الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي

Ministère de L'Enseignement Supérieur et de la Recherche Scientifique

N° Ref : ……

Université Abdelhafid BOUSSOUF- Mila

Institut des Sciences de la Nature et de la Vie

Département des Sciences Biologiques et Agricoles

Mémoire préparé en vue de l'obtention du diplôme de

Master

Domaine : Sciences de la Nature et de la Vie

Filière : Sciences Biologiques

Spécialité : Biochimie Appliquée

Thème :

Recherche de variants exomique par NGS

Présenté par :

- **ABDELALI Hanane**
- **MESSAADI Dalila**

Devant le jury :

- Dr BOURAS Ouassila	**MAA**	**Présidente**
- Dr GHOUT Agena	**MCA**	**Examinateur**
- Dr CHEKROUD Karim	**MCB**	**Promoteur**

Année Universitaire : 2025/2026

Remerciements

Au nom de Dieu, le Tout Miséricordieux, le Très Miséricordieux, et grâce à la volonté de Dieu Tout-Puissant, nous Lui rendons grâce et louange pour nous avoir accordé la force, la volonté et l'aide nécessaires afin de mener à bien ce travail, qui représente l'aboutissement d'un long parcours jalonné de doutes et d'efforts, mais surtout marqué par un soutien sincère et des rencontres enrichissantes ayant contribué à atteindre cet objectif.

Nous tenons à exprimer nos plus sincères remerciements et notre profonde gratitude à notre encadrant, le Professeur Chakroud Karim, pour son encadrement précieux, ses orientations rigoureuses et son soutien constant tout au long de la réalisation de ce travail. Nous lui témoignons notre profond respect et notre reconnaissance pour tout ce qu'il nous a apporté.

Nous adressons également nos vifs remerciements aux membres du jury de soutenance, pour avoir accepté d'examiner et d'évaluer ce travail, ainsi que pour le temps et les efforts consacrés à sa lecture et à son appréciation, et pour leurs remarques constructives qui ont contribué à son amélioration et à son enrichissement.

Nous exprimons également notre profonde gratitude à l'ensemble des enseignants qui nous ont accompagnés tout au long de notre parcours universitaire, chacun à son titre et à son niveau, pour leur savoir, leurs conseils et la qualité de la formation académique qu'ils nous ont dispensée, contribuant ainsi de manière essentielle à notre développement scientifique.

Nous remercions aussi toute personne ayant contribué de près ou de loin à ce travail, par un soutien, un encouragement ou un conseil, et qui a participé, d'une manière ou d'une autre, à la réalisation de ce mémoire.

Enfin, nous adressons nos plus sincères remerciements à nos familles, les familles Messaadi et Abdelali, qui ont été et demeurent le véritable pilier et le premier soutien tout au long de ce parcours. À nos chers parents, qui ont toujours cru en nous, même dans les moments les plus difficiles où la fatigue et le doute prenaient le dessus, nous exprimons toute notre reconnaissance pour leur amour inconditionnel, leur patience et leurs sacrifices que nous mesurons davantage chaque jour. Leur présence, leurs prières et leur soutien ont été notre plus grande force et la clé de notre persévérance. Sans eux, ce travail n'aurait jamais vu le jour.

Dédicace

C'est avec la grâce de Dieu Tout-Puissant que ce travail a été accompli. À Lui reviennent louange et gratitude, au début comme à la fin.

À celui qui a été un symbole de bonté, de vertu et de nobles valeurs, et qui a consacré sa vie à mon bonheur et à ma réussite, mon cher père Houcine, que Dieu te garde pour moi comme soutien et fierté dans ma vie. À celle dont le paradis repose sous ses pieds, qui m'a offert un amour et un sacrifice sans limites, à la source de mon cœur et au refuge de mon âme,

Ma chère mère Aziza, que Dieu te protège et prolonge ta vie. À mes frères qui ont été un soutien tout au long de ce parcours, Mouad, Rahim, Mohamed et Rafik, merci pour votre présence et votre soutien constant. À mes chères sœurs, compagnes de route et source de joie dans ma vie, Widad, Houda et Roufia, je vous adresse toute ma reconnaissance et mon affection.

À mes amies de l'université et de l'extérieur, qui ont partagé avec moi le chemin des études entre fatigue et espoir, et ont été un véritable soutien.

À toute personne qui m'a appris quelque chose ou m'a apporté une aide ou un mot bienveillant, j'exprime toute ma gratitude et ma reconnaissance.

À mon cher fiancé Zinou, à celui qui a été mon véritable soutien dans ce chemin et mon premier appui à chaque étape, à celui qui a cru en moi et qui m'a accompagnée dans les moments de faiblesse avant la force, sans sa compréhension et son soutien, je n'aurais pas pu réaliser ce travail, je lui adresse mes plus sincères remerciements et toute ma reconnaissance pour sa patience et son soutien jusqu'à l'achèvement de ce travail.

À tous ceux qui liront ce travail et en bénéficieront, je dédie le fruit modeste de mes efforts, en signe de respect et de gratitude.

Enfin, je prie Dieu de bénir ce travail, de le rendre utile et sincère, et de l'accepter favorablement.

Hanane

Dédicace

Tout dabord, je remercie le dieu , notre créateur de m'avoir donnée la force , la volonté et le courage afin d'acomplir ce travail modeste .

À l'homme de ma vie , mon précieux cadeau du dieu , qui doit ma vie , ma réussite et tout mon respect , je remercie pour ses efforts et ses conseils, mon cher papa (mohamed).

À la femme qui souffert sans me laisser souffrir , qui n'as jamais dit non face mes exigences qui nas épargné aucun effort pour me rendre heureuse , ma chère maman (Samia).

Je dédie ma chère grand-mére (Daloula) , qui ma enveloppée de son amour et m'a offert la plus grand part de ses sentiments effectueux .

À mes deus soeurs adorées qui m'ont soutenue dans tous les moments difficils et qui ont cru en ma réussite quand je doutais (Manal et Alaa) .

À mon chér frére , qui a toujours été une source de soutien et de force, toujours prêt à donner un coup de main (Abdelhakime).

À mes cousines , khawla , Mahassin, aya , kamar et romayssa.

À tous mes amis , Afaf , wahiba , Rawâa , nihad , khawla , Rana , chayma et fatima .

À notre encadrant (chekroud karim) .

À ma fammile messaadi .

Et enfin à mon âme qui a travaillé et rêvé , qui a peiné et lutté tout au long du chemin , et qui a toujours cru en la possibilité d'atteindre sa destination .

Dalila

Liste des abréviations

- **ADN** : Acide désoxyribonucléique
- **ATP** : Adénosine triphosphate
- **Bp** : paires de bases
- **dNTP** : désoxynucléotides triphosphates
- **ddNTP** : didésoxynucléotides triphosphates
- **DNBSEQ** : DNA Nanoball Sequencing
- **MGI** : MGI Tech (plateforme de séquençage)
- **Mb** : Mégabases
- **NGS** : Next Generation Sequencing
- **PCR** : Polymerase Chain Reaction
- **RNA / ARN** : Acide ribonucléique
- **SOLID** : Sequencing by Oligonucleotide Ligation and Detection
- **TE** : Transposable Elements
- **VNTR** : Variable Number Tandem Repeats
- **WES** : Whole Exome Sequencing
- **WGS** : Whole Genome Sequencing
- **Ml** : millilitre
- **µL** : microlitre
- **UDB** : ultra DNA Buffer
- **Ul** : microlitre
- **NF** : Nuclèase Free
- **FastQ** : FastQ format
- **VCF** : Variant call format
- **BAM** : Binary Alignment Map

- **SNP** : Single Nuclèotide Polymorphism
- **Indel** : insertion /deletion
- **CNV**: copy Number variation
- **Nt** : Nuclèotide
- **g** : gram
- **ng** : nanogram
- **Gb** : Gigabase
- **K** : killobase
- **M** : mèga

Liste des Figures

Figure N° 1 : Organisation de l'ADN dans la cellule humain 7
Figure N° 2 : Chéma représentative des exones et les introns. 9
Figure N° 3 : principe de technologie sanger. 26
Figure N° 4 : principe de la technologie de pyroséquençage 27
Figure N° 5 : principe de la technologie de illumina 29
Figure N° 6 : principe de technologie de Nanopore oxford. 31
Figure N° 7 : principe de technologie de solide 32
Figure N° 8 : Principe de génération des DNB dans la technologie DNBSEQ 33
Figure N° 9 : Schéma des étapes de préparation de la librairie d'ADN pour le séquençage NGS. 40
Figure N° 10 : Schéma du processus d'enrichissement des régions cibles par capture hybride (Hybrid Capture) en séquençage NGS. 41
Figure N° 11 : Schéma des étapes du séquençage NGS par technologie DNBSEQ (préparation de la librairie, formation des DNBs et séquençage) 42
Figure N° 12 : Tailles des séquences lues (Reads) après « Trimming » 57
Figure N° 13 : Nombre moyen de lectures (reads) par base 58
Figure N° 14 : %GC des lectures (reads) 58

Liste des tableaux

Tableau N° 1 : la taille des fragments isolés à partir de 100µL d'AND fragmenté en fonction de la quantité de billes magnétiques 43
Tableau N° 2 : mélange pour la réparation d'extrémité et de queue A 45
Tableau N° 3 : Le mélange de ligature de l'adaptateur 45
Tableau N° 4 : le mélange Amplification par PCR 47
Tableau N° 5 : Préparation du mélange d'hybridation 49
Tableau N° 6 : Préparation du mélange de capture 50
Tableau N° 7 : le mélange de réaction de circularisation 52

Table des Matières

Remerciements

Dédicace

Dédicace

Liste des abréviations

Liste des Figures

Liste des tableaux

Table des Matières

Introduction 1

Partie 1

Chapitre 01 : Le génome Humain

1. Le génome Humain 6

1.1. Définition 6

1.1.1. Le génome nucléaire 6

.1.1.2. Le génome mitochondrial humain 6

2. Organisation et composition du génome humain 7

.2.1 Organisation structurale du génome humain 7

2.1.1. Gènes 7

2.1.2. Chromatine 8

2.1.3. nucléosome 8

2.1.4. Chromosome 8

.2.2. Composition du génome 8

2.2.1. les séquences codantes 8

2.2.2. Les séquences non codantes 9

2.2.2.1. Les introns 9

2.2.2.2. Séquences régulatrices 9

2.2.3. Sequences répétées 9

2.2.3.1. Microsatellites 10

2.2.3.2. Mini satellites 10

2.2.3.3. Satellites 10

2.2.4. Séquences mobiles 10

2.2.4.1. Transposons 10

2.2.4.2. Rotrotransposons 10

3. L'Exome 11

Référence de chapitre 1 12

Chapitre 02 : Les maladies génétiques

1. Les maladies Génétique 15

2. Les types des maladies génétiques 16

2.1. Maladies mono géniques (mutation d'un seul gène) 16

2.2. Les maladies polygéniques (combinaison de gènes et de facteurs environnementaux) 16

2.3. Les maladies chromosomiques (anomalie dans les chromosomes) 16

3. La transmission des maladies génétique 17

4. Diagnostique des maladies génétiques 17

5. Analyse de l'exome : 18

6. Le traitement des maladies génétiques 18

7. Les limites de traitement des maladies génétiques 19

8. Les mutations génétiques 19

8.1. Les différents types des mutations 20

8.1.1. Mutation ponctuelles 20

8.1.2. Mutations spontanées et induites 20

8.1.3. Mutations somatiques et germinales 21

Références de chapitre 2 22

Chapitre 03 : Le Séquençage de l'ADN

Généralité sur le séquençage de l'ADN 25

1. Le séquençage bas débit 25
1.1. La technologie de Sanger 25
.1.1.1 Les étapes de la technologie de Sanger 26
1.2. La technologie de pyroséquençage 27
.1.2.1 Le principe de pyroséquençage 27
2. Le séquençage à haut débit (NGS) 28
2.1. Les technologies les plus utilisées actuellement 28
.2.1.1 La technologie « Illumina » 28
2.1.1.1. Le principe de illumina 29
2.1.2. Technolohgie « Xefoed Nano pore» 30
2.1.2.1. Le principe de Nanopore oxford 30
2.1.3. Technologie «SOLID» 31
2.1.3.1. Le principe de solide 31
2.1.4. Technologie « MGI (DNB SEQ)» 32
2.1.4.1. Le principe de la technologie de MGI (DNB SEQ) 32
3. Les applications liées au séquençage de l'ADN 33
3.1. Les applications médicales 33
3.2. Applications techniques 34
3.3. Les applications on microbiologie 34
Référence de chapitre 3 36

Partie 2 : Matériel et méthodes

1.Matériel et Méthodes: 40
2.Méthodes 40
2.1. Extraction et purification de l'ADN 42
2.2. Fragmentation de l'ADN 43
2.2.1. Sélection des Fragments d'ADN en fonction de leurs tailles : 43
2.2.2. Etapes de la sélection des fragments : 43

2.3. Réparation de l'extrémité des fragments et ajout de la queue « A » 44
2.4. Ligature de l'adaptateur 45
2.4.1. Mélange de ligature d'adaptateur 46
2.4.2. Programme de ligature d'adaptateur 46
2.4.3. Nettoyage de l'ADN ligaturé par un adaptateur 46
2.5. Amplification par PCR 47
2.5.1. Nettoyage du produit PCR 48
2.6. Hybridation et capture 49
2.7. Elution 51
2.7.1. PCR Post-Capture : 51
2.8. Dénaturation et circularisation simple brin 52
2.8.1. Dénaturation 52
2.8.2. Circularisation monocaténaire 52
2.8.3. Digestion 53
2.8.4. Nettoyage du produit de digestion 53
2.8.5. Contrôle qualité du produit de digestion 54
3. Séquençage sur la plateforme MGI 54
4. Données brutes : fichiers FASTQ 54
5. Analyse bioinformatique 55
5.1. Alignement des lectures 55
5.2. Appel des variants (Variant Calling) 55
5.3. Fichiers VCF 55
6. Analyse diagnostique 55
6.1. Annotation des variants 55
6.2. Interprétation clinique 55

Partie 3 : Résultats et discussion

1. Résultats 57

1.1. Matériel biologique de départ 57

1.2. Préparation de la librairie 57

1.3. Paramètre de séquençage 57

1.4. Analyse bioinformatique 58

1.5. Données brutes : fichiers FASTQ 58

1.6. Alignement des lectures 59

1.7. Appel des variants (Variant Calling) 59

1.8. Annotation des variants 59

1.9. Fichiers VCF 60

1.10. Interprétation clinique 60

2. Discussion 61

Conclusion 63

Rèfèrences

Annexes

Résumé

Abstract

ملخص

Introduction

Introduction

La génomique représente un domaine de la biologie contemporaine. L'avancement des technologies de séquençage a rendu possible le séquençage et l'analyse de différents organismes étudiés séparément. Le génome représente la totalité du matériel génétique d'un individu ou d'une espèce qui est codé dans son ADN. La génomique fait référence à l'analyse des génomes. Elle se rapporte au séquençage intégral du code génétique (ADN) des êtres vivants. Il s'agit d'une discipline qui étudie la façon dont l'intégralité du génome interagit avec le milieu environnant.

Dans ce contexte, L'analyse génomique consiste à séquencer d'immenses volumes d'ADN et par conséquent, elle génère une quantité considérable de données qui peuvent être stockées, structurées et accessibles. L'émergence des techniques de séquençage de nouvelle génération a marqué le début d'une autre révolution. Le développement du domaine technique de la bio-informatique, qui implique l'utilisation de la science et de la technologie informatiques pour gérer l'information biologique, est largement attribuable aux progrès dans le domaine de la génomique.

La génomiques se séparé en deux branches distinctes :

- La génomique structurale : l'analyse de la séquence et de structure des composants du génome, tels que les gènes, les régulateurs et les éléments mobiles, est appelée génomique structurale. l'objectif de la génomique structurale est d'approfondir notre compréhension fondamentale des macromolécules biologiques tout en réduisant les coûts moyens associés à la détermination de leur structure.
- La génomique fonctionnelle est un domaine qui englobe des recherches en biologie moléculaire et cellulaire, et se concentre sur la totalité de la structure, la fonction et le contrôle d'un gène. La génomique fonctionnelle se concentre sur l'analyse à grande échelle du génome pour comprendre la fonction de l'ADN (y compris les gènes). Elle s'appuie sur des données génomiques pour examiner l'expression des gènes et des protéines à une échelle globale, ainsi que leurs rôles fonctionnels. **(1)**

D'autre part, Le séquençage à haut débit pan génomique facilite l'identification des variations codants (destinées au séquençage de l'exome) et/ou non codants (liées au séquençage du génome) sur la totalité du génome. C'est un instrument d'une efficacité remarquable pour le diagnostic des maladies mono géniques, qui a transformé l'approche du diagnostic des pathologies génétiques rares. Si la maîtrise des outils techniques est acquise, pour l'heure, dans l'interprétation des nombreux variants identifiés et dans l'exploitation des données générées. Effectivement, chaque étude produit des dizaines de milliers de variations génétiques pour chaque individu. Bien que

certaines soient à l'origine de maladies, la grande majorité n'est pas responsable de telles affections, mais met en évidence la diversité interindividuelle propre à l'espèce humaine. **(2)**

Ces cinq dernières années, le séquençage du génome complet (WGS) et le séquençage de l'exome complet (WES) ont progressivement gagné en popularité pour établir le diagnostic étiologique dont on soupçonne des maladies génétiques. Le WGS et le WES, en autorisant l'analyse simultanée de la majorité ou de la totalité des gènes du diagnostic différentiel, offrent la possibilité d'établir un diagnostic exhaustif et rapide **(3)**. De plus, les avancées technologiques en matière de séquençage de nouvelle génération (NGS) ont considérablement simplifié le diagnostic de ces maladies monogéniques. Au fil de dernière décennie, les technologies du génome ont progressé mathématiquement et les méthodes de séquençage nouvelle génération sont maintenant couramment employées pour diagnostiquer les maladies héréditaires à un seul gène et à plusieurs gènes. Le taux de diagnostic du NGS entre 20 et 60 % selon divers éléments, y compris la génétique et l'hétérogénéité allélique de la pathologie, les critères de sélection des patients, l'aspect cliniques, les technologies de séquençage employées ainsi que le processus analytique du laboratoire. **(4)**

Ainsi**,** en l'espace d'une dizaine d'années, le NGS a permis l'identification de plus de 6000 gènes causant autant de maladies génétiques « rares », rendant possible le diagnostic extrêmement précis de pathologies autrefois non déterminées. Plus le cause de la maladie sont identifiées et plus le patient pourra bénéficier d'une pris en charge adaptée. **(5)**

Ce travail vise à analyser l'ensemble de l'exome en utilisant les techniques de séquençage de nouvelle génération (NGS), dans le but d'identifier les variants génétiques impliqués dans les maladies rares et d'en étudier l'intérêt dans le diagnostic en cadre médicale.

La partie thèorique organisèe en trois chapitres:

- Le premier : prèsente le gènome humain , sa strucutre et son organisation.
- Le deuxième : prèsente les maladies hèrèditaires , leurs origines , les types des mutations et le traitement proposè.
- Le troisième : dècrit les principales techniques de sèquençage , en concentre sur les techniques de sèquençage de nouvelle gènèration (NGS) , leurs principes et leurs applications dans diffèrents domaines .

Références de l'introduction

(1). Bellouber, A. bendouma, H. (2024). Analyse génomique de Caldicoprobacter algeriensis TH7C (Mémoire de Master, Université de Ghardaïa, Faculté des Sciences de la nature et de la vie et Sciences de la Terre, Département de Biologie, Algérie).p(15-16)

(2). Amiel, J. Héron, D. Isidore, B. (2024). Sèquençage de l'exome ou du gènome et donnèes additionnelles Quels enjeux pour la génétique médicale.p(377).

(3). Clark, M.M, Stark, Z., Franaes, L., Tann, T. Y., whit, M., Dimmock, D., Kings more, S.F. (2018). Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal micriarraay in chikdren with suspected genetic diseases. Npj Gnomic Medicine.p(1-2).

(4). Saglia, C., Bracciamà, V., Trott, L., Mioli, F., Corso Faini, A., Brach Del Prever, G. M., kalantari, S., luca, M., Roeo, C.M., Scolari, C., peruzzi, L., calvo, P,L., Mussa, A., Fenoglio, R. , Raccatello, D., Alberti, C., Carli, D., Aoroso, A., Deaglio, S., Vaisitti, T. (2023). Relevance (NGS) data re-analysis in the diagnosis of monogenic diseases leading to organ failure. BMC Médical Génomiques.16, 303, p(2).

(5). Jacob, A. (2020). Apport du séquençage d'exome pour le diagnostic et le traitement des maladies complexes (Thèse de doctorat, Université de Lille).p(41-42).

Partie 1

Chapitre 01 :

Le génome Humain

1. Le génome Humain

1.1. Définition

Le génome humain est défini à fois comme ″l'ensemble complet des gènes de chaque individu au double sens de matériel génétique (molécules d'ADN) et d'information génétique et comme l'ensemble des gènes qui constituent l'espèce humain″. Par conséquent, le génome humain est au sens large le génome collectif de l'espèce humaine, et comprend à la fois le matériel génétique et les données génétiques **(1)**. En effet, il se trouve principalement dans le noyau de chaque cellule, réparti sur 46 chromosome composés de la molécule d'ADN **(2)**. Les 46 chromosomes sont répartis en 23 paires. Pour chaque paire, un chromosome provient de la mère et l'autre du père. Seuls les gamètes, qui sont les cellules reproductrices, possèdent un nombre réduit de 23 chromosomes (une seule copie de chaque paire). **(3)**

Cette ADN est formée par l'agencement des quatre bases (ou lettre) de l'alphabet génétique (A, T, G, C), qui sont interprétées en « mots » de trois lettres. Il contient approximativement 20000 couples de gènes. Il est à la fois profondément similaire (nous avons tous les mêmes gènes situé sur les mêmes chromosomes) et profondément varie d'un individu à l'autre, car pour chaque gène, il y a des centaines de variations différentes. Ces variations sont à l'origine de l'unicité génétique de chaque individu (l'exception des jumeaux monozygotiques). **(2)**

1.1.1. Le génome nucléaire

Le noyau des cellules renferme l'ADN nucléaire, qui est de nature linéaire et se structure en chromatine à l'aide de protéines (histones). Comme indiqué précédemment, ces protéines sont impliquées dans l'emballage et la régulation des longues chaînes d'ADN. C'est à la condensation de cette chromatine que l'on obtient la structure allongée et compacte des chromosomes. L'homme n'est pas l'organisme avec l'ADN le plus étendu. Effectivement, certaines espèces de plantes possèdent jusqu'à trente fois plus d'ADN que l'être humain. **(4)**

1.1.2. Le génome mitochondrial humain

L'ADN mitochondrial (ADN-m) représente une partie du génome de l'homme, l'ADN-m est une petite molécule circulaire d'ADN double brin extranucléaire qui se transmet uniquement par vois maternelle **(5)**. La mitochondrie contient son propre génome, qui est un ADN circulaire et bicaténaire. En quantité absolue, l'ADN mitochondrial constitue approximativement 1% de l'ensemble de l'ADN cellulaire (environ 1000 à 10 000 reproductions par cellule). On observe une variation de 5 à 10 copies par mitochondrie. Chez l'humain, cet ADN se compose de 16 569 paires

de bases (Pb) et ne présente aucun intron. La séquence est totalement identifiée. Il est possible de distinguer les deux brins physiquement en un brin lourd (H) qui contient une grande quantité de bases puriques (G et A) et un brin léger (L) qui est riche en bases pyrimidiques (C et T). L'ADN mitochondrial abrite 37 gènes qui codent pour 13 protéines, 22 ARN de transfert et les ARN ribosomaux. **(6)**

2. Organisation et composition du génome humain

2.1. Organisation structurale du génome humain

2.1.1. Gènes

Le gène est l'unité fondamentale et fonctionnelle de la transmission héréditaire. Les gènes représentent des séquences d'ADN. Quelques gènes fonctionnent comme des modèles pour générer des molécules qui ressemblent à des protéines. Toutefois, bien que la majorité des gènes soient identiques au sein d'une même espèce, une minorité de gènes peut varier (moins de 1% du total). On désigne sous le terme d'allèles les différentes versions d'un même gène. Ces variations subtiles participent aux attributs physiques uniques à chaque individu. **(7)**

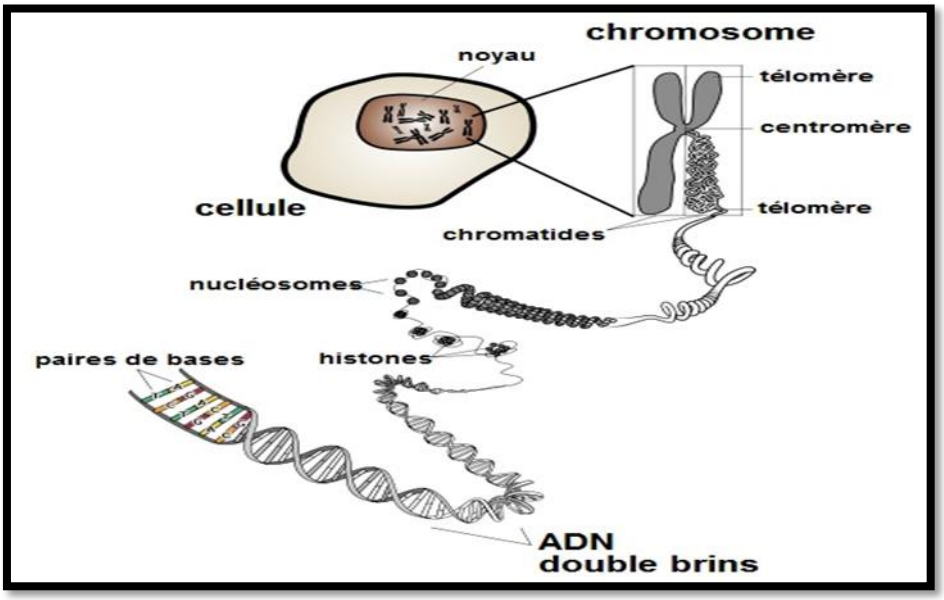


Figure N° 1 : Organisation de l'ADN dans la cellule humain.

Source: https://www.futura-sciences.com/sante/definitions/medecine-adn-87.

2.1.2. Chromatine

La chromatine est une structure macromoléculaire composée d'ADN, de protéines histones et de protéines non histones. Elle facilité l'agencement de la molécule d'ADN et le contrôle de l'expression des gènes dans les cellules eucaryotes. **(8)**

2.1.3. nucléosome

Eléments fondamentaux de la chromatine, se composent d'un octamère d'histones (comportant deux exemplaires de chaque histone H2A, H2B, H3 et H4) et enveloppent environ 147 paires de bases d'ADN autour de l'octamère. **(8)**

2.1.4. Chromosome

Les chromosomes sont formés d'une longue molécule de deux brins d'ADN enroulés en structure hélicoïdale **(3)**. Ainsi, le chromosome est une entité spécifique. L'ADN déplié mesure environ un mètre, tandis que sa taille ans le noyau se situe autour du µm. Cela signifie donc qu'il faut compacter la molécule d'ADN fondamentale pour qu'elle puisse s'adapter au noyau de la cellule. **(4)**

2.2. Composition du génome

2.2.1. les séquences codantes

- **Les exons**

Un exon est la portion du gène qui subsiste dans l'ARN **(9)** qui renferme les données essentielles pour la constitution de la protéine. L'exon fait référence à la partie du gène généralement codante, préservée lors de l'épissage, qui peut subir diverses modifications comme des rétentions ou des sauts susceptibles de modifier son contenu. mature, ce processus esr appelé « épissage de l'ARNm ». **(10)**

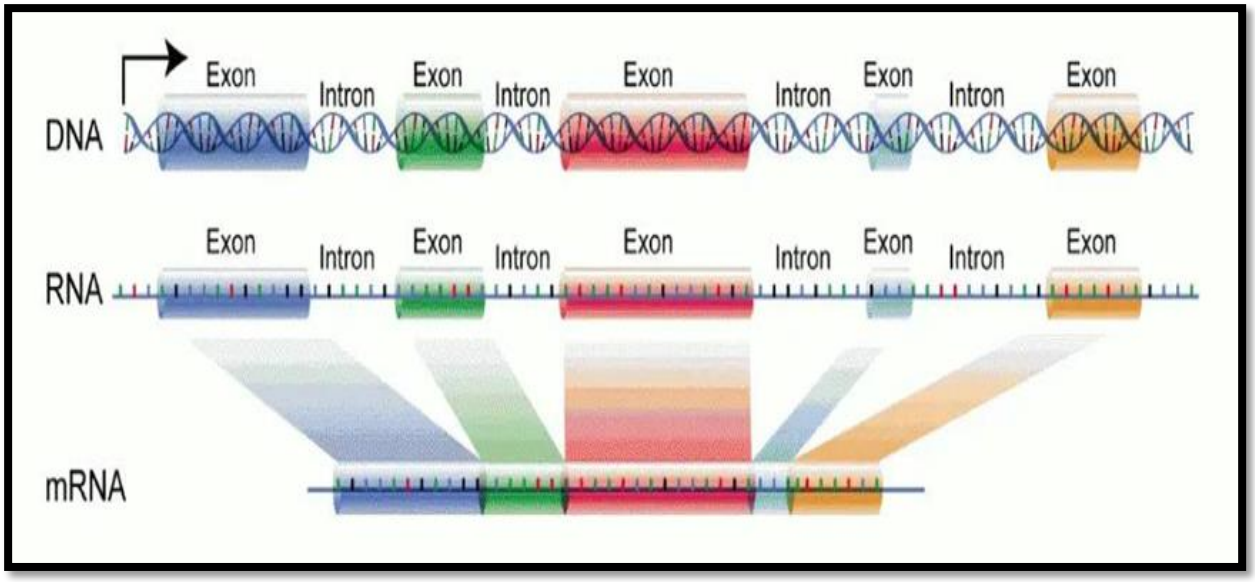


Figure N° 2 : Chéma représentative des exones et les introns.

Source:https://www.zmescience.com/medicine/genetic/intron-retention cancer 25012016.

2.2.2. Les séquences non codantes

2.2.2.1. Les introns

Un intron est une portion du gène qui sera éliminée lors de la transformation des ARN prémessagers nucléaires en ARN mature, ce processus esr appelé « épissage de l'ARNm ». **(9)**

2.2.2.2. Séquences régulatrices

Les séquences régulatrices sont des segments non codants de l'ADN qui ont un impact sur le niveau de transcription des gènes. Elle contient : **(8)**

- Promoteur
- Enhancers

2.2.3. Sequences répétées

On distingue deux sortes de séquences répétées : les séquences répétées en tandem et les séquences dispersées, selon que les répétitions de l'élément répété sont contiguës ou éparpillées. Les premières étant subdivisées en satellites, minisatillites et microsatellites selon la dimension du motif et le nombre de répétitions. Les VNTR (répétitions en tandem de nucléotides) sont formées de répétitions en tandem d'un motif unitaire dont a taille varie d'une paire de bases à plusieurs milliers de paires de bases et qui se répètent un nombre variable n. Un allèle est défini par chaque

valeur de n **(11)**. Pour l'être humain, la totalité des séquences répétées couvre approximativement 40% du génome (Genoscope). **(12)**

2.2.3.1. Microsatellites

Les microsatellites se réfèrent généralement à des séquences de moins de 9 nucléotides répétées entre 2 et 50 fois. **(11)**

2.2.3.2. Mini satellites

Les mini satellites représentent une catégorie spécifique de séquences répétées en tandem. Il s'agit de structures qui se composent de la répétition en tandem d'un motif élémentaire dont la longueur varie entre une dizaine et une centaine de nucléotides : de 7 à 100 paires de bases 10 et 50 paire de bases un mini satellite présente une dimension d'environ un kilo base. **(12)**

2.2.3.3. Satellites

La séquence α satellite, qui peut être répétée sur des longueurs atteignant plusieurs Mb, contitue la majorité de la famille, à ce groupe s'ajoutent les satellites β/γ, ainsi que les satellites 1, 2, 3. **(13)**

2.2.4. Séquences mobiles

Les éléments transposables (TEs) sont des séquences d'ADN capables de modifier leur position dans le génome **(14)** et constituent des répétitions dispersées qui représentent une fraction importante du génome chez les organismes supérieurs. L'apport des éléments transposables dans le génome humain a été révisé à au moins 50% en utilisant les versions récentes du séquençage du génome de référence et des annotations TE. **(15)**

2.2.4.1. Transposons

Les transposons ADN se déplacent par un mécanisme de « coupé et collé » controlé par la cellule elle-même et qui fait intervenir des enzymes codé par le même génome. **(16)**

2.2.4.2. Rotrotransposons

Les rétro transposons se déplacent grâce à un mécanisme de copié-collé semblalbe à celui des transposons via un intermédiaire ARN, un processus de rétro-transposition intervient pour reproduire un ADN à partir de fragments d'ARN. **(16)**

3. L'Exome

L'exome fait référence à l'intégralité des parties codantes du génome, également appelées exons. Ces régions codantes ne constituent que 1à 2%du génome global, ce qui représente environ 62 Mb d'information génétique, impliquées directement dans la physiopathologie des maladies génétiques. **(17)**

Ainsi, en raison de leur importance biologique et clinique, le séquençage de tout l'exome (WES) implique la capture, le séquençage et l'analyse de tous les exons des gènes codant pour des protéines dans le génome humain. Plutôt que d'examiner l'ensemble du génome qui compte environ 3 milliards de paires de bases, le séquençage cible (WES) se concentre sur les quelque 30 millions de paires de bases traduites en protéines fonctionnelles, où les mutations ont le plus chances d'entraîner des conséquences phénotypiques sévères directes. **(18)**

Référence de chapitre 1

(1). Kabata, F. thaldar, D. (2026), the human genome as the common heritage of humanity. Front. Genet. 14: 1282515. Doi: 10.3389/ fgene. 2023. 1282515.p(2).

(2). Manouvrier-Hanu, S. (2011), les diagnostics génétiques : enjeux personnels, familiaux et enjeux de société. Ethique et pratique médicale et siciales, CRHU de Lille, (77),p (33-36).

(3). Dor, E. (2023), la thérapie génique pour soigner les maladies génétiques. Sciences de l'Homme et société. Dumas 0453080, pp(7).

(4). Desgues Cousin, A. (2014). La SNP array appliquée à une sélection de patients au CHU d'Angers : Nouvelle méthode d'analyse de l'ADN (thèse de doctorat en pharmacie, Université d'Agenrs).

(5). Excoffier, L. et Roessli. D, (1990), origine of evolution de l'ADN mitochondrial humain : le paradigme perdu, Bulletins et Mémoires de la Société d'anthropologie de paris, Nouvelle serie, 2(1), p(25-26).

(6). lescuyer, P. (2002). Etude de l'expression des gènes nucléaires codant pour les sous-unités du complexe I mitochondrial humain (thèse de doctorat). Université jseph- Fourier (Grenoble I).

(7). Bousmat, Y. (2022). K-means, k-ers pour le regroupement et la comparaison de grands ensembles de séquences biologiques (mémoire de Master, Université du 8 Mai 1945- Guelma, Faculté des Mathématiques, d'informatique et des Sciences de la Matière).p(9).

(8). Cassinari, k. (2023). Etude de la régulation de l'expression des gènes pour l'interprétation de variations génomiques (thèse de doctorat, Normandie Université).p(25-28).

(9). Meziani, M. (2021-2022). Automatisation d'annotation de la séquence d'ADN du gène cox chez Saccharomyces cervisiae (Mémoire de Master). Université Frères Mentouri Constantine 1, Faculté des Sciences de la Nature et de la vie, Département de Microbiologie, Algérie.p(26).

(10). Clément le Choismier, M. (2023). L'exome en oncogénétique : hérédité manquante et stratégie moléculaire (Mémoire de médecine humaine et pathologie).

(11). kiando, S. (2016). Bases génétiques de la dysplasie fibromusculaire : une approche d'étude d'exome et de génétique épidémiologie (Thèse e doctorat, Université Sorbonne paris cité).p(22).

(12). Bérard, S. coparaison de séquences répétées en tandem et application à la génétique (Doctorat dissertation, Université Montellier 2).p(17).

(13). Ollion, J. (2014). Dynamique de l'organisation nucléaire des séquences d'ADN répétées centromérique humaines au cours du cycle cellulaire (Doctoral dissertions, Université pierre et Marie Curie-paris).p(58).

(14). Autio, M. I., Bin Amin, T., Perrin, A., Wong, J. Y. Foo, R.S-Y., Prabhakar, S. (2021). Transposable that have recently been mobile in the human genome.

(15). Ali, A. Han, k. Liang, P. Role of transposable Elements in Gene Regulation in the Human Genome.p(2).

(16). solyom, S., kazazian, H. H. jr. Mobile elements in the human genome: implications for disease. Review article.p(2).

(17). Lacoste Deixonne, (2016). Apport du séquençage haut débit dans l'amélioration de la prise en charge des maladies monogéniques (thèse de doctorat, université d'Aix-Marseille).p(31).

(18). Bertier, G. Hétu, M. joly, Y. (2016). Unsolved challeges of clinical whole-exome sequencing: A systematic literature review of end-users'views.p(2).

Chapitre 02 :

Les maladies génétiques

1. Les maladies Génétique

Les maladies génétiques sont des affections dues à des irrégularités dans les gènes ou les chromosomes. Il est possible de transmettre certaines anomalies génétiques à la descendance, ce qui fait que ces affections sont considérées comme des maladies familiales. Il s'agit de pathologies rares pouvant se manifester dès la naissance (ou éventuellement plus tard dans l'existence. On peut identifier ces anomalies génétiques à travers l'annotation du génome et l'évaluation du risque de contracter une maladie **(1)**. Ces anomalies peuvent être dues à simple changement de base, de petites insertions ou suppressions, des défauts d'épissage ou des variations de nombre de copies plus importantes qui pourraient être des translocation, des duplications ou des suppressions plus grandes (supérieures à 1 Mo). Cela pourrait même être une addition ou une suppression d'un chromosome entier ou d'une partie de chromosome. La perte de ce matériel génétique provoque des désordres spécifiques ou complexes, générés par les variations dans l'ADN individuel. Les anomalies peuvent être des phénotypes. Ainsi, l'héritage des caractéristique d'une génération à l'autre suit certaines lois proposées par George Mendel, communément appelées héritage mendélien. **(2)**

Les maladies génétiques, bien que rares, comptent plusieurs milliers d'individus et touchent globalement plus de 3% de la population. Elles sont le fruit d'une anomalie dans le capital génétique. Certaines se révèlent très tôt dans la vie (à partir de la phase embryonnaire si l'on parle de malformations), tandis que d'autres surgiront plus ou moins tôt durant l'enfance, ou à l'âge adulte**(3)**. De plus à l'échelle internationale, environ 6% des nouveau-nés présentent des anomalies génétique. Ces anomalies sont le résultat de mutations génomique. Néanmoins, les mutations qui ne sont pas présentes à la naissance peuvent également être à l'origine de maladies génétiques. **(4)**

Les maladies génétiques présentent une grande diversité de manifestations cliniques, c'est-à-dire une importante variabilité clinique ou phénotypique. Elles peuvent se manifester par de l'hypotonie, un développement psychomoteur retardé, une déficience intellectuelle, l'épilepsie, la neuro- régression, des anomalies congénitales, la petite taille, la microcéphalie, les immunodéficiences primaires, la schizophrénie, les troubles du comportement, le trouble déficitaire de l'attention avec hyperactivité, la démence, les mouvements anormaux et le cancer On trouve même des entités, comme la paralysie cérébrale infantile, pour laquelle un facteur génétique n'avait pas été identifié auparavant et il est désormais estimé que jusqu'à 20% des cas ont une origine génétique. Il est essentiel de souligner que les maladies génétiques peuvent se manifester à n'importe quelle étape de la vie, de la période prénatale jusqu'à l'âge adulte. **(5)**

2. Les types des maladies génétiques

Actuellement, on dénombre plus de 7000 maladies génétiques à l'échelle mondiale. Ces affections sont classées de manière différente en fonction des mutations qui les induisent. **(4)**

2.1. Maladies mono géniques (mutation d'un seul gène)

La maladie monogénique implique une mutation dans un seul gène **(4)**, les maladies monogéniques peuvent découler de mutations sporadiques (de novo) ou de modifications dans la séquence d'ADN transmises par l'un ou les deux parents. Le génotype à l'emplacement concerné a un rôle significatif.

Ainsi, si une seule version de l'allèle altéré suffit à provoquer la maladie, on parle d'effet dominant. En revanche, si deux copies de l'allèle muté sont requises pour manifester le phénotype, on parle d'un effet récessif. **(6)**

Par ailleurs, Les troubles monogénitiques présentent diverses modalités d'hérédité génétique, parmi lesquelles : l'hérédité autosomique dominante, où seule une copie du gène défectueux (provenant de l'un ou l'autre parent) est essentielle pour provoquer la condition ; l'hérédité autosomique récessive, qui nécessite deux copies du gène défectueux (une provenant de chaque parent) pour engendrer la condition, et l'hérédité liée à X, où le gène défectueux est situé sur le chromosome X de la mère. L'hérédité liée au chromosome X peut être dominante ou récessive. Parmi les troubles mono géniques, on peut citer la fibrose kystique, l'alpha et la bêta-thalassémie, l'anémie falciforme, le syndrome de Marfan, le syndrome du X fragile, la maladie de Huntington et l'hémochromatose. **(2)**

2.2. Les maladies polygéniques (combinaison de gènes et de facteurs environnementaux)

La maladie polygénique, de loin la plus fréquente, implique des mutations dans plusieurs gènes. L'étiologie des maladies polygéniques, maintenant préférentiellement désignées comme maladies multifactorielles, s'avère fréquemment difficile à établir en raison de leur dépendance à l'égard de mutations multiples affectant divers gènes, ainsi que de l'influence de facteurs environnementaux. Le diabète, l'asthme, les maladies cardiovasculaires et neurologiques, ainsi que certains cancers, peuvent être cités à titre d'exemples de maladies polygéniques. **(4)**

2.3. Les maladies chromosomiques (anomalie dans les chromosomes)

La maladie chromosomique, qui se manifeste par une irrégularité dans la structure d'un chromosome ou par l'existence d'un chromosome supplémentaire (ou à l'inverse, un chromosome manquant), est associée à une mutation chromosomique. La trisomie 21, qui est la maladie la plus

connue de ce groupe, résulte de la présence de trois chromosomes 21 au lieu des deux habituels. **(4)**

Lors d'un trouble chromosomique, on observe généralement une anomalie quantitative du matériel génétique, plutôt qu'une anomalie qualitative. Il y a une surcharge ou un manque de chromosome ou de segment de chromosome, ce qui entraîne une quantité incorrecte de gènes qu'il contient. Cela entraîne des ralentissements ou des accélérations de certaines fonctions métaboliques, généralement plusieurs fonctions seront affectées simultanément. **(7)**

3. La transmission des maladies génétique

Le type de transmission des maladies génétiques diffère en fonction de la mutation et de sa présence ou absence dans le patrimoine génétique des deux parents. Effectivement, l'enfant obtient, pour chaque paire de chromosomes, un chromosome du père et un chromosome de la mère. Par conséquent, il peut hériter ou non d'une mutation en fonction des chromosomes qu'il reçoit. En outre, une mutation peut être transmise par le génome des parents ou bien être de nouvelle apparition. C'est une mutation de dite « de novo », aucun des génomes parentaux ne porte la mutation et celle-ci se manifeste de manière aléatoire. Il faut donc souligner qu'une maladie génétique n'est pas forcément héréditaire. Effectivement, si les gamètes (cellules reproductrices) ne sont pas touchés par la mutation génétique, comme c'est le cas pour les maladies causées par des mutations induites, la maladie ne pourra être transmise. **(4)**

4. Diagnostique des maladies génétiques

On distingue plusieurs types au niveaux de diagnostic pour les troubles génétiques, parmi lesquels : clinique, biochimique, génétique (moléculaire ou cytogénétique).

Le diagnostic génétique débute généralement par l'examen de l'historique familial. L'accumulation d'un trouble au sein des familles indique une cause génétique, bien que certaines maladies infectieuses puissent également se regrouper dans les familles pour des raisons environnementales. Les maladies causées per une mutation dans un gène spécifique donnée (comme les syndromes de cancer familial, l'hémophilie ou les troubles de l'hémoglobine) suivent généralement des schémas d'hérédité classiques (Mendelien), ce qui permet de diagnostiquer une maladie héréditaire même en l'absence de tests de laboratoire appropriés.

Un diagnostic génétique nécessite une confirmation génétique de l'affection de l'individu en identifiant l'anomalie génétique (variant de séquence, variant du nombre de copies ou défaut d'imprédiction) à l'origine du trouble. Cela peut être un diagnostic moléculaire, généralement

signifiant un diagnostic effectué par le biais d'une analyse de gène unique ou d'un diagnostic cytogénétique bassé sur une microarray ou un caryotype. **(8)**

L'utilisation du séquençage de l'ADN dans le diagnostic médical a subi une transformation significative au fil des dernières décennies. L'émergence du séquençage de nouvelle génération (NGS) a grandement accéléré la procédure en rendent possible l'analyse simultanée de milliers, voire de millions de fragments d'ADN. Cette progression a permis d'aboutir à des diagnostics plus exacts et adaptés à chaque individu. **(9)**

5. Analyse de l'exome :

L'analyse de l'exome « Whole Exome Analysis / WES » est une technique de séquençage à haut débit qui repose sur l'étude simultanée de toutes les séquences codantes (et les régions introniques adjacentes) de l'ensemble des gènes du génome (environ 20 000), ce qui représente approximativement 1% du génome, soit près de 180 000 exons et 30 millions de paires de bases**(10)**. Cette méthode ciblant les régions codantes des protéines a permis d'identifier la défaillance sous-jacente, notamment chez les petites familles ou les patients uniques où l'analyse de liaison traditionnelle s'avère difficile. L'approche de séquençage parallèle en utilisant le NGS pour séquencer exomes, loci spécifiques ou génomes a permis d'explorer l'implication des variants rares dans différentes caractéristiques pathologiques. **(11)**

L'analyse de l'exome entier comprend plusieurs étapes, commençant par la collecte d'échantillons, le processus en laboratoire, le séquençage des courtes lectures et l'analyse bio-informatiques. Pour obtenir des données plus solides et reproductibles sur les échantillons, il est essentiel d'optimiser l'ensemble du processus de travail. **(12)**

6. Le traitement des maladies génétiques

Un nombre restreint des maladies génétiques peut être suffisamment bien traité pour permettre l'éducation, l'emploi, le mariage et parfois même la reproduction. Cependant, cela nécessite généralement une gestion coûteuse, onéreuse et à long terme, comme c'est le cas pour les troubles de l'hémoglobine, la phénylcétonurie ou la fibrose kystique. **(13)**

Les thérapies géniques peuvent cibler les cellules germinales (spermatozoïdes ou ovules) ou les cellules somatiques. La durée de l'expression du gène transféré est liée à la nature de la pathologie. Par exemple, dans le cas des maladies mono géniques, il faudrait prolonger la durée alors que pour les maladies multifactorielles (comme le cancer ou les maladies infectieuses), une durée plus courte serait préférable.

On distingue deux formes de transfert génétique : le « in vivo » et «l'ex-vivo». La première implique que le gène est administré directement à un tissu, tandis que dans seconde, les cellules sont prélevées chez le patient, le gène est introduit puis réinséré dans l'individu affecté. On peut classer les types de thérapie génique en ceux qui utilisent la thérapie médiée par des virus, les nanoparticules, et l'édition génique **(5)**. En provoquant une rupture de l'ADN double brin près du lieu où se situe une mutation, ces nucléases peuvent inciter la cellule à réparer les dommages par le biais d'une réparation par modèle en s'appuyant sur un ADN template distinct contenant la séquence souhaitée.**(14)**

Aujourd'hui, il existe de nombreux exemples de thérapie génique (principalement en utilisant des vecteurs viraux pour administrer un acide nucléique thérapeutique dans les cellules d'un patient) qui peuvent adopter divers approches :

1. Complémentation d'un gène défectueux (in vivo ou ex vivo).
2. Suppression ou correction d'un gène défectueux directement au sein de la cellule, techniques connues sous le nom d'édition génomique.
3. Injection de préparations d'ARNm pour la production d'une protéine fonctionnelle. **(9)**

7. Les limites de traitement des maladies génétiques

Bien que des progrès notables soient déjà réalisés, les chercheurs demeurent prudents concernant l'application de la thérapie génique et le risque potentiel d'effets secondaires à long terme. Il est essentiel d'assurer un suivi à long terme des patients sous traitement pour approfondir la compréhension des modes d'action et de l'efficacité de ces médicaments sur une durée étendue, dans le but d'assurer leur sécurité. Il y a donc un manque de recul dans l'observation de ces traitements. **(9)**

8. Les mutations génétiques

Une mutation désigne une modification de la séquence nucléotidique d'une petite portion du génome. La mutation, (un terme inventé par Hugo de Vries en 1900, en redécouvrant les principes de Mendel) désigne à la fois le processus par lequel un gène ou un chromosome se modifie structurellement et le résultat final de ce processus. **(15)**

Les mutations d'un organisme, issues du latin « Mutare », signifiant « Changer », se réfèrent à des modifications des séquences génétiques standards de son ADN. Ces erreurs sont dues à l'intervention d'agents physiques, chimiques ou à des fautes de réplication. Les maladies

génétiques héréditaires et le cancer ont pour cause les mutations, qui peuvent toucher l'ADN aussi bien mitochondrial que génomique.

Il existe deux types de mutations. En fonction de l''étendue de la séquence affectée, cela peut représenter une mutation ponctuelle ou une mutation de grande envergure. Seules les mutations qui se produisent dans les séquences codantes ou de régulation peuvent avoir un impact sur le phénotype. **(16)**

8.1. Les différents types des mutations

8.1.1. Mutation ponctuelles

On entend généralement par mutations ponctuelles des modifications relatives à des paires de bases individuelles de l'ADN ou à un petit ensemble de paires de bases contiguës. Selon leur nature moléculaire, les mutations ponctuelles sont classées. On distingue trois types de mutations ponctuelles : **(16)**

A. Substitutions des bases

Une base est ajoutée de manière incorrecte lors de la réplication et substitue la paire dans la position correspondant sur le brin complémentaire. **(15)**

- **Transition :** consiste à substituer une base par une autre de la même classe chimique, c'est-à-dire, remplacer une purine par une autre purine ou une pyrimidine par une autre pyrimidine.
- **Transvertion :** consiste à substituer une base d'une catégorie, c'est-à-dire, la substitution d'une pyrimidine par une purine ou vice versa. **(16)**

B. Délétion de base

Lors de la réplication, un ou plusieurs nucléotides sont souvent « omis » ou excisés, ce qui entraîne généralement un décalage du cadre de lecture.

C. Addition de base

Des nucléotides supplémentaires sont insérés dans l'ADN en cous de réplication, ce qui entraîne souvent un décalage du cadre de lecture. **(15)**

8.1.2. Mutations spontanées et induites

Les modifications dans le génome peuvent se produire de manière spontanée ou être provoquée.

Ces mutation surviennent suite à l'intervention de certains agents, connus sous le nom de mutagènes, qui rechaussent la fréquence d'apparition des mutations.

8.1.3. Mutations somatiques et germinales

- **Mutation somatique** : Autrement dit, les cellules qui ne sont pas sexuelle prendront fin au plus tard lors du décès de l'individu. Ainsi, elles ne sont pas héritées par la progéniture. Toutefois, si l'entité se reproduit par voie asexuée, une mutation somatique peut être, à condition que la cellule mutée contribue à la formation de ce nouvel organisme.
- **Mutation germinales :** Autrement dit, les mutations qui surviennent dans les cellules à l'origine des gamètes sont à l'inverse, descendants de l'individu. Elle devient donc héréditaire. **(16)**

8.1.4. Mutations chromosomiques

A. Aberrations portant sur un chromosome

- **Inversion :** une zone d'un chromosome est retournée et réinsérée.
- **Délétion :** une portion d'un chromosome est supprimée, entraînant l'absence de tous les gènes dans cette zone.
- **Duplication :** réplication d'un chromosome est répétée, ce qui entraîne une augmentation de la quantité de gènes dans cette zone.
- **Translocation :** une région d'un chromosome est anormalement liée à un autre chromosome.

B. Aberration portant sur deux chromosomes

- **Amplification génique :** l'augmentation du nombre de copies en tandem d'un locus est observée.
- **Répétition de tri-nucléotides en expansion :** le nombre habituel de séquences de triplets nucléotidiques est augmenté. **(15)**

Références de chapitre 2

(1). Boutelela, A. Merouani, H. K. (2023). Exploitation des techniques de Machine learning (apprentissage automatique) pour la prédiction de la maladie de l'anémie falciforme à partir des séquences d'ADN (mémoire de master, université Frères Mentouri Constantine 1).p(1).

(2). Rasool, U., Shrivastava, P. (2022). A review study on genetic disorder diseases. International journal of Health Sciences, 6(S2), 13574-13579. Https:// doi. Org/ 10.53730/ ijhs.v6nS2.8546.p(13574-13575).

(3). Manouvrier-Hanu, S. (2011). Les diagnostics génétiques : enjeux personnels, familiaux et enjeux de société. Actualité et Dossier en Santé publique (adsp), (77), p(34).

(4). Dor, E (2023). La thérapie génique pour soigner les maladies génétique (mémoire de master, université Sorbonne Nouvelle). HAL (dumas-04530380).

(5). Abarca-Barriga, H.H, trubnykova, M, Castro-Mujica, M. d C. (2021). Management of genetic diseases: present and future. Revista de la facultad de medicina Humana, 21(2), p(400).

(6). kiando, S. R. (2016).bases génétique de la dysplaise fibromusculaire : une approche d'étude d'exome et de génétique épidémiologique (thèse de doctorat, université paris Descartes, Sorbonne paris cité).

(7). Dutrillaux, B. (1980). La génétique et les rayonnements ionisants. 1re partie : Génétique et maladies héréditaires. Radioprotection, 15(3),p(117).

(8). Jones, W.D. (2017). Genetic and phenotypic investigations into developmental disorders (Doctoral dissertation, University of Cambridge). University of Cambridge.p(19-25).

(9). Garric, A. (2023). Le séquençage du génome des applications diagnostiques aux traitements de domain. Quels bénéfices concrets nous apporte le séquençage ? Entre espoir et vigilance (Master's thesis, Université de Lille, faculté d'ingénierie et de management de la Santé-ILIS).p(24-31).

(10). Krahn, M., levy, N., Bartoli, M. (2016). Le séquençage de nouvelle génération (Next-Génération Sequencing, ou NGS) appliqué au diagnostic de maladies monogéniques hétérogènes : Notions essentielles pour le dialogue entre cliniciens et généticiens. Cahiers de Myologie, 13,31-33. Publié par EDP Sciences.p(32).

(11). Wong, T. H. (2020). Whole Exome Sequencing in Alzheimer's disease and Front temporal lobar Degeneration (Thèse de doctorat, Erasmus University Rotterdam.p(5-7).

(12). Degli Esposti, C. Optimizing Whole Exome Sequencing for enhanced clinical investigations (Thèse de doctorat, University of Verona, Department of Biotechnology). University of Verona.p(18).

(13). Alwan, A.Modell, B. (2000). Community control of genetic and congenital disorders (EMRO technical publications series NO. 24).world health organization, regional office for the eastern Mediterranean.p(18).

(14). Roth, T. L. Marson, A. (2021); Genetic disease and therapy. Annual Review of pathology: Mechanisms of disease, 16, p(149-151)..

(15). Banoon, S.R., Salih, T. S., Ghasemian, A. (2022). Genetic mutatios and major human disorders: Areview. Egyptian journal of chemistryn, 65(2), p(571-589).

(16). Chihouai, S., Dafri, Z., yekhlaf, R. (2014). Vérification des caractères génétiques des souches d'ames salmonella typhimurium (mémoire de master, université 8 Mai 1945 Guelma, Faculté de sciences de la terre et de l'Univers, Département de SNV).

Chapitre 03 :

Le Séquençage de l'ADN

➢ Généralité sur le séquençage de l'ADN

Le séquençage de l'ADN est une méthode biochimique employée pour définir la disposition précise des quatre éléments chimiques connus sous le nom de bases adénine(A), thymine(T), cytosine(C) et guanine(G) dans la séquence d'ADN (acide désoxyribonucléique). Ainsi, le séquençage de l'ADN délivre des données génétiques contenues dans un fragment d'ADN spécifique, ce qui contribue à approfondir notre compréhension du secteur agricole, des relations évolutives entre les espèces, ou encore du diagnostic de maladies induites par des facteurs génétiques…**(1)**

Grâce aux avancées technologiques et aux outils informatiques contemporains, déterminer la séquence d'un acide nucléique est désormais d'une simplicité remarquable. Le séquençage complet de plusieurs génomes a été accompli, et nous connaissons actuellement plus de 90% du génome humain (qui compte environ 3,5 milliards de paires de bases) **(2)**. Par ailleurs, l'initiative de séquençage du génome humain a commencé en 1990 en tant qu'effort international financé, visant à séquencer le génome humain et ceux de divers organismes modèles de la génétique **(3)**. De plus, le séquençage de l'ADN est indispensable à la caractérisation des variants génétiques du génome humain. Le séquençage de Sanger a longtemps constitué la méthode standard. **(4)**

Aujourd'hui, diverses techniques de séquençage de l'ADN son en place pour établir l'ordre linéaire des bases A, C, G et T qui le constituent. souvent classées en deux catégories en fonction du débit de séquences produite : Haut et bas débit. **(5)**

1. Le séquençage bas débit

1.1. La technologie de « Sanger »

La méthode de Sanger est une technique de séquençage de l'ADN par synthèse enzymatique, qui repose sur l'emploi de didéoxynucleotides ddNTP (qui sont des analogues des nucléotides dATP, dCTP, dGTP, dTTP comportant un atome d'hydrogène H à la place du groupement hydroxyle OH sur le carbone 3 du ribose) ou terminateurs de chaîne **(6)**. Cette technique permettant la détermination de l'ordre des quatre nucléotides au sein d'un fragment d'ADN. Ce séquençage a permis aux scientifiques de séquencer une vaste gamme d'organisme et constituait la méthode privilégiée avant l'avènement des séquenceurs de nouvelle génération **(7)**. Plus précisément la méthode de Sanger utilise l'ADN polymérase pour réaliser une extension d'amorce en incorporant des nucléotides radio-marqués dépourvus de groupe 3'-OH, appelés didéoxynucléotides, dont l'incorporation interrompt l'allongement de l'ADN. **(8)**

La technique de séquençage par méthodes de Sanger a transformé la génétique médicale en déchiffrant et identifiant les sources génétiques de diverses maladies et cancers. Cette méthode, qualifiée de première génération de séquençage, a facilité l'identification d'un grand nombre de gènes responsables de maladies héréditaires, suite aux analyses de liaison génétique. Cette méthode, à la fois robuste et rentable, continue d'être largement employée dans les laboratoires. **(9)**

1.1.1. Les étapes de la technologie de « Sanger »

1. de l'ADN cible pour le séquençage ; génération d'un grand nombre de répliques de séquence d'ADN sélectionnée. Ensuite, l'ADN est dénaturé par la chaleur.
2. Réaction de séquence proprement dite : Consiste à utiliser, dans le mix réactionnel, un mélange de nucléotides normaux (dNTP) et de nucléotides dépourvus de terminaison 3'-OH (ddNTP) marqués par un fluorochrome. Ces derniers provoquent l'arrêt de la polymérisation une fois intégrés, générant ainsi des fragments d'amplicons de différentes tailles.
3. Nous procédons ensuite à la séparation des fragments synthétisés par électrophorèse capillaire en fonction de leur taille (les plus petits migrent plus rapidement).
4. Un détecteur capte la fluorescence au bout du capillaire, ce qui permet de révéler l'ordre des fragments synthétisés. Pour la lecture,un logiciel retrace le chromatogramme permettant la lecture des séquences. L'innovation du marquage par les fluorophores a permis le séquençage d'un plus grand nombre de brins d'ADN. **(7)**

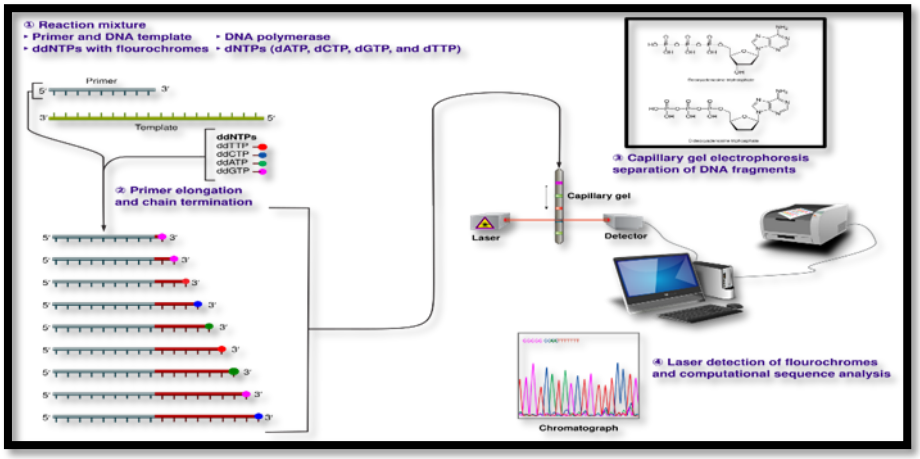


Figure N° 3 : principe de technologie sanger.

Source : Estevezj. (2012). Sanger sequencing. Illustration]. Wikimedia Commons.

1.2. La technolohie de « pyroséquençage »

Le pyroséquençage est une technique de séquençage qui repose sur l'identification du pyrophosphate libéré lors de la réaction de polymérisation de l'ADN. Les nucléotides sont incorporés successivement suivant un ordre préétabli. Quand un nouveau nucléotide est incorporé, l'ATP sulfurylase va exploiter le pyrophosphate libéré pendant la polymérisation pour produire de l'ATP. La luciférase se servira de cet ATP pour oxyder la luciférine en oxyluciférine et produire de la lumière. **(10)**

La méthode implique initialement la division de l'ADN en segment de 300 à 800 Pb. Par la suite, un fragment unique est attaché à une bille et une amplification par PCR en émulsion est effectuée pour générer des séquences identiques de manière clonale. **(11)**

1.2.1. Le principe de « pyroséquençage »

La synthèse du brin complémentaire commence à l'emplacement de l'amorce de l'ADN polymérase, avec le soutien de divers nucléotides introduits progressivement. Quand le nucléotide ajoute correspond au nucléotide prévu, il est intégré au segment d'ADN en cours de création par la polymérase et un groupement pyrophosphate ppi est libéré. Une ATP sulfurylase utilise ce groupement pour la synthèse de l'ATP. L'ATP généré est par la suite décomposé par une luciférase en présence de luciférine pour produire de l'oxyluciférine et un signal lumineux, qui est enregistré en temps réel par une caméra. **(7)**

Cette technologie a l'avantage de produire de longues séquences (330 bases en moyenne), qui facilitent l'assemblage de novo de génome sans utiliser de séquence de référence. **(12)**

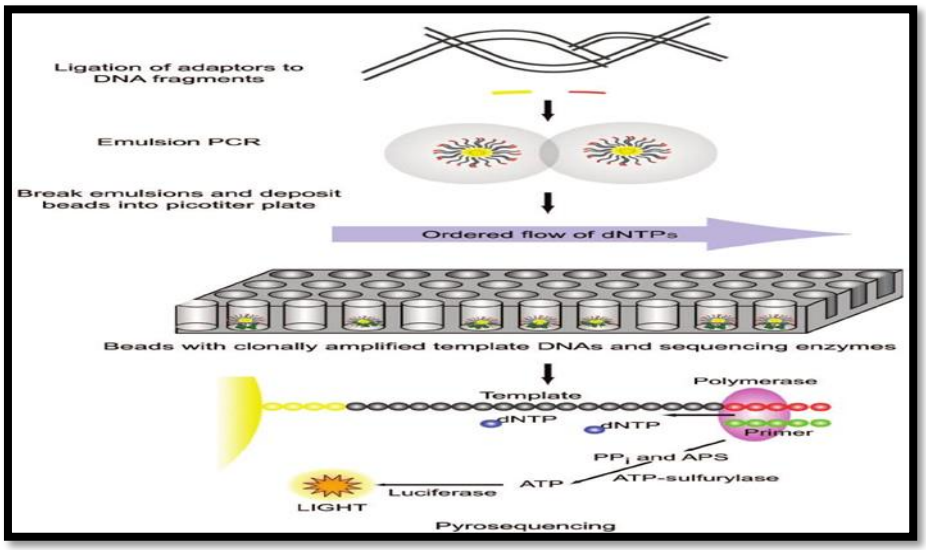


Figure N° 4 : principe de technologie de pyroséquençage.

Source : Voelkerding, K. V., Dames, S. A., & Durtschi, J. D. (2009). Next-generation sequencing: from basic research to diagnostics .

2. Le séquençage à haut débit (NGS)

Le séquençage de nouvelle génération (NGS) est la méthode utilisée pour identifier les séquences d'ARN et d'ADN grâce aux technologie et procédés contemporains **(1)**. Le séquençage NGS représente une série de technique de séquençage qui ont vu le jour dans la première moitié des années 2000. Ces méthodes de lecture d'ADN de manière massives et parallèle, capables de séquencer simultanément des centaines de milliers de fragments. **(4)**

Aujourd'hui, plusieurs entreprises de biotechnologies proposent des plateformes de séquençage de nouvelles générations, parmi les quelles : Illumina, Life Technologie , Nanopore Oxford et MGI. Le séquençage de nouvelle génération, permet une production rapide de données en séquençant d'énormes quantités d'ADN en parallèle. Cette parallélisation a permis d'accélérer la vitesse de séquençage et de réduire les coûts. **(6)**

Trois étapes communes caractérisent les technologies NGS :

- La constitution de banques de fragments d'ADN ou (librairies) : elles sont réalisées en fragmentant de façon aléatoire l'ADN en petits segments d'environ 150 à 600 paires de bases. l'utilisation d'endonucléases ou de sonication permet de réaliser cette fragmentation. Des oligonucléotides liés aux extrémités du fragment d'ADN sont utilisé pour l'adaptation avec des séquences d'amorces et des parqueurs.
- L'amplification : la banque est amplifiée grâce à la PCR ou à des méthodes d'amplification clonale, comme l'amplification en pont utilisée par la plateforme illumina et l'amplification employée par la plateforme ion torrent.
- Le séquençage : l'ADN est séquencée en fonction de la technologie (plateformes) employée, en utilisant diverses approches. **(13)**

2.1. Les technologies les plus utilisées actuellement

2.1.1. La technologie « Illumina »

La technologie de séquençage par synthèse d'illumina, qui repose sur la détection de signaux lumineux lors de l'incorporation successive de nucléotides, permet la production massive de données de séquençage. Le séquençage se réalise à l'aide d'instruments de type MySeq ou NovaSeq d'illumina **(14)**, dans cette technologie, les erreurs majeures proviennent souvent de la

substitution d'une base par une autre, dans cette méthode, l'amplification à examiner ne se fait pas en milieu liquide, mais sur un substrat solide. Cette technologie utilise une amplification en pont des fragments à séquencer (PCR en phase solide) sur une surface de verre connue sous le nom de flow Cell.**(8)**

2.1.1.1. Le principe de « Illumina »

Dans cette technique, l'amplification de l'échantillon à examiner ne se fait pas en milieu liquide, mais sur un substrat solide. On effectue alors la réaction de séquençage directement sur le support où l'ADN a été amplifié. en incorporant un mélange qui contient toutes les bases, chacune associée à un fluorophore distinct. Pour éviter l'ajout de bases supplémentaires à chaque cycle d'incorporation, l'extrémité de ces bases est protégée. Ainsi, une lecture laser permet de repérer en même temps toutes les positions intégrées. L'incorporation de base suivante est alors réalisée à la séparation des fluorophores. La lecture se fait de cette manière, cycle après cycle. **(15)**

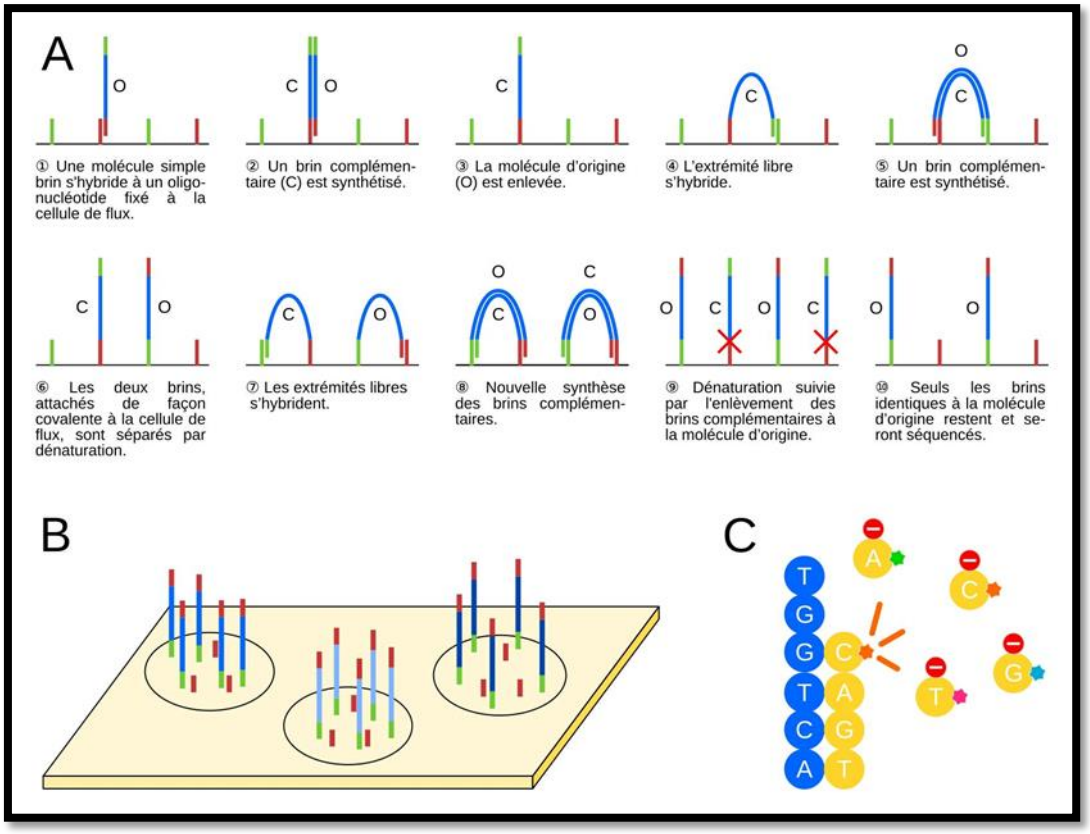


Figure N° 5 : principe de technologie illumina

(A). Amplification on pont des fragments d'ADN sur la cellule de flux.

(B). Amplification simultanée de plusieurs fragments d'ADN.

(C). Réaction de séquençage par synthèse avec incorporation de nucléotides fluorescents et détection par imagerie.

Source : Van Dijk, E., & Thermes, C. (2021, 8 janvier). La révolution de la génomique : les nouvelles méthodes de séquençage et leurs applications. Planet-Vie ENS.

2.1.2. Technologie « Oxford Nano pore»

La technologie de séquençage mise au point par Oxford Nanopore Technologie (ONT) se base sur l'idée d'une lecture directe de l'ADN natif, en temps réel. Cela s'opère via la détection des variations de courant ionique produites lors du mouvement des nucléotides à travers un nanopore, qui est intégré dans une membrane électroniquement chargée. Chaque association de plusieurs bases génère une « signature » perturbation du courant électrique, qui est par la suite transformée en séquence grâce à des algorithmes de base calling. **(16)**

La technologie de séquençage par nanopores est pleine d'espoir, car elle offre la possibilité d'identifier une séquence ADN à la résolution du nucléotide sans amplification préalable grâce à une lecture directe. Il est possible de déterminer les bases et leur statut de méthylation en temps réel avec une précision élevée (99,8%) et à moindre coût, car aucun système optique n'est requis et le traitement des données est minimal. La durée de lecture est également moins longue que celle des autres technologies. **(17)**

2.1.2.1. Le principe de « Oxford Nano pore»

Certaines molécules peuvent provoquer un signal électrique distinct lorsqu'elles interagissent avec des nanopores. Ces nanopores incorporés dans des membres hybrides polymères-lipidiques permettent à l'ADN introduit de les traverser, générant un signal électrique distinct qui est ensuite analysé par un séquenceur **(7)**. Ainsi, une seule molécule d'ADN passe à travers un pore crée par une protéine intégrée dans une bicouche lipidique sous l'influence d'un potentiel électrique. Une exonucléase découpe chaque nucléotide à l'entrée du nanopore, ce qui permet de les détecter électroniquement. Cette méthode n'exige aucune annotation préalable et nécessite seulement de petites quantités d'ADN. **(17)**

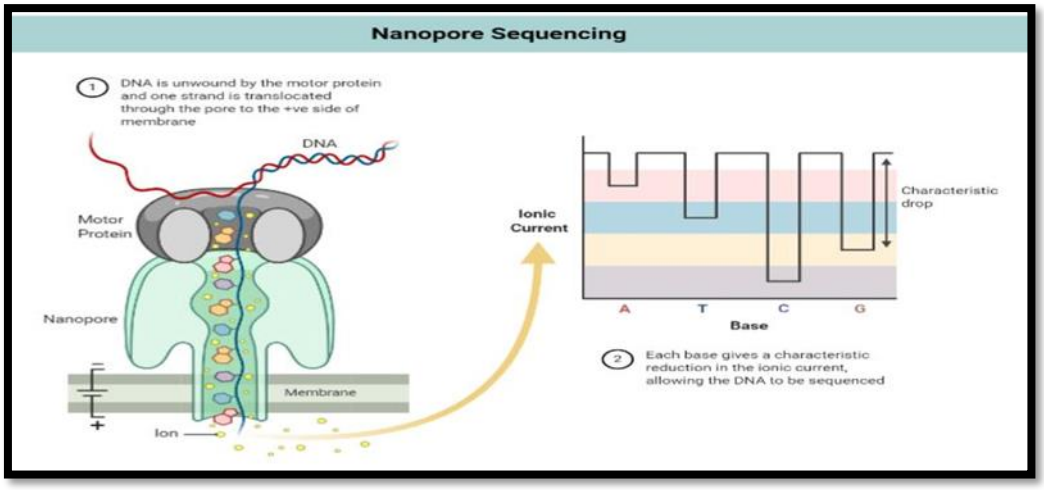


Figure N° 6 : principe de technologie de Oxford Nano pore .

Source : Glencross, F. (2025). Nanopore. Sequencing. BioRender.

2.1.3. Technologie «SOLID»

Depuis 2007, SOLID (sequencing by Oligonucleotide Ligation and Detection) représente la troisième plateforme de séquençage de nouvelle génération mise sur le marché par Applied Biosystems, aujourd'hui connus sous le nom de Life Technologies. La technologie s'appuie aussi sur une PCR en émulsion sur billes. Le séquençage se réalise non pas par synthèse, comme c'était le cas sur les précédentes plateformes, mais par ligation **(7)**. Ainsi, cette technique permet la lecture de jusqu'à un milliard et demi de séquences parallèles comportant 75 bases chacune. L'utilisation de la ligase couplée au système intégré de correction d'erreurs confère une grande fiabilité à cette technologie. **(18)**

2.1.3.1. Le principe de «SOLID»

La technologie Solid utilise une procédure d'amplification comparable à celle de Roche 454, cependant la méthode de séquençage adoptée diffère. Des billes sont placées sur une lamelle de verre et l'ordre est établi grâce à une hybridation et une ligation séquentielle quasi aléatoire d'oligonucléotides, chaque paire de bases étant détectable par un fluorophore**(19)**. Le séquençage Solide se distingue par sa technique de « double correction de base » pour les réactions de ligation, qui permet d'identifier chaque base à deux reprises. De plus, l'emploi des ligases et la modification de l'amorce de séquençage réduisent considérablement le risque d'erreurs de séquençage, faisant du Solid la plateforme la plus précise parmi les technologies de séquençage nouvelle génération. **(20)**

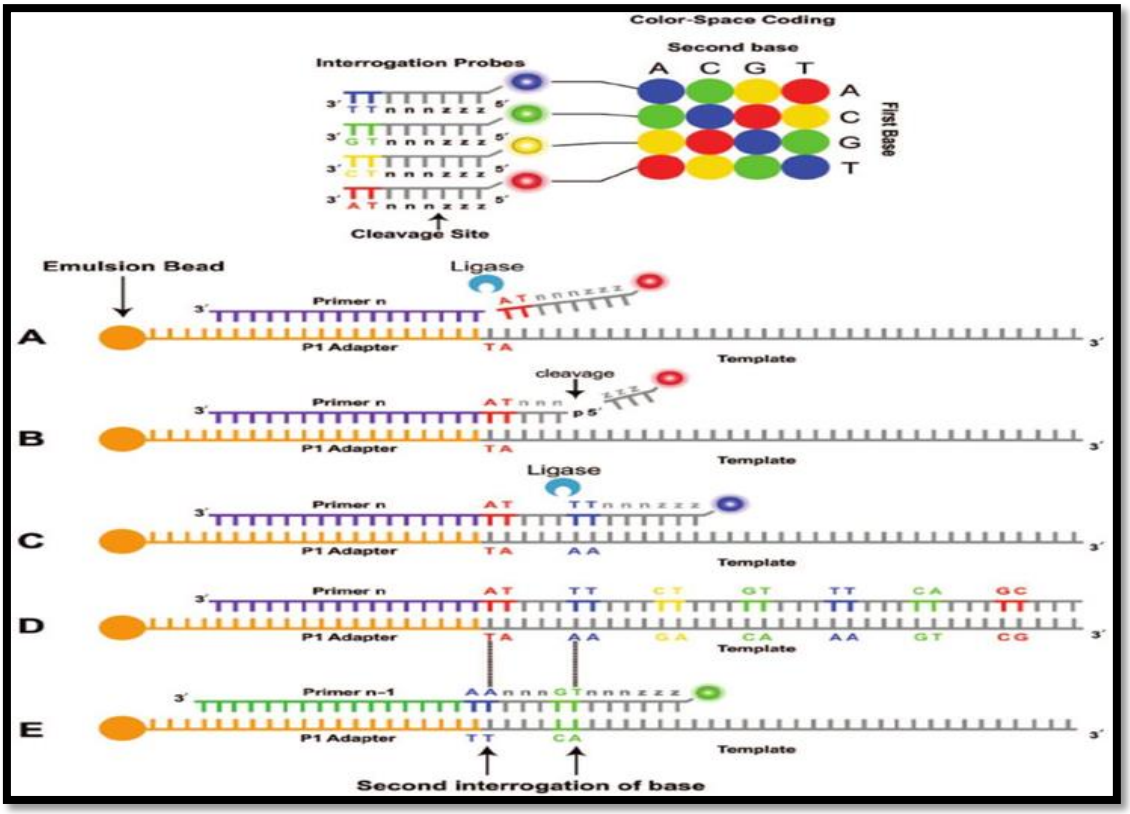


Figure N° 7 : principe de technologie de solide .

Source : Voelkerding, K. V., Dames, S. A., & Durtschi, J. D. (2009). Next-generation sequencing: from basic research to diagnostics.

2.1.4. Technologie « MGI (DNB SEQ) »

La technologie MGI DNBSEQ a été utilisée dans le cadre pratique de ce travail.

DNBSEQ est une technologie NGS qui utilise l'amplification par cycle déroulant pour produire des DNB ou DNA Nano balls. Ces DNBs sont chargés sur un réseau structuré et fortement photographiés grâce à un système de microscopie à fluorescence à très haute résolution. Le débit de DNBSEQ est influencé à la fois par la densité des DNB sur la puce (Flow Cell), et le coût temporel de chaque étape du flux de travail conçu. **(21)**

2.1.4.1. Le principe de la technologie MGI (DNB SEQ)

Selon MGI Tech, l'élaboration de la librairie DNBSEQ implique de convertir l'ADN en fragments simple brin circulaire (ssCirDNA), suivi d'une amplification en cercle roulant (RCA, pour Rolling Circle Amplification) qui requiert une polymérase et une amorce produisant une séquence continue et stable. Cette séquence génère une pelote d'ADN clonale nommée DNB « DNB Nano Ball », contenant environ 300 à 500 copies dérivées d'une seule copie initiale.

Des milliers de DNB sont simultanément insérés dans la puce de séquençage afin d'hybrider à des zones spécifiques de 220 nm de large espacées tous les 715 nm. Ces zones sont enrichies chacune par un DNB. Par la suite, l'identification des bases des séquences sera effectuée par ligation, incluant à chaque cycle l'intégration de dNTP fluorescents et de terminateurs qui seront stimulés par un laser afin déterminer la séquence. **(11)**

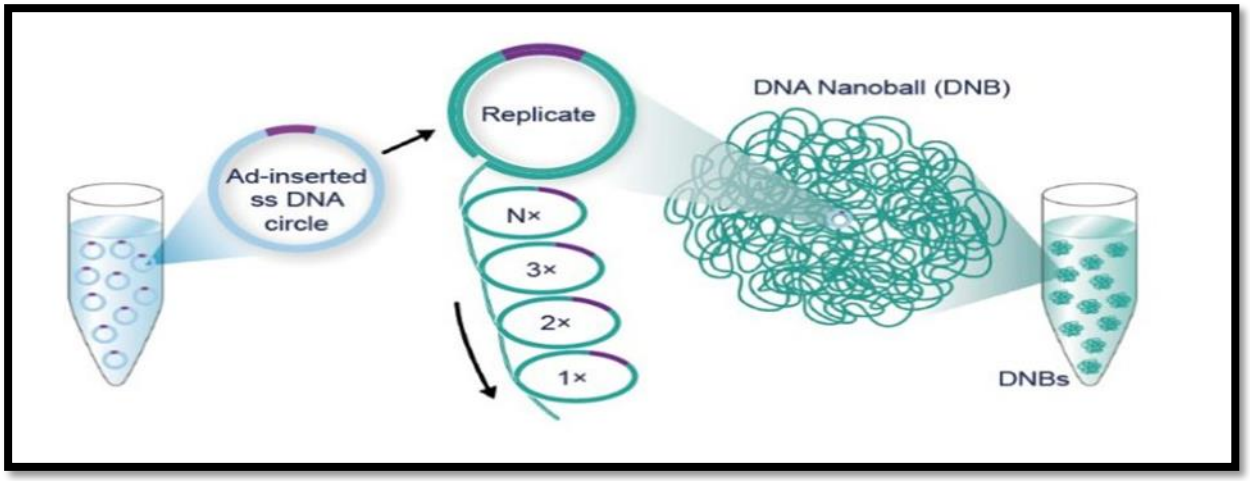


Figure N° 8 : Principe de génération des DNB dans la technologie DNBSEQ**.**

Source : MGI Tech, DNBSEQ™ Technology, https://global-mgitech.com.

3. Les applications liées au séquençage de l'ADN

L'analyse de la séquence d'un brin d'ADN, qui est propre et distincte à chaque individu, offre la possibilité d'examiner les données biologiques associées à chaque individu qu'elle recèle. Le séquençage possède donc de multiples applications dans le secteur médical. Parmi les nombreux exemples, on peut citer le diagnostic, les analyses génétiques (par exemple, les tests de paternité), ou encore la criminologie. **(5)**

3.1. Les applications médicales

- recherche sur les maladies complexes, en particulier le cancer.
- Analyse du statut mutationnel.
- Identification des variations du nombre de copies.
- Identification de plusieurs variations génétiques dans un seul test, offrant une perspective plus large du génotype.
- Analyse épigénétique.

- Prévention et détection précoce du cancer **(22)**

  - profilage génomique des individus porteurs de syndromes de susceptibilité au cancer héréditaire.

  - méthodes basées sur la biopsie liquide pour la détection précoce du cancer. **(23)**

3.2. Applications techniques

- Découverte de variants par le séquençage de régions ciblées ou de génomes complets.
- Assemblage de novo de génomes bactériens et de génomes bactériens et d'organismes eucaryotes inférieurs.
- Profilage à l'échelle du génome des marques épi génétiques et de la structure de la chromatine en utilisant :
  - Chip-seq
  - Méthyl-séquençage
  - DNase-seq
- Classification des espèces et / ou découverte de gènes par des études métagénomiques. **(24)**

3.3. Les applications on microbiologie

- découverte de médicaments ces modifications auront un impact bénéfique sur toutes les applications de la technologie du séquençage dans la découverte de médicaments basée sur les oligonucléotides.
- Les TGSTs pourraient également revêtir une importance particulière dans l'étude des mécanismes d'action des antibiotiques, de leurs effets sur les patients et de la résistance aux antibiotiques grâce au séquençage complet du génome microbien. **(25)**

3.4 Application métagénomique

- L'analyse méta génomique des organismes dans un environnement spécifique, comme le microbiome intestinale, a d'immenses répercussions. Des recherches ont été menées pour déterminer le lien entre le microbiome intestinal et de nombreuses maladies telles que le diabète, le syndrome de l'intestin irritable, l'obésité, …

On peut recourir au séquençage de nouvelle génération ciblé de la région du cluster 16s-23s RNA pour caractériser les bactéries à partir d'échantillons cliniques, permettant ainsi la détection et l'analyse de l'ensemble de la population microbienne présente dans un échantillon clinique. **(26)**

Référence de chapitre 3

(1). MELIANI, I. BOUDEMAGH, I. (2021). Traitement des données générées par le séquençage NGS du génome du SARS-Cov-2. Mémoire de Master, Université Frères Mentouri Constantine 1, Algérie.p(8).

(2). Kamoun, P., Lavoire, A., Vermeil, H., Darmoun, M., Demotes-Mainaed, j. Biochimie et biologie moléculaire clinique. Paris : Médecine-Sciences flammarion.

(3). Klug, W., Cummings, M. R., Spencer, C.A. (2007). Génétique (8e éd)/ Pearson Education.

(4). Kiando, S. R. (2016). Bases génétique fibromusculaire : une approche d'étude d'exome et de génétique épidémiologique (Thèse de doctorat, Université Sarbonne Paris Cité).p(34-36).

(5). Garric, A. (2026). Le séquençage du génome : des applications diagnostique aux traitements de demain. Quels bénéfices concrets nous apporte le séquençage ? Où se place la limite éthique ? Entre espoir et vigilance (Mémoire de Master, Université de Lille, Faculté d'Ingénierie et de Management de la Santé- ILIS).p(11).

(6). Héliou, A. (2017). Analyse des séquences génomiques : Identification des ARNs circulaires et calcul de l'information négative (Thèse de doctorat, Université Paris- Saclay, école .Polytechnique).

(7). Garnaud, C. (2013). Détection de mutations associées à la résistance aux antifongiques chez Candida spp. Par pyroséquençage haut débit (Thèse de doctorat en pharmacie, Université joseph Fourier, Faculté de Pharmacie de Grenoble, France).

(8). Karek, S. (2019). Séquençage : intérêt et impacte en amélioration des plantes – cas du blé (Mémoire de master, Université des fréres Mantouri Constantine 1, Faculté des Sciences de la nature et de la vie, Algérie)p(13-14).

(9). Largarde, A. (2021). Apport du séquençage à haut débit associé aux identifiants moléculaires uniques dans la prise en charge des tumeurs de la loge hypophysaire (Thèse de doctorat en génétique, Université d'Aix- Marseille, France).p(18-20).

(10). Bruto, M ; (2010). Etude de la plasticité génomique chez Streptomyces ambofaciens : assemblage et analyse comparative du génome des souches ATCC23877 et DSM40697 (Mémoire de master en microbiologie, Université Henri poincaré Nancy 1, France).

(11). Andreani, J. (2024). Le séquençage de nouvelle génération en microbiologie : exemles d'applications dans le contexte de la biologie médicale. Sciences pharmaceutiques. Université. Disponible sur HAL : dumas -0463519.

(12). Collet, A. (2013). L'utilisation d'une plateforme de séquençage à haut débit dans l'étude des gènes BRCA2. Sciences du vivant.p(25).

(13). Amouche, S. Benlamri, A. S. (2023). Une approche de novo pour l'assemblage des génomes (Mémoire de master en bioinformatique, Université Frères Mentouri Constantine 1, Faculté des Sciences de la Nature et de la Vie, Algérie).p(11-12).

(14). Faucher, M. (2019). Le séquençage de l'exome pour le diagnostic des anomalies du développement chez 110 patients dans le cadre du protocole FIND (Thèse d'exercice en médecine, Université Claude Bernard Loyon 1, France).p(26).

(15). Le Crom, S., Calude, J.-R., Domenjoud, L., Fattal, D. (2011). Le séquençage à haut débit : méthodes et enjeux en médecine, pharmacologie et toxicologie (Concept Paper, Version 1, 8 décembre 2011). Direction de l'évaluation des Médicaments et des produits Biologiques, Département de Toxicologie, Afssas, France.

(16). Tardy, C. (2026). Analyse comparative des méthodes de séquençage long-read pour décrypter les génotypes complexes dans le locus de la dystrophie facio-scapulo-humérale (Thèse, Sciences du Vivant). HAL Open Sciences.

(17). Sengenès, J. (2012). Développement de méthodes de séquençage de seconde génération pour l'analyse des profils de méthylation de l'ADN (Thèse de doctorat, Université Pierre et Marie Curie Paris).p(23-24).

(18). Agence française de sécurité sanitaire des produits de santé (Afssaps). (2011). Le séquençage à haut débit : méthodes et enjeux en médecine, pharmacologie et toxicologie (Concept paper, France.p(3).

(19). Vannier, T. (2017). Dynamique de la structure des génomes et de leur biogéographie dans l'océan : analyses comparatives des données métagénomiques du projet Tara Oceans pour l'étude de la microalgue Bathycoccus et des communautés planctoniques globales (Thèse de doctorat, Université Paris-Saclay, préparée à l'Université d'évry Val d'Essonne, école Doctorat n° 577, Bioinfomatique). Université Paris- Saclay.

(20). Wang, Z. High-throughput Sequencing Technology and its application in Epigenetics studies. Nanjing Agricultural University, Nanjing, China. Publié par EDP Sciences.p(3).

(21). Li, j., Zhai, Z., Zhang, H., Su, Z., liu, Y., Chen, H., li, Y., Shen, M. (2024). Deep learning enables the use of ultra-high-density array in DNBSEQ. Scientific Reports, 14, 27847.

(22). Deneke, Y., Getahun, T., Desa, G. (2019). Review on Next Generation DNA Sequencing: Historical Emergence, Advantages and Applications. Global jornal of Biotechnology, Biochemistry, 14(1),p(5-6).

(23). Patrik, K., Chdwuick, F. (2025). Genomic Sequencing, personalized Medicine. Cân Mind Eng Live, 331,p(17).

(24). Metzke, M. L. (2010). Applications of Next-Generation Sequencing. Watlik Reviews in Genetics, Volume 1. Macmillan publishing limited.p(42-44).

(25). Aman, S., Yusuf, A. (2020). Review on different types of sequencing technologies and their application; Merit and demerit. Global Veterinaria, 22 (1), p(9-18).

(26). Bhaskaram, S., Saikumaer, C. (2022). A review of next generation sequencing methods and its applications in laboratory diagnosis. Journal of pure and Applied Microbiology, 16(2),p(825-833)

Partie 2 :

Matériel et méthodes

1. Matériel et Méthodes:

❖ Instruments

- Séquenceur d'ADN NGS : DNB seq G99, MGI
- Spectrophotomètre à fluorescence : QUBIT4, Thermofisher scientific
- Thermocycleur : NEOS 48 q PCR, LINEAR
- Sonicateur : M220 Focused-ultrasonicator, COVARIS
- PSM 080 : TMLab
- Rack magnétique : MagSorb 16, BIOSAN

2. Méthodes

La préparation de la librairie de l'ADN pour le séquençage passe par plusieurs étapes représentées par le schéma suivant :

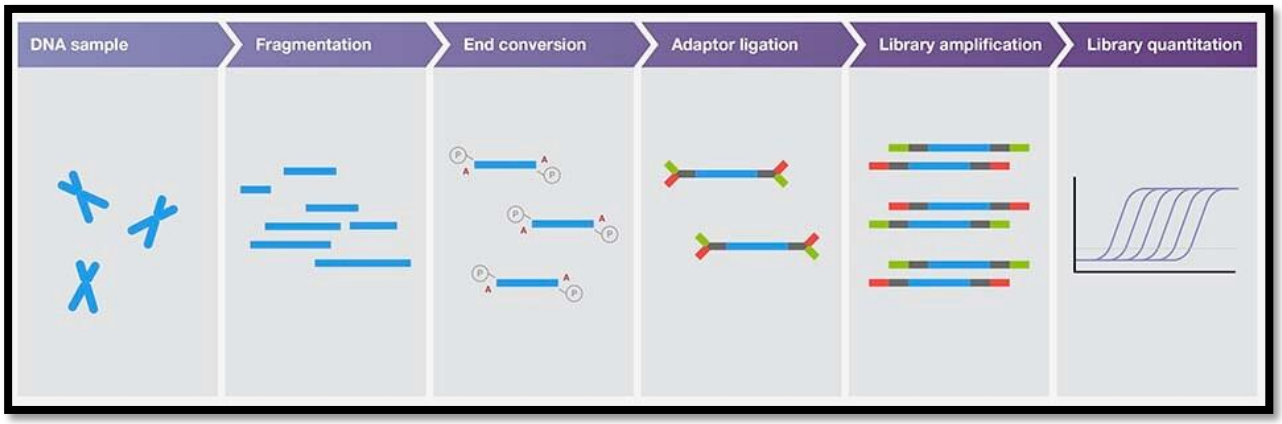


Figure N° 9 : Schéma des étapes de préparation de la librairie d'ADN pour le séquençage NGS.

Une sélection des séquences cibles est ensuite réalisée par hybridation et capture :

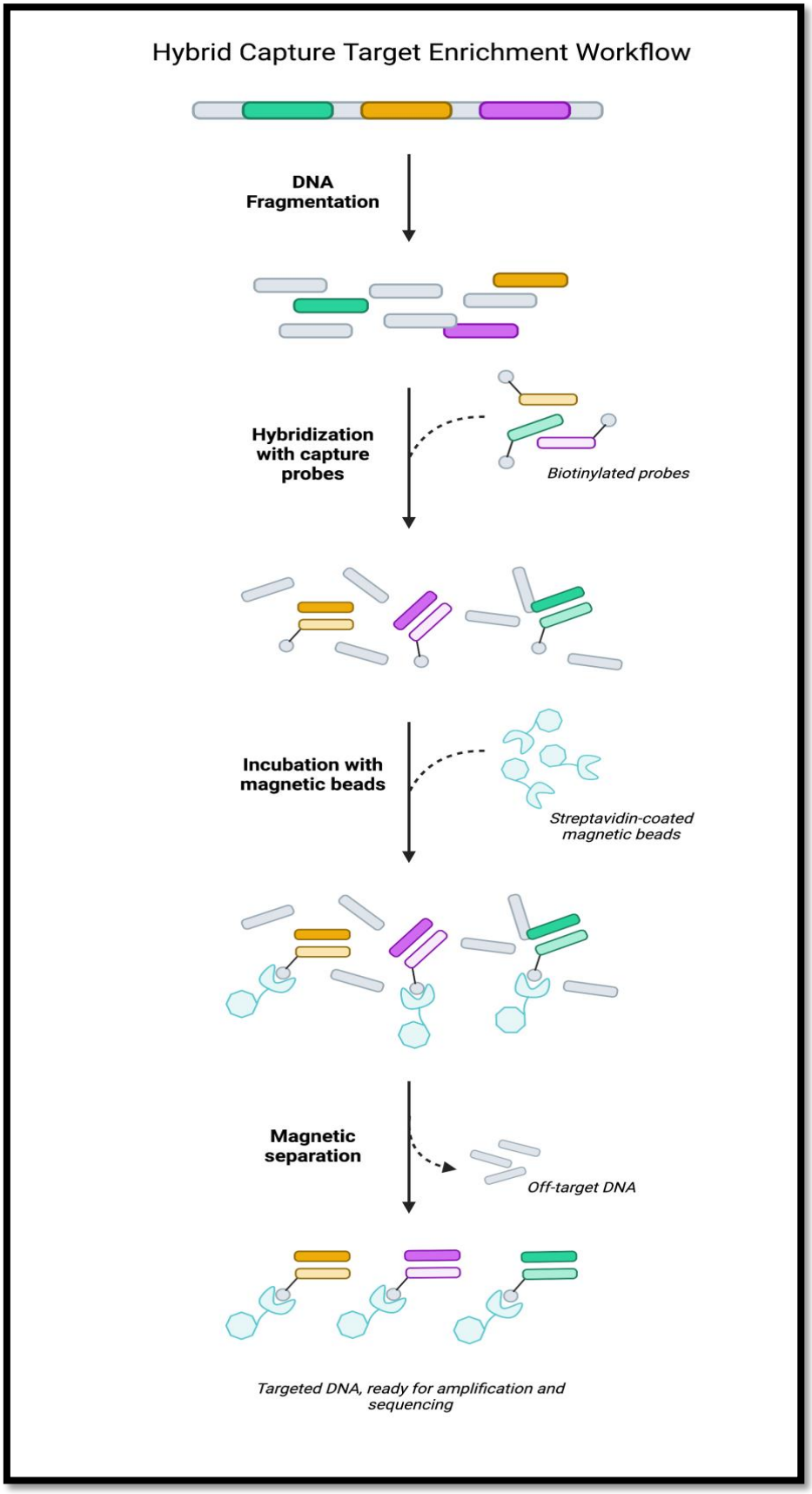


Figure N° 10 : Schéma du processus d'enrichissement des régions cibles par capture hybride (Hybrid Capture) en séquençage NGS.

Une fois les fragments cibles sélectionnés, vient l'étape de la formation des DNBs (DNA nano-bals), leur charge sur les puces et le séquençage comme indiqué sur le schéma suivant :

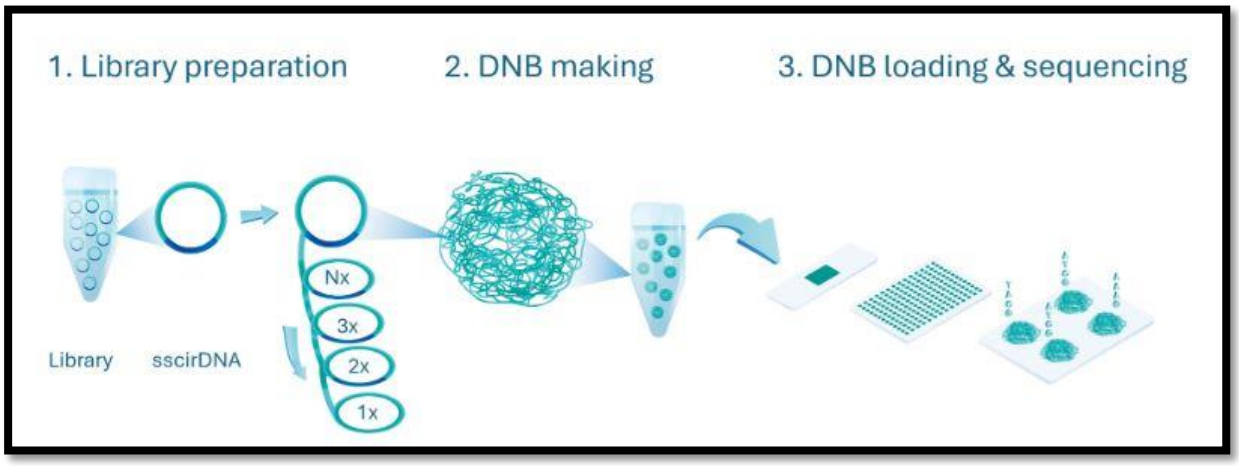


Figure N° 11 : Schéma des étapes du séquençage NGS par technologie DNBSEQ (préparation de la librairie, formation des DNBs et séquençage)

En pratique, voici les étapes réalisées depuis l'échantillon et jusqu'au séquençage final :

- **Echantillon biologique :**

3-5 mL de sang total de patient sont prélevés dans un tube hépariné. Le sans est centrifugé pendant 10 min à 1000g. 200µL de la couche correspondant aux cellules lymphocytaires est récupérée dans un tube eppendorf.

2.1. Extraction et purification de l'ADN

L'extraction est réalisée à partir de 200µL d'échantillon biologique humain.

Lyse cellulaire : Les cellules sont lysées à l'aide de tampons contenant des agents chaotropiques et des détergents (ex. : guanidinium thiocyanate + Triton X-100).

Ajout des billes magnétiques : Les billes sont ajoutées au lysat. L'ADN s'y fixe en présence du sel chaotropique.

Séparation magnétique : L'échantillon est placé à proximité d'un aimant. Les billes (avec l'ADN lié) sont attirées contre la paroi du tube. Le surnageant est éliminé.

Lavages : Les billes sont lavées plusieurs fois avec des tampons d'éthanol pour éliminer les impuretés (protéines, sels, détergent).

Élution de l'ADN : L'ADN est ensuite libéré des billes en utilisant un tampon faiblement salin ou de l'eau ultra-pure, les billes sont retirées à l'aide de l'aimant.

2.2. Fragmentation de l'ADN

Le sonicateur Covaris M 220 est utilisé pour fragmenter un volume de 55 µL d'échantillon d'ADN génomique. Les fragments produits ont une taille comprise entre 300 et 500 pb.

2.2.1. Sélection des Fragments d'ADN en fonction de leurs tailles :

La sélection de la taille des fragments d'ADN est réalisées par billes magnétiques et ce en jouant sur la proportion des billes par rapport au volume de l'échantillon d'ADN. Deux sélections successives permettent d'isoler les fragments à taille souhaitées. **Le tableau N°01** montre la taille des fragments isolés à partir de 100µL d'ADN fragmenté en fonction de la quantité de billes magnétiques :

Tableau N° 1 : la taille des fragments isolés à partir de 100µL d'ADN fragmenté en fonction de la quantité de billes magnétiques

Taille cible (ADN pb)	**180**	**230**	**280**	**335**	**420**	**550**
1ère selections par billes mag	100	90	80	70	60	50
2ème selections par billes mag	50	20	20	20	20	20

Dans notre cas, nous avons ciblé des fragments de 420 pb ce qui correspond à un mix de 100µL d'échantillons + 60 µL de billes pour la première sélection, et 100µL d'échantillon + 20 µL de billes pour la seconde sélection.

2.2.2. Etapes de la sélection des fragments :

Transférer l'ADN fragmenté dans un nouveau tube 1.5mL puis centrifuge brièvement. Compléter jusqu'à 100µL avec du tampon TE.

Ajouter 60 µL de billes magnétiques dans le tube à échantillon. Mélanger au vortex jusqu'à ce que toutes les billes soient en suspension.

Incuber à pendant 5 minutes.

Centrifuger brièvement le tube et le placer sur le support magnétique jusqu'à ce que le liquide soit clair. Transférer ensuite délicatement le surnageant dans un nouveau tube à centrifuger de 1,5 ml. Conserver le surnageant et jeter les billes.

Ajouter des billes d'ADN propres au tube à centrifuger contenant le surnageant. Mélanger au vortex jusqu'à ce que toutes les billes soient en suspension. Incuber à température ambiante pour 5minutes.

Centrifuger brièvement le tube et le placer sur le support magnétique jusqu'à ce que le liquide soit clair. Retirer et éliminer soigneusement le surnageant.

Tout en maintenant le tube à centrifuger sur le support magnétique, ajoutez de l'éthanol à 80% dans chaque tube pour laver les billes et la paroi du tube. Attendez. Retirez et jetez soigneusement le surnageant.

Répétez l'étape et essayez de retirer tout le liquide du tube.

Conservez le tube sur le support magnétique. Ouvrez le bouchon du tube et laissez sécher les billes à l'air libre jusqu'à ce qu'elles ne soient plus humides ni brillantes. Aucune fissure ne doit apparaître à la surface des billes.

Retirez le tube du support magnétique et ajoutez 32 µL de TE pour éluer l'ADN. Mélanger au vortex jusqu'à ce que toutes les billes soient en suspension.

Incuber le tube température ambiante pour 5 minutes.

Centrifuger brièvement le tube et le placer sur le support magnétique pendant 2 à 5 minutes jusqu'à ce que le liquide soit clair. Transférer délicatement. 30 µL du surnageant dans un nouveau tube à centrifuger de 1,5 ml.

- **Quantification de l'ADN**

La Quantification est réalisée avec le spectro-fluoromètre Qubit**.**

2.3. Réparation de l'extrémité des fragments et ajout de la queue « A »

Transférez l'échantillon d'ADN dans un nouveau tube PCR de 0,2 ml et ajoutez du tampon TE pour un volume total de 39,5 µL.

Préparez le mélange de réparation d'extrémité et de queue A sur la glace :

Tableau N° 2 : mélange pour la réparation d'extrémité et de queue A

	Composants	**Volume**
1	ERAT Buffer	6 µL
2	Mélange d'enzyme ERAT	4,5 µL
	Total	**10,5 µL**

Transfert 10,5 µL du mélange dans le tube PCR de 0,2 ml de l'étape précédente. Vortexer 3 fois et centrifuger brièvement.

Placez le tube PCR de 0,2 ml dans le thermocycleur et exécutez le programme.

	Température	**Temps**
1	Couvercle chauffant (105°C)	Oui
2	4°C	30 minutes
3	65°C	15minutes
4	20 °C	Pause

Centrifuger brièvement pour recueillir la solution au fond du tube.

2.4. Ligature de l'adaptateur

Ajouter 5 µL de l'adaptateur UDB au tube d'échantillon. Mélangez 3 fois au vortex, centrifuger brièvement et placer sur la glace.

Préparez le mélange de ligature de l'adaptateur. Agiter 6 fois au vortex puis centrifuger brièvement et placer sur la glace.

Tableau N° 3 : Le mélange de ligature de l'adaptateur

	Composant	**Volume**
1	ADN ligase	4Ul
2	Tampon de ligature	21uL
	Total	**25uL**

2.4.1. Mélange de ligature d'adaptateur

Pipeter lentement 25 µL du mélange de ligature de l'adaptateur dans le tube d'échantillon. Vortexer 6 fois, centrifuger brièvement et placer sur la glace.

Le mélange de ligature de l'adaptateur est très visqueux. Pipeter lentement et avec précaution.

Placez le tube PCR dans le thermocycleur. Exécutez le programme dans les conditions suivantes.

	Température	**Temps**
1	Couvercle chauffant (50 C°)	Oui
2	23 C°	30 minutes
3	4 C°	Pause

2.4.2. Programme de ligature d'adaptateur

Une fois le programme terminé, centrifugez brièvement le tube PCR et placez-les sur la glace.

Ajouter 20 µL Tampon TE, pour un volume total de 100 µL et transférer dans un nouveau tube à centrifuger de 1,5 ml.

2.4.3. Nettoyage de l'ADN ligaturé par un adaptateur

Ajouter 50 µL de billes magnétiques propres dans chaque tube à échantillon. Mélanger au vortex jusqu'à ce que toutes les billes soient en suspension.

Incuber le tube à échantillon à température ambiante pendant 5 minutes

Centrifuger brièvement le tube à échantillon et le placer sur le support magnétique pendant 2 à 5 minutes jusqu'à ce que le liquide soit clair. Retirer et éliminer soigneusement le surnageant.

Gardez le tube sur le support magnétique, ajoutez 200 µL d'éthanol à 80 % dans le tube pour laver les billes et la paroi du tube. Attendre 30 secondes puis retirez et jetez soigneusement le surnageant.

Répétez l'étape de lavage une fois. Essayez d'éliminer tout le liquide du tube. S'il reste du liquide sur la paroi du tube, centrifugez-le brièvement et placez-le sur le support magnétique pour séparation. Éliminez tout le liquide à l'aide d'une pipette à faible volume.

Laissez le tube sur le support magnétique. Ouvrez le bouchon du tube et laissez sécher les billes à l'air libre. Température ambiante jusqu'à ce qu'il n'y ait plus d'humidité ni de brillance à la surface des perles. Aucune fissure ne doit apparaître à la surface des perles.

Retirez le tube du support magnétique et ajoutez 40 µL de tampon TE pour éluer l'ADN. Mélanger au vortex jusqu'à ce que toutes les billes soient en suspension.

Incuber le tube à température ambiante pour 5 minutes.

Centrifuger brièvement le tube et le placer sur le support magnétique pendant 2 à 5 minutes jusqu'à ce que le liquide soit clair. Transférer délicatement 38 µL du surnageant dans un nouveau tube PCR de 0,2 mL.

2.5. Amplification par PCR

Réaliser le mélange selon le tableau suivant :

Tableau N° 4 : le mélange Amplification par PCR

	Composants	**Volume**
1	produit de ligature	38 µl
2	Mélange d'enzymes PCR	50 µl
3	Mélange d'amorces PCR UDB	12 µL
	Total	**100 µL**

Placez le tube PCR dans le thermocycleur. Exécutez le programme dans les conditions suivantes :

	Température	Temps	Cycles
1	Couvercle chauffant 105°C	Oui	
2	95°C	3 min	1
3	98°C	20 sec	×20
4	60°C	15 sec	×20
5	72°C	30 sec	×20
6	72°C	10 minutes	1
7	4°C		

Une fois le programme terminé, centrifugez brièvement le tube et transférez-le dans un nouveau tube à centrifuger de 1,5 ml.

2.5.1. Nettoyage du produit PCR

Ajouter 100 µL de billes magnétiques dans chaque tube d'échantillon. Mélanger avec un vortex jusqu'à ce que toutes les billes soient en suspension.

Incuber le tube à échantillon à Température ambiante pour 5 minutes.

Centrifuger brièvement le tube à échantillon et le placer sur le support magnétique pendant 2 à 5 minutes jusqu'à ce que le liquide soit clair. Retirer et éliminer soigneusement le surnageant.

Gardez le tube sur le support magnétique, ajoutez 200 µL d'éthanol à 80 % dans chaque tube pour laver les billes et la paroi du tube. Attendre 30 secondes, puis retirez et jetez soigneusement le surnageant.

Répétez l'étape de lavage une fois. Essayez d'éliminer tout le liquide du tube. S'il reste du liquide sur la paroi du tube, centrifugez-le brièvement et placez-le sur le support magnétique pour séparation.

Éliminez tout le liquide à l'aide d'une pipette à faible volume.

Laissez le tube sur le support magnétique. Ouvrez le bouchon du tube et laissez sécher les billes à l'air libre à température ambiante jusqu'à ce qu'il n'y ait plus d'humidité ni de brillance à la surface des billes. Aucune fissure ne doit apparaître à la surface des billes.

Retirez le tube du support magnétique et ajoutez 32 µL de tampon TE pour éluer l'ADN. Mélanger au vortex jusqu'à ce que toutes les billes soient en suspension.

Incuber le tube à température ambiante pour 5 minutes.

Centrifuger brièvement le tube et le placer sur le support magnétique pendant 2 à 5 minutes jusqu'à ce que le liquide soit clair. Transférer délicatement. 30 µL du surnageant dans un nouveau tube à centrifuger de 1,5 ml.

Quantifier l'ADN double brin avec le kit d'analyse Qubit dsDNA-HS. Le rendement des produits PCR doit être supérieur à 1 pmol.

2.6. Hybridation et capture

Cette étape permet de sélectionner (hybridation) et d'isoler (capture) les fragments d'ADN correspondant aux séquences exomiques.

Préparation du mélange d'hybridation :

Préparer le mélange suivant dans un tube à PCR de 0,2 mL :

Tableau N° 5 : Préparation du mélange d'hybridation

Composant	**Volume**
Hyb Buffer 1	10µL
Hyb Buffer 1	0,4µL
Hyb Buffer 3	4µL
Hyb Buffer 4	5,6µL
Total	**20**

Incuber le mélange à 65°C pendant 5 minutes.

Préparation du mélange de capture :

Préparer le mélange de capture suivant dans un nouveau tube 0.2 mL.

Tableau N° 6 : Préparation du mélange de capture

Composant	Volume
NF water	1,5µL
BLOCK 5	0,5µL
MGI Exome V5Probe	5µL
Total	**7µL**

Placer le mélange dans un thermocyleur avec le programme suivant :

Température	Time
105°C Heated lid	On
65°C	2min
65°C	Hold

- **Hybridation et capture**

Les deux mélanges (Hybridation et capture) doivent être gardés à une température de 65°C

Transférer 20µL du mélange d'hybridation aux 7 µL du mélange de capture. Mélanger à l'aide d'une pipette.

Placer le tube dans le thermocyleur et lancer le programme suivant pendant 24 heures :

Temperature	Time
105°C	On
65°C	2min
65°C	Hold

2.7. Elution

Pour la réaction d'hybridation, ajouter 1,8mL de Wash Buffer II dans un nouveau tube de 2mL. Placer le mélange dans le thermo-mixeur réglé à 65°C. Puis ajouter 50µL de Billes magnétiques et mélanger l'ensemble.

Ajouter 200µL de Binding buffer au tube contenant le mélange. Placer le tube contre le rack magnétique pour 2-5 minutes et jeter le liquide transparent.

Ajouter 200µL de Binding buffer et resuspendre les billes.

Placer le mélange dans un agitateur rotatif et faire pivoter pendant 30 minutes à température ambiante.

Placer à nouveau les tubes contre un rack magnétique, laisser agir pendant 2-5 minutes et éliminer le surnageant.

Réaliser un lavage avec 500 µL de Wash Buffer.

Réaliser un second lavage avec le Wash Buffer II préchauffé à 65°C. Mélanger à l'aide de l'agitateur magnétique (1000rpm) pendant 10 secondes. Incuber à 65°C pendant 7 minutes.

Placer contre le rack magnétique et éliminer le surnageant.

Répéter cette étape une seconde fois.

Ajouter 44µL de NF water au tube et re-suspendre le mélange et transférer le mélange (avec les billes magnétiques) dans un nouveau tube 0,2 mL.

2.7.1. PCR Post-Capture :

Ajouter 65µL de mixture PCR au tube de l'étape précedente (44µL avec billes magnétiques), bien mélanger et lancer la PCR avec le programme suivant :

Temperatur	**Time**	**Cycles**
105°C Heatide lid	On	-
95°C	3min	1
98°C	20sec	×
60°C	15sec	×
72°C	30sec	×
72°C	10min	1
4°C	Hold	-

A l'issu de la PCR, placer le tube contre le rack magnétique et récupéré le surnageant dans une nouveau tube 1,5mL .

2.8. Dénaturation et circularisation simple brin

Le kit de circularisation MGIEasy (réf. : 1000020570) est utilisé pour la préparation des DNB.

2.8.1. Dénaturation

En fonction de la concentration des produits PCR, ajoutez 300 ng des produits PCR dans un nouveau tube PCR de 0,2 ml. Si le volume est inférieur à 48 µL, ajoutez du tampon TE pour obtenir un volume total de 48 µL.

Placez le tube PCR dans le thermocycleur. Exécutez le programme dans les conditions suivantes:

	Temperature	**Temps**
1	Couvercle chauffant à 100°C	Oui
2	95°C	3 mintes
3	4°C	10 mintes

Après la réaction, centrifuger brièvement le tube et placer sur la glace .

2.8.2. Circularisation monocaténaire

Préparez le mélange de réaction de circularisation dans un nouveau tube PCR de 0,2 mL sur la glace. Vortexez 3 fois puis centrifuger brièvement et placer sur la glace :

Tableau N° 7 : le mélange de réaction de circularisation

	Réactif	**Volume**
1	Tampon d'attelle a doubles code-barres	11,6 uL
2	Ligase rapide d'ADN	0,5 uL
	Total	**12,1 µL**

Ajouter 12,1 µL du mélange réactionnel de circularisation dans le tube à échantillon. Vortexer 3 fois puis centrifuger brièvement et placer sur la glace.

Placez le tube PCR dans le thermocycleur et exécutez le programme dans les conditions suivantes:

	Températures	**Temps**
1	Couvercle chauffant 42c°	Oui
2	37C°	10 Minutes
3	4C°	Pause

Une fois le programme terminé, placez le tube PCR sur la glace puis centrifuger brièvement et passer immédiatement à l'étape suivante.

2.8.3. Digestion

Préparer le mélange de digestion dans un tube PCR de 0,2 mL sur la glace . Vortexez 3 fois, centrifuger brièvement et placer sur la glace.

Ajouter 4 µL du mélange de digestion dans le tube à échantillon. Vortexer 3 fois, centrifuger brièvement, puis placer sur la glace.

Placez le tube PCR dans le thermocycleur. Exécutez le programme dans les conditions suivantes:

	Température	**Temps**
1	Couvercle chauffant 42C°	Oui
2	37C°	10 minutes
3	4C°	Pause

Une fois le programme terminé, centrifugez brièvement le tube et ajoutez immédiatement 7,5µL de tampon d'arrêt de digestion dans le tube à échantillon. Mélanger au vortex 3 fois puis centrifuger brièvement et placer sur la glace.

2.8.4. Nettoyage du produit de digestion

Mélangez soigneusement les billes magnétiques de nettoyage d'ADN. Ajoutez 130 µL de billes d'ADN propres dans le tube à échantillon. Mélanger au vortex jusqu'à ce que toutes les billes soient en suspension.

Incuber à température ambiante pour 5 minutes.

Centrifuger brièvement le tube et le placer sur le support magnétique pendant 5 minutes jusqu'à ce que le liquide soit clair. Retirer et jeter soigneusement le surnageant.

Réaliser un nettoyage avec 160 µL d'éthanol à 80 %.

Placez-le sur le support magnétique pour séparation. Éliminez tout le liquide à l'aide d'une pipette à faible volume.

Conservez le tube sur le support magnétique. Ouvrez le bouchon du tube et laissez sécher les billes à l'air libre à température ambiante jusqu'à ce qu'elles ne présentent plus d'humidité ni de brillance. Aucune fissure ne doit apparaître à la surface des billes.

Retirez le tube du support magnétique et ajoutez 25 µL de tampon TE pour éluer l'ADN. Mélanger au vortex jusqu'à ce que toutes les billes soient en suspension.

Incuber à température ambiante pour 5 minutes.

Centrifuger brièvement le tube et le placer sur le support magnétique pendant 5 minutes jusqu'à ce que le liquide soit clair. Transférer délicatement 24 µL du surnageant dans un nouveau tube à centrifuger de 1,5 mL.

2.8.5. Contrôle qualité du produit de digestion

Quantifier le ssCir avec le kit d'analyse Qubit ssDNA. Les produits finaux de la digestion enzymatique (ssDNA,ng) / les produits d'entrée de la PCR (dsADN, 300 ng) doivent être ≥ 7%.

3. Séquençage sur la plateforme MGI

Les DNBs sont chargés sur des flow cells spécifiques au séquenceur MGI G-99.

Le séquençage est réalisé en utilisant la technologie DNBSEQ, qui repose sur la synthèse d'ADN par ancrage de sondes combinatoires. Cette méthode permet une lecture efficace et précise des séquences d'ADN.

4. Données brutes : fichiers FASTQ

Les données de séquençage sont initialement générées sous forme de fichiers FASTQ, qui contiennent les séquences d'ADN lues ainsi que les scores de qualité associés à chaque base. Ces fichiers sont essentiels pour les analyses bioinformatiques ultérieures.

5. Analyse bioinformatique

5.1. Alignement des lectures

Les lectures des fichiers FASTQ sont alignées sur un génome de référence (GRCh38) en utilisant un outils interne au serveur du séquenceur équivalent à BWA ou Bowtie2. Cette étape génère des fichiers BAM, qui contiennent les informations d'alignement.en.mgi-tech.com

5.2. Appel des variants (Variant Calling)

Les fichiers BAM sont ensuite utilisés pour identifier les variations génétiques (SNPs, indels, etc.) par rapport au génome de référence. Cette étape est réalisée avec un outils interne au serveur du séquenceur équivalent à GATK, Sentieon ou DeepVariant. Les résultats sont enregistrés dans des fichiers VCF (Variant Call Format).

5.3. Fichiers VCF

Les fichiers VCF contiennent des informations détaillées sur les variants identifiés, y compris leur position dans le génome, le type de variant, la qualité de l'appel et d'autres annotations pertinentes. Ces fichiers sont essentiels pour les analyses ultérieures et la recherche clinique.

6. Analyse diagnostique

6.1. Annotation des variants

Les variants identifiés sont annotés pour déterminer leur impact potentiel sur les gènes et les protéines. Des outils comme SnpEff ou VEP (Variant Effect Predictor) sont utilisés pour cette annotation.

6.2. Interprétation clinique

Les variants annotés sont ensuite interprétés en fonction de leur pertinence clinique, en se référant à des bases de données telles que ClinVar ou gnomAD. Cette étape est cruciale pour identifier les variants pathogènes potentiels et établir des liens avec des phénotypes ou des maladies spécifiques. L’annotation des variants et l’interprétation clinique sont pris en charge par l’équipe médicale.

Partie 3 :

Résultats et discussion

1. Résultats

1.1. Matériel biologique de départ

Extrait d'ADN, sang humain hépariné :

Concentration : 25 ng/µL,

Quantité totale : 0,7 µg

1.2. Préparation de la librairie

Quantité d'ADN utilisée : 50ng

Kit de préparation : MGIEasy Fast FS Exome Library Prep

1.3. Paramètre de séquençage

2x100 pb

Contrôle qualité du séquençage : Q30 = 92,03 %

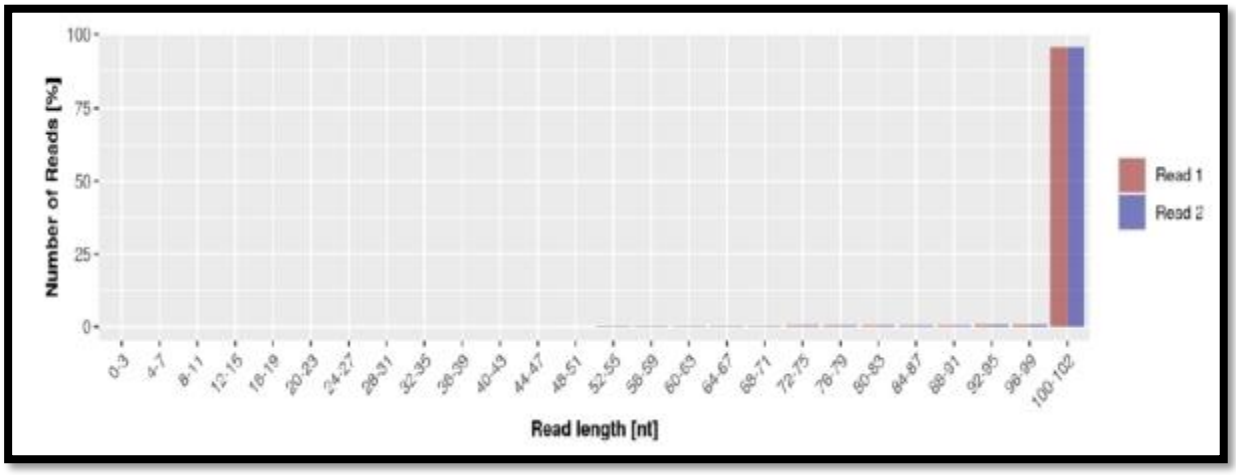


Figure N° 12 : Tailles des séquences lues (Reads) après « Trimming »

La figure n°12 indique que plus de 98% des séquences lues ont une taille comprise entre 98 et 102nt.

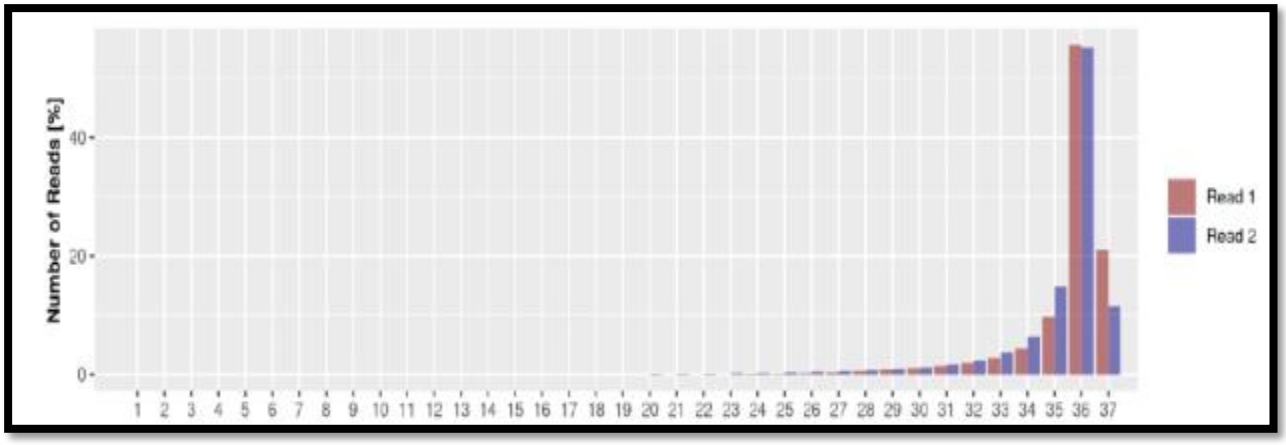


Figure N° 13 : Nombre moyen de lectures (reads) par base

La figure n°13 montre que plus de 80% des bases ont été lues 34-37 fois en moyenne.

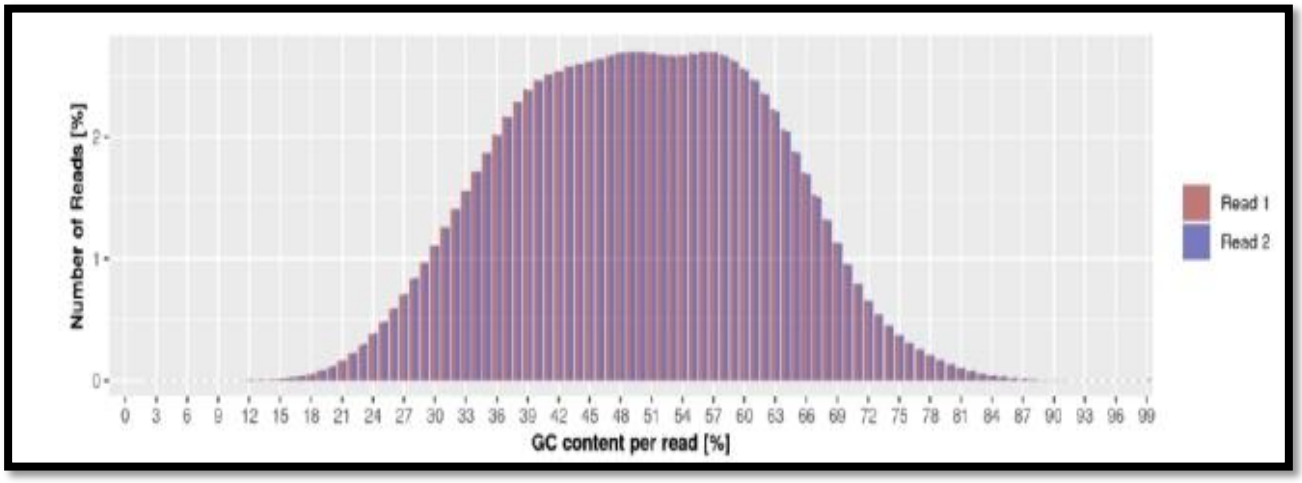


Figure N° 14 : %GC des lectures (reads)

La figure n°14 montre une distribution gaussienne du pourcentage GC des séquences lues centrée à 50%.

1.4. Analyse bioinformatique

Le démultiplexage des séquences et l'ajustage des adaptateurs (trimming) ont été réalisé avec un module associé au serveur interne du séquenceur G-99.

1.5. Données brutes : fichiers FASTQ

Les données de séquençage sont initialement générées sous forme de fichiers FASTQ, qui contiennent les séquences ADN lues ainsi que les scores de qualité associés à chaque base.

Ces fichiers sont essentiels pour les analyses bioinformatiques ultérieures.

Taille de fragments lus (reads) : 43,05 Millions

Nombre de bases lues : 16,278 Gb

1.6. Alignement des lectures

Les lectures des fichiers FASTQ sont alignées sur un génome de référence (GRCh38) via un module interne au serveur du séquenceur G-99. Cette étape génère un fichier BAM (Binary Alignment Map) contenant les informations d'alignement.

1.7. Appel des variants (Variant Calling)

Les fichiers BAM sont ensuite utilisés pour identifier les variations génétiques (SNPs, indels et CNVs) par rapport au génome de référence. Des outils tels que GATK, Sentieon ou DeepVariant sont couramment utilisés pour cette étape. Les résultats sont enregistrés dans des fichiers VCF (Variant Call Format).

Un logiciel interne au serveur du séquenceur G-99 a été utilisé pour la détection des variants.

Appel des variants a été effectué dans les régions cibles avec +/- 30 paires de bases.

Les variants ont été annotés à partir de diverses bases de données publiques. Les variations du nombre de copies (CNV ou copy number variations) ont été détectées en comparant le nombre de lectures chevauchant les régions génomiques cibles « Coverage » avec le nombre attendu dans une cohorte échantillons de référence. Pour analyser les CNVs sur les chromosomes X et Y, le sexe du patient (si non spécifié par le client) a été estimé à partir de la couverture afin de déterminer le nombre attendu de copies.

La qualité des fichiers FASTQ a été analysée avec FastQC (version 0.11.5-cegat) (Andrews 2010). Les tracés ont été créés avec ggplot2 (Wickham 2009) dans R (version 3.6.1) (R Core Team 2015).

Les résultats de cette étape d'analyse sont les suivants :

Nombre SNVs : **48604**

Nombre d'Indels : **9528**

Nombre de CNVs : **491**

1.8. Annotation des variants

Les variants identifiés sont annotés pour déterminer leur impact potentiel sur les gènes et les protéines. Les outils SnpEff et VEP (Variant Effect Predictor) sont utilisés pour cette annotation.

1.9. Fichiers VCF

A l'issu de l'analyse bioinformatique, des fichiers VCF (Variant call format) sont générés contenant des informations détaillées sur les variants identifiés, y compris leur position dans le génome, le type de variant, la qualité de appel et autres annotations pertinentes. Les fichiers VCF peuvent être consultés avec un outil d'édition de texte tels que bloc note ou Word et peuvent contenir un nombre de variants allant de 50000 jusq'à 10000.

Ces fichiers sont essentiels pour les analyses ultérieures et la recherche clinique. Un exemple d'un fichier VCF est annexé à ce manuscrit.

1.10. Interprétation clinique

Cette partie de l'analyse est réalisée par le médecin généticien. Les variants annotés sont interprétés en fonction de leur pertinence clinique, en se référant à des bases de données telles que ClinVar ou gnomAD. Cette étape est cruciale pour identifier les variants pathogènes potentiels et établir des liens avec des phénotypes ou des maladies spécifiques.

2. Discussion

L'analyse de l'exome est un test important pour le diagnostic de nombreuses maladies rares et héréditaires. Dans de nombreux cas, les informations fournies par ce test suffisent pour poser le diagnostic et permettent l'établissement d'un conseil génétique adéquat.

Dans ce travail, nous avons réalisé le séquençage d'un exome dans le but de fournir au médecin généticien une liste de variants génétiques susceptibles d'être à l'origine de la pathologie du malade.

L'ADN extrait à partir d'un prélèvement sanguin du malade à subit une sérié de modifications afin de le préparer pour le séquençage dans la plateforme MGI G-99. Les différents points de contrôle qualité réalisés autorisent le passage d'une étape à la suivante.

La quantité d'ADN de départ (0,7 µg) et sa qualité évaluée par le Qubit sont satisfaisantes pour la suite des expérimentations.

Le score finale du séquençage (Q30=93,03%) est supérieur au seuil de validation des résultats fixé par le fabriquant MGI qui est de l'ordre de 90%.

Le nombre de « reads » (43,05M) et de bases lues (16,278Gb) correspond à la moyenne observée habituellement avec la technologie utilisée. Plus de 80% des bases ont été lues à plus de 34x ce qui est considéré comme acceptable.

La courbe de distribution du contenu en %GC est également dans les normes avec une forme gaussienne centrée à 50%.

Le nombre de variantes totaux (60k) est comparable au nombre obtenu habituellement par des analyses WES similaires.

L'ensemble des analyses bioinformatiques ont été réalisés au sein du serveur intégré au séquenceur G-99.

Le fichier VCF final contenant une liste de variants de plus de 55000 variants a été livré au médecin généticien qui prendra en charge la dernière étape de L'analyse et établir un compte rendu final. Tout le travail technique a été effectué sous la supervisons du médecin généticien compte tenu de l'aspect médical de l'analyse.

Conclusion

Conclusion

Ce travail s'inscrit dans le domaine de la génomique et du diagnostic moléculaire, et a pour objectif le séquençage d'un exome humain dans un cadre médical à la recherche de variants responsables d'une maladie héréditaire à l'aide des technologies de séquençage de nouvelle génération (NGS), une approche moderne devenue incontournable pour l'analyse des variations génétiques impliquées dans les maladies héréditaires.

Ce travail a permis de décrire de manière globale les différentes étapes du processus de séquençage de l'exome, depuis la préparation des échantillons d'ADN et la construction des librairies, jusqu'au séquençage et à l'analyse bioinformatique des données obtenues. Cette démarche met en évidence l'importance de l'intégration entre les techniques de biologie moléculaire et les outils informatiques dans l'étude du génome humain.

Le séquençage de l'exome représente aujourd'hui un outil essentiel en génétique médicale, permettant de cibler les régions codantes du génome et de faciliter la détection des variations génétiques responsables de nombreuses pathologies, ce qui contribue au développement du diagnostic moléculaire et de la médecine de précision.

Cependant, cette technologie présente encore certaines limites, notamment liées à la complexité des données générées et aux difficultés d'interprétation de certains variants, ce qui nécessite une analyse rigoureuse et multidisciplinaire.

En perspective, l'évolution continue des technologies de séquençage et des outils bioinformatiques devrait permettre d'améliorer la précision des analyses et d'élargir davantage les applications cliniques du séquençage de l'exome.

Références

Références

Références de l'introduction

(1). Bellouber, A. bendouma, H. (2024). Analyse génomique de Caldicoprobacter algeriensis TH7C (Mémoire de Master, Université de Ghardaïa, Faculté des Sciences de la nature et de la vie et Sciences de la Terre, Département de Biologie, Algérie).p(15-16).

(2). Amiel, J. Héron, D. Isidore, B. (2024). Quels enjeux pour la génétique médicale.p(377).

(3). Clark, M.M, Stark, Z., Franaes, L., Tann, T. Y., whit, M., Dimmock, D., Kings more, S.F. (2018). Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal micriarraay in chikdren with suspected genetic diseases. Npj Gnomic Medicine.p(1-2).

(4). Saglia, C., Bracciamà, V., Trott, L., Mioli, F., Corso Faini, A., Brach Del Prever, G. M., kalantari, S., luca, M., Roeo, C.M., Scolari, C., peruzzi, L., calvo, P,L., Mussa, A., Fenoglio, R. , Raccatello, D., Alberti, C., Carli, D., Aoroso, A., Deaglio, S., Vaisitti, T. (2023). Relevance (NGS) data re-analysis in the diagnosis of monogenic diseases leading to organ failure. BMC Médical Génomiques.p(2).

(5). Jacob, A. (2020). Apport du séquençage d'exome pour le diagnostic et le traitement des maladies complexes (Thèse de doctorat, Université de Lille).p(41-42).

Référence de chapitre 1

(1). Kabata, F. thaldar, D. (2026), the human genome as the common heritage of humanity. Front. Genet. 14: 1282515. Doi: 10.3389/ fgene. 2023. 1282515.p(2).

(2). Manouvrier-Hanu, S. (2011), les diagnostics génétiques : enjeux personnels, familiaux et enjeux de société. Ethique et pratique médicale et siciales, CRHU de Lille, (77), p(33-36).

(3). Dor, E. (2023), la thérapie génique pour soigner les maladies génétiques. Sciences de l'Homme et société. Dumas 0453080, pp(7).

(4). Desgues Cousin, A. (2014). La SNP array appliquée à une sélection de patients au CHU d'Angers : Nouvelle méthode d'analyse de l'ADN (thèse de doctorat en pharmacie, Université d'Agenrs).

(5). Excoffier, L. et Roessli. D, (1990), origine of evolution de l'ADN mitochondrial humain : le paradigme perdu, Bulletins et Mémoires de la Société d'anthropologie de paris, Nouvelle serie, 2(1),p(25-26).

(6). lescuyer, P. (2002). Etude de l'expression des gènes nucléaires codant pour les sous-unités du complexe I mitochondrial humain (thèse de doctorat). Université jseph- Fourier (Grenoble I).

(7). Bousmat, Y. (2022). K-means, k-ers pour le regroupement et la comparaison de grands ensembles de séquences biologiques (mémoire de Master, Université du 8 Mai 1945- Guelma, Faculté des Mathématiques, d'informatique et des Sciences de la Matière).p(9).

(8). Cassinari, k. (2023). Etude de la régulation de l'expression des gènes pour l'interprétation de variations génomiques (thèse de doctorat, Normandie Université).p(25-28).

(9). Meziani, M. (2021-2022). Automatisation d'annotation de la séquence d'ADN du gène cox chez Saccharomyces cervisiae (Mémoire de Master). Université Frères Mentouri Constantine 1, Faculté des Sciences de la Nature et de la vie, Département de Microbiologie, Algérie.p(26).

(10). Clément le Choismier, M. (2023). L'exome en oncogénétique : hérédité manquante et stratégie moléculaire (Mémoire de médecine humaine et pathologie).

(11). kiando, S. (2016). Bases génétiques de la dysplasie fibromusculaire : une approche d'étude d'exome et de génétique épidémiologie (Thèse e doctorat, Université Sorbonne paris cité).p(22).

(12). Bérard, S. coparaison de séquences répétées en tandem et application à la génétique (Doctorat dissertation, Université Montellier 2).p(17).

(13). Ollion, J. (2014). Dynamique de l'organisation nucléaire des séquences d'ADN répétées centromérique humaines au cours du cycle cellulaire (Doctoral dissertions, Université pierre et Marie Curie-paris).p(58).

(14). Autio, M. I., Bin Amin, T., Perrin, A., Wong, J. Y. Foo, R.S-Y., Prabhakar, S. (2021). Transposable that have recently been mobile in the human genome.

(15). Ali, A. Han, k. Liang, P. Role of transposable Elements in Gene Regulation in the Human Genome.p(2).

(16). solyom, S., kazazian, H. H. jr. Mobile elements in the human genome: implications for disease. Review article.p(2).

(17). Lacoste Deixonne, (2016). Apport du séquençage haut débit dans l'amélioration de la prise en charge des maladies monogéniques (thèse de doctorat, université d'Aix-Marseille).p(31).

(18). Bertier, G. Hétu, M. joly, Y. (2016). Unsolved challeges of clinical whole-exome sequencing: A systematic literature review of end-users'views.p(2).

(19). https://www.futura-sciences.com/sante/definitions/medecine-adn-87.

(20). https://www.zmescience.com/medicine/genetic/intron-retention cancer 25012016**.**

Références de chapitre 2

(1). Boutelela, A. Merouani, H. K. (2023). Exploitation des techniques de Machine learning (apprentissage automatique) pour la prédiction de la maladie de l'anémie falciforme à partir des séquences d'ADN (mémoire de master, université Frères Mentouri Constantine 1).p(1).

(2). Rasool, U., Shrivastava, P. (2022). A review study on genetic disorder diseases. International journal of Health Sciences, 6(S2), 13574-13579. Https:// doi. Org/ 10.53730/ ijhs.v6nS2.8546**.**p(13574-13575).

(3). Manouvrier-Hanu, S. (2011). Les diagnostics génétiques : enjeux personnels, familiaux et enjeux de société. Actualité et Dossier en Santé publique (adsp), (77), p(34).

(4). Dor, E (2023). La thérapie génique pour soigner les maladies génétique (mémoire de master, université Sorbonne Nouvelle). HAL (dumas-04530380).

(5). Abarca-Barriga, H.H, trubnykova, M, Castro-Mujica, M. d C. (2021). Management of genetic diseases: present and future. Revista de la facultad de medicina Humana, 21(2),p(399-400).

(6). kiando, S. R. (2016).bases génétique de la dysplaise fibromusculaire : une approche d'étude d'exome et de génétique épidémiologique (thèse de doctorat, université paris Descartes, Sorbonne paris cité).

(7). Dutrillaux, B. (1980). La génétique et les rayonnements ionisants. 1re partie : Génétique et maladies héréditaires. Radioprotection, 15(3),p(117).

(8). Jones, W.D. (2017). Genetic and phenotypic investigations into developmental disorders (Doctoral dissertation, University of Cambridge). University of Cambridge.p(19-25).

(9). Garric, A. (2023). Le séquençage du génome des applications diagnostiques aux traitements de domain. Quels bénéfices concrets nous apporte le séquençage ? Entre espoir et vigilance (Master's thesis, Université de Lille, faculté d'ingénierie et de management de la Santé-ILIS).p(24-31).

(10). Krahn, M., levy, N., Bartoli, M. (2016). Le séquençage de nouvelle génération (Next-Génération Sequencing, ou NGS) appliqué au diagnostic de maladies monogéniques hétérogènes : Notions essentielles pour le dialogue entre cliniciens et généticiens. Cahiers de Myologie, 13,31-33. Publié par EDP Sciences.p(32).

(11). Wong, T. H. (2020). Whole Exome Sequencing in Alzheimer's disease and Front temporal lobar Degeneration (Thèse de doctorat, Erasmus University Rotterdam. P(5-7).

(12). Degli Esposti, C. Optimizing Whole Exome Sequencing for enhanced clinical investigations (Thèse de doctorat, University of Verona, Department of Biotechnology). University of Verona.p(18).

(13). Alwan, A.Modell, B. (2000). Community control of genetic and congenital disorders (EMRO technical publications series NO. 24).world health organization, regional office for the eastern Mediterranean.p(18).

(14). Roth, T. L. Marson, A. (2021); Genetic disease and therapy. Annual Review of pathology: Mechanisms of disease, 16,p(149-151).

(15). Banoon, S.R., Salih, T. S., Ghasemian, A. (2022). Genetic mutatios and major human disorders: Areview. Egyptian journal of chemistryn, 65(2), p(571-589).

(16). Chihouai, S., Dafri, Z., yekhlaf, R. (2014). Vérification des caractères génétiques des souches d'ames salmonella typhimurium (mémoire de master, université 8 Mai 1945 Guelma, Faculté de sciences de la terre et de l'Univers, Département de SNV).

Référence de chapitre 3

(1). MELIANI, I. BOUDEMAGH, I. (2021). Traitement des données générées par le séquençage NGS du génome du SARS-Cov-2. Mémoire de Master, Université Frères Mentouri Constantine 1, Algérie.p(8).

(2). Kamoun, P., Lavoire, A., Vermeil, H., Darmoun, M., Demotes-Mainaed, j. Biochimie et biologie moléculaire clinique. Paris : Médecine-Sciences flammarion.

(3). Klug, W., Cummings, M. R., Spencer, C.A. (2007). Génétique (8e éd)/ Pearson Education.

(4). Kiando, S. R. (2016). Bases génétique fibromusculaire : une approche d'étude d'exome et de génétique épidémiologique (Thèse de doctorat, Université Sarbonne Paris Cité).p(34-36).

(5). Garric, A. (2026). Le séquençage du génome : des applications diagnostique aux traitements de demain. Quels bénéfices concrets nous apporte le séquençage ? Où se place la limite éthique ? Entre espoir et vigilance (Mémoire de Master, Université de Lille, Faculté d'Ingénierie et de Management de la Santé- ILIS).p(11).

(6). Héliou, A. (2017). Analyse des séquences génomiques : Identification des ARNs circulaires et calcul de l'information négative (Thèse de doctorat, Université Paris- Saclay, école .Polytechnique).

(7). Garnaud, C. (2013). Détection de mutations associées à la résistance aux antifongiques chez Candida spp. Par pyroséquençage haut débit (Thèse de doctorat en pharmacie, Université joseph Fourier, Faculté de Pharmacie de Grenoble, France).

(8). Karek, S. (2019). Séquençage : intérêt et impacte en amélioration des plantes – cas du blé (Mémoire de master, Université des fréres Mantouri Constantine 1, Faculté des Sciences de la nature et de la vie, Algérie).p(13-14).

(9). Largarde, A. (2021). Apport du séquençage à haut débit associé aux identifiants moléculaires uniques dans la prise en charge des tumeurs de la loge hypophysaire (Thèse de doctorat en génétique, Université d'Aix- Marseille, France).p(18-20).

(10). Bruto, M ; (2010). Etude de la plasticité génomique chez Streptomyces ambofaciens : assemblage et analyse comparative du génome des souches ATCC23877 et DSM40697 (Mémoire de master en microbiologie, Université Henri poincaré Nancy 1, France).

(11). Andreani, J. (2024). Le séquençage de nouvelle génération en microbiologie : exemles d'applications dans le contexte de la biologie médicale. Sciences pharmaceutiques. Université. Disponible sur HAL : dumas -0463519.

(12). Collet, A. (2013). L'utilisation d'une plateforme de séquençage à haut débit dans l'étude des gènes BRCA2. Sciences du vivant.p(25).

(13). Amouche, S. Benlamri, A. S. (2023). Une approche de novo pour l'assemblage des génomes (Mémoire de master en bioinformatique, Université Frères Mentouri Constantine 1, Faculté des Sciences de la Nature et de la Vie, Algérie).p(11-12).

(14). Faucher, M. (2019). Le séquençage de l'exome pour le diagnostic des anomalies du développement chez 110 patients dans le cadre du protocole FIND (Thèse d'exercice en médecine, Université Claude Bernard Loyon 1, France).p(26).

(15). Le Crom, S., Calude, J.-R., Domenjoud, L., Fattal, D. (2011). Le séquençage à haut débit : méthodes et enjeux en médecine, pharmacologie et toxicologie (Concept Paper, Version 1, 8 décembre 2011). Direction de l'évaluation des Médicaments et des produits Biologiques, Département de Toxicologie, Afssas, France.

(16). Tardy, C. (2026). Analyse comparative des méthodes de séquençage long-read pour décrypter les génotypes complexes dans le locus de la dystrophie facio-scapulo-humérale (Thèse, Sciences du Vivant). HAL Open Sciences.

(17). Sengenès, J. (2012). Développement de méthodes de séquençage de seconde génération pour l'analyse des profils de méthylation de l'ADN (Thèse de doctorat, Université Pierre et Marie Curie Paris).p(23-24).

(18). Agence française de sécurité sanitaire des produits de santé (Afssaps). (2011). Le séquençage à haut débit : méthodes et enjeux en médecine, pharmacologie et toxicologie (Concept paper, France.p(3).

(19). Vannier, T. (2017). Dynamique de la structure des génomes et de leur biogéographie dans l'océan : analyses comparatives des données métagénomiques du projet Tara Oceans pour l'étude de la microalgue Bathycoccus et des communautés planctoniques globales (Thèse de doctorat, Université Paris-Saclay, préparée à l'Université d'évry Val d'Essonne, école Doctorat n° 577, Bioinfomatique). Université Paris- Saclay.

(20). Wang, Z. High-throughput Sequencing Technology and its application in Epigenetics studies. Nanjing Agricultural University, Nanjing, China. Publié par EDP Sciences.p(3).

(21). Li, j., Zhai, Z., Zhang, H., Su, Z., liu, Y., Chen, H., li, Y., Shen, M. (2024). Deep learning enables the use of ultra-high-density array in DNBSEQ. Scientific Reports, 14, 27847.

(22). Deneke, Y., Getahun, T., Desa, G. (2019). Review on Next Generation DNA Sequencing: Historical Emergence, Advantages and Applications. Global jornal of Biotechnology, Biochemistry, 14(1), 01-07.p(5-6).

(23). Patrik, K., Chdwuick, F. (2025). Genomic Sequencing, personalized Medicine. Cân Mind Eng Live, 331,p(17).

(24). Metzke, M. L. (2010). Applications of Next-Generation Sequencing. Watlik Reviews in Genetics, Volume 1. Macmillan publishing limited. P(42-44).

(25). Aman, S., Yusuf, A. (2020). Review on different types of sequencing technologies and their application; Merit and demerit. Global Veterinaria, 22 (1),p(9-18).

(26). Bhaskaram, S., Saikumaer, C. (2022). A review of next generation sequencing methods and its applications in laboratory diagnosis. Journal of pure and Applied Microbiology, 16(2), p(825-833).

(27). Estevezj. (2012). Sanger sequencing. Illustration]. Wikimedia Commons.

(28). Voelkerding, K. V., Dames, S. A., & Durtschi, J. D. (2009). Next-generation sequencing: from basic research to diagnostics.

(29). Van Dijk, E., & Thermes, C. (2021, 8 janvier). La révolution de la génomique : les nouvelles méthodes de séquençage et leurs applications. Planet-Vie ENS.

(30). Glencross, F. (2025). Nanopore. Sequencing. BioRender.

(31).MGI Tech, DNBSEQ™ Technology, https://global-mgitech.com

Annexes

Annexes

Annexe – Pipeline bioinformatique pour l'analyse du séquençage de l'exome (WES)

📌Objectif :

Analyser les données brutes issues du séquençage (fichiers FASTQ) pour obtenir des fichiers de variants annotés (VCF), interprétables pour le diagnostic.

Étapes du pipeline :

Étape	Outil	Commande / Description
1. Contrôle qualité des lectures	FastQC	fastqc sample_R1.fastq sample_R2.fastq
2. Nettoyage des lectures (trimming)	Trimmomatic	trimmomatic PE sample_R1.fastq sample_R2.fastq [...]
3. Alignement au génome de référence (hg38)	BWA-MEM	bwa mem -t 8 hg38.fa sample_R1_trimmed.fq sample_R2_trimmed.fq > sample.sam
4. Conversion SAM → BAM	Samtools	samtools view -Sb sample.sam > sample.bam
5. Tri et indexation des BAM	Samtools	samtools sort sample.bam -o sample_sorted.bamsamtools index sample_sorted.bam
6. Marquage des duplicats	Picard	picard MarkDuplicates I=sample_sorted.bam O=sample_dedup.bam M=metrics.txt
7. Appel de variants	GATK HaplotypeCaller	gatk HaplotypeCaller -R hg38.fa -I sample_dedup.bam -O sample_raw.vcf
8. Filtrage des variants	GATK VariantFiltration	gatk VariantFiltration -R hg38.fa -V sample_raw.vcf -O sample_filtered.vcf [...]
9. Annotation des variants	ANNOVAR / VEP	table_annovar.pl sample_filtered.vcf humandb/ -buildver hg38 -out sample_annotated

Structure des fichiers en sortie :

- Sample_fastqc.html : rapport qualité
- Sample_sorted.bam / .bai : fichier aligné
- Sample_filtered.vcf : fichier de variants filtrés
- Sample_annotated.txt : variants annotés cliniquement

Remarque :

- Le pipeline est généralement automatisé via un script shell ou un workflow (Snakemake, Nextflow).
- Il est essentiel d'utiliser une base de données de référence à jour (ClinVar, gnomAD) pour l'interprétation clinique.

Exemple fichier VCF :

```
##fileformat=VCFv4.2

##ALT=<ID=*,Description="Represents allele(s) other than observed.">

##FILTER=<ID=HARDTOVALIDATE,Description="QD < 2.0 || FS > 60.0 || MQ < 40.0 || HaplotypeScore > 13.0 ||
MappingQualityRankSum < -12.5 || ReadPosRankSum < -8.0">

##FILTER=<ID=LOWQUAL,Description="Set if not true: %QUAL>10">

*

*

*

*

##samtoolsVersion=1.12+htslib-1.12

##source=HaplotypeCaller
```

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	63327465_S33.FREEBAYES	63327465_S33.GATK	63327465_S33.SAMTOOLS
chrM	73	.	A	G	15853.10	PASS	AB=0;ABP=0;AC=6;AC1=2;AF=1.00;AF1=1;AN=6;AO=683;CIGAR=1X;DP=1785;DP4=0,0,319,1;DPB=683;DPRA=0;EPP=35.4425;EPPR=0;ExcessHet=3.0103;FQ=-281.989;FS=0;GQ_MEAN=1481;GTI=0;LEN=1;MEANALT=1;MLEAC=2;MLEAF=1;MQ0F=0.0106762;MQM=54.757;MQMR=0;MQSB=1;NCC=0;NS=1;NUMALT=1;ODDS=951.444;PAIRED=0.814056;PAIREDR=0;PAO=0;PQA=0;PQR=0;PRO=0;QA=24174;QD=32.09;QR=0;REGION=.;RO=0;RPL=217;RPP=200.131;RPPR=0;RPR=466;RUN=1;SAF=379;SAP=20.8939;SAR=304;SGB=-0.693147;SOR=1.089;SRF=0;SRP=0;SRR=0;TYPE=snp;VDB=7.48586e-13;set=Intersection;technology.illumina=1	GT:AD:AO:DP:GQ:PL:QA:QR:RO	1/1:0,683:683:683:.:2062,2056,0:24174:0:0	1/1:0,494:.:494:99:15867,1481,0	1/1:.:.:320:99:255,255,0
chrM	195	.	T	C	36371.10	PASS	AB=0;ABP=0;AC=6;AC1=2;AF=1.00;AF1=1;AN=6;AO=1475;BaseQRankSum=2.214;CIGAR=1X;DP=4200;DP4=0,0,689,154;DPB=1479;DPRA=0;EPP=4.81373;EPPR=7.35324;ExcessHet=3.0103;FQ=-281.989;FS=0;GQ_MEAN=3777;GTI=0;LEN=1;MEANALT=2;MLEAC=2;MLEAF=1;MQ=60;MQ0F=0.000722022;MQM=59.9675;MQMR=60;MQRankSum=0;MQSB=0.99208;NCC=0;NS=1;NUMALT=1;ODDS=2036.78;PAIRED=0.935593;PAIREDR=1;PAO=0;PQA=0;PQR=0;PRO=0;QA=49984;QD=27.74;QR=21;REGION=.;RO=2;RPL=820;RPP=43.0905;RPPR=3.0103;RPR=655;RUN=1;ReadPosRankSum=-0.807;SAF=756;SAP=5.02572;SAR=719;SGB=-0.693147;SOR=0.257;SRF=1;SRP=3.0103;SRR=1;TYPE=snp;VDB=0.0101606;set=Intersection;technology.illumina=1	GT:AD:AO:DP:GQ:PL:QA:QR:RO	1/1:2,1475:1475:1479:.:44951,4428,0:49984:21:2	1/1:1,1310:.:1311:99:36385,3777,0	1/1:.:.:843:99:255,255,0

chr1	1451674	.	T	C	21.04	PASS	AC=1;AC1=1;AF=0.500;AF1=0.499998;AN=2;BQB=0.450401;DP=12;DP4=3,2,0,3;FQ=24.0321;MQ=60;MQ0F=0;MQB=0.900802;MQSB=0.900802;PV4=0.196429,0.104859,1,1;REGION=.;RPB=0.150134;SGB=-0.511536;VDB=0.0295886;set=SAMTOOLS	GT:DP:GQ:PL	./.	./.	0/1:8:54:51,0,167
chr1	1451703	.	G	T	39.64	PASS	AB=0.555556;ABP=3.25157;AC=2;AC1=1;AF=0.500;AF1=0.5;AN=4;AO=5;BQB=0.372138;CIGAR=1X;DP=18;DP4=1,3,0,4;DPB=9;DPRA=0;EPP=13.8677;EPPR=5.18177;FQ=44.981;GTI=0;LEN=1;MEANALT=1;MQ=60;MQ0F=0;MQB=0.992367;MQM=60;MQMR=60;MQSB=1;NS=1;NUMALT=1;ODDS=9.12696;PAIRED=1;PAIREDR=1;PAO=0;PQA=0;PQR=0;PRO=0;PV4=1,0.157798,1,1;QA=150;QR=139;REGION=.;RO=4;RPB=0.124046;RPL=0;RPP=13.8677;RPPR=11.6962;RPR=5;RUN=1;SAF=0;SAP=13.8677;SAR=5;SGB=-0.556411;SRF=1;SRP=5.18177;SRR=3;TYPE=snp;VDB=0.0320192;set=FREEBAYES-SAMTOOLS;technology.illumina=1	GT:AD:AO:DP:GQ:PL:QA:QR:RO	0/1:4,5:5:9:.:111,0,101:150:139:4	./.	0/1:.:.:8:75:72,0,93
chr1	1451703	.	G	GAGGGACGGGCCGGTGTTGGTGAGGGCGTCTGGTCGTCCGT	171.60	PASS	AC=1;AF=0.500;AN=2;BaseQRankSum=-0.253;DP=9;ExcessHet=3.0103;FS=0;GQ_MEAN=29;MLEAC=1;MLEAF=0.5;MQ=60;MQRankSum=0;NCC=0;QD=28.6;REGION=.;ReadPosRankSum=-0.21;SOR=2.303;set=GATK	GT:AD:DP:GQ:PL	./.	0/1:1,5:6:29:179,0,29	./.
chr1	1453199	.	G	A	570.50	PASS	AB=0.731707;ABP=22.1298;AC=2;AC1=1;AF=0.500;AF1=0.5;AN=4;AO=30;BQB=0.952821;CIGAR=1X;DP=125;DP4=14,11,19,7;DPB=41;DPRA=0;EPP=26.4622;EPPR=7.94546;FQ=82.0157;GTI=0;LEN=1;MEANALT=1;MQ=2	GT:AD:AO:DP:GQ:PL:	0/1:11,30:30:41:.:669,0,104:1006:285:11	./.	0/1:.:.:51:99:242,0,109

						5;MQ0F=0.511905;MQB=0.0199485;MQM=35.2667;MQMR=34.6364;MQSB=0.849282;NS=1;NUMALT=1;ODDS=5.1341;PAIRED=1;PAIREDR=1;PAO=0;PQA=0;PQR=0;PRO=0;PV4=0.249161,1,1,1;QA=1006;QR=285;REGION=.;RO=11;RPB=0.368424;RPL=14;RPP=3.29983;RPPR=3.20771;RPR=16;RUN=1;SAF=14;SAP=3.29983;SAR=16;SGB=-0.692976;SRF=2;SRP=12.6832;SRR=9;TYPE=snp;VDB=0.0101692;set=FREEBAYES-SAMTOOLS;technology.illumina=1	QA:QR:RO			
chr4	3225478	.	T	C	136.64	PASS	AB=0.454545;ABP=3.20771;AC=3;AC1=1;AF=0.500;AF1=0.5;AN=6;AO=5;BQB=0.406082;BaseQRankSum=-0.131;CIGAR=1X;DP=33;DP4=2,3,1,3;DPB=11;DPRA=0;EPP=3.44459;EPPR=8.80089;ExcessHet=3.0103;FQ=56.0154;FS=0;GQ_MEAN=144;GTI=0;LEN=1;MEANALT=1;MLEAC=1;MLEAF=0.5;MQ=60;MQ0F=0;MQB=0.974597;MQM=60;MQMR=60;MQRankSum=0;MQSB=0.924584;NCC=0;NS=1;;RPB=0.730948;RPL=4;RPP=6.91895;RPPR=8.80089;RPR=1;RUN=1;ReadPosRankSum=0.175;SAF=1;SAP=6.91895;SAR=4;SGB=-0.556411;SOR=1.179;SRF=2;SRP=4.45795;SRR=4;TYPE=snp;VDB=0.286646;set=Intersection;technology.illumina=1	GT:AD:AO:DP:GQ:PL:QA:QR:RO	0/1:6,5:5:11:.:109,0,149:154:199:6	0/1:6,5:.:11:99:144,0,168	0/1:.:.:9:86:83,0,146

Résumé

Le séquençage de l'exome constitue une approche essentielle en génomique et en diagnostic moléculaire, ciblant les régions codantes du génome qui contiennent la majorité des variants pathogènes. Cette approche permet l'identification précise des variations génétiques responsables de nombreuses maladies héréditaires, ce qui en fait un outil indispensable en pratique clinique. Les avancées des technologies de séquençage de nouvelle génération (NGS) ont considérablement amélioré la rapidité, la précision et l'accessibilité de cette méthode. Dans ce contexte, ce travail

vise à analyser l'exome par les techniques NGS afin d'identifier des variants génétiques associés aux maladies héréditaires à des fins diagnostiques.

Cette étude s'inscrit dans le domaine de la recherche biomédicale et du diagnostic génétique, et repose sur le séquençage complet de l'exome à l'aide de la technologie NGS sur la plateforme MGI DNBSEQ G99. Le protocole comprend l'extraction de l'ADN, suivie de sa fragmentation et de la préparation des librairies (réparation des extrémités, ligation des adaptateurs et amplification par PCR), puis une étape d'enrichissement des régions exoniques par hybridation spécifique. Les fragments obtenus sont ensuite circularisés pour former des DNBs (DNA Nanoballs), puis séquencés et analysés afin d'identifier et d'annoter les variants génétiques.

Les résultats du séquençage de l'exome chez le patient ont montré des données de haute qualité, avec un score Q30 supérieur à 92 %, 43,05 millions de lectures et une couverture suffisante de la majorité des régions exoniques. Les analyses de qualité ont également montré une conformité aux standards de référence en termes de longueur des reads et de contenu GC. L'analyse a permis d'identifier un nombre important de variants génétiques (SNPs, Indels et CNVs), ce qui est cohérent avec les résultats attendus pour ce type d'analyse. Ces résultats indiquent la présence de variants génétiques chez le patient, potentiellement liés à sa pathologie, nécessitant une interprétation clinique par un spécialiste. Ces données confirment l'efficacité du NGS dans la détection des variations génétiques et son importance dans le diagnostic moléculaire, tout en soulignant la nécessité d'une interprétation clinique complémentaire.

Mots-clés

Séquençage de l'exome, séquençage de nouvelle génération (NGS), variants génétiques, maladies héréditaires, diagnostic moléculaire, analyse bioinformatique, plateforme MGI DNBSEQ G99, SNPs, Indels, CNVs, fichiers VCF, médecine de précision.

Abstract

Exome sequencing is a fundamental approach in genomics and molecular diagnostics, focusing on the coding regions of the genome that contain the majority of pathogenic variants. This approach enables the precise identification of genetic variations responsible for many hereditary diseases, making it an essential tool in clinical practice. Advances in next-generation sequencing (NGS) technologies have significantly improved the speed, accuracy, and accessibility of this method. In this context, this work aims to analyze the exome using NGS technologies to identify genetic variants associated with hereditary diseases for diagnostic purposes.

This study is part of the field of biomedical research and genetic diagnosis, and is based on whole exome sequencing using NGS technology on the MGI DNBSEQ G99 platform. The protocol includes DNA extraction, followed by fragmentation and library preparation (end repair, adapter ligation, and PCR amplification), then an enrichment step targeting exonic regions through specific hybridization. The resulting fragments are then circularized to form DNA Nanoballs (DNBs), which are subsequently sequenced and analyzed to identify and annotate genetic variants.

The results of exome sequencing in the patient showed high-quality data, with a Q30 score above 92%, 43.05 million reads, and sufficient coverage of most exonic regions. Quality analyses also showed compliance with reference standards in terms of read length and GC content. The analysis identified a significant number of genetic variants (SNPs, Indels, and CNVs), which is consistent with expected results for this type of analysis. These findings indicate the presence of genetic variants in the patient that may be related to the disease condition, requiring clinical interpretation by a specialist. These results confirm the effectiveness of NGS in detecting genetic variations and its importance in molecular diagnosis, while highlighting the need for complementary clinical interpretation.

Keywords

Exome sequencing, next-generation sequencing (NGS), genetic variants, hereditary diseases, molecular diagnosis, bioinformatics analysis, MGI DNBSEQ G99 platform, SNPs, Indels, CNVs, VCF files, precision medicine.

ملخص

يُعدّ تسلسل الإكسوم نهجًا أساسيًا في علم الجينوم والتشخيص الجزيئي، حيث يركّز على المناطق المُرمِّزة من الجينوم التي تحتوي على معظم الطفرات المُمرضة. يتيح هذا النهج التعرف الدقيق على المتغيرات الجينية المسؤولة عن العديد من الأمراض الوراثية، مما يجعله أداةً ضرورية في الممارسة السريرية. وقد ساهمت التطورات في تقنيات التسلسل من الجيل الجديد (NGS) في تحسين سرعة ودقة هذه المقاربة وإمكانية الوصول إليها. في هذا السياق، يهدف هذا العمل إلى تحليل الإكسوم باستخدام تقنيات NGS لتحديد المتغيرات الجينية المرتبطة بالأمراض الوراثية لأغراض تشخيصية.

يندرج هذا البحث ضمن مجال البحث البيوطبي والتشخيص الجيني، ويعتمد على تسلسل الإكسوم الكامل باستخدام تقنية NGS عبر منصة MGI DNBSEQ G99. يتضمن البروتوكول استخراج الحمض النووي، ثم تجزئته وتحضير المكتبات (إصلاح النهايات، ربط المحولات، والتضخيم بتفاعل PCR)، يليها إغناء المناطق الإكسونية عن طريق التهجين النوعي. بعد ذلك تُحوَّل الشظايا إلى شكل دائري لتكوين كرات الحمض النووي النانوية (DNBs)، ثم يتم تسلسلها وتحليل البيانات لتحديد المتغيرات الجينية وتعليقها (توصيفها).

أظهرت نتائج تسلسل الإكسوم لدى المريض بيانات ذات جودة عالية، حيث تجاوزت قيمة Q30 نسبة 92%، مع عدد قراءات بلغ 43,05 مليون وعمق تغطية كافٍ لأغلب المناطق الإكسونية. كما بيّنت تحاليل الجودة توافقًا مع المعايير المرجعية من حيث طول القراءات ونسبة GC. وقد مكّن تحليل البيانات من تحديد عدد معتبر من المتغيرات الجينية (SNPs، Indels، وCNVs)، وهو ما يتماشى مع ما هو متوقع في هذا النوع من التحاليل. وتشير هذه النتائج إلى وجود متغيرات جينية لدى

المريض قد تكون مرتبطة بحالته المرضية، مما يستدعي تفسيرها سريريًا من طرف المختص. تؤكد هذه المعطيات فعالية تقنية NGS في الكشف عن المتغيرات الجينية وأهميتها في دعم التشخيص الجزيئي، رغم ضرورة استكمالها بالتحليل السريري لتحديد دلالة هذه المتغيرات.

الكلمات المفتاحية :

المتغيرات الجينية، الأمراض الوراثية، التشخيص الجزيئي، التحليل (NGSتسلسل الإكسوم، التسلسل من الجيل الجديد (والطب ،VCFملفات ،CNVsو Indelsو SNPsالطفرات من نوع MGI DNBSEQ G99، البيوانفورماتيكي، منصة الدقيق.