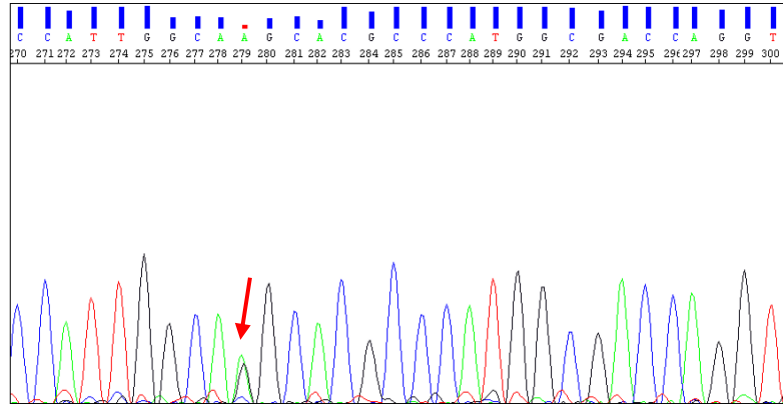


Figure 43 :

Electrophérogramme d'un patient qui présente un gliome avec une mutation R172K du gène IDH2

## 2-2 Cancers colorectaux :

La majorité des tumeurs colorectales sont des adénocarcinomes qui surviennent au cours d'une série de changements histopathologiques bien caractérisés [1, 2] (*figure 1*). Ces adénocarcinomes sont issues dans près de 70 % des cas de la transformation d'un polype adénomateux préexistant. Des filiations alternatives sont décrites : voie des adénomes festonnés, voie des adénomes plans et la voie des cancers développés sur maladies inflammatoires chroniques de l'intestin [3]. Il est maintenant bien établi qu'au moins une quinzaine de gènes identifiés à travers les études des formes familiales et sporadiques, sont susceptibles d'être impliqués dans la carcinogenèse colique.

Certains sont des oncogènes (*KRAS*,  $\beta$ -*caténine*), d'autres sont des gènes suppresseurs de tumeurs (*APC*, *p53*, *DCC*, *SMAD4/DPC4*, *SMAD2/MADR*, *TGR\_RII*, *IGF2R*, *BAX*, *hMLH1*, *hMSH2*, *hPMS1*, *hPMS2*, *hMSH6*, *PTEN/MMAC1*). Les altérations de ces gènes, oncogènes et gènes suppresseurs de tumeurs sont sous-tendues par trois grands mécanismes : l'instabilité chromosomique, l'instabilité des microsatellites et les mécanismes épigénétiques (hyper-méthylation des îlots CpG des promoteurs). Si les deux grands mécanismes d'instabilité génétique (instabilité chromosomique et instabilité des microsatellites) ont été identifiés depuis une quinzaine d'années, le rôle des modifications épigénétiques dans le cancer est de découverte plus récente et son importance est de plus en plus reconnue.

L'épigénétique regroupe tous les phénomènes influençant l'expression des gènes qui ne sont pas liés à une modification de la séquence d'ADN. Les phénomènes épigénétiques les plus étudiés sont la méthylation de l'ADN et les modifications chimiques des histones.

Les différentes modifications épigénétiques interagissent entre elles et influencent la machinerie transcriptionnelle pour contrôler l'expression des gènes [4]. Les progrès dans la compréhension des mécanismes d'instabilité (génétique et épigénétique) et l'acquisition des données des altérations moléculaires ont permis une avancée vers l'individualisation de sous-groupes de tumeurs plus homogènes quant à leur voie de progression, leur pronostic et leur éventuelle réponse au traitement.

Exemples de mutation du gène MLH1 :

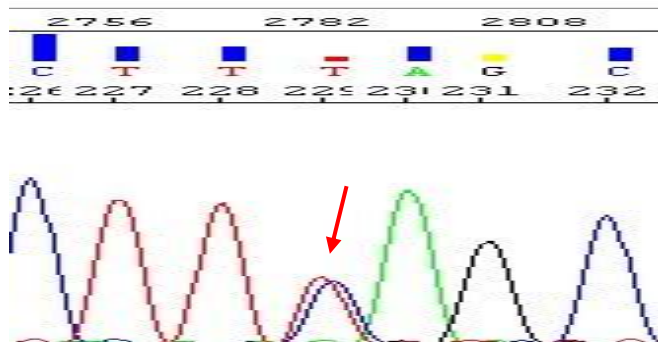


Figure 44 :

Electrophérogramme du gène *MLH1*  
Polymorphisme c.93G>A à l'état hétérozygote

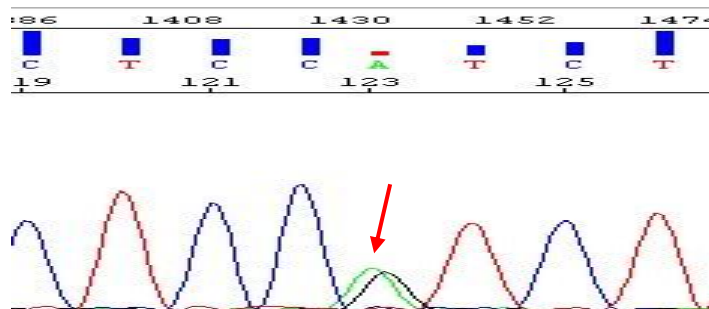


Figure 45 :

Electrophérogramme du gène *MLH1*  
Polymorphisme c.655A>G à l'état hétérozygote

# DISCUSSION

Le séquençage de l'ADN par la méthode de synthèse enzymatique de Sanger, a été utilisé pendant plus de 30 ans pour lire le code génétique des organismes vivants. La réalisation la plus emblématique de cette approche est sans aucun doute le décryptage du génome humain, projet phare de la génétique à la fin des années 2000. La dernière génération des séquenceurs à capillaires, utilisant la technique Sanger, permet aujourd'hui de lire jusqu'à 2 millions de bases en une demi-journée.

Cependant en 2007 sont apparus sur le marché des machines dotées de débits de 50 à 1 000 fois supérieures. Ces séquenceurs de « nouvelle » génération ont permis de s'affranchir d'un certain nombre de biais de la méthode Sanger comme la nécessité de cloner l'ADN à séquencer. C'est grâce notamment à la lecture de plusieurs millions de séquences en parallèle que ces nouveaux séquenceurs à « haut débit » ont pu révolutionner les analyses en génomique [66].

Avec des coûts en baisse et un champ d'application très étendu, ces méthodes vont dans les années à venir complètement modifier la façon dont les scientifiques envisagent les études en génomique. En outre, la mise à disposition par les fabricants, d'appareils abordables et automatisés, va rendre ces méthodes incontournables.

## **I. Champs d'application des NGS :**

### **1- Maladies génétiques mendéliennes**

#### **2-3 Tests préconceptionnels et dépistage des hétérozygotes**

Dans notre pays, le dépistage préconceptionnel des couples à risque de maladies héréditaires monogéniques s'effectue dans le cadre d'un conseil génétique s'il existe un cas index dans la famille. Mais dans quelques populations présentant une forte prévalence de certaines affections héréditaires le dépistage des hétérozygotes est devenu depuis longtemps systématique :  $\beta$  thalassémies dans les pays méditerranéens et maladie de Tays-Sachs dans la communauté juive

Ashkénase des États-Unis. À Chypre, le dépistage de la  $\beta$  thalassémie a fait baisser le nombre de cas de 51 en 1974 à simplement 5 entre 1991 et 2001 ; aucun cas n'a été observé entre 2002 et 2007. Aux États-Unis 1379 couples à risque de Tays-Sachs ont été identifiés (aboutissant soit à la non réalisation de l'union projetée, soit à un diagnostic prénatal, soit à un diagnostic préimplantatoire) et l'incidence de la maladie a diminué de 90 % ; dans cette communauté juive le programme de dépistage a été étendu à 16 autres maladies récessives. Aux États-Unis le dépistage des hétérozygotes est recommandé pour des maladies fréquentes comme la mucoviscidose ou l'amyotrophie spinale ; la firme « Genzyme genetics » propose ainsi un test pour 98 des mutations les plus fréquentes du gène CFTR (un couple à risque sur 1 400).

En France, le dépistage néonatal systématique concerne cinq maladies : mucoviscidose, phénylcétonurie, hypothyroïdies, hyperplasies congénitales des surrénales et drépanocytose ; il s'agit d'un dépistage biologique des homozygotes. L'apparition des techniques de séquençage à haut débit est susceptible d'entraîner le passage à un dépistage néonatal génotypique, identifiant également de ce fait les hétérozygotes et éventuellement étendu à plusieurs centaines de maladies dont les gènes et les mutations pathogènes sont connus et ceci pour moins de 400 dollars [68]. Les NGS rendent ainsi possible le diagnostic préconceptionnel généralisé, ce qui soulève de multiples questions tant éthiques que socio-économiques. Sur un tel marché potentiel se positionnent dès à présent des firmes commercialisant des DTC (Direct Test to Consumer). Ainsi « AmbyScreen » propose le dépistage de 75 maladies pédiatriques à transmission récessive ou liée à l'X. « 23 and me » offre un test plus étendu incluant, outre le génotypage de mutations responsables de maladies monogéniques, la recherche de polymorphismes associés à des maladies dites plurifactorielles et un profil pharmacogénétique, et cela pour 99 dollars [67].

## **2-4 Diagnostic prénatal et préimplantatoire**

Les diagnostics prénataux ne peuvent être pratiqués dans notre pays que dans un cadre réglementaire strict par des centres pluridisciplinaires et des laboratoires agréés. En France, dans le cas du diagnostic préimplantatoire seuls quatre centres sont autorisés par l'Agence de Biomédecine : Paris, Strasbourg, Montpellier et Nantes. Les indications diagnostiques sont restreintes puisque la loi n'autorise l'interruption de grossesse que pour une affection d'une particulière gravité et reconnue comme incurable. Le rapport 2010 de l'Agence fait état de 55 000 diagnostics prénataux en cytogénétique (essentiellement recherche de trisomie) dont près de 4 000 positifs et de 2 750 en génétique moléculaire dont 544 positifs. Le diagnostic prénatal est effectué sur biopsie de trophoblaste à 12 semaines d'aménorrhée (entraînant 1 à 2 % d'avortements), sur cellules amniotiques à 16 semaines (0,2 à 0,5 % d'avortements), sur sang fœtal après 20 semaines (3 % d'avortements) ou sur tissu fœtal au troisième trimestre de la grossesse. Un diagnostic préimplantatoire nécessite une stimulation ovarienne hormonale, la collecte d'ovules suivie d'insémination artificielle, une biopsie embryonnaire au troisième jour (J3), un diagnostic en 24-48 heures (caryotype avec hybridation in situ ou recherche d'une mutation) et un transfert in utero de 1 à 3 embryons sains à J4/J5. Chacune de ces étapes comporte un pourcentage d'échecs important et en définitive le nombre de grossesses ainsi obtenu est faible [69].

Les pertes fœtales liées aux prélèvements invasifs pratiqués pour le diagnostic prénatal sont estimées entre 300 et 600 par an, d'où l'intérêt majeur d'un diagnostic non invasif par analyse des acides nucléiques fœtaux circulant dans le sang maternel.

La présence d'ADN fœtal ou cffDNA (cell free fetal DNA) dans le sang maternel a été démontrée en 1997 et sa concentration estimée à plus de 10 % de l'ADN total

circulant ; il disparaît rapidement du sang maternel après l'accouchement, contrairement aux cellules mononuclées fœtales qui pourraient persister plusieurs années. Détectable dès 5 à 6 semaines d'aménorrhée, il a une origine trophoblastique ou placentaire, et provient sans doute de cellules en apoptose ; il est constitué majoritairement de petits fragments dont la taille moyenne est de 143 bases, ce qui correspond à la longueur d'ADN enroulé sur un nucléosome. La première application de l'analyse de l'ADN fœtal du sang maternel a été la détermination du sexe du fœtus et potentiellement des maladies liées à l'X : il s'agit d'un examen inscrit à la nomenclature des actes de biologie (B500). En 2011 la HAS a également donné un avis favorable à la prise en charge de la détermination du rhésus fœtal chez les femmes rhésus négatif : actuellement environ 6000 examens par an, mais il s'agit d'un « marché » potentiel de 160 à 180 000 tests annuels. En utilisant des « primers » spécifiques il était également possible de faire différents diagnostics ciblés ( $\beta$  thalassémie, hémophilie A...). L'apparition des techniques de séquençage à haut débit, en séquençant à la fois l'ADN circulant maternel et l'ADN fœtal et en déterminant leurs proportions relatives (par détermination des SNPs pour reconstituer les haplotypes), a ouvert de nouvelles perspectives : diagnostic des maladies génétiques et surtout des aneuploïdies [70, 71]. La recherche d'une trisomie repose sur la mise en évidence d'un petit excès de matériel provenant de ce chromosome par rapport à l'ADN maternel ; il implique un séquençage aléatoire des petits fragments d'ADN avec une grande profondeur suivi d'une analyse bio-informatique complexe.

Les premiers résultats récemment publiés sont très bons pour la trisomie 21 (sensibilité 100 % et 0,1 % de faux positifs) et la trisomie 18 (sensibilité 100 % et 0,3 % de faux positifs), sensiblement moins performants en cas de trisomie 13 (sensibilité de 91,4 % et 0,9 % de faux positifs). Bien entendu les possibilités offertes

par les NGS dans ce domaine vont se multiplier ; elles devraient pouvoir également s'appliquer dans l'avenir au diagnostic préimplantatoire, mais aussi à la détection des petites délétions et des variations du nombre de copies ou CNV (« Copy Number Variation »), restreignant ainsi le champ de la cytogénétique classique [72].

## **2-5 Diagnostic des maladies mendéliennes**

L'apparition des techniques de séquençage à haut débit a fait changer d'échelle la stratégie diagnostique des maladies monogéniques, en substituant à la recherche ponctuelle des mutations connues ou les plus fréquentes dans un ou quelques gènes, la détection simultanée de quasiment toutes les mutations dans des centaines de gènes et ceci avec une réduction considérable des temps d'analyse et donc des coûts.

En pratique médicale le séquençage NGS est habituellement mis en œuvre, non pas sur le génome entier pour des raisons de prix et surtout des difficultés d'interprétation, mais soit sur un groupe de gènes dont les mutations sont à l'origine d'un même phénotype pathologique (paraplégie spasmodique : plus de 50 gènes, cardiomyopathies : plus de 15 gènes, rétinopathies...) soit sur l'exome [73, 74]. Un bon exemple illustrant ces deux situations est celui des paraparésies spastiques familiales : il s'agit d'un groupe de maladies neurologiques rares (prévalence de 1 à 10 pour 100 000) caractérisées par une spasticité des membres inférieurs en raison d'un syndrome pyramidal, elles se présentent sous deux grands types de formes : des formes pures sans signes neurologiques additionnels et avec une imagerie cérébrale (IRM, scanner) normale et des formes complexes associées avec différents autres signes neurologiques et dans certains cas une imagerie cérébrale anormale. Les paraparésies spastiques sont également hétérogènes sur le plan génétique avec des formes à transmission autosomique dominante ou récessive et quelques formes à transmission liée à l'X, et cela sous la dépendance d'une

cinquantaine de gènes dont 24 ont été identifiés à ce jour. Dans les formes à transmission dominante les anomalies des gènes SPG4 et SPG3A expliquent plus de la moitié des cas ; dans les formes récessives le gène SPG11 est le plus fréquemment en cause (21 %) mais plus de 60 % des cas sont encore d'origine inconnue. Sur le plan physiopathologique la maladie est due à une dégénérescence progressive du faisceau pyramidal ; l'identification partielle des gènes impliqués éclaire les mécanismes en cause qui peuvent être classés en cinq grands groupes d'anomalies : trafic intracellulaire, transport axonal, myélinisation, fonctions mitochondriales et métabolisme des lipides complexes [75]. Dans un cadre de recherche, en l'absence d'anomalie d'un gène SPG déjà identifié, la stratégie utilisée combine habituellement des études de liaison et le séquençage haut débit d'une région chromosomique ou surtout de l'exome complet [76, 77]. L'exclusion des gènes connus fait appel de plus en plus au séquençage à moyen débit qui permet de les explorer simultanément par des techniques d'amplification multiplexe ou de capture.

L'étape suivante est le séquençage de l'exome qui caractérise en fait par individu 8 à 10 000 variants non synonymes dont 2 à 300 variants d'épissage et 80 à 100 variants générant un codon stop dont une vingtaine à l'état homozygote, mais dans la quasi-totalité des cas ces gènes avec perte de fonction ne sont pas responsables du phénotype observé, ils appartiennent le plus souvent à des familles multigéniques dont les membres ont une grande similitude de séquence et des fonctions potentiellement redondantes [78].

Quoiqu'il en soit la complexité des résultats bruts obtenus explique la nécessité d'étapes supplémentaires bio-informatiques d'interrogation des banques de données et de tri successif des variants, enfin de validation biofonctionnelle de la

mutation retenue. Néanmoins le séquençage de l'exome n'est pas toujours couronné de succès : la mutation pathologique est-elle localisée dans un intron ou dans une zone régulatrice à distance des séquences codantes ? S'agit-il d'une expansion de séquences répétées ? Cette situation d'échec peut relever d'un séquençage du génome entier, dont l'interprétation reste très délicate. Les connaissances acquises en recherche peuvent maintenant être utilisées pour améliorer le diagnostic. Dans des pathologies très hétérogènes comme les paraparésies spastiques, la tendance actuelle est au développement des outils de moyen débit qui couvrent l'ensemble des gènes connus responsables d'une ou d'un groupe de maladies. Même si l'aspect technique est bien contrôlé (couverture de la région, profondeur et qualité du séquençage), le défi reste d'identifier le variant causal parmi les nombreux variants détectés. En outre il est possible que certaines formes résultent non pas de l'anomalie d'un seul gène mais de di voire de trigénisme.

L'analyse directe de l'exome en première intention n'est donc pas encore une option, en raison de son coût et de la couverture inégale des exons ; néanmoins, ces limitations risquent d'évoluer rapidement dans un avenir proche.

Un autre exemple de l'intérêt du NGS est celui des MODY (Maturity Onset Diabetes of the Young), formes monogéniques les plus fréquentes du diabète de type II : dues à un dysfonctionnement des cellules  $\beta$  des îlots de Langerhans responsable d'une sécrétion insuffisante d'insuline, ces diabètes non auto-immuns sont diagnostiqués avant 25 ans et sont à transmission autosomique dominante. Les anomalies d'une trentaine de gènes (HNF1 $\alpha$ , HNF4 $\alpha$ , ABCC8, KCNJ 11 ou Kir 6-2, glucokinase, etc.) ont été caractérisées, mais environ 30 % des cas de MODY relèvent encore de causes non identifiées ; le diagnostic moléculaire relève

donc également d'une stratégie en deux temps : séquençage du groupe de gènes connus, puis éventuellement séquençage de l'exome. En pratique médicale l'identification du gène en cause permet une meilleure prise en charge du patient ; ainsi si les MODY par mutation des facteurs de transcription (HNF) présentent un risque de complications vasculaires identique à celui du diabète de type I et doivent être traités par sulfamides hypoglycémifiants ou insuline, à l'inverse les MODY par mutation de la glucokinase présentent peu de risques de ce type et n'ont habituellement pas besoin d'un traitement médicamenteux mais simplement d'un régime.

## **2- Maladies dites multifactorielles**

La grande majorité des maladies communes (cardiovasculaires, nutritionnelles, neurodégénératives, etc.), responsables d'une part importante de la morbidité et de la mortalité dans la population générale, relève de l'interaction entre facteurs génétiques multiples et facteurs environnementaux ou comportementaux. Les études d'association (« Genome Wide Association Studies » ou GWAS), utilisant les polymorphismes nucléotidiques connus (« Single Nucleotide Polymorphism » ou SNP) et menées sur de grandes cohortes, ont finalement donné des résultats décevants en ne caractérisant qu'une faible proportion de ces facteurs et conduisant ainsi à la notion d'une « hérédité manquante ». L'explication du relatif échec des études GWAS est que les SNPs analysés correspondaient à des polymorphismes trop fréquents (allèle mineur d'au moins quelques pour cent), alors que l'effet pathogène recherché est vraisemblablement le résultat de mutations ou variants beaucoup plus rares [79, 80]. Ainsi, dans le cas du diabète de type II dont 95 % des cas sont polygéniques avec une interaction importante de l'environnement, les premières études GWAS et par puces ADN ont caractérisé une soixantaine de polymorphismes mais n'expliquant au total qu'environ 10 % de l'hérédité ; les variants rares et «

forts » restent donc à identifier : l'un des gènes concernés est celui du récepteur de la mélatonine MTNR1B, les études fonctionnelles établissant par ailleurs un lien entre cette hormone épiphysaire et le métabolisme du glucose. Dans le cas de l'obésité, deux variants non synonymes (R67C et R270H) de GPR 120 c'est-à-dire du récepteur des acides gras à longue chaîne localisé principalement sur l'intestin et le tissu adipeux viennent d'être identifiés, l'importance de ce gène dans la physiopathologie de l'obésité étant prouvée par le fait que les souris invalidées deviennent obèses.

La caractérisation de ces variants « forts », considérés comme d'apparition récente dans l'évolution de l'espèce humaine et donc potentiellement différents d'un groupe ethnique à l'autre, devrait maintenant se généraliser en bénéficiant de la nouvelle technologie de séquençage à haut débit, mais ce travail n'en est encore qu'à son début ; appliqué à l'étude de cohortes correctement annotées, provenant donc de populations et d'ethnies diverses, il ouvre de nouvelles voies de recherche pour comprendre les interactions et effets de multiples gènes, et pour aider aux études physiopathologiques et à l'identification de nouveaux biomarqueurs. Les résultats attendus devraient définir, dans ces pathologies communes, les bases d'une médecine dite personnalisée [81, 82].

### **3- Oncologie**

Dans les cancers dits sporadiques les évènements mutationnels ont lieu au niveau de l'ADN somatique, tandis que dans les cancers dits héréditaires, transmis dans les familles de façon mendélienne dominante ou récessive, les mutations sont constitutionnelles de l'ADN et transmises par un gamète parental. Dans l'un et l'autre cas l'apport des NGS est considérable : dans le cas des cancers héréditaires en permettant l'exploration simultanée et rapide d'un vaste « panel » de gènes au moindre coût, dans les affections malignes somatiques en conduisant à une

dissection moléculaire des tumeurs susceptible de générer des indications thérapeutiques spécifiques [83, 84].

#### **4-1 Oncogénétique constitutionnelle**

En oncogénétique constitutionnelle, le séquençage haut débit des exons offre donc l'intérêt d'une exploration systématique d'un large éventail de gènes : 21 gènes (BRCA1 et 2, CHEK2, PTEN, STK11, TP53, BALB2, etc.) dans les cancers familiaux du sein ou de l'ovaire, 6 gènes (MLH1, MSH2, MSH6, PMS2, APC, MUTYH) dans les cas de polypose adénomateuse et de cancer héréditaire du colon [85-87]. À ce jour cette stratégie n'a encore conduit à l'identification que d'un petit nombre de nouveaux gènes de prédisposition héréditaire aux cancers ; l'un des exemples est celui du gène PALB2, gène de prédisposition au cancer du pancréas : le séquençage comparatif des exomes germinale et tumorale a en outre montré la présence d'une mutation constitutionnelle et d'une mutation somatique acquise (c'est-à-dire l'existence d'un mécanisme en deux temps identique à celui de la théorie des « anti-oncogènes » de Knudson dans le rétinoblastome, appelés ainsi parce que la protéine Rb que la mutation rend déficiente inhibe la multiplication cellulaire) [88].

#### **4-2 Hémopathies malignes**

Dans les hémopathies malignes de très nombreuses translocations chromosomiques sont observées ; elles conduisent soit à la dérégulation d'un oncogène (lymphome de Burkitt, etc.) soit à l'expression d'une protéine de fusion anormale dont les deux meilleurs exemples sont la leucémie myéloïde chronique (réarrangement BCR/ABL) et la leucémie aiguë à promyélocytes (réarrangements PML/RARA ou PLZF/RARA). La leucémie myéloïde chronique a bénéficié au début des années 90 d'une des premières thérapeutiques ciblées par l'imatinib, inhibant spécifiquement la tyrosine kinase Bcr/Abl. Du fait de sa spécificité et de sa tolérance, l'imatinib a rapidement supplanté toute autre forme de traitement. La

majorité des patients a une réponse cytogénétique complète et permanente ; cependant 25 % des malades sont résistants soit primitivement soit secondairement en raison de mutations ponctuelles dans le gène de fusion. Ces échecs ont conduit à la mise au point de nouveaux inhibiteurs (dasatinib, nilotinib, etc.) [89, 90]. De même PML/RARA est ciblé par les agents thérapeutiques de la leucémie aiguë à promyélocytes : l'acide rétinoïque, l'arsenic et les agents stabilisant l'AMPc. De nombreuses autres hémopathies ont également été étudiées [91], et peuvent être citées en exemple : les leucémies myéloïdes aiguës qui présentent une grande diversité clinique et pronostique, sont également caractérisées par une hétérogénéité importante des mutations [92, 93]. Elles ne résultent pas d'une instabilité génomique mais seraient plutôt associées à un faible nombre de mutations oncogéniques « driver », la grande majorité étant des mutations « passagers », qui jouent un rôle passif dans le développement de la maladie. Le défi actuel est donc d'identifier les mutations contribuant réellement à la leucémogénèse. Une étude portant sur 108 génomes a montré des mutations portant sur 23 gènes parmi lesquels des mutations de 3 membres du complexe cohésine (STAG2, SMC3, SMC1A) [94]. Le séquençage génome entier a également montré des mutations des gènes IDH1 et IDH2, ainsi que DNMT3A agissant par la voie de la méthylation [91].

L'étude des myélodysplasies révèle de façon inattendue des mutations de la machinerie d'épissage de l'ARN, tels que U2AF35, ZRSR2, SRSF2, SF3B1 [95]. Les mutations de SF3B1 sont associées aux anémies réfractaires avec sidéroblastes en couronne, alors que les mutations de SRSF2 sont plutôt associées aux leucémies myélomonocytaires chroniques. Les gènes du spliceosome semblent très intéressants à étudier pour améliorer le diagnostic et la classification pronostique des syndromes myélodysplasiques.

Les leucémies lymphoïdes ont également fait l'objet de travaux de séquençage : concernant les leucémies lymphoblastiques, le séquençage génome entier de 12 cas de LAL-ETP (early T-cellprecursor) a été réalisé ainsi que l'évaluation de la fréquence des mutations somatiques identifiées dans 94 cas de LAL-T [96]. Cette étude montre que les LAL-ETP sont caractérisées par les mutations de gènes impliqués dans 3 grandes voies : la voie de signalisation de RAS et des récepteurs de cytokines, le développement hématopoïétique et les modifications des histones. De plus, de nouvelles cibles mutées ont été identifiées : DNM2, ECT2L et RELN [91].

Les leucémies lymphoïdes chroniques (LLC) sont les plus fréquentes des leucémies de l'adulte, elles apparaissent habituellement dans la sixième décennie de la vie avec une prédominance masculine ; leur étiologie est inconnue mais l'observation de formes familiales fait supposer l'existence de facteurs génétiques de prédisposition.

Dans environ un tiers des cas la maladie a une évolution très lente ne nécessitant pas de traitement, le décès survenant pour une autre cause ; dans un autre tiers l'évolution très lente est suivie d'une brutale aggravation ; enfin le dernier tiers se présente d'emblée sous une forme grave. Dans 80 % des cas des anomalies cytogénétiques sont retrouvées, elles peuvent être associées à des pronostics différents favorables (13qdel, trisomie 12) ou défavorables (11qdel, 17qdel). La mise en œuvre du séquençage à haut débit a montré que le spectre mutationnel des LLC était très limité (10 à 20 gènes) par rapport à ce qui est habituellement observé dans les tumeurs malignes.

Cinq gènes sont plus fréquemment mutés : SF3B1 (intervenant dans l'épissage des ARN), TP53, NOTCH1 (activant la différenciation lymphoïde), MYD88 (activant la voie des Tolllikerecepteurs) et ATM. Les mutations de SF3B1 sont présentes dans

environ 10 % des cas, elles ne sont pas spécifiques des LLC car également observées dans différentes autres affections malignes (myélodysplasies, cancers du sein, mélanomes....) ; elles sont souvent associées à des délétions en 11q et de mauvais pronostic. Les mutations de NOTCH1, présentes dans 5 à 10 % des cas, sont associées à une trisomie 12 et résistantes au traitement. Les mutations de MYD88, également retrouvées dans 85 % des cas de maladie deWaldenström, sont souvent associées à des délétions 13q et à des mutations des IGHV. En fait le séquençage à haut débit montre une hétérogénéité tumorale avec une organisation clonale et sous clonale, souvent secondaire à la chimiothérapie ; en définitive il permet de faire un bilan mutationnel complet des LLC, d'affiner leur pronostic et surtout d'adapter leur traitement. Les leucémies myéloïdes aiguës qui présentent une grande diversité clinique et pronostique, sont également caractérisées par une hétérogénéité importante des mutations [91, 92].

### **4-3 Tumeur malignes**

La classification habituelle des tumeurs repose sur les seules données morphologiques et cytologiques de l'anatomie pathologique ; la mise en œuvre successive des différentes techniques de biologie moléculaire a montré que cette classification était bien entendu insuffisante pour rendre compte de la très grande diversité d'expressions géniques à l'intérieur d'un même type de cancer, diversité correspondant souvent à des différences importantes de pronostic et de réponse à la chimiothérapie.

Pour rendre compte de cette diversité et de ses conséquences, le séquençage par les nouvelles méthodes à haut débit tend à s'imposer comme nouvelle méthode d'analyse systématique des tumeurs [97].

Par rapport au génome constitutionnel, le nombre de variations somatiques est très variable d'un type de tumeur à l'autre : les plus mutagènes sont les mélanomes alors que des tumeurs de l'enfant pourtant très agressives comme les rhabdomyosarcomes le sont peu. Les sarcomes sont des tumeurs rares mais d'une considérable variabilité d'un cas à l'autre ; ils sont caractérisés par une très grande fréquence des anomalies chromosomiques, en particulier des translocations conduisant à la formation de gènes chimériques codant souvent pour des facteurs de transcription. Les neuroblastomes ont des évolutions très différentes d'un malade à l'autre avec des cas de régression et de guérison spontanée, et à l'inverse des formes très agressives et métastatiques ; ces différences sont associées à des anomalies chromosomiques différentes.

Une des grandes difficultés concernant ces mutations est de faire la distinction entre celles qui sont importantes pour la progression tumorale (mutations « drivers ») et celles qui sont seulement accessoires (mutations « passagers »). L'intérêt du séquençage est également thérapeutique : identifier parmi les mutations principales celles susceptibles d'un ciblage médicamenteux (mutations « druggable »). Les gènes « druggables » codent soit des récepteurs membranaires (habituellement des tyrosine ou des sérine kinases) qui peuvent être inhibés par des anticorps monoclonaux agissant sur leur face extracellulaire (ex : Trastuzumab sur HER2, etc.) ou par des produits chimiques agissant sur la face intracellulaire (ex : Imatinib sur cKIT et RET, Lapatinib sur HER2, etc.), soit des facteurs de voies de signalisation intracytoplasmique contrôlant la prolifération cellulaire (ex : Vémurafemib sur RAF, Everolimus sur mTOR, etc.). Par ailleurs le séquençage met en évidence l'extrême hétérogénéité des tumeurs de même type et conduit à des thérapies ciblées : un exemple est celui du cancer colorectal au stade IV où l'existence de mutations de K-RAS rend inefficace le traitement par un anticorps monoclonal anti-EGFR ; à l'inverse

dans le mélanome métastasé la mutation V600E du gène BRAF est une indication au traitement par le verumafinib. En définitive les polymorphismes mutationnels d'un nombre restreint de gènes (APC, BRAF, EGFR, KRAS, HER2, C-KIT, PIK3CA, etc.) sont susceptibles de moduler la thérapeutique de différents types de tumeurs (mélanomes, glioblastomes, cancers du sein, du poumon, de l'estomac, etc.). Jusqu'à présent l'approche est ainsi ciblée sur certains gènes, témoin d'une médecine dite stratifiée (K-RAS muté ou non, BRAF muté ou non, etc.).

Par ailleurs, étant donné la variabilité des réponses individuelles aux médicaments anti-cancéreux, le séquençage haut débit mérite également d'être utilisé pour caractériser, sur l'ADN normal, les polymorphismes génétiques de l'ensemble des acteurs protéiques de leur métabolisme [98], avec trois grands types d'indications :

1) faire la distinction entre répondeurs et non répondeurs,

2) identifier les sujets à risques pour la survenue d'effets indésirables, liés en général à des mutations inhibant les enzymes impliqués dans leur métabolisme (ex : risque de neutropénie par l'irinotecan en fonction du polymorphisme de l'UGT1A1, hypersensibilité à la carbamazépine et HLA-A 3101, allopurinol et HLA-B 5801, etc.) [99],

3) individualiser la posologie (ex : doses de coumadine en fonction des polymorphismes du CYP2D6 et de la vitamineKépoxy-réductase VKORC1 [100], adaptation de la posologie en tricyclines en fonction du nombre de copies de CYP2D6).

En définitive le développement des NGS devrait rapidement mener à une médecine dite personnalisée impliquant le séquençage des génomes tumoraux et

constitutionnels, et conduisant à l'administration, en fonction des mutations identifiées et des caractéristiques pharmacogénétiques individuelles, de thérapies ciblées et éventuellement combinées. Les premières études réalisées montrent que ce type de médecine personnalisée, fondée sur l'analyse génomique des tumeurs, se traduit par un gain en termes d'absence de récurrences ou de survie, par rapport aux patients traités par les méthodes standard. Il reste néanmoins à évaluer la capacité de notre système de santé à en supporter les charges financières.

Assez paradoxalement, l'industrie pharmaceutique, confrontée à une augmentation importante des coûts en recherche et développement du fait en particulier d'une réglementation de plus en plus contraignante, et à une diminution des retours sur investissements, est conduite à un changement de stratégie ; elle tend à remplacer les essais randomisés classiques (patients recevant ou non le médicament après tirage au sort) par la comparaison de l'efficacité du médicament dans deux groupes de patients présélectionnés selon qu'ils sont porteurs ou non du marqueur, passant ainsi d'une approche probabiliste à une approche déterministe.

## **II. Limitations actuelles des méthodes de séquençage à haut débit :**

Les méthodes de séquençage à haut débit de première génération ont maintenant atteint la maturité technologique nécessaire à leur adoption large par la communauté scientifique et médicale. Les fabricants de matériels ne s'y sont pas trompés, et lancent d'ailleurs en 2011 chacun des versions de leurs matériels dédiés aux applications de diagnostic. C'est le cas de Roche avec le GS Junior, d'Illumina avec le MiSeq et d'AppliedBiosystems avec le Ion PersonalGenome Machine. Dans le même temps la seconde génération d'appareils commence à arriver sur le marché.

Elle ambitionne de résoudre deux des limitations actuelles des premières générations de machines à haut débit.

D'une part la nécessité d'amplifier l'échantillon à analyser, les machines de seconde génération seront capables de lire directement la séquence d'ADN. D'autre part le temps d'acquisition des appareils actuels est assez lent car il se fait par cycle. Les nouveaux séquenceurs seront eux capables d'effectuer la lecture de l'ADN en temps réel.

Cependant face à cette escalade technologique et à l'ébullition autour des applications potentielles, il faut garder à l'esprit que le séquençage à haut débit doit faire face encore à plusieurs limitations. Tout d'abord, c'est une technologie encore jeune. Il reste de nombreuses zones d'ombre en ce qui concerne ces approches. Les biais techniques ne sont pour le moment qu'au début de leur caractérisation. On manque encore de recul sur les capacités de ces outils à quantifier précisément les événements biologiques complexes (épissage alternatif, site de fixation à l'ADN des facteurs de transcription...).

Ensuite, le séquençage à haut débit produit des quantités très importantes de données dont la gestion n'est pas totalement maîtrisée et encore moins automatisée. Les résultats générés se chiffrent aujourd'hui en téraoctets, la question de savoir quoi garder et pendant combien de temps se pose de façon cruciale. Ce point est d'autant plus important que les études sur les effets de molécules thérapeutiques ont lieu pendant plusieurs années et qu'elles nécessitent un grand nombre d'expériences. Des questions importantes se posent également sur le traitement des résultats obtenus. Si par exemple dans le cas des puces à ADN on dispose de plus de 15 ans de recul technologique et de méthodologies statistiques bien déterminées pour corriger les données brutes et obtenir des résultats fiables, ce n'est pas du tout

le cas en ce qui concerne les résultats du séquençage à haut débit. Les questions de normalisation ne sont pas tranchées, et les modèles statistiques à appliquer aux données ne sont toujours pas déterminés sans ambiguïtés. Pour terminer, dans le domaine du séquençage du génome des patients, des questions éthiques importantes se posent. Il est maintenant possible d'accéder au génome complet d'un individu. Tous les aspects éthiques de ces outils doivent donc être envisagés et surtout débattus pour y offrir une réponse appropriée.

### **III. Au Maroc :**

Au Maroc, il est recommandé d'importer les NGS essentiellement pour des panels de pathologies fréquentes au Maroc tel que le retard mental, autisme, myopathies, les épilepsies, prédispositionaux cancers... et en estimant une cinquantaine de maladies concerné par cette approche NGS-panels « de gènes ».

Ce qui permettra à notre pays un diagnostic rapide pour un maximum de patients et leurs familles, pour une meilleure prise en charge et un conseil génétique adéquat.

## Manuel d'installation du CD-Rom

### 1. Objet :

Ce document a pour but de décrire l'installation de l'application « Séquençage d'ADN ».

### 2. Livrables :

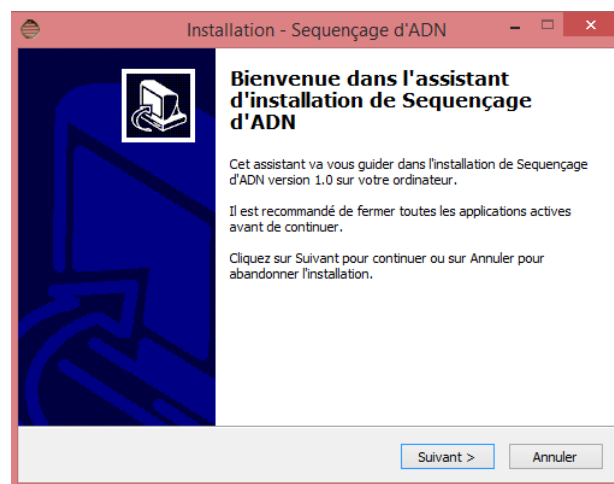
L'installation est fournie sous forme d'un fichier Setup nommé Setup.exe

### 3. Configuration minimale requise :

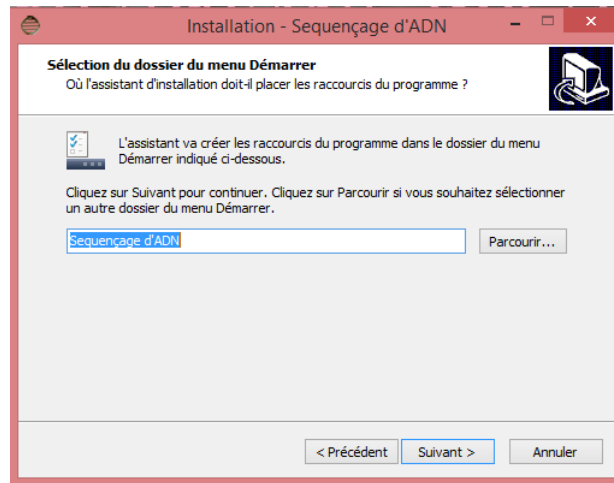
Win XP, PIV 1.4GHz, 256 Mo de RAM, carte graphique NVIDIA GeForce 4 Ti ou ATI Radeon 9600, 4.5 Go d'espace libre sur le DD.

### 4. Procédures d'installation :

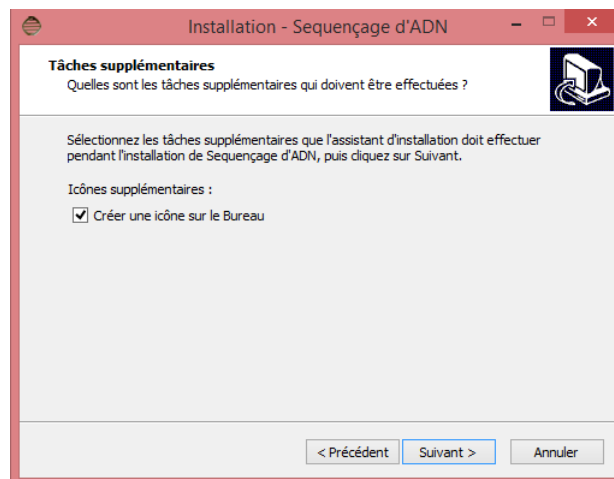
- 1) Ouvrir le CD-Rom
- 2) Cliquer sur Setup.exe
- 3) Cliquer sur « suivant »



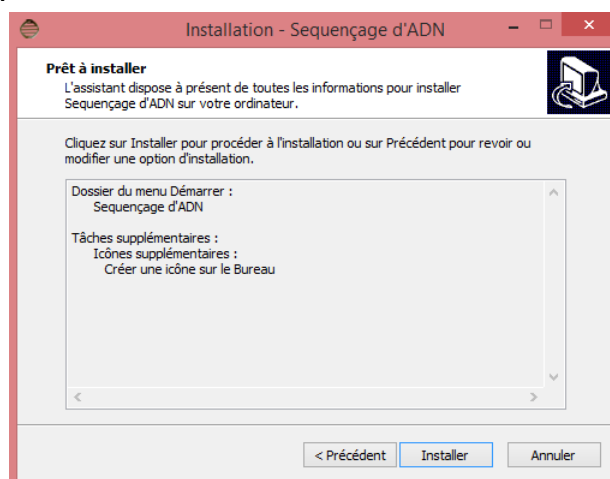
- 4) Cliquer sur « suivant »



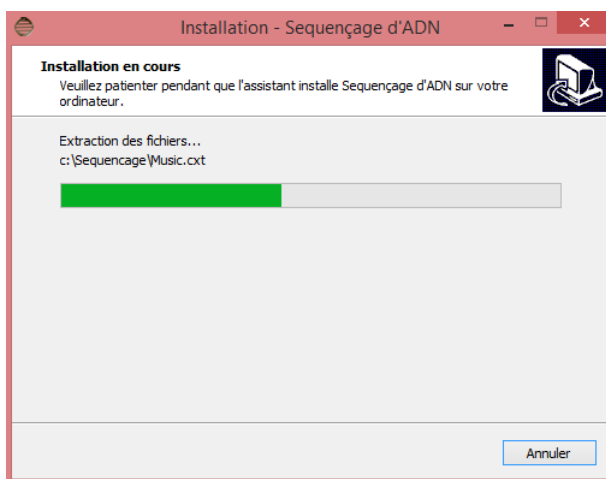
5) Cliquer sur « suivant »



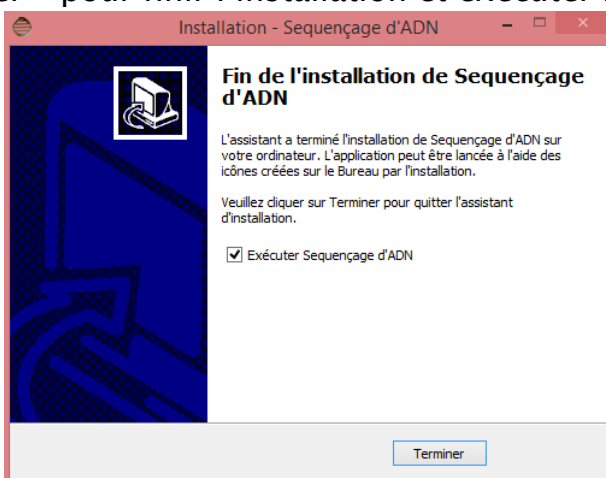
6) Cliquer sur « Installer »



7) Attendez un instant pour finir l'installation



8) Cliquer sur « Terminer » pour finir l'installation et exécuter le programme



# RESUMES

## Résumé

Le séquençage de l'ADN constitue une méthode dont le but est de déterminer la succession linéaire des bases A, C, G et T prenant part à la structure de l'ADN. La lecture de cette séquence permet d'étudier l'information biologique contenue par celle-ci.

Étant donné l'unicité et la spécificité de la structure de l'ADN chez chaque individu, la séquence de l'ADN permet de nombreuses applications dans le domaine de la médecine, comme, par exemple, le diagnostic, les études génétiques, l'étude de paternité, la criminologie, la compréhension de mécanismes physiopathologiques, la synthèse de médicaments, les enquêtes épidémiologiques.

A travers cette thèse, nous avons abordé la technique du séquençage d'ADN sur tous les plans, théorique et pratique :

- ✓ Théorique : en abordant les différentes méthodes de la technique, classique ou moderne, avec la mise au point de l'expérience de notre service de génétique médicale et oncogénétique dans ce domaine.
- ✓ Pratique : un CD-ROM intitulé « technique de séquençage d'ADN », destiné aussi bien aux étudiants en médecine qu'aux résidents en génétique médicale et doctorants en science, qui aborde de manière simplifiée et pratique les différents étapes de la technique de séquençage de l'ADN et aussi donne une petite aperçu sur les nouvelles générations de séquenceurs.

## Abstract

The DNA sequencing is a method that aims to determine the linear sequence of bases A, C, G and T taking part in the structure of DNA. Reading of this sequence is used to study the biological information contained by it.

Because the uniqueness and specificity of the DNA structure in each individual, the DNA sequence allows numerous applications in the field of medicine, such as, for example, diagnosis, genetic studies, paternity study criminology, understanding of pathophysiology, drug synthesis, epidemiological surveys.

Through this thesis, we discussed the technique of DNA sequencing on all levels, theoretical and practical:

- ✓ Theoretical: addressing the different methods of the art, classic or modern, with the development experience of our medical genetics and cancer genetics service in this area.
- ✓ Practical: a CD-ROM entitled "Technical DNA sequencing" destined both medical student to residents in medical genetics and PhD students in science, which addresses in a simplified and practical manner the various stages of the technique DNA sequencing and also gives a small glimpse of the new generation of sequencers.

## ملخص

تقنية تسلسل الحمض النووي هي ال تقنية التي تهدف إلى تحديد التسلسل الخطي للقواعد الأزوتية A، C، G و T التي تدخل في بنية الحمض النووي . وتستخدم قراءة هذا التسلسل لدراسة المعلومات البيولوجية الواردة به.

بسبب تفرد وخصوصية بنية الحمض النووي من فرد لآخر، فإن تقنية تسلسل الحمض النووي تسمح بالعديد من التطبيقات في مجال الطب، على سبيل المثال، التشخيص، الدراسات الوراثية، علم الجريمة، دراسة الأبوة، فهم الفيزيولوجيا المرضية، تركيب الأدوية والاستقصاءات الوبائية.

من خلال هذه الأطروحة، ناقشنا تقنية تسلسل الحمض النووي على جميع المستويات، النظرية والعملية:

- ✓ النظرية: معالجة أساليب مختلفة من التقنيات، الكلاسيكية أو الحديثة، مع التركيز على تجربة مصلحتنا، مصلحة الأمراض الوراثية، في هذا المجال.
- ✓ العملية: قرص مدمج بعنوان "تقنية تسلسل الحمض النووي" خصص لكل من طلاب الطب والأطباء المقيمين في علم الوراثة الطبية وطلبة الدكتوراه في العلوم، والذي يتناول بطريقة مبسطة وعملية المراحل المختلفة للتقنية تسلسل الحمض النووي وأيضا يعطي لمحة صغيرة من الجيل الجديد من آلات تسلسل الحمض النووي.

## Liste des figures

Figure 1 : Fred Griffith (1877–1941)

Figure 2 : Oswald T.Avery (1877–1955)

Figure 3 : Maclyn McCarty avec Francis Crick et James D.Watson

Figure 4 : Phoebus Aaron Levene (1869–1940)

Figure 5 : Francis Crick, James D.Watson, Rosalind Franklin

Figure 6 : Quelques étapes importantes démontrant l'évolution des progrès du séquençage

Figure 7 : Schémas de la composition d'un nucléotide

Figure 8 : Les différentes bases azotées

Figure 9 : Dégradation des bases azotées

Figure 10 : Différentes formes des bases azotées

Figure 11 : Spectre d'absorption de différentes bases azotées

Figure 12 : Liaison N-osidique

Figure 13 : Les liaisons esters

Figure 14 : Liaison phosphodiester

Figure 15 : Charge des bases azotées en fonction du pH

Figure 16 : Exemples d'enzymes de restriction

Figure 17 : Spectre d'absorption de l'ADN

Figure 18 : Appariement pyrimidines-purines selon Watson et Crick

Figure 19 : Modèle de Watson et Crick de l'ADN (conformation B)

Figure 20 : Principe de la méthode Maxam et Gilbert

Figure 21 : Structure des didésoxyribonucleotides

Figure 22 : Principe du séquençage par la méthode de Sanger

Figure 23 : Séquençage automatique sur des séquenceurs à 1 et 4 fluorophores

Figure 24 : Evolution des coûts du séquençage

Figure 25 : Aperçu de la technologie de séquençage 454

Figure 26 : Amplification en pont de la technologie Illumina

Figure 27 : Aperçu de la technologie de séquençage Illumina

Figure 28 : Aperçu de la technologie de séquençage SOLiD

Figure 29 : Chronologie de la commercialisation des séquenceurs haut-débit

Figure 30 : Détection de nucléotides par l'utilisation d'un nanopore

Figure 31 : Mécanisme de la fluorescence par la technique du transfert d'énergie par résonance

Figure 32 : Représentation de nombre de séquences réalisés au sein de notre service

Figure 33 : Electrophérogramme de l'exon 2 du gène GJB2 Mutation c.35delG à l'état homozygote

Figure 34 : Electrophérogramme de l'exon 2 du gène GJB2 Mutation c.35delG à l'état hétérozygote

Figure 35 : Electrophérogramme de l'exon 2 du gène GJB2 Mutation c.175G>C à l'état hétérozygote

Figure 36 : Electrophérogramme de l'exon 3 du gène PTN11 Mutation c. 182A>G à l'état hétérozygote

Figure 37 : Electrophérogramme de l'exon 3 du gène PTN11 Mutation c.182A>G à l'état hétérozygote

Figure 38 : Electrophérogramme de l'exon 3 du gène PTN11 Mutation c.188A>G à l'état hétérozygote

Figure 39 : Electrophérogramme de l'exon 8 du gène PTN11 Mutation c.903A>G à l'état hétérozygote

Figure 40 : Electrophérogramme d'un patient qui présente un gliome avec un gène IDH1 normal

Figure 41 : Electrophérogramme d'un patient qui présente un gliome avec une mutation R132H du gène IDH1

Figure 42 : Electrophérogramme d'un patient qui présente un gliome avec un gène IDH2 normal

Figure 43 : Electrophérogramme d'un patient qui présente un gliome avec une mutation R172K du gène IDH2

Figure 44 : Electrophérogramme du gène MLH1 Polymorphisme c.93G>A à l'état hétérozygote

Figure 45 : Electrophérogramme du gène MLH1 Polymorphisme c.655A>G à l'état hétérozygote

## Liste des tableaux

Tableau 1 : Nomenclatures des différents nucléotides

Tableau 2 : Exemple de quelques nucléases avec leurs spécificités

Tableau 3 : Caractéristiques des technologies de séquençage

Tableau 4 : Cycle utilisé au niveau du thermocycleur

Tableau 5 : Nombre de séquence séquencé dans la période entre 02/2014 et 04/2015

Tableau 6 : Les différentes pathologies explorées au sein de notre service

Tableau 7 : Fréquence de la mutation 35delG dans la population étudiée

Tableau 8 : Fréquence des mutations c.598G>A et c.175G>C dans la population étudiée

# Références

- [1] J. Lamoril, N. Ameziane, J.-C. Deybach, P. Bouizegarene et M. Bogard, «Les techniques de séquençage de l'ADN :une révolution en marche. Première partie,» *Immuno-analyse et biologie spécialisée*, vol. 23, n° 15, pp. 260–279, 23 october 2008.
- [2] M. Ilie, É. Long, V. Hofman, V. Lespinet, O. Bordone, K. Washetine, V. Gavric-Tanga et P. Hofman, «Les méthodes de séquençage de « nouvelle génération » (NGS) et le cancer broncho-pulmonaire : principales technologies, applications et limites actuelles en pathologie.» *REVUE FRANCOPHONE DES LABORATOIRES*, n° 1458, pp. 51–58, Janvier 2014.
- [3] Avery, O.T., McLeod, C.M. & McCarthy M. (1944) Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* Type III *J. Exp. Med.* 79, 137–158.
- [4] Chargaff, E. (1950) Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia* 6, 201.
- [5] Hershey, A.D. (1946) Spontaneous mutations in bacterial viruses. *Cold Spring Harbor Symposia on Quantitative Biology* XI, 33–37.
- [6] Hershey, A.D. (1946) Mutations of bacteriophage with respect to type of plaque. *Genetics* 31, 620–640.
- [7] Hershey, A.D. & Chase, M. (1952) Independent functions of viral proteins and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* 36, 39–56.
- [8] Watson, J.D. & Crick, F.H.C. (1953) A structure for desoxyribose nucleic acid. *Nature* 171, 737–738.
- [9] Watson, J.D. & Crick, F.H.C. (1953) Genetical implications of the structure of deoxyribonucleic acid. *Nature* 171, 964–967.
- [10] Watson J.D. (1968) *La double hélice*, Hachette/Pluriel.
- [11] *Recombinant DNA*, (1992) Watson, J.D., Gilman, M., Witkowski, J. & Zoller, M. Scientific American Books.
- [12] *Molecular biology of the cell*, (1994) Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J.D. Ed. Garland Publishing.

- [13] Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171, 737–738.
- [14] Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T. et al. (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature*, 452, 872–876.
- [15] Pushkarev, D., Neff, N.F. and Quake, S.R. (2009) Single-molecule sequencing of an individual human genome. *Nat Biotechnol*, 27, 847–852.
- [16] Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R. et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456, 53–59.
- [17] Ahn, S.M., Kim, T.H., Lee, S., Kim, D., Ghang, H., Kim, D.S., Kim, B.C., Kim, S.Y., Kim, W.Y., Kim, C. et al. (2009) The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res*, 19, 1622–1629.
- [18] Kim, J.I., Ju, Y.S., Park, H., Kim, S., Lee, S., Yi, J.H., Mudge, J., Miller, N.A., Hong, D., Bell, C.J. et al. (2009) A highly annotated whole-genome sequence of a Korean individual. *Nature*, 460, 1011–1015.
- [19] Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J., Guo, Y. et al. (2008) The diploid genome sequence of an Asian individual. *Nature*, 456, 60–65.
- [20] Ley, T.J., Mardis, E.R., Ding, L., Fulton, B., McLellan, M.D., Chen, K., Dooling, D., Dunford–Shore, B.H., McGrath, S., Hickenbotham, M. et al. (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*, 456, 66–72.
- [21] Mardis, E.R., Ding, L., Dooling, D.J., Larson, D.E., McLellan, M.D., Chen, K., Koboldt, D.C., Fulton, R.S., Delehaunty, K.D., McGrath, S.D. et al. (2009) Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Eng! J Med*, 361, 1058–1066.

- [22] (2010) Human genome: Genomes by the thousand. *Nature*, 467, 1026–1027.
- [23] Mardis, E.R. (2006) Anticipating the 1,000 dollar genome. *Genome Biol*, 7, 112.
- [24] Sboner, A., Mu, X.J., Greenbaum, D., Auerbach, R.K. and Gerstein, M.B. (2011) The real cost of sequencing: higher than you think! *Genome Biol*, 12, 125.
- [25] Glenn, T.C. (2011) Field guide to next-generation DNA sequencers. *MolEcolResour*, 11, 759–769.
- [26] Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z. et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437, 376–380.
- [27] Metzker, M.L. (2010) Sequencing technologies – the next generation. *Nat Rev Genet*, 11, 31–46.
- [28] Green, R.E., Krause, J., Ptak, S.E., Briggs, A.W., Ronan, M.T., Simons, J.F., Du, L., Egholm, M., Rothberg, J.M., Paunovic, M. et al. (2006) Analysis of one million base pairs of Neanderthal DNA. *Nature*, 444, 330–336.
- [29] Noonan, J.P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Paabo, S., Pritchard, J.K. et al. (2006) Sequencing and analysis of Neanderthal genomic DNA. *Science*, 314, 1113–1118.
- [30] Wall, J.D. and Kim, S.K. (2007) Inconsistencies in Neanderthal genomic DNA sequences. *PLoS Genet*, 3, 1862–1866.
- [31] Mellmann, A., Harmsen, D., Cummings, C.A., Zentz, E.B., Leopold, S.R., Rico, A., Prior, K., Szczepanowski, R., Ji, Y., Zhang, W. et al. (2011) Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One*, 6, e22751.
- [32] Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M. et al. (2011)

- Anintegrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475, 348–352.
- [33] Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D. and Church, G.M. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, 309, 1728–1732.
- [34] Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G. et al. (2010) Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science*, 327, 78–81.
- [35] Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X. et al. (2008) The potential and challenges of nanopore sequencing. *Nat Biotechnol*, 26, 1146–1153.
- [36] Clarke, J., Wu, H.C., Jayasinghe, L., Patel, A., Reid, S. and Bayley, H. (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol*, 4, 265–270.
- [37] Butler, T.Z., Pavlenok, M., Derrington, I.M., Niederweis, M. and Gundlach, J.H. (2008) Single-molecule DNA detection with an engineered MspA protein nanopore. *ProcNatiAcadSci U S A*, 105, 20647–20652.
- [38] Min, S.K., Kim, W.Y., Cho, Y. and Kim, K.S. (2011) Fast DNA sequencing with a graphene-based nanochannel device. *Nat Nanotechnol*.
- [39] Hall, A.R., Scott, A., Rotem, D., Mehta, K.K., Bayley, H. and Dekker, C. (2010) Hybrid pore formation by directed insertion of alpha-haemolysin into solid-state nanopores. *Nat Nanotechnol*, 5, 874–877.
- [40] Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B. et al. (2009) Real-time DNA sequencing from single polymerase molecules. *Science*, 323, 133–138.
- [41] Korlach, J., Bjornson, K.P., Chaudhuri, B.P., Cicero, R.L., Flusberg, B.A., Gray, J.J., Holden, D., Saxena, R., Wegener, J. and Turner, S.W. (2010) In

- Nils, G. W. (ed.), *Methods in Enzymology*. Academic Press, Vol. Volume 472, pp. 431–455.
- [42] Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J. and Turner, S.W. (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods*, 7, 461–465.
- [43] Song, C.X., Clark, T.A., Lu, X.Y., Kislyuk, A., Dai, Q., Turner, S.W., He, C. and Korlach, J. (2011) Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nat Methods*, 9, 75–77.
- [44] Pennisi, E. (2010) Genomics. Semiconductors inspire new sequencing technologies. *Science*, 327, 1190.
- [45] Meyerson, M., Gabriel, S. and Getz, G. (2010) Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet*, 11, 685–696.
- [46] Lantos, J.D., Artman, M. and Kingsmore, S.F. (2011) Ethical considerations associated with clinical use of next-generation sequencing in children. *J Pediatr*, 159, 879–880 e871.
- [47] Greenbaum, D., Sboner, A., Mu, X.J. and Gerstein, M. (2011) Genomics and privacy: implications of the new reality of closed data for the field. *PLoSComputBiol*, 7, e1002278.
- [48] Gaynor, A.M., Nissen, M.D., Wilely, D.M., Mackay, I.M., Lambert, S.B., Wu, G., Brennan, D.C., Storch, G.A., Sloots, T.P. and Wang, D. (2007) Identification of a novel polyomavirus from patients with acuterespiratory tract infections. *PLoSPathog*, 3, e64.
- [49] Chiu, R.W., Chan, K.C., Gao, Y., Lau, V.Y., Zheng, W., Leung, T.Y., Foo, C.H., Xie, B., Tsui, N.B., Lun, F.M. et al. (2008) Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallelgenomic sequencing of DNA in maternal plasma. *Proc Natl AcadSci U S A*, 105, 20458–20463.

- [50] Cirulli, E.T. and Goldstein, D.B. (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet*, 11, 415–425.
- [51] Isidor, B., Lindenbaum, P., Pichon, O., Bezieau, S., Dina, C., Jacquemont, S., Martin-Coignard, D., Thauvin-Robinet, C., Le Merrer, M., Mandel, J.L. et al. (2011) Truncating mutations in the last exon of NOTCH2 cause a rare skeletal disorder with osteoporosis. *Nat Genet*, 43, 306–308.
- [52] Lupski, J.R., Reid, J.G., Gonzaga-Jauregui, C., Rio Deiros, D., Chen, D.C., Nazareth, L., Bainbridge, M., Dinh, H., Jing, C., Wheeler, D.A. et al. (2010) Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med*, 362, 1181–1191.
- [53] Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A. et al. (2010) Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*, 42, 30–35.
- [54] O'Roak, B.J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J.J., Girirajan, S., Karakoc, E., Mackenzie, A.P., Ng, S.B., Baker, C. et al. (2011) Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet*, 43, 585–589.
- [55] Kan, Z., Jaiswal, B.S., Stinson, J., Janakiraman, V., Bhatt, D., Stern, H.M., Yue, P., Haverty, P.M., Bourgon, R., Zheng, J. et al. (2010) Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature*, 466, 869–873.
- [56] Meyerson, M., Gabriel, S. and Getz, G. (2010) Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet*, 11, 685–696.
- [57] Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D. et al. (2011) Tumour evolution inferred by single-cell sequencing. *Nature*, 472, 90–94.
- [58] Harris, S.R., Feil, E.J., Holden, M.T., Quail, M.A., Nickerson, E.K., Chantratita, N., Gardete, S., Tavares, A., Day, N., Lindsay, J.A. et al.

- (2010) Evolution of MRSA during hospital transmission and intercontinentalspread. *Science*, 327, 469–474.
- [59] Chin, C.S., Sorenson, J., Harris, J.B., Robins, W.P., Charles, R.C., Jean-Charles, R.R., Bullard, J., Webster, D.R., Kasarskis, A., Peluso, P. et al. (2011) The origin of the Haitian cholera outbreak strain. *N Engl J Med*, 364, 33–42.
- [60] Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T. et al. (2010) A human gut microbial gene catalogue established by metagenomics sequencing. *Nature*, 464, 59–65.
- [61] Ozsolak, F. and Milos, P.M. (2011) RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet*, 12, 87–98.
- [62] Ingolia, N.T., Ghaemmaghami, S., Newman, J.R. and Weissman, J.S. (2009) Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science*, 324, 218–223.
- [63] Park, P.J. (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*, 10, 669–680.
- [64] Laird, P.W. (2010) Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet*, 11, 191–203.
- [65] Schwarzenbach, H., Hoon, D.S. and Pantel, K. (2011) Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer*, 11, 426–437.
- [66] Metzker, M.L. (2010) Sequencing technologies – the next generation. *Nat Rev Genet*, 11, 31–46.
- [67] Le Gall J.Y., Ardaillou R.—Diffusion et validation des tests génétiques en France. *Bull. Acad. Natle Med.* 2009, 193(9), 2093–19.
- [68] Bell C. J., Dinwiddie D C., Miller N. A. et al. — Carrier testing for severe childhood recessive diseases by next generation sequencing. *Sci. Transl. Med.*, 2011, 3, 65ra4.

- [69] Steffann J., Frydman N., Burlet Ph. et al. Le diagnostic pré-implantatoire couplé au typage HLA: l'expérience française. *Bull. Acad. Natle Med.*, 2011, 195(4-5), 1015-22.
- [70] Fan H. C., Gu W., Wang J. et al. Non-invasive prenatal measurement of the fetal genome. *Nature*. 2012, 487, 320-24.
- [71] Snyder M.W., Simmons L.E., Kitzman J.O. et al.—Noninvasive fetal genome sequencing: a primer. *Prenatal diagnosis*. 2013, 33, 547-54.
- [72] Dugoff L.—Application of genomic technology in prenatal diagnosis. *N. Engl. J. Med.*, 2012, 367, 2249-50. *Bull. Acad. Natle Méd.*, 2014, 198, no 1, 101-117, séance du 7 janvier 2014
- [73] De Ligt J., Willemsen M. H., Van Bon B.W.M. et al. — Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.*, 2012, 367, 1921-29.
- [74] Saunders C. J., Miller N. A., Soden S. E. et al. — Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci. Transl. Med.*, 2012, 4, 77-89.
- [75] Tesson C., Nawara M. N., Salih M.A.M. et al. — Alteration of fatty-acid-metabolizing enzymes affects mitochondrial form and function in hereditary spastic paraplegia. *Am. J. Hum. Genet.*, 2012, 91, 1051-64.
- [76] Klebe S., Lossos A., Azzedine H. et al.—KIF1A missense mutations in SPG30, an autosomal recessive spastic paraplegia: distinct phenotypes according to the nature of the mutations. *Eur. J. Hum. Mut.*, 2012, 20, 645-49.
- [77] Goizet C., Depienne C., Benard G. et al. — REEP1 mutations in SPG31: frequency, mutational spectrum, and potential association with mitochondrial morpho-functional dysfunction. *Hum. Mut.*, 2011, 32, 1118-27.
- [78] MacArthur D.G., Balasubramanian S., Frankish A. et al. — A systematic survey of loss-of-function variants in human protein-coding genes. *Science*. 2012, 335, 823-28.
- [79] Jordan B. — Rare is frequent. *Med. Sci.*, 2012, 28, 893-96.

- [80] Tennessen J.A., Bigham A.W., O'Connor T.D. et al. — Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 2012, 337, 64–69.
- [81] Jordan B. — Le véritable génome personnel. *Med. Sci.* 2011, 27, 553–56.
- [82] Roberts N.J., Vogelstein J.T., Parmigiani G. et al. — The predictive capacity of personal genome sequencing. *Sci. Transl. Med.* 2012, 4, 132–36.
- [83] Dancey J.E., Bedard Ph.L., Onetto N. et Hudson T.J. — The genetic basis for cancer treatment decisions. *Cell*. 2012, 148, 409–19.
- [84] Longo D.L. — Tumor heterogeneity and personalized medicine. *N. Engl. J. Med.* 2012, 366, 956–57.
- [85] Pritchard C.C., Smith C., Salipante S.J. et al. — Coloseq provides comprehensive Lynch and polyposis syndrome mutational analysis using massive parallel sequencing. *J. Mol. Diagn.* 2012, 14, 357–65.
- [86] Walsh T., Casadei S., Lee M.K. et al. — Mutations in 12 genes for inherited ovarian, fallopian tube and peritoneal carcinoma identified by massively parallel sequencing. *PNAS*. 2011, 108, 18032–37.
- [87] Walsh T., Lee M.K., Casadei S. et al. — Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *PNAS*. 2010, 107, 12629–33.
- [88] Jones S., Hruban R.H., Kamiyama M. et al. — Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science*. 2009, 234, 17.
- [89] Schiffer C.A. — Bcr–Abl tyrosine kinase inhibitors for chronic myelogenous leukemia. *N. Engl. J. Med.* 2007, 357, 258–64.
- [90] Goldman J.M. — Ponatinib for chronic myeloid leukemia. *N. Engl. J. Med.* 2012, 367, 2148–49.
- [91] Marceau–Renault A., Renneville A., Nibourel O. et al. — Le séquençage de nouvelle génération (NGS) a-t-il déjà sa place

- dans nos laboratoires d'onco-hématologie ? *Hématologie*. 2013, 19, 112-22.
- [92] Patel J.P., Gönen M., Figuerora M.E. et al. — Pronostic relevance of integrated genetic profiling in acute myeloid leukemia. *N. Engl. J. Med.* 2012, 366, 1079-89.
- [93] The cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 2013, 368(22) 2059-74. *Bull. Acad. Natle Méd.*, 2014, 198, no 1, 101-117, séance du 7 janvier 2014
- [94] Welch J.S., Ley T.J., Link D.C. et al. — The origin and evolution of mutations in acute myeloid leukemia. *Cell.* 2012, 150 (2), 264-78.
- [95] Yoshida K., Sanada M., Shiraishi Y. et al. — Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature.* 2011, 478 (7367), 64-9.
- [96] Zhang J., Ding L., Holmfeldt L. et al. — The genetic basis of early T-cell precursor acutelymphoblastic leukaemia. *Nature.* 2012, 481 (7380), 157-63.
- [97] Vogelstein B., Papadopoulos N., Velculescu V.E. et al. — Cancer genome landscapes. *Science.* 2013, 339, 1546-58.
- [98] Wang L., McLeod H.L., Weinshilboum R.M. — Genomics and drug response. *N. Engl. J. Med.* 2011, 364, 1144-53.
- [99] Becquemont L. — HLA: A pharmacogenomics success story. *Pharmacogenomics.* 2010, 11, 277-81.
- [100] Schwarz U.I., Ritchie M.D., Bradford Y. et al. — Genetic determinants of response to warfarin during initial anticoagulation. *N. Engl. J. Med.* 2008, 358, 999-1007.