

Année 2019

THESE N° : 01/19 CSVS

Centre d'Étude Doctorale des Sciences de la vie et de la Santé  
Formation Doctorale : Biologie médicale, Pathologie Humaine et  
Expérimentale et Environnementale  
Option : Bio-Informatique / Biotechnologie

**THÈSE**  
En vue de l'obtention du  
**DOCTORAT**

Intitulé de thèse

**Analyses bio-informatique de données génomiques  
et transcriptomiques dans deux pathologies :  
Tuberculose et maladie de Parkinson**

**Mme Lamiaa LAHLOU – OUZZINE**

Présentée et soutenue le 04 Octobre 2019

**Jury :**

**Professeur Taoufiq DAKKA**

Faculté de Médecine et de Pharmacie, Université Mohammed V, Rabat

**Président**

**Professeur Azeddine IBRAHIMI**

Faculté de Médecine et de Pharmacie, Université Mohammed V, Rabat

**Directeur de thèse**

**Professeur Abdelhamid BENZAOUZ**

Institut des maladies neurodégénératives (IMN), Université de Bordeaux

**Co-encadrant**

**Professeur Alexandre FAVEREAUX**

Institut Interdisciplinaire de Neurosciences (IINS), Université de Bordeaux

**Co-encadrant**

**Professeur Rachid EL JAUDI**

Faculté de Médecine et de Pharmacie, Université Mohammed V, Rabat

**Rapporteur**

**Professeur Wafaa REGRAGUI**

Faculté de Médecine et de Pharmacie, Université Mohammed V, Rabat

**Rapporteur**

**Professeur Leila SBABOU**

Faculté de sciences, Université Mohammed V, Rabat

**Rapporteur**

# AVANT- PROPOS

Cette thèse a été réalisée au sein du Centre d'Étude Doctoral Science de la vie et de la Santé de la faculté de médecine et pharmacie de Rabat.

J'ai réalisé ma thèse sous la direction du Professeur **IBRAHIMI Azeddine**, directeur du Laboratoire de Biotechnologie Médicale (Médbiotech) et directeur du centre de recherche BioInova à la Faculté de Médecine et de Pharmacie de Rabat, Université Mohammed V.

Au cours de ces années de thèse, deux volets ont été étudiés, le premier comporte l'analyse des données génomiques issues du séquenceur Illumina Miseq pour la maladie Tuberculose qui était encadré et supervisé par Professeur **IBRAHIMI Azeddine** et le deuxième volet se focalise sur l'analyse de données transcriptomiques pour la maladie de parkinson, qui s'est déroulé sous l'encadrement du professeur **BENAZZOUZ Abdelhamid**, chef de l'équipe Neurochimie, stimulation cérébrale profonde et maladie de Parkinson au sein de l'institut des maladies neurodégénératives à Bordeaux, et directeur de recherche à INSERM et avec la collaboration du professeur **FAVEREAUX Alexandre**, professeur d'enseignement supérieur à Institut Interdisciplinaire de Neurosciences (IINS) de l'université de Bordeaux, dans le cadre de la bourse UNIVERS octroyée par l'ambassade de France au Maroc.

## Structure de recherche :

### **Au Maroc**

Laboratoire de Biotechnologie (MedBiotech) – Faculté de Médecine et Pharmacie RABAT

### **En France**

Institut des Maladies Neurodégénératives (IMN) - UMR CNRS 5293– Université de Bordeaux

Institut Interdisciplinaire de Neurosciences (IINS) - UMR CNRS 5293– Université de

Bordeaux

## Dédicaces

*Aucun mot ne pourrait exprimer des sentiments comme la gratitude,*

*l'amour, le respect ou la reconnaissance.*

*C'est avec toute l'ardeur de mes sentiments que je dédie cette thèse*

*Tous les mots du monde ne pourraient suffire pour exprimer mes sentiments  
envers vous.*

***A mon cher PAPA***

*Vous avez toujours été mon école de patience, de confiance, d'espoir et d'amour  
Vous êtes et vous resterez pour moi ma référence, la lumière qui illumine mon  
chemin.*

*Vous avez toujours eu confiance en moi et vous m'avez offert l'encouragement et  
le soutien tout au long de mes années d'étude*

***A ma chère MAMAN***

*Vous avez guetté mes pas, et m'avez couvé de tendresse, ta prière et ta  
bénédictioin m'ont été d'un grand secours pour mener à bien mes études.  
Vous m'avez aidé et soutenu pendant de nombreuses années avec à chaque fois  
une attention renouvelée.*

***A mes chers et respectueux parents***

*Aucune dédicace, ne pourrait exprimer la profondeur des sentiments que  
j'éprouve pour vous, vos sacrifices innombrables et votre dévouement firent pour  
moi un encouragement.*

*Vous vous êtes toujours préoccupées de mon avenir, tant de sacrifices, tant  
d'affection, un engagement sans retenu, un amour sans pareil  
Ce travail est le résultat de l'esprit de sacrifice dont vous avez fait preuve, de  
l'encouragement et le soutien que vous ne cessez de manifester, j'espère que vous*

*y trouverez les fruits de votre semence et le témoignage de ma grande fierté de  
vous avoir comme parent.*

*J'espère avoir répondu aux espoirs que vous avez fondés en moi et réalisé  
aujourd'hui l'un de vos rêves.*

*Je vous rends hommage par ce modeste travail en guise de ma reconnaissance  
éternelle et de mon infini amour.*

*J'implore Dieu, tout puissant, de vous accorder une bonne santé, une longue vie  
et beaucoup de bonheur.*

*Je vous aime*

## *A mon mari et mon amour*

*Pour l'amour et l'affection qui nous unissent.*

*Pour les sacrifices que tu as faits et pour ton soutien moral que tu n'as cessé  
d'offrir*

*Pour la grande patience, l'encouragement et la confiance*

*Je ne saurais exprimer ma profonde reconnaissance pour le soutien continu dont  
tu as toujours fait preuve.*

*Tu m'as toujours soutenu et supporté dans mes moments de doute et d'angoisse.  
Je te remercie pour ta rigueur et la qualité de ton approche scientifique qui m'a  
permis d'aller jusqu'au bout et plus encore pour l'amour que tu m'as démontré  
pendant cette période.*

*Tu m'as toujours encouragé, incité à faire de mon mieux, ton soutien m'a  
permis de réaliser le rêve tant attendu.*

*Tu es un modèle d'honnêteté, de loyauté et de force de caractère.*

*J'espère te combler et te rendre toujours heureux.*

*Je te dédie ce travail qui est aussi le tien, avec mes vœux de réussite, de  
prospérité et de bonheur, que ce travail soit le témoignage de ma reconnaissance  
et de mon amour sincère et fidèle. Je prie Dieu le tout puissant de préserver  
notre attachement mutuel, et d'exaucer tous nos rêves.*

***Je t'aime***

***A ma chère sœur LAÏLA***

*Pour ta bonté, ton précieux soutien, tes conseils et ton encouragement tout au long de mes années d'étude.*

*Je t'exprime mon amour, mon respect et ma fierté.*

*Tu es et tu resteras toujours un bon exemple de persévérance et de réussite.*

*Je ne pourrais jamais imaginer la vie sans toi, tu comptes énormément pour moi, tu es la sœur qui assure son rôle comme il faut*

*Puisse Dieu, tout puissant te combler de santé, de bonheur et te procure une longue vie.*

***Je t'aime***

***A mon beau-frère ABDELJALIL***

*Je trouve en toi le conseil du frère et le soutien de l'ami.*

*Que ce travail soit l'expression de mon estime pour toi et que Dieu t'accorde santé, bonheur et succès.*

***A Mon adorable neveu ZAKARIA,***

*Je te dédie ce travail en témoignage de mon profond amour*

*Que dieu te bénisse et te protège*

***Je t'aime***

*A ma grande mère OUADIA*

*Je vous dédie cette thèse pour vos attentions particulières, vos prières et votre  
amour inconditionnel.*

*Que Dieu vous donne bonne santé et longue vie parmi nous.*

*A la mémoire de mon grand-père maternel, mon grand-père paternel et la plus  
belle grand-mère paternelle*

*Puisse Dieu vous avoir en sa sainte miséricorde et que ce travail soit une prière  
pour vos âmes.*



*A Mon beau-père Mohammed et ma belle-mère Aïcha*

*Mon beau-frère Driss*

*ET A Mes belles sœurs : Latifa, Khadija, Meryem et Fatima*

*Je profite de la présente occasion pour vous remercier pour tout le soutien, la sympathie et l'amour que vous m'accordez.*

*Vous êtes un modèle de générosité et de force.*

*Vous m'avez accueilli les bras ouverts.*

*Chaque ligne de cette thèse vous exprime la reconnaissance, le grand respect, et mon estime envers vous.*

*Que Dieu le tout puissant vous comble de santé, de bonheur et vous prouve une longue vie pleine de joie.*

*A mes tantes et oncles paternels et leurs conjoints*

*A mes tantes et oncles maternels et leurs conjoints*

*A mes cousins et cousines*

*Les expressions me trahissent, et ne peuvent exprimer mon attachement, mon  
amour et ma gratitude pour vous.*

*Qu'il me soit permis de vous exprimer à travers ce travail mon grand respect*

*A Madame Wafae chedghane et Monsieur Abdlatif Raïssouni*

*Je ne trouverai jamais l'expression forte pour vous exprimer mon affection.  
Trouvez ici l'assurance de mon profond respect et de mon fidèle attachement.*

*A mes chères amies : JIHANE, SAFAA, ZINEB*

*En souvenir des moments heureux passés ensemble, avec mes vœux sincères de  
réussite, bonheur, santé et de prospérité.*

*A tous ceux qui j'ai omis involontairement de citer*

# Remerciements

Pour souscrire à la tradition, je vais me soumettre au difficile exercice de la rédaction des remerciements de thèse. Exprimer sa gratitude aux personnes qui vous ont entouré et soutenu au cours de ces années est tout naturel. Il ne s'agit aucunement d'une simple formalité, mais de sincères sentiments qui proviennent du fond du cœur. Cependant, la difficulté de cette tâche réside essentiellement dans le fait de n'oublier personne. Ainsi, je commence par remercier tous ceux dont le nom n'apparaît pas sur ces pages par omission et qui m'ont aidée d'une manière ou d'une autre dans mon parcours.

Cette thèse est le fruit de collaborations multiples entre le Laboratoire de Biotechnologie Médicale (MedBiotech) de la Faculté de Médecine et de Pharmacie de Rabat de l'université Mohammed V, le neurocampus de l'université de Bordeaux à travers deux équipes de recherche : l'Institut des Maladies Neurodégénératives (IMN) et l'Institut Interdisciplinaire de Neurosciences (IINS), ainsi le laboratoire de bactériologie de l'hôpital d'instruction militaire Mohammed V de Rabat.

Nombreuses sont les personnes qui ont contribué, à divers titres, à l'élaboration de ce travail. Je voudrais profiter de l'occasion qui m'est donnée ici pour les remercier.

Je remercie tout d'abord mon directeur de thèse le **Pr. Azeddine IBRAHIMI**, professeur d'enseignement supérieur, directeur du Laboratoire de Biotechnologie Médicale (Médbiotech) et directeur du centre de recherche BioInova à la Faculté de Médecine et de Pharmacie de Rabat, Université Mohammed V, d'avoir suggéré et dirigé ce travail de thèse. Toujours disponible, je le remercie pour les discussions que nous avons eues : elles m'ont beaucoup appris et m'ont aidé à intégrer le domaine médical et de la biologie ainsi à avancer dans ce domaine nouveau pour moi. Son dynamisme et son enthousiasme ont également été une source de motivation et de persévérance dans les moments de doute. Je lui suis extrêmement reconnaissante de ses conseils, son soutien et ses encouragements tout au long de ces années. Son sens de la rigueur et son souci de perfection ont contribué l'accomplissement de ce travail.

Je le remercie également pour la confiance qui m'a témoignée au cours de ces années.

Les mots me manquent pour exprimer tout le plaisir que j'ai eu à travailler sous sa direction.

Ses qualités humaines et professionnelles sont connues et susciteront toujours mon admiration.

La gentillesse et la bienveillance avec lesquelles il a guidé mes pas dans ce travail ont suscité ma bonne volonté de donner de mon mieux.

Veillez trouver dans ce travail le témoignage de ma gratitude et mon profond respect.

Je tiens à remercier sincèrement **Pr BENAZZOUZ Abdelhamid**, chef de l'équipe Neurochimie, stimulation cérébrale profonde et maladie de Parkinson au sein de l'institut des maladies neurodégénératives à bordeaux et directeur de recherche à INSERM, de m'avoir chaleureusement accueilli dès le premier jour et de m'avoir donné l'opportunité d'effectuer une partie de ma thèse sous son encadrement et au sein de son équipe.

Je le remercie également pour sa confiance dès le départ, pour son encadrement, pour sa disponibilité sans faille malgré un emploi du temps plus que chargé et surtout pour sa patience ainsi pour ses précieux conseils, ses connaissances scientifiques et ses encouragements tout au long des années que j'ai passé au sein de son équipe.

Veillez trouver ici toute ma reconnaissance et mon profond respect.

Je tiens tout particulièrement à remercier **Pr FAVEREAUX Alexandre**, qui m'a judicieusement guidée durant mes travaux de recherche en analyse de données transcriptomiques et il a su orienter mes recherches aux bons moments et m'a permis de les conduire à terme. Qu'il trouve également ici l'expression de ma reconnaissance pour la qualité de sa collaboration, sa rigueur scientifique et sa compétence qui susciteront à jamais mon respect et mon admiration.

J'exprime mes vifs remerciements à **Mr EL JAUDI Rachid**, professeur d'enseignement supérieur à la faculté de médecine et pharmacie Rabat, à **Mme REGRAGUI Wafaa** professeur d'enseignement supérieur à la faculté de médecine et pharmacie Rabat et **Mme SBABOU Leila**, professeur habilité à la faculté des sciences de Rabat, d'avoir accepté de rapporter sur ce travail. Je les remercie de l'intérêt qu'ils ont éprouvé pour cette recherche, de leurs pertinentes remarques et conseils qui se sont révélés très enrichissants. Par ailleurs, je voudrais leur exprimer ma gratitude pour leur réactivité et la célérité de leur réponse.

A Mr Le président de jury ainsi que les membres de jury, malgré vos multiples occupations vous me faites l'honneur d'évaluer mon travail en acceptant de participer à mon jury. Je vous exprime ma profonde gratitude. Je remercie **Professeur DAKKA Taoufiq** Vice Doyen chargé des Affaires Spécifiques à la faculté de médecine et pharmacie, de me faire l'honneur de

présider le jury de cette soutenance de thèse malgré vos nombreuses occupations. Veuillez trouver ici toute ma reconnaissance et mes sincères remerciements.

A **Mme REGRAGUI Wafaa** et **Mme SBABOU Leila** pour m'avoir accordé du temps et d'avoir accepté de participer à ce jury en tant que rapportrices de cette thèse malgré leurs emplois du temps plus que chargé. Veuillez trouver ici le témoignage de mes respectueuses considérations.

A Monsieur le **Pr. Rachid ELJAUDI**, Professeur Agrégé à la Faculté de Médecine et de Pharmacie de Rabat. Je vous suis reconnaissante d'avoir accepté de participer à ce jury et de bien vouloir juger ce travail en tant rapporteur. Veuillez trouver ici l'expression de ma profonde reconnaissance.

A Monsieur le **Pr. Jamal TAOUFIK**, Professeur d'Enseignement Supérieur et Directeur du Centre d'Études Doctorales des Sciences de la Vie et de la Santé à la Faculté de Médecine et de Pharmacie de Rabat, Université Mohammed V, pour m'avoir inscrit au sein de cet établissement et pour votre aide administrative.

La réalisation de ce travail n'aurait pas été possible sans la contribution et le soutien de nombreuses personnes.

**A l'équipe du Laboratoire MedBiotech**, je tiens à remercier les personnes qui ont participé à ce travail, je vous renouvelle ma gratitude pour vos efforts et vous souhaitez la réussite que vous méritez.

Je remercie le **Dr LEMRISS Sanaâ**, Chef du Département de Biosécurité PCL3 au sein du LRAM de m'avoir orientée et initiée aux principes du séquençage à haut débit et le domaine de la génomique. Veuillez trouver ici l'expression de ma profonde reconnaissance.

**A l'équipe Neurochimie, stimulation cérébrale profonde et maladie de Parkinson** au sein de l'institut des maladies neurodégénératives où j'ai pu commencer un nouveau domaine dans un cadre particulièrement agréable, grâce à l'ensemble des membres de l'équipe. Un immense merci pour leur aide, leur gentillesse, leur infinie patience..., et leur bonne humeur.

Je tiens à remercier **Pr DE DEURWAEDERE Philippe** pour son aide autant d'un point de vue expérimental que pour la réflexion scientifique. Sa patience, ses conseils et son expertise qui m'ont énormément aidée.

Je remercie également **Mr NAUDET Frédéric**, ingénieur au sein de l'équipe IMN6 du neurocampus de m'avoir initié aux principes de la chirurgie des rongeurs spécifiquement des

rats et pour sa patience durant la formation, ainsi je remercie **Mme OUBROU Layna** pour son aide et sa patience.

A l'équipe de l'**Institut Interdisciplinaire de Neurosciences (IINS)**, je remercie **Mme BOUALI-BENAZZOZ Rabia** pour son accueil, son soutien, ses conseils, ses encouragements ainsi que ses connaissances scientifiques durant la partie expérimentale ainsi que la période de rédaction.

Un grand merci à mes collègues au sein du neurocampus de bordeaux pour leur soutien et leurs encouragements.

J'exprime ma profonde sympathie à mes collègues du laboratoire MedBiotech notamment mes collègues de la promotion doctorale. Je tiens à remercier également, **Mme ARROUCHI Housna**, **Mme LAKHLILI Wiam** et **Mme LEMRISS Hajar** pour leur accueil chaleureux les premiers jours de mon arrivée au laboratoire MedBiotech.

Mes remerciements s'adressent également aux étudiants du master Biotechnologie Médicale et Bio-informatique et particulièrement aux étudiants que j'ai encadrés pour leur projet de fin d'étude :

A **Mr ALOUANE Tarek**, le premier étudiant que j'ai encadré pour son stage de fin d'étude, je le remercie pour les échanges scientifiques.

A **Mr. BENDAHOU Mohammed Amine**, un étudiant du master de la faculté de médecine dentaire de Rabat, pour le partage des connaissances.

Je remercie les membres de l'équipe génomique : **Meryem**, **Souad** et **Oussama** pour leur bonne humeur, leur écoute aussi bien scientifique qu'amicale ! Je leur souhaite une bonne continuation dans leurs travaux de recherche.

Un merci particulier à **Mme LAAMARTI Meryem** avec qui j'ai partagé des moments de galère et de joie ainsi de discussions scientifiques

Mes remerciements vont à tous mes amis et en particulier à **Mr EL MASSAOUDI Abdelhakim** pour son amitié depuis mon premier jour à la faculté de médecine et pharmacie, et pour son soutien fraternel ainsi j'ai également une pensée pour **Mme EL GUASSABI Sara** qui m'a offert son amitié cette dernière année. J'espère que notre amitié durera encore plus longtemps.

Je tiens à remercier l'ambassade de France à Rabat pour leur soutien financier.

Que serais-je devenue sans ma famille et leur soutien ?

Merci à vous papa, maman, ma sœur et mon beau-frère Abdeljalil de m'avoir soutenue moralement et financièrement et encouragée de manière inconditionnelle en m'offrant les meilleures conditions afin d'en être là aujourd'hui.

Malgré l'éloignement géographique vous avez su être toujours près de moi.

Des remerciements tous particuliers à mes beaux-parents pour leur soutien et leurs prières dans les moments difficiles.

Mes remerciements les plus amoureux vont à SAID, la personne avec qui je partage ma vie. Merci d'être entré dans ma vie, pas forcément dans la période la plus sereine, d'avoir accepté le peu de disponibilité induits par cette thèse, d'avoir su écouter mes lamentations et de m'avoir apaisée dans les moments les plus difficiles.

Enfin, que toutes les personnes qui y ont contribué de près ou de loin trouvent ici ma sincère reconnaissance et mes remerciements, ainsi je remercie le corps administratif de la faculté de médecine et pharmacie RABAT et le corps administratif de l'université de BORDEAUX.

**Un très grand Merci à Tous !**

*LAHLOU - OUZZINE Lamiaa*



*Si nous savions ce que nous faisons, cela ne s'appellerait pas de la recherche, n'est-ce pas ?*

*Albert Einstein*

# Tables de matières

AVANT- PROPOS .....	2
Dédicaces .....	3
Remerciements.....	12
Tables de matières.....	18
Liste des figures .....	22
Liste des tableaux .....	25
Abréviations.....	27
Les publications et communications .....	29
Chapitre 1 : Partie bibliographique .....	31
<b>I. État de l'art du séquençage à haut débit.....</b>	<b>32</b>
1. Technologies de séquençages.....	32
1.1. Séquençage de 1 <sup>ère</sup> génération .....	32
1.2. Séquençage de 2 <sup>ème</sup> génération.....	32
1.3. Séquençage de 3 <sup>ème</sup> génération.....	38
2. Plateformes de séquençages .....	39
2.1. Roche 454.....	39
2.2. Ion torrent .....	40
2.3. Illumina.....	41
<b>II. Bio-informatique.....</b>	<b>45</b>
1. Généralités.....	45
2. Objectifs de la Bio-informatique .....	45
3. Les applications de la Bio-informatique .....	46
4. Les techniques "omiques" .....	48
4.1. La Génomique.....	48
4.2. L'épigénomique .....	49
4.3. La protéomique .....	50
4.4. La transcriptomique.....	50
<b>III. Étude de cas.....</b>	<b>52</b>
1. Cas n°1 : La maladie de la tuberculose .....	52
1.1. La Mycobacterium Tuberculosis .....	52
1.1.1 Caractéristiques du genre Mycobacterium.....	52
1.1.2 Caractéristiques de Mycobacterium tuberculosis .....	53
1.1.3 La Mycobacterium tuberculosis résistante.....	54
1.2. La maladie de la tuberculose .....	61

1.2.1.	Les types de Tuberculose.....	62
1.2.2.	Les symptômes de la maladies tuberculose .....	63
1.2.3.	Facteurs à risque.....	64
1.2.4.	Statistiques sur la propagation de la maladie tuberculose .....	65
1.2.5.	Traitement de la Tuberculose.....	68
Objectifs du travail sur la maladie Tuberculose.....		70
2.	Cas n°2 : La maladie de Parkinson.....	71
2.1.	Généralités .....	71
2.2.	Symptômes .....	71
2.2.1.	Symptômes moteurs .....	71
2.2.2.	Symptômes non-moteurs .....	72
2.3.	Les causes de la MP .....	74
2.4.	Traitement.....	75
2.5.	Les causes de la MP .....	76
2.6.	Traitement.....	78
2.7.	Épigénétique de la MP .....	79
2.5.1.	L'épigénétique et les miRNAs .....	79
2.5.2.	Les miRNAs, de régulateurs importants de l'expression de gènes .....	81
2.5.3.	L'épigénétique et le Parkinson.....	85
Objectifs du travail sur la MP.....		87
CHAPITRE 2 : Workflow d'analyse génomique pour la Maladie de Tuberculose.....		89
<b>II.</b>	<b>Partie expérimentale (Biologique).....</b>	<b>93</b>
1.	Culture.....	93
2.	Extraction et dosage d'ADN .....	93
3.	Préparation de librairie.....	93
4.	Séquençage des 4 souches MTB.....	94
4.1.	Séquençage par Illumina-MiSeq.....	94
4.2.	Applications et méthodes utilisées pour le séquençage de MTB.....	95
<b>III.</b>	<b>Analyse génomique des 4 MTB .....</b>	<b>96</b>
1.	Le prétraitement.....	97
1.1.	Contrôle qualité.....	97
1.2.	Filtration - Trimming.....	104
2.	Assemblage.....	106
2.1.	Assemblage par référence – Mapping .....	106
2.2.	Assemblage de NOVO .....	108
2.3.	Le contrôle qualité de l'assemblage .....	110
3.	Annotation .....	112

3.1.	L'annotation structurale ou syntaxique .....	114
3.2.	L'annotation fonctionnelle.....	115
3.3.	L'annotation relationnelle.....	115
3.4.	Annotations des 4 MTB.....	115
4.	Analyse des gènes de résistance .....	116
5.	La Comparaison génomique des quatre MTB.....	123
6.	Analyse des gènes de virulence .....	127
7.	L'arbre phylogénétique.....	130
	Discussion sur la partie Tuberculose.....	131
	CHAPITRE 3 : Workflow d'analyse Transcriptomique pour la Maladie de Parkinson.....	135
<b>I.</b>	<b>Partie expérimentale (Biologique)</b> .....	136
1.	Les modèles animaux de la MP .....	136
1.1.	Animaux .....	136
1.2.	Lésion unilatérale.....	136
2.	Études comportementales.....	138
2.1.	Test de Cylindre.....	138
2.2.	Test de Von Frey : Test de l'allodynie mécanique .....	142
3.	Analyse biochimique de la déplétion monoaminergique .....	146
3.1.	Traitement des tissus cérébraux .....	146
3.2.	Analyse par HPLC des surnageants .....	147
3.3.	Détection électrochimique.....	147
3.4.	Analyses statistiques .....	147
<b>II.</b>	<b>Analyse biomoléculaire</b> .....	163
1.	Extraction d'ARN total .....	163
2.	Les technologies de détection et quantification des miRNAs.....	177
3.	Séquençage RNA-Seq.....	181
3.1.	Préparation de librairie avec Illumina .....	181
3.2.	Exécution du séquençage RNA-Seq .....	186
4.	Analyse transcriptomique (Bio-informatique).....	187
4.1.	Contrôle qualité et filtrage.....	187
4.2.	Le référencement des miRNAs.....	190
4.3.	La prédiction des miRNAs par miRPro .....	190
4.3.1.	Pipeline d'analyse miRPro .....	191
4.3.2.	Résultats miRPro.....	200
5.	Validation des résultats par RT QPCR .....	210
5.1.	Analyse statistique du résultat RT PCR .....	215
5.2.	Prédiction des gènes .....	217

Discussion de la partie de la MP.....	223
Conclusion générale .....	230
Résumé .....	235
Abstract.....	236
ملخص.....	237
Bibliographie.....	238
Annexe 1 .....	267
Annexe 2.....	270
Annexe 3.....	272

# Liste des figures

Figure 1:Principe du pyroséquençage .....	35
Figure 2:Représentation du diagramme du séquençage par terminateur réversible (Ambardar et al. 2016) .....	36
Figure 3:Représentation du diagramme du changement de pH impliqué dans le séquençage par détection d'ions hydrogene (Ambardar et al. 2016) .....	37
Figure 4 : Représentation de digramme des réactions enzymatiques impliquées dans le séquençage par hybridation et ligation (Ambardar et al. 2016).....	38
Figure 5:Workflow de Roche 454 GSFLX: construction de librairie (Mardis 2008).....	40
Figure 6:Illumina Flow cell .....	42
Figure 7: Principe de la technologie Illumina.....	43
Figure 8:Système de séquençage Illumina (Illumina 2017) .....	43
Figure 9: Composition chimique des antituberculeux de la première ligne.....	58
Figure 10:Taux d'incidence estimé de la tuberculose en 2016 .....	65
Figure 11:Distribution des cas de TB par âge et par sexe, dans le monde, 2017 .....	66
Figure 12:Distribution de la mortalité de TB par âge et par sexe, dans le monde, 2017 .....	66
Figure 13:Distribution proportionnelle des cas de tuberculose par région, Maroc, 2017 .....	67
Figure 14:Distribution des cas de TB par âge et par sexe, Maroc 2017.....	68
Figure 15:Précurseur miR-let-7 .....	81
Figure 16: Représentation des différentes localisations des gènes de miRNAs (Ul Hussain 2012).....	84
Figure 17:Workflow de préparation de librairie avec le kit Nextera XT .....	94
Figure 18:Protocole de séquençage paired-end .....	95
Figure 19: Résultat du séquençage sur le séquenceur Illumina MiSeq.....	96
Figure 20: Processus de l'étape Phasing .....	100
Figure 21:Représentation du Graph de De Bruijn.....	109
Figure 22: Graphique de longueur cumulative (Quast).....	111
Figure 23:Graphique du contenu GC des 4 MTB et référence (Quast).....	112
Figure 24: Workflow d'annotation génomique.....	114
Figure 25:Cartographie des 4 souches MTB .....	124
Figure 26:Position du gène katG sur les 4 souches et la référence H37Rv .....	125
Figure 27:Position du gène RpoB sur les 4 souches et la référence H37Rv.....	125

Figure 28:Position du gène gyrA sur les 4 souches et la référence H37Rv.....	126
Figure 29:Position du gène gyrB sur les 4 souches et la référence H37Rv.....	126
Figure 30:Position du gène tlyA sur les 4 souches et la référence H37Rv.....	127
Figure 31: L'arbre phylogénétique des 4 souches MTB et la référence H37Rv .....	130
Figure 32:Localisation du faisceau médian du télencéphale (MFB) par ses coordonnées stéréotaxiques sur une coupe frontale de cerveau de rat. ....	137
Figure 33:Schéma chronologique du déroulement des expériences .....	138
Figure 34: Test de cylindre; un rat parkinsonien au cours de l'essai au cylindre (Rattka et al. 2016).....	139
Figure 35:Taux de réussite au test de cylindre (Groupe 6-OHDA) .....	140
Figure 36:Taux de réussite au test de cylindre (Groupe Alpha-Syn).....	141
Figure 37:Application de la pointe de Von Frey sur la patte arrière du rat (Ferrier, Marchand, et Balayssac 2016).....	142
Figure 38:Taux de réussite du test Von Frey chez le groupe 6-OHDA .....	144
Figure 39:Taux de réussite du test Von Frey chez le groupe alpha-Syn.....	145
Figure 40:Les structures du cerveau prélevé pour HPLC .....	146
Figure 41:contenu du tissu Striatum - DL du groupe 6-OHDA en Dopamine .....	149
Figure 42:Contenu du tissu Striatum - DL du groupe 6-OHDA en Dopac et HVA .....	150
Figure 43:Contenu du tissu Striatum - DL du groupe 6-OHDA en NA, 5HT, 5HIAA.....	151
Figure 44:Contenu du tissu Striatum - VL du groupe 6-OHDA en Dopamine .....	153
Figure 45:Contenu du tissu Striatum - VL du groupe 6-OHDA en Dopac et HVA .....	154
Figure 46:Contenu du tissu Striatum - VL du groupe 6-OHDA en NA, 5HT et 5HIAA .....	155
Figure 47:Contenu du tissu Cortex Frontale en 5HT et NA.....	156
Figure 48:Contenu du tissu Striatum DL en Dopamine pour le groupe Alpha-Syn .....	157
Figure 49:Contenu du tissu Striatum DL en Dopac et HVA pour le groupe Alpha-Syn.....	158
Figure 50:Contenu du tissu Striatum DL en NA, 5HT et 5HIAA pour le groupe Alpha-Syn.....	159
Figure 51:Contenu du tissu Striatum VL - Dopamine pour le groupe Alpha-Syn .....	160
Figure 52:Contenu du tissu Striatum VL - Dopac et HVA pour le groupe Alpha-Syn .....	161
Figure 53:Contenu du tissu Striatum VL - NA, 5HT et 5HIAA pour le groupe Alpha-Syn..	162
Figure 54:pitulatif du résultat du Bioanalyzer pour le groupe 6OHDA.....	167
Figure 55:Récapitulatif du résultat du Bioanalyzer pour le groupe Alpha_Syn.....	172
Figure 56:Conversion de miRNA matures en ADNc et détection ultérieure. ....	178
Figure 57:Organigramme de préparation des échantillons ARN avec NEXTflex .....	183
Figure 58:BluePippine.....	184

Figure 59:Résultat du bioanalyseur Agilet de l'ADNc (librairie) pour le groupe 6-OHDA..	185
Figure 60:Qualité de librairie pour RNAseq.....	186
Figure 61:Les principaux algorithmes de mirPro pour l'exploration et la quantification des variants de miRNAs (Shi et al. 2015).....	193
Figure 62:Schéma des étapes d'analyse bio-informatique des données issues de RNA-Seq .	199
Figure 63:Diagramme de Venn pour les miRNAs communs entre le groupe 6-OHDA et Alpha-Syn .....	201
Figure 64:volcano plot groupe 6-OHDA.....	204
Figure 65:volcano plot groupe Alpha-Syn .....	205
Figure 66:snRNA U6.....	211
Figure 67: Analyse de la courbe de dissociation.....	212
Figure 68:Courbe de dissociation des miRNA .....	212
Figure 69:La différence d'expression des miRNAs dans le groupe 6-OHDA en utilisant un RTPCR.....	216
Figure 70:La différence d'expression des miRNAs dans le groupe Alpha-Syn en utilisant un RTPCR.....	216
Figure 71:Pipline d'analyse génomique et transcriptomique.....	233
Figure 72:Séquenceur Illumina MiSeq.....	267
Figure 73:Le logiciel Illumina Experiment Manager.....	267



# Liste des tableaux

Tableau 1:Espèces mycobactériennes majeures .....	52
Tableau 2:Antituberculeux de première et de deuxième intention, gènes impliqués dans leur activation .....	61
Tableau 3:Les caractéristiques principales des trois types de petits ARNs non codants exprimés chez les mammifères .....	82
Tableau 4:Per Base Sequence Quality des 4 MTB avec FastQC.....	99
Tableau 5: Per Sequence Quality Scores des 4 MTB.....	102
Tableau 6:Per Base Sequence Content des 4 MTB.....	103
Tableau 7:Per Base Sequence Quality des 4 fichiers Reverse avec FastQC après filtrage .....	105
Tableau 8:Résultat d'alignement avec Mauve avec MTB H37rV.....	107
Tableau 9:Résultat d'assemblage des 4 souches MTB .....	111
Tableau 10:Les résultats d'annotation des 4 MTB .....	117
Tableau 11:Récapitulatif des médicaments avec les résistances des 4 MTB .....	122
Tableau 12:Gènes de virulence pour MTB1_M .....	128
Tableau 13:Gènes de virulence pour MTB2_M .....	128
Tableau 14:Gènes de virulence pour MTB13_M .....	129
Tableau 15:Gènes de virulence pour MTB20_M .....	129
Tableau 16:Le pourcentage de mapping des 4 souches avec la référence H37Rv.....	130
Tableau 17:Les coordonnées des deux sites d'injection de l'Alpha-Synucleine .....	137
Tableau 18:Résultat du test de student pour le test de cylindre pour le groupe 6-OHDA.....	140
Tableau 19:Résultat du test de student pour le test de cylindre pour le groupe Alpha-Syn .....	141
Tableau 20:Résultat du test de student pour le test de Von Frey pour le groupe 6-OHDA.....	144
Tableau 21:Résultat du test de student pour le test de Von Frey pour le groupe Alpha-Syn .....	145
Tableau 22:Résultat du test de Wilcoxon du groupe 6-OHDA - DL pour la dopamine .....	149
Tableau 23:Résultat du test de Wilcoxon du groupe 6-OHDA - DL pour Dopac et HVA .....	150
Tableau 24:Résultat du test de Wilcoxon du groupe 6-OHDA - DL pour NA, 5HIAA et 5HT.....	151
Tableau 25:Résultat du test de Wilcoxon du groupe 6-OHDA - VL pour Dopamine .....	153
Tableau 26:Résultat du test de Wilcoxon du groupe 6-OHDA - VL pour Dopac et HVA .....	154
Tableau 27:sultat du test de Wilcoxon du groupe 6-OHDA - VL pour NA, 5HT et 5HIAA.....	155
Tableau 28:Résultat du test de Wilcoxon du groupe 6-OHDA - Cortex Frontal pour 5HT et NA.....	156
Tableau 29:Résultat du test de Wilcoxon du groupe Alpha-Syn - Striatum DL pour Dopamine.....	157
Tableau 30:Résultat du test de Wilcoxon du groupe Alpha-Syn - Striatum DL pour Dopac et HVA	158
Tableau 31:Résultat du test de Wilcoxon du groupe Alpha-Syn - Striatum DL pour NA, 5HT et 5HIAA.....	159

Tableau 32:Résultat du test de Wilcoxon du groupe Alpha-Syn - Striatum VL pour Dopamine.....	160
Tableau 33:Résultat du test de Wilcoxon du groupe Alpha-Syn - Striatum VL pour Dopac et HVA	161
Tableau 34:Résultat du test de Wilcoxon du groupe Alpha-Syn - Striatum VL pour NA, 5HT et 5HIAA.....	162
Tableau 35:La concentration et le rapport de pureté obtenus par le Nanodrop-one .....	165
Tableau 36:Comparaison des caractéristiques principales des différentes méthodes de détection et de quantification des miRNAs (de Planell-Saguer and Rodicio, 2013).....	180
Tableau 37:Liste des adaptateurs et des primers fournis par le kit Bioo Scientific .....	182
Tableau 38:Résultat FastQC des échantillons - RNA-Seq .....	188
Tableau 39:Les caractéristiques du génome de référence rattus norvegicus .....	195
Tableau 40:Liste des fichiers obtenus par miRPro .....	198
Tableau 41:Résultat miRPro pour le groupe 6-OHDA .....	200
Tableau 42:Résultat miRPro pour le groupe Alpha-Syn.....	200
Tableau 43:Résultat de DESeq2 pour le Cortex du groupe 6OHDA .....	206
Tableau 44:Résultat de DESeq2 pour le groupe 6OHDA coté Droit VS coté Gauche .....	207
Tableau 45:Résultat de DESeq2 pour le striatum du groupe 6OHDA.....	208
Tableau 46:Résultat de DESeq2 pour le Striatum pour le groupe Alpha-Syn.....	209
Tableau 47:Résultat de DESeq2 pour le cortex pour le groupe Alpha-Syn .....	209
Tableau 48:Résultat de DESeq2 pour le groupe Alpha-Syn.....	210
Tableau 49:Description des miRNAs étudiés en QPCR pour groupe 6OHDA .....	213
Tableau 50:Description des miRNAs étudiés en QPCR pour groupe Alpha-Syn .....	214
Tableau 51: miRNAs sélectionnés pour les deux groupes .....	217
Tableau 52:Les paramètres de TargetScan.....	218
Tableau 53:rédiction des gènes par TargetScan pour le groupe 6-OHDA .....	220
Tableau 54:Prédiction des gènes par TargetScan pour le groupe Alpha-Syn.....	222
Tableau 55:Les miRNAs impliqué dans la MP et décrits dans la littérature .....	226

# Abréviations

<b>5HT</b>	Sérotonine
<b>6-OHDA</b>	6-Hydroxydopamine
<b>ADN</b>	Acide Désoxyribonucléique
<b>ADNc</b>	Acide Désoxyribonucléique Complémentaire
<b>ARN</b>	Acide Ribonucléique
<b>ARNr</b>	ARN Ribosomique
<b>ARNsn</b>	Small Nuclear RNA
<b>ARNt</b>	ARN De Transfert
<b>ATP</b>	
<b>CDS</b>	Coding Dna Sequence
<b>DA</b>	Dopamine
<b>DBG</b>	De-Bruijn-Graph
<b>DDBJ</b>	DNA Data Bank Of Japan
<b>ddNTP</b>	Didésoxyribonucléotide
<b>dNTP</b>	Désoxyribonucléotides
<b>DO</b>	Dopac
<b>EMBL</b>	European Molecular Biology Laboratory
<b>emPCR</b>	PCR En Emulsion
<b>ENA</b>	European Nucleotide Archive
<b>FC</b>	Fold Chanfe
<b>GC</b>	Guanine Cytosine
<b>HPLC</b>	High Performance Liquid Chromatography
<b>INSDC</b>	International Nucleotide Sequence Database Collaboration
<b>L-Dopa</b>	Levodopa
<b>MDR</b>	Tuberculose Multirésistante
<b>MFB</b>	Medial Forebrain Bundle
<b>MIC</b>	Minimum Inhibitory Concentration
<b>miRNA</b>	Micro RNA
<b>MP</b>	Maladie De Parkinson
<b>MTB</b>	Mycobacterium Tuberculosis
<b>NAc</b>	Noyau Accumbens

<b>NCBI</b>	National Center For Biotechnology Information
<b>NGS</b>	Next Generarion Sequencing
<b>NST</b>	Noyau Sous-Thalamique
<b>OLC</b>	Overlap–Layout–Consensus
<b>OMS</b>	Organisation Mondiale De La Santé
<b>PCR</b>	Polymerase Chain Reaction
<b>PPi</b>	Pyrophostphate
<b>RIN</b>	Rna Integrity Number
<b>RISC</b>	RNA- Induced Silencing Complexe
<b>RTPCR</b>	Reverse Transcriptase Pcr
<b>SGS</b>	Seconde Génération De Séquençage
<b>SNC</b>	Système Nerveux Central
<b>SNP</b>	Single Nucleotide Polymorphism
<b>UTR</b>	Untranslated Region
<b>WGS</b>	Whole Genome Sequencing
<b>XDR</b>	Extensively Drug Resistant

# Les publications et communications

- **Lahlou - Ouzzine L**, Naudet F, Bouali-Benazzouz R, De Deurwaedere P, Ibrahimi A, Favereaux A, Benazzouz A. **Identification of the expression of miRNAs in two animals models of Parkinson's diseases.** (En cours de préparation)
- **Lahlou - Ouzzine L**, Bendahou MA, El Mrimar N, Bssaibis F, Ben Lahlou Y, Zegmout A, El Hafidi N, ElJaoudi R, Frikh M, Lemnouar A, Elouennass M, Ibrahimi A. **A genomic comparison of four strains of Mycobacterium Tuberculosis from four Moroccan patients.** (En cours de preparation)
- **Lahlou L**, El Mrimar N, Laamarti M, Alouane T, Bendahou MA, Bssaibis F, Ben Lahlou Y, Zegmout A, El Hafidi N, ElJaoudi R, Frikh M, Lemnouar A, Elouennass M, Ibrahimi A. **Whole-Genome Shotgun Sequences of Three Multidrug-Resistant Mycobacterium tuberculosis Strains Isolated from Morocco.** Genome Announc. 2017 Nov 16;5 (46). pii: e01275-17. doi: 10.1128/genomeA.01275-17.
- **Lahlou L**, El Mrimar N, Alouane T, Laamarti M, Karti S, Benhrif O, El Mesbahi H, Lemriss H, Bssaibis F, Maleb A, El Rarit S, Zegmout A, El Jaoudi R, Frikh M, Lemnouar A, Dakka T, Elouennass M, Ibrahimi A. **Whole-Genome Shotgun Sequence of Multidrug-Resistant Mycobacterium tuberculosis MTB13\_M Isolated from Morocco.** Genome Announc. 2017.
- Alouane T, Uwingabiye J, Lemnouer A, **Lahlou L**, Laamarti M, Karti S, Benhrif O, El Misbahi H, Frikh M, Benlahlou Y, Bssaibis F, El Abbassi S, Kabbage S, Maleb A, Elouennass M, Ibrahimi A. **First Whole-Genome Sequences of Two Multidrug-Resistant Acinetobacter baumannii Strains Isolated from a Moroccan Hospital Floor.** Genome Announc. 2017
- H. Lemriss, S. Lemriss, P. Martins-Simoes, M. Butin, **L. Lahlou**, J.-P. Rasigade, A. Kearns, O. Denis, M. Deighton, A. Ibrahimi, F. Laurent, S. El Kabbaj. **Genome Sequences of Four Staphylococcus capitis NRCS-A Isolates from Bioinformatique Geographically Distant Neonatal Intensive Care Units.** Genome Announc. 2015 Aug 6;3 (4)
- H. Lemriss, DUMONT Y, S. Lemriss, P. Martins-Simoes, M. Butin, **L. Lahlou**, J.-P. Rasigade, A. Kearns, O. Denis, M. Deighton, A. Ibrahimi, F. Laurent, S. El Kabbaj. **Genome Sequences of Multiresistant Staphylococcus capitis Pulsotype NRCS-A**

**and Methicillin-Susceptible S. capitis Pulsotype NRCS-C.** Genome Announc. 2016 jun 9;4 (3)

- Communication : **Lahlou L**, El Mrimar N, Alouane T, Laamarti M, Karti S, Benhrif O, El Mesbahi H, Lemriss H, Bssaibis F, Maleb A, El Rarit S, Zegmout A, El Jaoudi R, Frikh M, Lemnouar A, Dakka T, Elouennass M, Ibrahimi A. **Whole-Genome Shotgun Sequence of Multidrug-Resistant Mycobacterium tuberculosis MTB13\_M Isolated from Morocco.** 4ème édition de la journée Internationale de Biotechnologie Médicale, Le 20 Décembre 2016.
- Communication : **Lahlou L, Moussa A, Dakka T, Ibrahimi A.** **New statistical and computational analysis platform of affymetrix microarray data: WAMDA,** Agadir School on Numerical Methods in Engineering, Le 21 Octobre 2015

# Chapitre 1 : Partie bibliographique

L'objectif de ce chapitre est de donner une vue globale sur le séquençage de nouvelle génération ainsi le contexte bio-informatique en citant la définition et le principe et les applications, ensuite une description détaillée des deux cas étudiés : la maladie de tuberculose et la maladie de parkinson.

# I. État de l'art du séquençage à haut débit

## 1. Technologies de séquençages

### 1.1. Séquençage de 1<sup>ère</sup> génération

Le séquençage de l'ADN a été découvert dans la fin des années 1970. Deux méthodes ont été développées indépendamment, l'une par l'équipe de Walter Gilbert (Maxam et Gilbert 1977), aux États-Unis, et l'autre par celle de Frederick Sanger (Sanger, Nicklen, et Coulson 1977), en Grande-Bretagne. La méthode de Maxam et Gilbert est basée sur des stratégies de dégradation chimique sélective de l'ADN et la méthode Sanger consiste à utiliser une stratégie de synthèse enzymatique sélective de l'ADN. La méthode la plus utilisée en routine est celle de Sanger. C'est une méthode de synthèse enzymatique du brin complémentaire de l'ADN dont on cherche à déterminer la séquence à l'aide d'une ADN polymérase et de nucléotides « terminateurs de chaîne » ddNTP (didésoxyribonucléotides) qui ont un atome d'hydrogène à la place du groupement OH sur le carbone 3' du ribose. Il s'en suit une migration par électrophorèse sur gel de polyacrylamide, puis le gel est autoradiographie afin de détecter les fragments radioactifs. Toutefois, c'est l'automatisation de cette méthode à la fin des années 80, avec le développement des marquages fluorescents et de l'électrophorèse capillaire, qui a ouvert la voie du séquençage à haut débit (L. M. Smith et al. 1986).

Le séquençage automatisé de Sanger est actuellement considéré comme la "première génération" des technologies de séquençage d'ADN (Rizzo et Buck 2012). Récemment, une nouvelle génération de séquenceurs à très haut débit est apparue. Ces techniques permettent de séquencer, en quelques jours (heures), plusieurs giga bases d'ADN et sont regroupées sous le nom de méthodes « NGS » pour Next (ou New) Generation Sequencing.

### 1.2. Séquençage de 2<sup>ème</sup> génération

Depuis 2005, l'apparition de nouvelle méthode de séquençage à haut débit appelé nouvelle génération de séquençage (NGS : Next Generation Sequencing) ou seconde génération de séquençage (SGS). Le système 454 Roche GS FLX, en voie de disparition, était la première plate-forme de séquençage de nouvelle génération disponible dans le commerce en 2004, suivi par l'analyseur de génomes Illumina en 2006, le séquenceur SOLiD en 2007 et l'Ion Torrent en 2010 (Ambardar et al. 2016).



Les différents types de plateformes de séquençage diffèrent principalement par les chimies de séquençage qui entraînent des différences de débit, de longueur de reads, de taux d'erreur, de couverture génomique, de coût et de temps d'exécution (Ambardar et al. 2016).

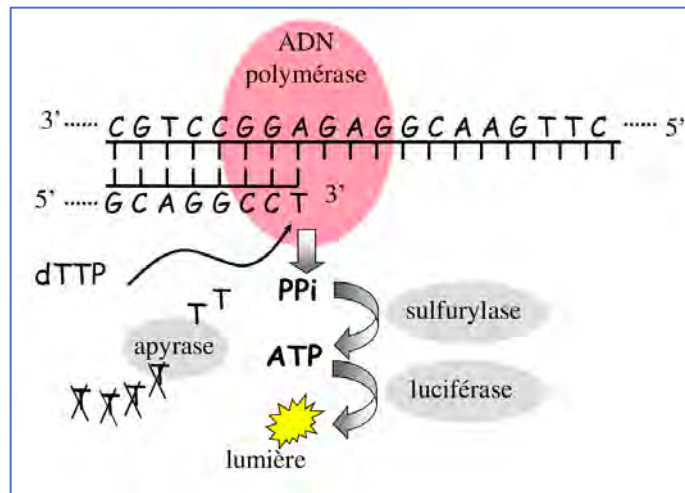
Toutes les plateformes de séquençage de seconde génération impliquent deux étapes primordiales :

- **Préparation du modèle (Template Preparation)** comporte trois étapes :
  - **Extraction d'ADN** dont les protocoles ne sont pas universels mais dépendent du type et de la source de l'échantillon ainsi l'objectif de l'étude (Metzker 2010) (Fierer et al. 2012).
  - **Préparation de librairie de l'ADN (Library preparation)** pour le séquençage implique la fragmentation de l'ADN extraite en petits fragments aléatoires qui se chevauchent suivi du polissage final de l'ADN fragmentée et ligation des adaptateurs. L'ADN extraite est fragmentée entre 150pb et 800pb selon la plateforme de séquençage utilisée (Hingamp et al. 2013) (Caruccio 2011). La préparation de librairie à partir de l'ARN se fait en capturant l'ARNm, amorçage aléatoire et synthèse d'ADN complémentaire (ADNc) suivis du polissage finale et la ligation des adaptateurs (Frias-Lopez et al. 2008) (Lesniewski et al. 2012). La taille moyenne de l'ARN est plus petite, donc la fragmentation n'est pas nécessaire pour le séquençage de l'ARN (transcriptome).
  - **L'amplification du modèle** nécessite la sélection des fragments d'ADN après la préparation de librairie sont soumis à une amplification clonale. Cette amplification implique une amplification en phase solide des fragments d'ADN et contribue au développement d'un signal détectable puissant pendant le séquençage. Un seul fragment d'ADN à séquencer est lié à des billes, à des surfaces d'ion ou à des cellules de flux. Selon la plateforme de séquençage utilisée, la PCR en émulsions ou le bridge PCR est utilisé pour amplifier les fragments d'ADN ancrés en millions de fragments de modèle séparé dans l'espace (Fakruddin et Chowdhury 2012) (Shao et al. 2011).
  - **Séquençage** : la différence entre les technologies SGS réside dans le principe du séquençage de chaque base, il existe deux approches :
- **Séquençage par synthèse (Sequencing by Synthesis (SBS))** est adopté par trois principaux types de chimie de séquençage qui sont les suivants :

**a. Pyroséquençage:** Après la préparation du modèle par emPCR (Shao et al. 2011), les billes amplifiées par clonage, restants dans les puits de la "picotitre<sup>1</sup> plate PTP", sont prêtes pour le pyroséquençage. En pyroséquençage, le pyrophosphate (PPi) libéré au cours de la réaction de polymérisation de l'ADN est détecté et utilisé en tant qu'indicateur de l'incorporation de bases spécifiques. Elle implique la détection de la base incorporée dans l'ADN fille par l'ADN polymérase, avec alimentation et élimination, en séquence, des quatre bases, ainsi que la détection du pyrophosphate libéré par une cascade d'enzymes émettant de la lumière (Chowdhury et al. 2012) (Marzorati et al. 2013). Le simple brin d'ADN lié aux billes, est hybridé par le séquençage des amorces et incubé avec l'ADN polymérase, "ATP sulfurylase", "Luciférase", "Apyrase", "Adenosine 5' phosphate (APS) ", "luciférine" et dNTP dans le puit de la plaque (Figure 1) (Matthieu FALQUE 2011) . Sur l'incorporation complémentaire dNTP par l'ADN polymérase, le pyrophosphate (PPi) est libéré et convertit en ATP par ATP sulfurylase en utilisant l'adenosine 5' phosphosulfate. En présence d'ATP, la luciférase convertit la luciférine en oxyluciférine générant de la lumière visible. La lumière produite par la réaction de la luciférase est détectée et mesurée par une photodiode à avalanche, un tube photomultiplicateur ou par un appareil photo à dispositif à couplage de charge (avec ou sans plaque à microcanaux). Le coût du séquençage en pyroséquençage par rapport aux autres technologies de séquençage à haut débit est relativement élevé. Bien qu'il soit pionnier des technologies de séquençage de deuxième génération, le pyroséquençage a déjà été écarté du marché, ce qui indique le rythme auquel les technologies de séquençage évoluent.

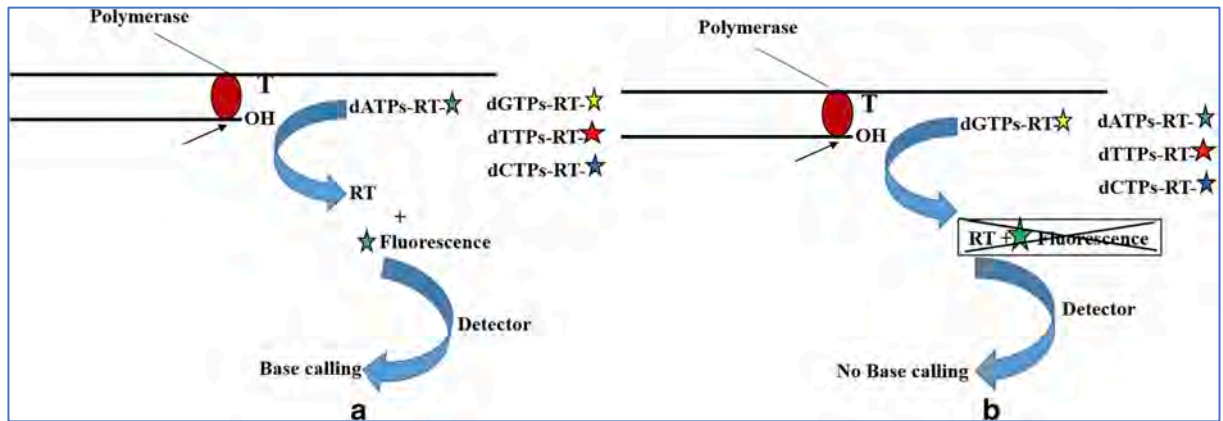
---

<sup>1</sup> Picotiter plate (PTP) ou plaque picotitre, version miniaturisée des plaques microtitres, ce sont des plaques contenant de nombreux puits utilisés comme des éprouvettes miniatures. Ici les puits mesurent 44 µm.



**Figure 1:Principe du pyroséquenceage**

- a. Séquenceage** par terminaison réversible avec la plateforme Illumina, les fragments d'ADN des banques sont soumis à une amplification clonale par "Bridge PCR" (PCR en pont) (Shao et al. 2011) suivie d'un séquenceage par terminaison réversible à l'aide de nucléotide terminateur réversible (RT), ce nucléotide RT est protégé au niveau des groupe 3'-OH (2-Cyanoéthyle) et marqué par fluorescence (Berglund, Kiialainen, et Syvänen 2011). Lors de l'ajout d'un mélange de nucléotides de RT à la cellule d'écoulement, l'ADN polymérase incorpore des nucléotides modifiés dans le brin d'ADN en cours de synthèse. Chaque cycle de séquenceage par terminateur réversible se compose de trois étapes : i) incorporation du nucléotide RT complémentaire par l'ADN polymérase mutante au brin d'ADN lié à la Flow Cell, ii) détection des différents signaux fluorescents pour les quatre bases, iii) restauration du groupe 3'OH libre en coupant le fragment de terminaison et la molécule reportée. La répétition de ce cycle conduit au séquenceage du modèle d'ADN. Au cours du séquenceage, le nucléotide RT est incorporé et pris en image, suivie par l'élimination des fluorophores et les bases terminales sont activées par "de-protection" au niveau du groupe 3'OH, permettant ainsi une nouvelle incorporation de nucléotide (Figure 2).



**Figure 2: Représentation du diagramme du séquençage par terminateur réversible (Ambardar et al. 2016)**

a) lors de l'incorporation complémentaire de nucléotides. b) le nucléotide n'est pas incorporé.

La compétition naturelle entre les quatre nucléotides présents au cours de chaque cycle de séquençage réduit le biais inhérent par rapport au pyroséquençage dans lequel un seul type de nucléotide est disponible à la fois pour l'appariement (Chen et al. 2012) (Berglund, Kiialainen, et Syvänen 2011). L'erreur de séquençage des homopolymères (comme dans le cas du pyroséquençage) est résolue par cette technique en raison de l'incorporation d'une base unique à la fois, car l'ajout d'un autre terminateur de base doit d'abord être éliminé.

**b. Séquençage** par détection d'hydrogène ions est basé sur la détection des ions hydrogène libérés lors de l'incorporation de chaque nucléotide et ne dépend pas de bases modifiées, d'enzymes ou de la détection optique (Eid et al. 2009). Le modèle d'ADN obtenu après la préparation de la librairie et l'amplification clonale est lié aux particules brevetées Ion Sphère dans les micro-puits de manière à ce que chaque micro-puit contienne une seule de sphère d'ions. Un seul type de dNTP est ajouté au micro-puit à la fois et son incorporation est détectée par libération d'ion hydrogène qui déclenche un capteur ionique ISFET (Ion Sensitive Field Effect Transistor). Le changement de pH est détecté par une couche de micro-puit sensible qui convertit le signal chimique en signal numérique, mesuré en quelques secondes (Figure 3). Par rapport à d'autres méthodes de séquençage, où la détection est indirecte à l'aide de scanners laser, la détection est indépendante de ces dispositifs et donc directe (Mascher et al. 2013).

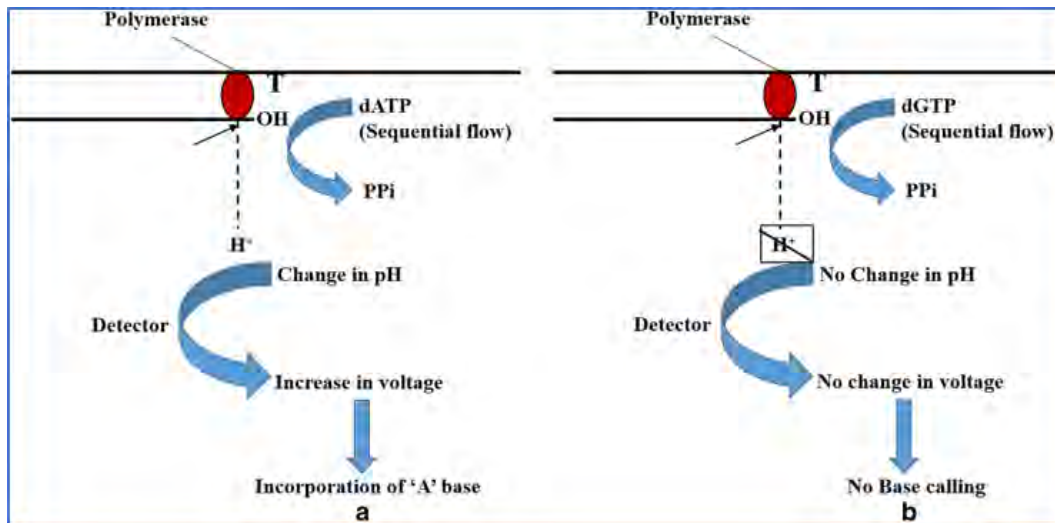
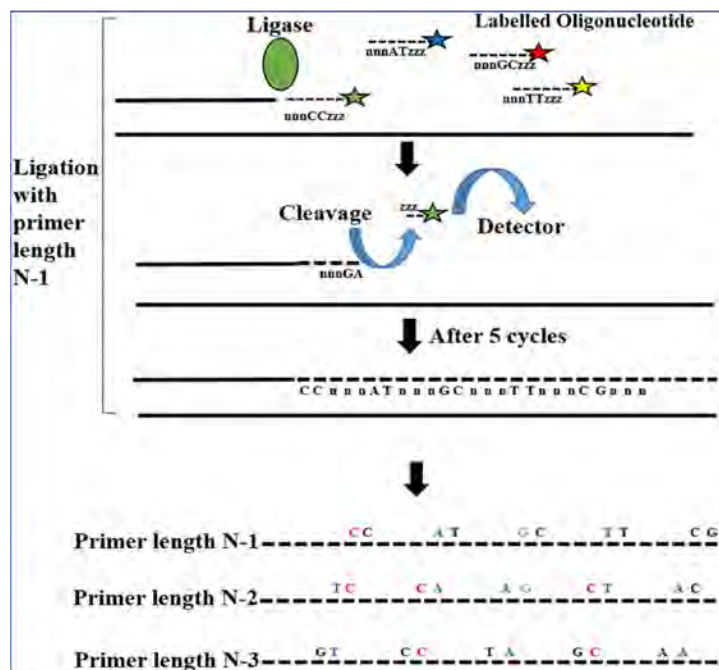


Figure 3: Représentation du diagramme du changement de pH impliqué dans le séquençage par détection d'ions hydrogène (Ambardar et al. 2016)

a) lors de l'incorporation complémentaire de nucléotides. b) le nucléotide n'est pas incorporé.

- Séquençage par hybridation et ligation (SBL)** est la base de plateforme de séquençage de détection de ligation d'oligonucléotides (SOLiD) développée par Biosystèmes appliqués. SBL dépend de la spécificité de l'ADN ligase pour le "mismatch" des paires de bases au lieu de l'ADN polymérase dans le cas de SBS. La préparation de librairie comprend la fragmentation du modèle et sa fixation avec la séquence d'adaptateur connue. Les adaptateurs attachés au fragment d'ADN à séquencer sont ensuite attachés à des billes et amplifiés de manière clonale par emPCR (Shao et al. 2011). Ces ADN amplifiés par clonage liés aux billes sont déposés sur une plaque de verre où les billes sont liées de manière covalente (Liu et al. 2012) (Meldrum, Doyle, et Tothill 2011). La chimie de séquençage comprend l'hybridation et la ligation de diverses sondes codées à une ou deux bases avec le modèle. La sonde fait une longueur de huit ou neuf bases et consiste une ou deux bases suivies de trois bases dégénérées et de trois bases universelles attachées à un marqueur fluorescent. Une amorce courte est ajoutée avec le pool mélangé de sonde oligonucléotidique fluorescente qui hybride/annèle à l'ADN cible ayant une séquence complémentaire. Les

sondes sont ligaturées avec les amorces en utilisant un ADN ligase et sont détectées par imagerie, par fluorescence et les sondes non ligaturées sont lavées. Les sondes oligonucléotidiques ont des liaisons clivables attachées à un marqueur fluorescent qui peut être clivé après la détection, en préparant ainsi le système pour un autre cycle de ligation (Figure 4). Ce cycle est répété plusieurs fois pour séquencer l'ADN cible complet. A la fin d'un tour, les séquences de bases ne seront connues qu'à certaines positions autres que les bases dégénérées. Pour séquencer les positions "skipped", un autre cycle de séquençage se produit avec une amorce ayant une ou plusieurs bases plus courtes que l'amorce précédente (Meldrum, Doyle, et Tothill 2011) (Liu et al. 2012).



**Figure 4 : Représentation de digramme des réactions enzymatiques impliquées dans le séquençage par hybridation et ligation (Ambardar et al. 2016)**

### 1.3. Séquençage de 3<sup>ème</sup> génération

Les plates-formes de séquençage de deuxième génération présentent généralement deux limites : **i)** la longueur de "read" court qui doit être assemblée à l'aide de divers outils. **ii)** le biais de PCR introduit par amplification clonale, pour la détection du signal d'incorporation de base. La troisième génération (TGS) de la technologie NGS à haut débit ou le séquençage d'une molécule en temps réel (SMRT) a été mise au point pour remédier à ces limitations (Schadt, Turner, et Kasarskis 2010b). Le TGS comprend cinq séquenceurs présentés par les

plates-formes en temps réel à molécule unique: "*Helicos' Genetic Analysis System*", "*Pacific Biosciences*" et "*Oxford Nanopore*" (Rusk 2009). Les deux autres plateformes TGS sont "*Complete Genomics*" de l'institut de génomique de Pékin et "*GnuBio*" de "*BioRad*".

## 2. Plateformes de séquençages

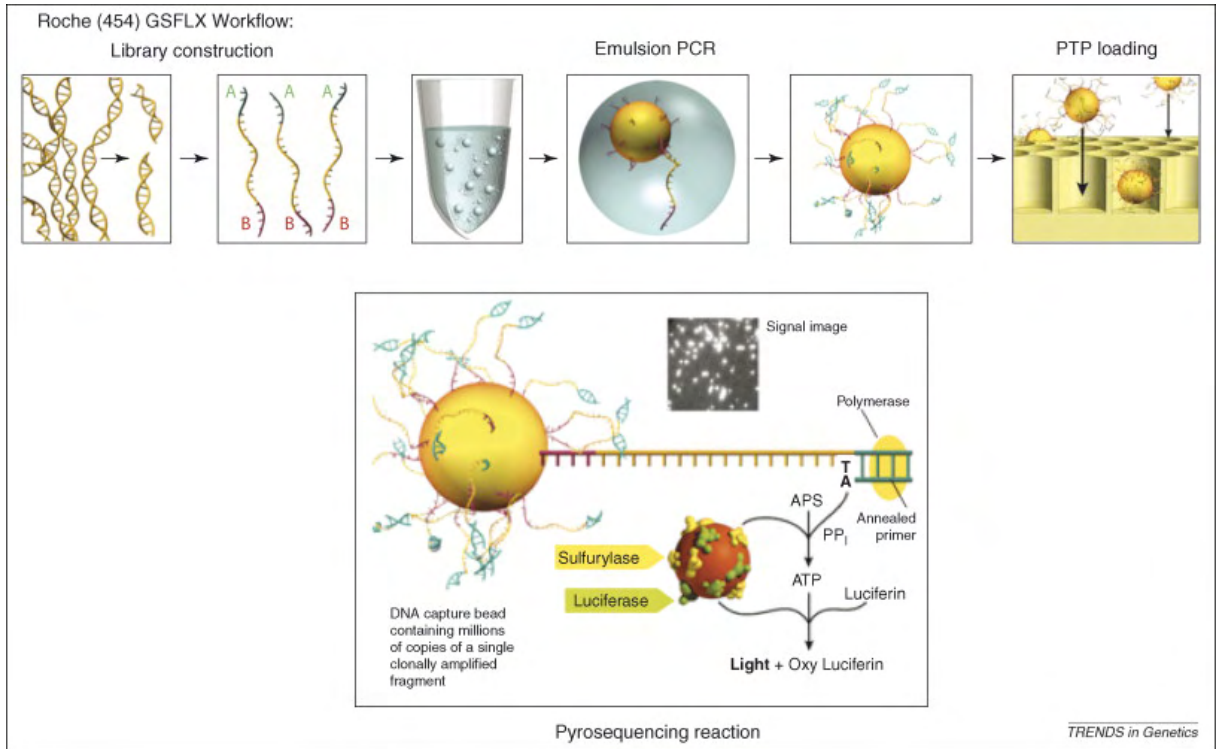
### 2.1. Roche 454

Roche 454 est une plate-forme développée par la société Roche Diagnostics, qui est basée sur la technologie de pyroséquençage avec une fragmentation d'ADN par PCR en émulsion. La préparation de librairie commence par un fractionnement aléatoire (Shotgun) de l'ADN en fragments de 300 à 800 pb pour obtenir une banque d'ADN simple brin. Ensuite, l'ajout des adaptateurs spécifiques des extrémités 3' et 5'. Les fragments ayant deux adaptateurs différents s'hybrident sur une bille de quelques microns de diamètre. Puis l'amplification à émulsion des billes avec les produits de PCR et de l'huile. Le fragment d'ADN fixé sur la bille est amplifié et les nouveaux produits PCR s'ancrent également sur la bille. Après un nombre de cycle de PCR optimum, ces billes sont purifiées et déposées sur une plaque picotitre, dont le diamètre des puits ne permet qu'une seule bille à la fois, suivi de l'ajout des enzymes de séquençage et envoi des nucléotides individuels les uns après les autres. Le signal chimie luminescent est enregistré par une caméra CCD (Metzker 2010). La lecture est effectuée en simultanée à chaque cycle du pyroséquençage et obtenir la séquence correspondante à chaque bille déposée (Figure 5).

Le format de fichier binaire utilisé pour coder les résultats du pyroséquençage à partir de la plate-forme 454 Roche c'est un fichier "Standard Flowgram File" d'extension "sff<sup>2</sup>" qui peut être visualisés, édité et convertis au format FASTQ avec sffinfo (Newbler tools).

---

<sup>2</sup> Le fichier SFF (Standard Flowgram File, extension. sff) est un fichier texte utilisé pour stocker les séquences biologiques associées aux technologies Roche (séquenceurs 454) et Life Technologies (Ion torrent). Il contient les séquences, les données de qualité et les données de flux permettant un retraitement des données sans perte de signal. Ce fichier est humainement illisible.



**Figure 5: Workflow de Roche 454 GSFLX: construction de librairie (Mardis 2008)**

## 2.2. Ion torrent

La plate-forme Ion Torrent utilise également une méthodologie de séquençage par synthèse. La principale différence de cette plate-forme par rapport aux autres techniques de séquençage de nouvelle génération réside dans le fait qu'au lieu d'utiliser l'optique pour détecter l'incorporation de nucléotides, elle détecte l'incorporation de nucléotide en surveillant les changements du pH. Sur le plan conceptuel, les principales étapes sont très similaires à la technologie Illumina. La première différence principale est qu'au lieu d'utiliser une amplification en "bridge" pour amplifier par clonage les fragments individuels d'ADN, elle utilise la PCR en émulsion. Ici, chaque brin individuel est compartimenté dans une perle d'émulsion où la PCR se produit pour amplifier par clonage des fragments de la séquence cible. Cela permet essentiellement à des milliers de PCR individuelles de se produire simultanément. Une fois que les segments de séquences ont été amplifiés par clonage, ils sont chargés sur une puce d'ordinateur où chaque perle d'émulsion sera chargée dans un micropuits sur la puce. Pour l'étape de séquençage, l'un des quatre nucléotides est inondé dans les micropuits. Ce cycle se répète jusqu'à ce que la séquence de modèle soit complètement



séquencée. Ensuite, les données brutes sont traitées et analysées de manière similaire à la plate-forme Illumina. Ion Torrent a également plusieurs plates-formes utilisant leur technologie. La différence entre les plates-formes réside dans le débit et l'automatisation. La machine avec le débit le plus faible est la machine à génome personnel, la suivante est le proton, et il existe maintenant le S5 XL, qui est plus flexible en ce sens qu'il peut exécuter des débits plus faibles et plus élevés et qu'il est plus automatisé. Trois puces couramment utilisées avec des capacités différentes sont les 314 (1,2 million de puits, 1 échantillon), 316 (6 millions de puits, 2 à 7 échantillons) et 318 (12 millions de puits, 8 à 10 échantillons). L'un des principaux avantages de la technologie Ion Torrent par rapport à la technologie Illumina est que les exigences en matière de quantité d'ADN sont beaucoup plus faibles. Cela rend la technologie Ion Torrent idéale pour les échantillons de tumeur solide où l'ADN de tumeur disponible est généralement inférieur à celui utilisé dans les hémopathies malignes et le dépistage de défauts constitutionnels (Wanger et al. 2017).

### 2.3. Illumina

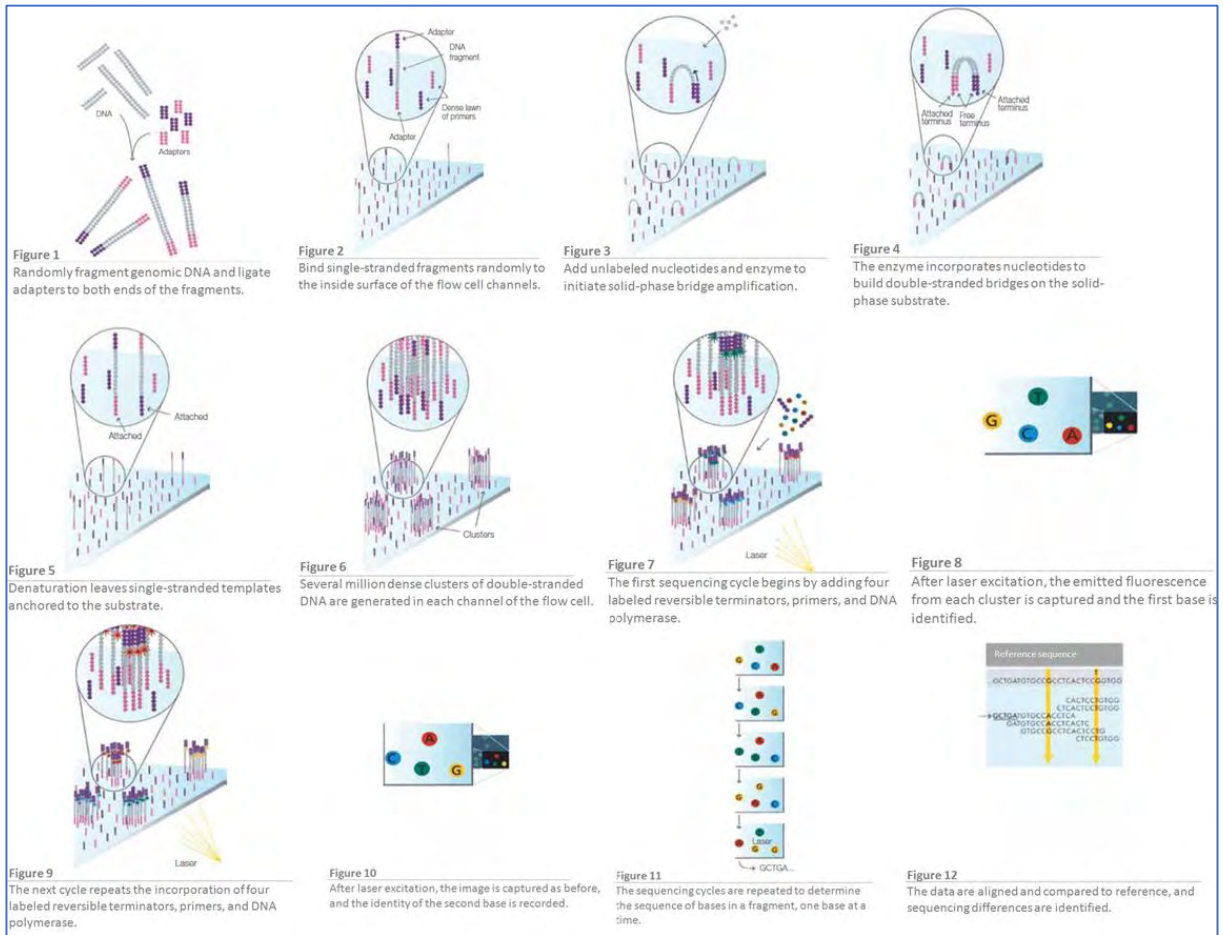
La technologie de "séquençage par synthèse" actuellement utilisée par Illumina a été initialement développée par Shankar Balasubramanian et David Klenerman de l'Université de Cambridge. Ils ont fondé la société Solexa en 1998 pour commercialiser leur méthode de séquençage. Illumina a ensuite acheté Solexa en 2007 et s'est appuyée sur la technologie d'origine pour l'améliorer rapidement. La technologie de séquençage Illumina exploite la formation de réseaux clonaux et la technologie exclusive de terminateur réversible pour un séquençage rapide et précis à grande échelle. Le système de séquençage innovant et flexible permet un large éventail d'applications en génomique, transcriptomique et épigénomique. Les "Templates" (modèles) de séquençage sont immobilisées sur une surface en verre appelé "Flow cell " (Figure 6) conçue pour présenter l'ADN de manière à faciliter l'accès aux enzymes tout en garantissant une stabilité élevée de la Template liée à la surface et une faible liaison non spécifique de nucléotides marqués par fluorescence.



**Figure 6: Illumina Flow cell**

L'amplification en phase solide (figures 7) crée jusqu'à 1 000 copies identiques de chaque molécule modèle, à proximité immédiate (diamètre d'un micron ou moins).

Du fait que ce processus ne fait pas appel à la photolithographie, au maculage mécanique ni au positionnement des billes dans des puits, des densités de l'ordre de dix millions d'agrégats mono-moléculaires par centimètre carré sont obtenues. Un système de séquençage à quatre canaux est utilisé pour le séquenceur Illumina HiSeq et MiSeq, dans lequel chaque base est détectée par une image individuelle. Illumina a mis au point le dernier NextSeq 500, doté de la technologie SBS à deux canaux et ne nécessitant que deux images pour déterminer les quatre appels de base. Cela réduit le temps de capture d'image, le nombre de cycles, le coût du séquençage et le temps requis pour le traitement des données, tout en offrant une qualité et une précision (Reuter, Spacek, et Snyder 2015).



**Figure 7: Principe de la technologie Illumina**

La technologie Illumina est très flexible et évolutif. Des systèmes de séquençage sont disponibles pour chaque méthode et échelle d'étude, des petits laboratoires aux grands centres génomiques. Les instruments Illumina NGS vont du système MiniSeq™ de laboratoire, avec sortie allant de 1,8 à 7,5 Gb pour les études de séquençage ciblées, au système NovaSeq™ 6000, qui peut générer des reads impressionnantes de 6 Tb et 20 B en 2 jours pour des études à l'échelle de la population (figure 8) (Illumina 2017) (Schadt, Turner, et Kasarskis 2010a)



**Figure 8: Système de séquençage Illumina (Illumina 2017)**

Illumina propose des plates-formes NGS innovantes qui offrent une qualité de données et une précision exceptionnelle à grande échelle (Illumina, 2017). Des configurations d'exécution flexibles sont également intégrées dans la conception des séquenceurs Illumina NGS. Par exemple, le système HiSeq® 2500 offre deux modes de fonctionnement et un séquençage à un ou deux flowcell tandis que la série de systèmes de séquençage NextSeq® offre deux types de flowcell pour répondre aux différents besoins. La série HiSeq 3000/4000 utilise la même technologie de Flowcell à motifs que les instruments HiSeq X pour un séquençage économique et rentable à l'échelle de la production. La nouvelle série de systèmes NovaSeq réunit les toutes dernières performances imagerie avec la prochaine génération de la technologie de flowcell à motifs Illumina pour Augmenter considérablement le débit. Cette flexibilité permet aux chercheurs de configurer des analyses adaptées à leurs besoins spécifiques, avec l'instrument de leur choix (Schadt et al., 2010). Le séquenceur MiSeq® d'Illumina, qui allie rapidité, simplicité et efficacité pour l'analyse vos petits génomes, gènes d'intérêt, amplicons et RNA-seq. Le MiSeq® génère en 56h de run, jusqu'à 25 millions de reads avec 15 Gb de données de sortie, pour des reads allant jusqu'à 300 bp à chaque extrémité des fragments d'ADN séquençage dit en paired-end 2 x 300 pb (Illumina 2017).

## II. Bio-informatique

### 1. Généralités

La bio-informatique est une science hybride née de la rencontre entre de la biologie et l'informatique. C'est une branche à part entière de la biologie en plein essor (Dardel et Képès 2002) qui est très interdisciplinaire et qui utilise des techniques et des concepts issus de l'informatique, de la statistique, des mathématiques, de la chimie, de la biochimie, de la physique et de la linguistique. Elle associe des données biologiques à des techniques de stockage, de distribution et l'analyse d'informations afin de prendre en charge de nombreux domaines de la recherche scientifique, y compris le domaine biomédical (Cohen-Boulakia et Valduries 2016). L'utilisation du terme bio-informatique est utilisé pour la première fois en 1970 (Hogeweg 2011), en référence à l'étude des processus d'information dans les systèmes biotiques.

Cette science multidisciplinaire a pour but de résoudre un problème scientifique liée à la biologie. Depuis quelques années, l'acquisition des données en biologie a connu une explosion quantitative grâce au développement de nouvelles techniques biologiques qui permettent la détermination de l'ordre d'enchaînement des nucléotides pour un fragment d'ADN (Hogeweg 2011). Pour stocker, traiter et analyser ces données, plus complexes, l'utilisation de l'informatique qui a connu le développement de nouvelles technologies surtout au niveau de la capacité de stockage, la puissance et performance des logiciels d'analyses pour la recherche scientifique cette évolution technologique de la bio-informatique permet de mieux comprendre et mieux connaître les phénomènes et processus biologiques, dans le but de faire de nouvelles découvertes scientifiques qui peuvent améliorer la qualité de vie des personnes malades grâce à la mise en place de nouveaux traitements médicaux plus efficaces (Ranganathan 2009).

### 2. Objectifs de la Bio-informatique

La Bio-informatique comprend deux sous domaines, le premier consiste au développement des outils et les bases de données et le deuxième sous domaine c'est l'application de ces outils développés pour générer des connaissances biologiques permettant de mieux comprendre les systèmes vivants. Ces deux sous domaines sont complémentaires. Le développement de logiciel pour l'analyse séquentielle, structurelle et fonctionnelle, ainsi que le traitement et le

stockage de bases de données biologiques. Les analyses de données biologiques génèrent souvent de nouveaux problèmes et défis qui, à leur tour, nécessitent le développement de nouveaux outils (Xiong 2006).

La bio-informatique a pour but d'intégrer des données d'origines très diverses pour modéliser les systèmes vivants afin de comprendre et prédire leurs comportements (analyse du génome, modélisation de l'évolution d'une population animale, modélisation moléculaire, reconstruction d'arbres phylogénétiques...) L'informatique est devenue un apport fondamental à la biologie moléculaire. Les moyens informatiques sont utilisés pour le stockage ou la gestion des données mais également pour l'interprétation de ces données (Damien Imbs et Mohamed Sayed Hassan 2017).

Il a de nombreuses applications pratiques dans différents domaines de la biologie et de la médecine.

### 3. Les applications de la Bio-informatique

La bio-informatique est non seulement devenue essentielle pour la recherche fondamentale en génomique et en biologie moléculaire, mais elle a également un impact majeur sur de nombreux domaines de la biotechnologie et des sciences biomédicales. Elle a des applications dans la conception de médicaments basés sur la connaissance, l'analyse ADN médico-légale et la biotechnologie agricole. Les études informatiques sur les interactions protéine-ligand fournissent une base rationnelle pour l'identification rapide de nouvelles pistes pour les médicaments synthétiques. La connaissance des structures tridimensionnelles des protéines permet de concevoir des molécules capables de se lier au site récepteur d'une protéine cible avec une grande affinité et spécificité. Cette approche basée sur l'informatique réduit considérablement le temps et le coût nécessaires au développement de médicaments plus puissants, avec moins d'effets secondaires et moins de toxicité que l'approche traditionnelle par essais et erreurs.

Dans le domaine de la justice, les résultats de l'analyse phylogénétique moléculaire ont été acceptés comme preuve par les tribunaux pénaux. Certaines statistiques bayésiennes sophistiquées et des méthodes d'analyse de l'ADN fondées sur la probabilité ont été appliquées à l'analyse de l'identité judiciaire.

Actuellement, la génomique et la bio-informatique sont sur le point de révolutionner notre système de santé en développant une médecine personnalisée. Le séquençage génomique à

haute vitesse, associé à une technologie informatique sophistiquée, permettra au médecin d'une clinique de séquencer rapidement le génome du patient, de détecter facilement les mutations potentiellement néfastes, de diagnostiquer rapidement et de traiter efficacement les maladies.

Les outils bio-informatiques sont également utilisés en agriculture. Les bases de données sur le génome des plantes et les analyses du profil d'expression génétique ont joué un rôle important dans la mise au point de nouvelles variétés de cultures plus productives et plus résistantes aux maladies (Xiong 2006).

#### ▪ **Les banques de données Biologiques**

La base de la bio-informatique est la création et la maintenance de bases de données d'informations biologiques. Cependant, en raison du succès du projet sur le génome humain, une tâche plus immédiate et urgente nécessite l'analyse des informations de séquence. Ce domaine de la biologie informatique implique généralement la recherche des gènes dans les séquences d'ADN de divers organismes, l'élaboration de méthodes permettant de prédire la structure et / ou la fonction de séquences, le regroupement de séquences de protéines dans des familles de séquences apparentées et l'alignement de protéines similaires afin d'évaluer les relations évolutives. L'origine des banques de données biologiques remonte à l'utilisation des premiers ordinateurs par des cristallographes ou des biochimistes.

Depuis plus de 30 ans, la collaboration internationale pour la base de données sur les séquences de nucléotides (INSDC ; <http://www.insdc.org/>) s'est engagée à capturer, à préserver et à donner accès à une séquence complète de nucléotides du domaine public et à des métadonnées associées permettant la découverte en biomédecine, biodiversité et sciences biologiques. Depuis 1987, la banque de données génétiques du Japon (DDBJ) de l'Institut national de génétique de Mishima, au Japon; European Nucleotide Archive (ENA) de l'Institut européen de bio-informatique (EMBL-EBI) du Laboratoire européen de biologie moléculaire à Hinxton, Royaume-Uni; et GenBank au Centre national d'information sur la biotechnologie (NCBI), à la Bibliothèque nationale de médecine, aux Instituts nationaux de la santé de Bethesda, dans le Maryland, aux États-Unis, ont collaboré pour permettre l'accès aux données de séquences de nucléotides dans des formats normalisés à la communauté scientifique mondiale (Karsch-Mizrachi, Takagi, et Cochrane 2018).

## 4. Les techniques "omiques"

L'ajout "omiques" à un terme moléculaire implique une évaluation complète ou globale d'un ensemble de molécules (<http://omics.org/>). Le domaine de "omique" repose en grande partie sur les progrès technologiques qui ont rendu possible une analyse rentable et à haut débit des molécules biologiques (Hasin, Seldin, et Lusic 2017).

Les bio-informaticiens et les biologistes moléculaires ont figuré parmi les premiers scientifiques à appliquer les techniques "omiques". Parmi les premiers utilisateurs, il y avait des bio-informaticiens à Cambridge, au Royaume-Uni, où se trouvaient de nombreux premiers laboratoires de bio-informatique tels que le centre MRC, le centre Sanger et le EBI (*European Bioinformatics Institute*). Par exemple, le centre MRC a réalisé les premiers projets sur le génome et le protéome. Les techniques "omiques" comprennent plusieurs approches tel que l'épigénomique, la génomique, la transcriptomique, la protéomique et la métabolomique (Fröhlich 2017).

### 4.1. La Génomique

La première discipline "omique" apparue était la génomique, elle est basée sur l'étude de gènes entiers, par opposition à la génétique qui interrogeait des variantes individuelles ou des gènes individuels (Hasin, Seldin, et Lusic 2017). La génomique est une science récente en pleine expansion, qui permet l'étude de l'ensemble de l'information génétique d'un organisme afin de comprendre le fonctionnement de l'organisme humain et sur certaines maladies héréditaires (Makarova et Koonin 2007). C'est une discipline qui implique le séquençage de grandes quantités d'ADN, une technique utilisée lors des diagnostics de maladies et produit une énorme quantité de données. La génomique utilise une combinaison d'ADN recombinant, de méthodes de séquençage d'ADN et de bio-informatique pour séquencer, assembler et analyser la structure et la fonction des génomes.

L'analyse génomique se compose de trois étapes primordiales : le séquençage de l'ADN à haut débit, l'assemblage et l'annotation.

- a. **L'assemblage** est une étape d'analyse *in silico* nécessaire après le séquençage, elle consiste à reconstituer un génome en identifiant les reads chevauchants, et les groupés en contigs<sup>3</sup> ou scaffolds<sup>4</sup>. L'assemblage peut être globalement divisé en deux approches

---

<sup>3</sup> Contigs : Séquence sans gaps obtenue par chevauchement des séquences courtes générées par le séquenceur.

<sup>4</sup> Scaffolds : séquence avec gaps constituée par plusieurs contigs ordonnés.



: l'assemblage de novo, pour les nouveaux génomes qui sont séquencés pour la première fois et ne ressemblent à aucun séquencé dans le passé, et l'assemblage par référence, qui utilise la séquence existante d'un organisme étroitement apparenté comme référence lors de l'assemblage (Miller, Koren, et Sutton 2010).

b. **L'annotation** du génome consiste à identifier les emplacements des gènes et toutes les régions codantes et déterminer la fonction de ces gènes par une approche *in silico*, en utilisant des outils bio-informatiques dotés de fonctionnalités spécifiques (Koonin et Galperin 2003), qui distinguent trois types d'annotation :

- **L'annotation syntaxique (structurale)** permet d'identifier les objets génétiques, l'essentiel de ces objets sont les séquences codant des protéines, des molécules d'ARN (ARNt, ARNr, ARNsn, ...), les pseudo-gènes, les retrotransposons, certaines séquences répétées et séquences régulatrices de l'expression des gènes peuvent également être détectées grâce à leurs motifs structuraux particuliers (de Sá et al. 2018).

- **L'annotation fonctionnelle** se base sur la recherche d'homologie des objets génétiques identifiés considérés inconnus avec des séquences annotées et répertoriées dans les banques de données biologiques (NCBI, EMBL, DDBJ), afin de collecter plusieurs informations utiles pour prédire et comprendre leurs fonctions (de Sá et al. 2018).

- **L'annotation relationnelle** détermine plusieurs relations susceptibles d'exister entre les éléments génétiques prédits tel que l'homologie, l'interaction physique et l'implication dans un processus biologique (même voie métabolique, même voie de transport,...) (de Sá et al. 2018).

#### 4.2. L'épigénomique

L'épigénomique est l'étude de l'ensemble complet de modifications épigénétiques sur le matériel génétique d'une cellule, appelé épigénome. Le domaine est analogue à la génomique et à la protéomique. Les modifications épigénétiques sont des modifications réversibles de l'ADN d'une cellule ou d'histones affectant l'expression des gènes sans modifier la séquence de l'ADN. La maintenance épigénomique est un processus continu qui joue un rôle important dans la stabilité des génomes eucaryotes en participant à des mécanismes biologiques cruciaux tels que la réparation de l'ADN (Russell PJ 2010).

### 4.3. La protéomique

La protéomique est l'étude des protéomes<sup>5</sup>, qui sont les collections de protéines exprimées dans les cellules. Alors que les génomes sont essentiellement invariants dans différentes cellules d'un organisme, les protéomes varient d'une cellule à l'autre, avec le temps et en fonction des stimuli environnementaux et du stress (Liebler 2002). Cette approche permet d'identifier de manière globale les protéines extraites d'une culture cellulaire, d'un tissu ou d'un fluide biologique, leur localisation dans les compartiments cellulaires, leur quantité et leurs modifications post-traductionnelles<sup>6</sup>. La protéomique tente de déterminer toutes les propriétés biochimiques et physiques des protéines, telles que le poids moléculaire, le pHi, les séquences de protéines et les structures tridimensionnelles (Kratchmarova et al. 2002).

### 4.4. La transcriptomique

La transcriptomique est l'un des domaines les plus développés de l'ère post-génomique, qui repose sur l'analyse du transcriptome<sup>7</sup> qui comprend l'ARN messenger, l'ARN de transfert, l'ARN ribosomal et d'autres ARN non codants. La Transcriptomique se concentre sur l'expression des gènes au niveau de l'ARN et offre des informations à l'échelle du génome sur la structure et la fonction des gènes afin de révéler les mécanismes moléculaires impliqués dans des processus biologiques spécifiques. La première technologie utilisée est basée sur les puces à ADN "MicroArray", qui permet d'analyser un répertoire large de transcrits non exhaustif. La seconde approche, est le séquençage à haut débit du transcriptome (RNA-seq) qui a progressivement permis d'étudier un plus grand nombre d'informations, tel que l'expression de gènes, l'analyse des ARN non-codants (miRNA, lncRNA, etc) et le mécanisme d'épissage alternatif.

Après le séquençage du génome, l'analyse du transcriptome permet de comprendre l'expression du génome au niveau de la transcription, ce qui fournit des informations sur la structure des gènes, la régulation de l'expression des gènes, la fonction du produit du gène et la dynamique du génome. L'analyse du transcriptome révélera en outre le réseau de régulation

---

<sup>5</sup> Protéomes : Ensemble des protéines d'une cellule, d'un tissu, d'un organite, d'un organe ou d'un organisme à un temps "t" et sous des conditions spécifiques.

<sup>6</sup> Une modification post-traductionnelle est une modification chimique d'une protéine, réalisée le plus souvent par une enzyme, après sa synthèse ou au cours de sa vie dans la cellule.

<sup>7</sup> Le transcriptome est défini comme l'ensemble des transcrits présents dans une cellule à un moment donné et dans des conditions données

des processus biologiques et finira par donner des indications sur le diagnostic des maladies, le traitement clinique et l'amélioration des cultures.

### III. Étude de cas

#### 1. Cas n°1 : La maladie de la tuberculose

##### 1.1. La Mycobacterium Tuberculosis

###### 1.1.1 Caractéristiques du genre Mycobacterium

Mycobacterium est un genre d'Actinobacteria, étant donné sa propre famille, Mycobacteriaceae. Plus de 190 espèces, réparties en espèces saprophytes<sup>8</sup> ou commensales et des espèces pathogènes font partis de ce genre.

Le genre Mycobacterium comprend des agents pathogènes connus pour causer des maladies graves chez les mammifères, notamment la tuberculose (*Mycobacterium tuberculosis*) et la lèpre (*Mycobacterium leprae*) chez l'homme (Tableau 1). Ces mycobactéries sont des micro-organismes d'une importance écologique et médicale.

**Tableau 1:Espèces mycobactériennes majeures**

<b>Espèces pathogènes pour l'homme</b>	<b>Espèces opportunistes pour l'homme et/ou pathogènes pour les animaux</b>	<b>Espèces saprophytes</b>
<i>M. bovis</i>	<i>M. africanum</i>	<i>M. duvalii</i>
<i>M. tuberculosis</i>	<i>M. avium</i>	<i>M. flavescens</i>
<i>M. leprae</i>	<i>M. chelonae</i>	<i>M. gadium</i>
<i>M. ulcerans</i>	<i>M. farcinogenes</i>	<i>M. gastri</i>
	<i>M. fortuitum</i>	<i>M. gilvum</i>
	<i>M. intracellulare</i>	<i>M. gordonae</i>
	<i>M. kansasii</i>	<i>M. nonchromogenicum</i>
	<i>M. lepraemurium</i>	<i>M. parafortuitum</i>
	<i>M. marinum</i>	<i>M. phlei</i>
	<i>M. paratuberculosis</i>	<i>M. smegmatis</i>
	<i>M. scrofulaceum</i>	<i>M. terrae</i>
	<i>M. szulgai</i>	<i>M. triviale</i>
	<i>M. xenopi</i>	<i>M. vaccae</i>

<sup>8</sup> Espèces saprophytes c'est des micro-organismes qui vivent aux dépens de matières organiques inertes, par opposition au parasite, et qui ne sont généralement pas pathogène chez l'homme

**Le nom d'espèce est précédé du nom de genre : M. signifiant Mycobacterium**

La définition du genre Mycobacterium repose sur trois critères dont l'acido-alcool résistance des bacilles, le contenu en GC de l'ADN compris entre 61 et 71 %, et la synthèse d'acides mycoliques de 60 à 90 C libérant des esters de pyrolyse de 22 à 26 C (Vincent 1995). Ce genre peut être divisé en 3 grands groupes :

- **Complexe *Mycobacterium tuberculosis*** comprend *M. tuberculosis*, *M. bovis*, *M. africanum*, *M. microti* et *M. canetti*. Ce complexe provoque la maladie tuberculose chez l'homme, mais la bactérie *Mycobacterium tuberculosis* est la cause principale de la majorité des cas tuberculeux.
- **Mycobacterium non tuberculeuse (MNT)** comprend toutes les autres mycobactéries potentiellement pathogènes pour l'être humain. Ce groupe provoque parfois (rare) des symptômes similaires à celle de la tuberculose, mais il n'est pas transmissible, ainsi que pour les personnes dont le système immunitaire est intact ou le tissu pulmonaire est sain, les MNT ne sont pas pathogènes.
- ***Mycobacterium leprae*** c'est la cause de la lèpre.

#### 1.1.2 Caractéristiques de *Mycobacterium tuberculosis*

- a. **Histoire :** Mycobactérium tuberculosis découverte par Robert Koch en 1882 (Worboys 1990) (Steve M.Blevins et Michael S.Bronze 2010), c'est la bactérie responsable de la maladie tuberculose. Cette bactérie appartient au genre des mycobactéries (*Mycobacterium* spp.)
- b. **Morphologie :** *M. Tuberculosis* est un bacille aérobie strict, fin, droit légèrement incurvé, assez long entre 2µm à 5µm. Au microscope, il faut utiliser la méthode *Ziehl-Neelsen*<sup>9</sup> révélant son alcool-acido résistance (propriété de toutes les *Mycobacterium*), ces bactéries sont difficilement colorable à cause d'une couche cireuse externe, elles appartiennent au type Gram positif (+) étant donné leur absence de membrane externe (I. Smith 2003).
- c. **Caractéristiques génétiques :** Afin d'améliorer notre compréhension de la biologie du bacille tuberculeux, le génome de la souche **H37Rv** la plus utilisée, a été

---

<sup>9</sup> La coloration de Ziehl-Neelsen est une méthode de coloration permettant l'identification des mycobactéries au microscope. Elle fait partie des colorations qui mettent en évidence l'acido-alcool-résistance, caractère fondamental des mycobactéries, en prenant en compte la difficulté de pénétration des colorants.

entièrement séquencé en 1998 (S. T. Cole et al. 1998). Ce génome contient un chromosome circulaire de 4 411 529 pb (4,42 Mb) et présente une teneur moyenne en %GC de 65,6%. L'analyse bio-informatique a permis d'identifier 50 gènes codants pour des espèces d'ARN stable et 3924 gènes codants pour des protéines (Stewart T Cole 1999).

**d. La culture :** La culture des *M.Tuberculosis* exige des milieux spéciaux tel que des milieux sélectifs en micro-aérobi ou en aérobie stricte, ainsi que la *M.Tuberculosis* est un bacille à croissance très lente ( de 2 à 6 semaines), alors le temps de régénération est compris entre 20 et 38 heures. La culture de cette bactérie permet d'identifier les espèces et de tester leur sensibilité aux antibiotiques. La culture primaire se fait sur un milieu solide, celui de *Lowenstein-Jensen* et *Coletsos*, ce sont des milieux à base d'œufs, elle est obtenue en moyenne après 2 à 4 semaines d'incubation. La culture en milieux liquides améliore la croissance soit en moyenne 5 à 14 jours, selon la richesse de l'échantillon clinique. Dans le suivi de patients, la culture présente l'avantage de mettre en évidence le caractère viable de *M. tuberculosis* et de permettre la pratique d'un antibiogramme phénotypique en vue d'adapter le traitement (Dunn, Starke, et Revell 2016). La culture peut également fournir des tests de résistance aux médicaments, cela signifie des tests pour déterminer si une personne est résistante à des médicaments antituberculeux. Les tests de résistance aux médicaments antituberculeux consistent à cultiver des bactéries tuberculeuses en présence d'antituberculeux. Si les bactéries se développent, cela signifie qu'elles résistent aux médicaments. Si les bactéries ne se développent pas, cela signifie que les médicaments sont efficaces et que les bactéries ne résistent pas.

### 1.1.3 La *Mycobacterium tuberculosis* résistante

La tuberculose est un grave problème de santé publique dans le monde. Sa situation est aggravée par la présence de souches multi résistantes de *Mycobacterium tuberculosis*, l'agent responsable de la maladie. Les souches de *Mycobacterium tuberculosis* sont réparties selon le type de résistance (Palomino et Martin 2014):

- **Monorésistance :** résistance à un seul antituberculeux de première ligne de résistance.

- **Multirésistance (MDR)** : résistance à l'isoniazide et à la rifampicine, au moins.
- **Ultrarésistance (XDR)** : multirésistance avec l'ajout de la résistance aux fluoroquinolones et l'un des antituberculeux de la deuxième ligne de résistance (capréomycine, kanamycine et amikacine).
- **Totalrésistance (TDR)** : résistance au-delà de MDR et XDR, due à une mauvaise gestion des médicaments.

Les souches de *M. Tuberculosis* (MDR) présentant une résistance in vitro à l'isoniazide et à la rifampicine ont représenté environ 480000 cas incidents et 250 000 décès en 2015. Les souches de XDR-TB présentent une résistance supplémentaire à la fois aux fluoroquinolones et aux agents injectables de deuxième intention. À ce jour, elles seraient à l'origine de la maladie dans 106 pays (« WHO | World Health Organization » 2018).

#### a. La première ligne des antituberculeux

**Rifampicine** : Il est un composé naturel isolé de *Streptomyces mediterranei*<sup>10</sup> et introduit en 1972 en tant qu'agent antituberculeux. Il s'agit de l'un des antibiotiques antituberculeux les plus efficaces. La rifampicine est active contre les bacilles qui sont en croissance et les bacilles qui ne le sont pas (métabolisme lent) (Mitchison D.A 1979). Le mode d'action de la rifampicine est de se lier à la sous-unité  $\beta$  de l'ARN polymérase, en inhibant l'élongation de l'ARN messager (Blanchard 1996). La majorité des souches de *M. tuberculosis* résistants à la rifampicine portent des mutations dans le gène *rpoB*<sup>11</sup> qui code pour la sous-unité  $\beta$  de l'ARN polymérase. Environ 96% des souches isolées de *M. Tuberculosis* résistantes à la rifampicine ont des mutations dans des codons entre la position 507 à 533 (81 pb) du gène *rpoB*. La mono-résistance à la rifampicine est assez rare et presque toutes les souches résistantes à la rifampicine sont également résistantes à d'autres médicaments, en particulier à l'isoniazide.

**Isoniazid** : Il a été introduit en 1952 en tant qu'agent antituberculeux. L'Isoniazid et la rifampicine sont à la base du traitement de la maladie Tuberculose, mais contrairement à la rifampicine, l'isoniazide n'est actif que contre les bacilles de

---

<sup>10</sup> ***Amycolatopsis mediterranei*** est une bactérie à Gram positif de l'ordre des Actinomycetales. C'est à partir de ce microorganisme que les antibiotiques de la famille des rifamycines ont été découverts<sup>2</sup>.

<sup>11</sup> **rpoB** (DNA-directed RNA polymerase subunit beta): Gene ID: 888164

réplication métaboliquement actif. C'est un précurseur du médicament qui doit être activé par l'enzyme *catalase/peroxydase*<sup>12</sup> **KatG** (Y. Zhang et al. 1992), codée par le gène **KatG**<sup>13</sup>. L'isoniazide agit en inhibant la synthèse des acides mycoliques qui sont des constituants essentiels de la paroi mycobactérienne, codée par **inhA**<sup>14</sup>. La résistance à ce médicament a été associée à des mutations de plusieurs gènes, tels que **katG**, **inhA**, **ahpC**<sup>15</sup>, et **kasA**<sup>16</sup>. Les deux principaux mécanismes moléculaires de la résistance à l'isoniazide sont associés à des mutations géniques dans **katG** et **inhA** (Rawat, Whitty, et Tonge 2003) Parmi ceux-ci, la mutation génique la plus répandue a été identifiée comme étant **S315T** dans **katG**. Cette mutation a été associée de manière constante à un niveau élevé de résistance (MIC > 1 µg / mL) à l'isoniazide et se produit plus fréquemment dans les souches MDR (Sensoy et al. 2016). La deuxième mutation la plus fréquente se produit dans la région promotrice d'**inhA**, provoquant une surexpression d'**InhA** (Palomino et Martin 2014).

**Ethambutol** : Il a été introduit en 1966 dans le traitement de la tuberculose et fait actuellement partie du schéma thérapeutique actuel. La résistance à l'Ethambutol a été liée aux mutations du gène **embB**<sup>17</sup>, ces mutations situées à la position **embB306** sont les plus courantes dans la plupart des études réalisées (Telenti et al. 1997), mais d'autres études ont montré que les mutations de l'**embB306** n'étaient pas nécessairement associées à une résistance à l'Ethambutol mais liées à une prédisposition à développer une résistance. Selon les études, environ 30% de souches résistantes à l'Ethambutol ne présentent aucune mutation dans l'**embB** (Palomino et Martin 2014).

**Pyrazinamide** : Elle a été introduite dans le traitement de la tuberculose au début des années 50 et fait désormais partie du schéma thérapeutique standard du traitement de

---

<sup>12</sup> Une **catalase-peroxydase** est une oxydoréductase qui catalyse les réactions :

(1) donneur + H<sub>2</sub>O<sub>2</sub> ⇌ donneur oxydé + 2 H<sub>2</sub>O ;

(2) 2 H<sub>2</sub>O<sub>2</sub> ⇌ O<sub>2</sub> + 2 H<sub>2</sub>O.

<sup>13</sup> **katG**: catalase-peroxydase. Gene ID: 885638

<sup>14</sup> **inhA**: NADH-dependent enoyl-[ACP] reductase. Gene ID :886523

<sup>15</sup> **ahpC**: alkyl hydroperoxide reductase subunit AhpC. Gene ID: 885717

<sup>16</sup> **kasA** 3-oxoacyl-ACP synthase 1. Gene ID : 887269

<sup>17</sup> **embB** arabinosyltransferase B. Gene ID:886126



la maladie Tuberculose. La pyrazinamide est un analogue du *nicotinamide*<sup>18</sup> et son introduction a permis de réduire la durée du traitement à six mois. Il a pour caractéristique d'inhiber les bacilles semi-dormants résidant dans des environnements acides tels que ceux présents dans les lésions tuberculeuses. Le mécanisme d'action proposé du pyrazinamide implique la conversion du pyrazinamide en acide pyrazinoïque, ce qui perturbe le transport membranaire par l'énergétique de la membrane bactérienne. Chez les souches de M.Tuberculosis résistants à la pyrazinamide, des mutations dans le gène *pncA*<sup>19</sup> sont les plus fréquentes, ces mutations sont focalisées dans tout le gène. Des études ont montré que quelques souches résistantes à la pyrazinamide n'ont aucune mutation au niveau du gène *pncA* (Palomino et Martin 2014).

**Streptomycin :** Elle est un glycoside d'aminocyclitol actif contre les bacilles en croissance active et son mode d'action consiste à inhiber l'initiation de la traduction dans la synthèse des protéines. Plus spécifiquement, la streptomycine agit au niveau de la sous-unité 30S du ribosome au niveau de la protéine ribosomale S12 et de l'ARNr 16S codé par les gènes *rpsL*<sup>20</sup> et *rrs*<sup>21</sup>, respectivement. Les mutations dans *rpsL* et *rrs* sont les principaux mécanismes de résistance à la streptomycine, mais représentent 60% à 70% de la résistance retrouvée. Parmi les mutations signalées dans *rpsL*, une substitution dans le codon 43 de la lysine à l'arginine est la forme la plus fréquente. Cette mutation produit une résistance élevée à la streptomycine. Également, des mutations dans le gène *gidB*<sup>22</sup> qui code pour 7-méthylguanosine méthyltransférase, causent une faible résistance à la streptomycine (Palomino et Martin 2014).

---

<sup>18</sup> **Le nicotinamide** : une vitamine hydrosoluble et fait partie du groupe de vitamines B

<sup>19</sup> **pncA**: pyrazinamidase/nicotinamidase. Gene ID :888260

<sup>20</sup> **rpsL** :30S ribosomal protein S12. Gene ID : 888259

<sup>21</sup> **rrs**: 16S ribosomal RNA.Gene ID: 2700429

<sup>22</sup> **gidB**: 16S rRNA (guanine (527)-N (7))-methyltransferase RsmG. Gene ID :886243

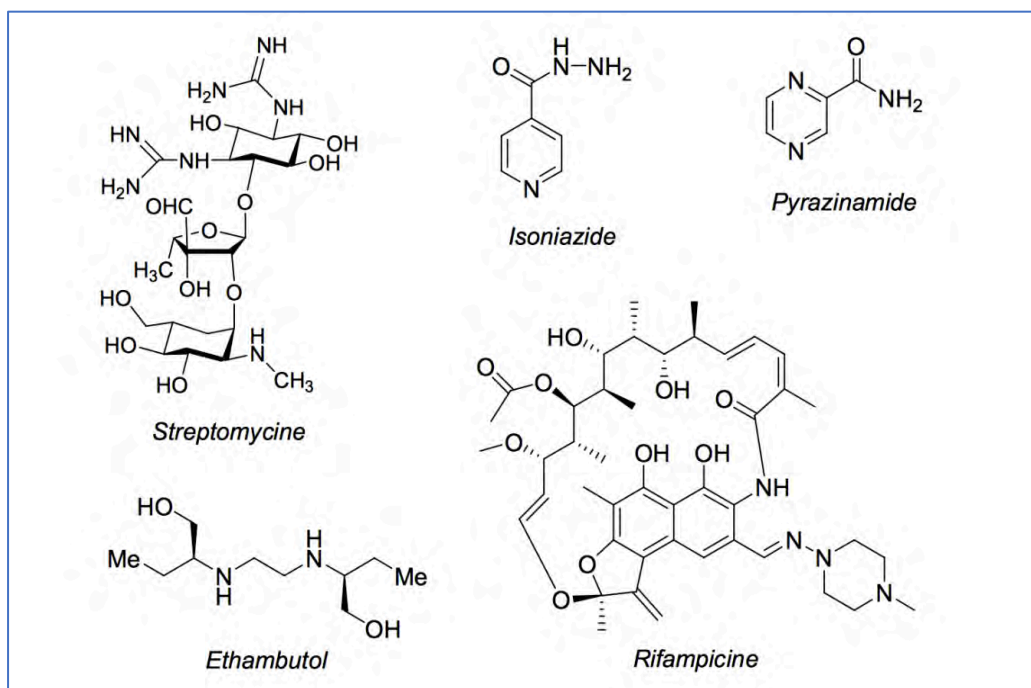


Figure 9: Composition chimique des antituberculeux de la première ligne

### b. La deuxième ligne des antituberculeux

**Fluoroquinolones :** Ils sont actuellement utilisés en tant que médicaments de deuxième intention dans le traitement de la TB-MDR. Des quinolones de nouvelle génération telles que la *moxifloxacin*<sup>23</sup> et la *gatifloxacin*<sup>24</sup> sont en cours d'essais cliniques et proposées comme antibiotiques de première intention dans le but de raccourcir la durée du traitement de la tuberculose (Fàbrega et al. 2009). Le mode d'action des fluoroquinolones consiste à inhiber la *topoisomérase*<sup>25</sup> II (ADN gyrase) et la *topoisomérase* IV, deux enzymes essentielles à la viabilité bactérienne. Ces protéines sont codées par les gènes *gyrA*<sup>26</sup>, *gyrB*<sup>27</sup>, *parC*<sup>28</sup> et *parE*<sup>29</sup>, respectivement (Aubry et al. 2004). Chez M.Tuberculosis, seule la *topoisomérase de type II* (ADN

<sup>23</sup> La **moxifloxacin** est une molécule antibiotique, de la classe des fluoroquinolones.

<sup>24</sup> La **gatifloxacin** est un médicament antibiotique de la classe des quinolones, dit de quatrième génération.

<sup>25</sup> Les **ADN topoisomérases** (de type I : ADN topo-isomérases et de type II : ADN gyrases) sont des enzymes qui contrôlent la structure topologique de l'ADN en générant des coupures transitoires dans celui-ci et en catalysant le passage des segments d'ADN à travers ces coupures avant de les refermer.

<sup>26</sup> **gyrA** : DNA gyrase subunit A. Gene ID:887105

<sup>27</sup> **gyrB** : DNA gyrase subunit B. Gene ID:887081

<sup>28</sup> **parC**: DNA topoisomerase IV subunit A. Gene ID: 947499

<sup>29</sup> **parE**: DNA topoisomerase IV subunit B. Gene ID:947501

gyrase) est présente et constitue donc la seule cible d'activité de la fluoroquinolone. Le principal mécanisme de développement de la résistance à la fluoroquinolone chez *M. tuberculosis* est constitué de mutations chromosomiques dans la région de *gyrA* ou *gyrB*, déterminante de la résistance à la quinolone (Maruri et al. 2012). Les mutations les plus fréquentes trouvées se situent aux positions 90 et 94 du *gyrA*, mais des mutations aux positions 74, 88 et 91 ont également été détectées. Chez *M. Tuberculosis*, la présence d'un polymorphisme naturel en position 95 dans *gyrA* n'est pas lié à la résistance à la fluoroquinolone, car ce polymorphisme se trouve également dans les souches sensibles à la fluoroquinolone (Sun et al. 2008).

**Kanamycin, Capreomycin, Amikacin, Viomycin :** La kanamycine et l'amikacine sont des aminosides, la capréomycine et viomycine sont des antibiotiques à peptides cycliques, et ces quatre antibiotiques ont le même mécanisme d'action en inhibant la synthèse des protéines. La kanamycine et l'amikacine inhibent la synthèse des protéines par modification au niveau de l'ARNr 16S. Les mutations les plus courantes observées dans les souches résistantes à la kanamycine se situent aux positions 1400, 1401 et 1483 du gène *rrs*, conférant un haut niveau de résistance à la kanamycine et à l'amikacine (Alangaden et al. 1998). Les mutations dans la région promotrice du gène *eis*<sup>30</sup> codant pour *aminoglycoside*<sup>31</sup> *acétyltransférase*<sup>32</sup> une provoque une faible résistance à la kanamycine (Zaunbrecher et al. 2009). Des mutations aux positions -10 et -35 du promoteur *eis* ont conduit à une surexpression de la protéine et à un faible niveau de résistance à la kanamycine mais pas à l'amikacine. Ces mutations ont été retrouvées dans 80% des isolats cliniques présentant une faible résistance à la kanamycine (Campbell et al. 2011) (Zaunbrecher et al. 2009).

La capréomycine et la viomycine ont une structure similaire et se lient au même site dans le ribosome, à l'interface des petites et des grandes sous-unités (Stanley et al. 2010) (Palomino et Martin 2014). Des mutations du gène *tlyA*<sup>33</sup> ont également été associées à la résistance à la capréomycine et à la viomycine (Johansen et al. 2006).

---

<sup>30</sup> **eis** : enhanced intracellular survival protein. Gene ID :885903

<sup>31</sup> **Aminoglycosides** : constituent une famille d'antibiotiques actifs sur certains types de bactéries. Ils comprennent l'amikacine, l'isépamicine, la gentamicine, la kanamycine, la streptomycine ...

<sup>32</sup> **acétyltransférase** (ou transacétylase) est une enzyme de type transférase qui transfère un groupe acétyle.

<sup>33</sup> **TlyA** : un ARNr méthyltransférase spécifique de la 2'-O-méthylation du ribose dans l'ARNr. Gene ID :885396

**Ethionamide :** Il est un dérivé de l'acide *isonicotinique*<sup>34</sup> dont la structure est similaire à celle de l'isoniazide. C'est également un précurseur qui nécessite une activation par une monooxygénase codée par le gène *ethA*<sup>35</sup>. *EthA* est régulée par le répresseur transcriptionnel *ethR*<sup>36</sup>. La résistance à l'éthionamide se produit par des mutations dans *etaA* / *ethA*, *ethR* ainsi que par des mutations dans *inhA*, qui provoquent une résistance à la fois à l'isoniazide et à l'éthionamide (Brossier et al. 2011).

**Para-Amino Salicylic Acid :** Il était l'un des premiers médicaments antituberculeux utilisés dans le traitement de la maladie, avec l'isoniazide et la streptomycine, l'acide para-amino-salicylique ou PAS est maintenant considéré comme un médicament de deuxième intention dans le schéma thérapeutique de la TB-MDR. Diverse mutations faux-sens dans la *dihydrofolate synthase* codant pour le *folC*<sup>37</sup> cause une résistance au PAS liée aux patients tuberculeux (F. Zhao et al. 2014). Moins de 40% des souches résistantes au PAS présentent des mutations dans la *thyA*, ce qui indique qu'il pourrait exister d'autres mécanismes de résistance au PAS (Mathys et al. 2009).

**Thiacetazone :** c'est un médicament ancien utilisé dans le traitement de la tuberculose en raison de son activité *in vitro* favorable contre *M. tuberculosis* et de son coût très faible. Il présente toutefois des problèmes de toxicité, en particulier chez les patients co-infectés par le VIH. Il appartient au groupe 5 des médicaments de l'OMS et agit en inhibant la synthèse de l'acide mycolique (Andini et Nash 2006).

**Macrolides :** Ils sont plus fréquemment recommandés pour le traitement d'autres infections à mycobactéries en raison de leur activité limitée contre *M. tuberculosis*. Parmi eux, la *clarithromycine*<sup>38</sup> qui est considérée comme faisant partie des médicaments du groupe 5 de l'OMS. La résistance intrinsèque aux macrolides a été

---

<sup>34</sup> L'acide **isonicotinique** est un composé aromatique constitué d'un noyau de pyridine (azine) substitué par un groupe carboxyle en position 4.

<sup>35</sup> **ethA** : monooxygenase EthA. Gene ID : 886175

<sup>36</sup> **ethR** : HTH-type transcriptional repressor EthR. Gene ID : 886189

<sup>37</sup> **folC** : foylpolypolyglutamate synthase. Gene ID: 85902

<sup>38</sup> **La clarithromycine** est un antibiotique macrolide utilisé pour traiter les pharyngites, les angines, les sinusites aiguës maxillaires, en association avec d'autres antibiotiques pour traiter l'infection à *Helicobacter pylori*, les complications bactériennes aiguës de la bronchite chronique, les pneumonies, les infections cutanées, et, chez les patients porteurs de VIH, pour prévenir et traiter le *Mycobacterium avium* complex disséminé.

attribuée à l'expression de *emr37*, gène codant pour une méthylase sur un site spécifique de l'ARNr 23S, bloquant la liaison de l'antibiotique (Andini et Nash 2006). Le tableau 2 récapitule les médicaments antituberculeux de première et de deuxième intention actuellement utilisés et ciblés.

**Tableau 2: Antituberculeux de première et de deuxième intention, gènes impliqués dans leur activation**

Drug	Gene
Isoniazid	<i>katG, inhA</i>
Rifampicin	<i>rpoB</i>
Pyrazinamide	<i>pncA, rpsA</i>
Ethambutol	<i>embB</i>
Streptomycin	<i>rpsL, rrs, gidB</i>
Quinolones	<i>gyrA, gyrB</i>
Capreomycin	<i>rrs, tlyA</i>
Kanamycin/Amikacin	<i>rrs</i>
Ethionamide	<i>ethA</i>
Para-aminosalicylic acid	<i>thyA, folC</i>

### 1.2. La maladie de la tuberculose

La tuberculose (TB) est une maladie infectieuse contagieuse dont la transmission est interhumaine se fait par voie aérienne, elle est une maladie microbienne due au complexe *Mycobacterium Tuberculosis* ou bacille de Koch (décrit en 1882 par Robert Koch), *M.bovis*, *M.africanum*, *M.microti* et *M.pinnipedii*.

Les bactéries responsables de la tuberculose se propagent d'une personne à l'autre par de minuscules gouttelettes libérées dans l'air par la toux et les éternuements. La TB est une maladie mortelle mais curable qui constitue jusqu'à nos jours l'une des maladies les plus meurtrières dans le monde et particulièrement dans les pays en voie de développement dont les déterminants sont essentiellement représentés par les conditions socio-économique défavorable (World Health Organisation 2018).

Elle est en effet responsable d'environ 1,7 million de décès par an selon l'Organisation Mondiale de la Santé (OMS), la tuberculose se situe en deuxième position après le VIH/sida.

### 1.2.1. Les types de Tuberculose

Pour la plupart des personnes en bonne santé, le système immunitaire est capable de détruire les bactéries responsables de la tuberculose mais dans certains cas les bactéries infectent le corps,

Les médecins distinguent deux types d'infection tuberculeuse : latente et active.

**TB latente** : les bactéries infectent le corps mais ne provoquent aucun symptôme (état inactif) et ils ne sont pas contagieux, mais ils peuvent devenir actifs.

**TB active** : la bactérie provoque des symptômes en quelques semaines, mois ou même quelques années et peut être transmise à d'autres personnes saines (I. Smith 2003).

Le risque de transformation de la TB latente vers une TB active, concerne en particulier les personnes dont les défenses immunitaires sont affaiblies par des facteurs comme la malnutrition, le diabète, le tabagisme, mais aussi le VIH, ou certains médicaments.

On peut répartir la maladie de la tuberculose sur deux catégories de patients :

#### a. La tuberculose de l'adulte

Il existe plusieurs types de tuberculose qui atteignent les personnes adultes :

**La tuberculose pulmonaire** : La tuberculose atteint les poumons dans 70 à 80 % des cas, parfois en association avec d'autres localisations. Les bactéries détruisent les tissus pulmonaires, créant ainsi des cavités. La maladie reste localisée dans les poumons.

**La tuberculose extra-pulmonaire** : Ce type est souvent associé à une atteinte pulmonaire, ainsi qu'à d'autres parties du corps, comme les atteintes osseuses (os), l'urogénétales digestives (intestinale, hépatique), les reins, les ganglions lymphatiques, les méninges ou le système nerveux central.

**La tuberculose miliaire ou disséminée** : C'est une forme rare de maladie active qui se produit lorsque la bactérie de la tuberculose pénètre dans le sang. Sous cette forme, les bactéries se propagent rapidement dans tout le corps dans de nombreux nodules minuscules et affectent plusieurs organes à la fois. Cette forme de tuberculose peut être rapidement fatale. Il s'agit donc d'une tuberculose pulmonaire et extra-pulmonaire.

#### b. La tuberculose de l'enfant

La tuberculose est une cause importante de morbidité et de mortalité chez les enfants des pays d'endémie tuberculeuse. Le fardeau de la tuberculose et de la tuberculose chez l'enfant représente une transmission récente et continue dans la communauté. Le diagnostic est

généralement clinique et simple dans de nombreux cas, mais peut être difficile, en particulier dans les groupes les plus exposés aux maladies sévères et disséminées, tels que les nourrissons, les jeunes enfants et les enfants mal nourris ou infectés par le VIH. Selon les estimations de l'OMS, environ 1 million d'enfants ont contracté la tuberculose et 233 000 sont décédés de la tuberculose en 2017 (80 % des enfants < de 5 ans ) (« WHO | World Health Organization » 2018).

### 1.2.2. Les symptômes de la maladies tuberculose

Les symptômes de la tuberculose varient en fonction de la partie du corps touchée.

La tuberculose évolue généralement lentement et il faut parfois plusieurs semaines pour que les symptômes commencent à apparaître et pourraient ne pas apparaître avant plusieurs mois, voire plusieurs années après la première infection (Y. Zhou et al. 2019). Ces symptômes peuvent avoir différentes causes et ne sont pas toujours un signe de tuberculose, et ils sont plus ou moins associés entre eux.

Pour le cas le plus fréquent, la tuberculose pulmonaire se manifeste par :

- Une fièvre traînante, avec souvent des sueurs nocturnes ;
- Une toux chronique avec des crachats épais, comportant parfois quelques filets de sang ;
- Une perte d'appétit, un amaigrissement
- Des douleurs au niveau de la poitrine ;
- Une fatigue extrême ;
- Accélération du rythme respiratoire.

Sans traitement, la tuberculose peut se propager à d'autres parties du corps par la circulation sanguine, c'est la tuberculose miliaire qui se déclenche avec les symptômes généraux de la tuberculose pulmonaire ainsi que des symptômes spécifiques à la partie du corps atteinte (« WHO | World Health Organization » 2018).

- **Atteinte abdominale** : douleurs abdominales à la colonne vertébrale et à différentes parties osseuses touchant surtout la hanche et le genou, diarrhée, la destruction des articulations.
- **Atteinte neurologique et méningée** : maux de tête, troubles physiques, ainsi que la méningite tuberculeuse atteint surtout les enfants et les adultes.

- **Atteinte urinaire** : touchant le foie et les reins, peut altérer les fonctions de la filtration des déchets et entraîner la formation de sang les urines.
- **Atteinte cardiaque**

### 1.2.3. Facteurs à risque

Les facteurs à risque de la maladie tuberculose sont bien établis, tels que les personnes dont le système immunitaire est affaibli à cause d'une maladie du système immunitaire, comme l'infection par le virus d'immunodéficience humaine (VIH), ainsi que les enfants moins de 5ans et les personnes plus de 60 ans. Le diabète, l'alcool, la malnutrition, la fumée de tabac, l'utilisation de médicaments immunosuppresseurs et la pollution de l'air intérieur sont des facteurs qui affectent une plus grande partie de la population et accélèrent la progression de la maladie.

Des facteurs socioéconomiques, comportementaux et les risques de mixité sociale (avec le surpeuplement) sont accrus, augmentent la susceptibilité à l'infection et la transmission de la maladie.

Des groupes spécifiques tels que les personnes qui travaillent ou résident dans un milieu où vivent ou circulent des patients avec une tuberculose en phase active (les hôpitaux, les prisons, les centres d'accueil) ainsi que les personnes manipulant des bactéries en laboratoire (Narasimhan et al. 2013).



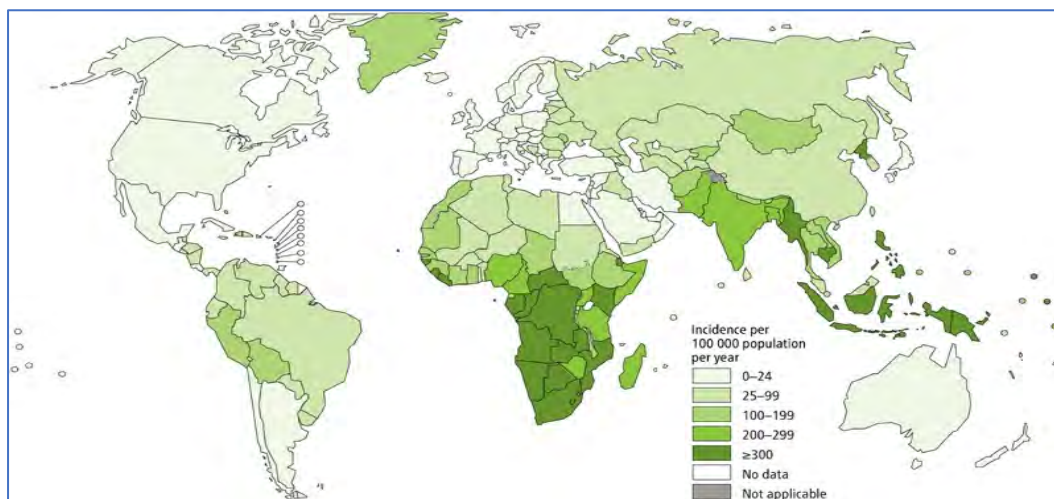
#### 1.2.4. Statistiques sur la propagation de la maladie tuberculose

##### ▪ La tuberculose dans le monde

L'organisation mondiale de santé et Éducation à la santé mondiale (GHE<sup>39</sup> Global Health Organisation) estime à 10,0 millions le nombre de nouveaux cas de tuberculose (également appelée tuberculose active) en 2017. La tuberculose concerne tous les pays et tous les groupes d'âge, mais dans l'ensemble, les meilleures estimations pour 2017 étaient que 90% étaient des adultes (âgés de 15 ans ou plus), 64% étaient des hommes et 9% des personnes vivant avec le VIH (dont 72% en Afrique). Les deux tiers se trouvaient dans huit pays : Inde (27%), Chine (9%), Indonésie (8%), Philippines (6%), Pakistan (5%), Nigéria (4%) et Afrique du Sud (3%).

Seuls 6% des cas étaient dans la Région européenne de l'OMS et la Région des Amériques de l'OMS, qui comptent chacun 3% des cas (Figure 9) (Global Health Education 2018)).

En 2017, environ un million et dix mille enfants ont contracté la tuberculose. La gravité des épidémies nationales varie considérablement. En 2017, il y avait moins de 10 nouveaux cas pour 100 000 habitants dans la plupart des pays à revenu élevé. Il y en avait 150 à 400 dans la plupart des 30 pays fortement touchés par la tuberculose et plus de 500 dans quelques pays, dont le Mozambique, les Philippines et l'Afrique du Sud (Global Health Education 2018)

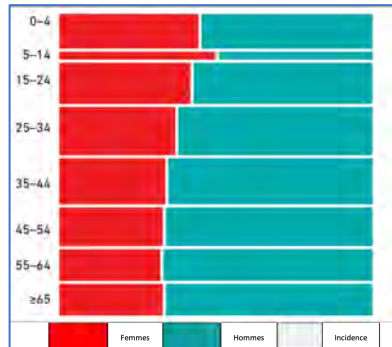


(Figure 10).

**Figure 10: Taux d'incidence estimé de la tuberculose en 2016**

<sup>39</sup> GHE est un organisme de bienfaisance enregistré (n° 1146692) basé au Royaume-Uni. Les initiales GHE représentent Global Health Education. GHE a été créé en tant qu'association pour améliorer la santé par l'éducation, en particulier en ce qui concerne les maladies telles que la tuberculose qui ont un impact mondial.

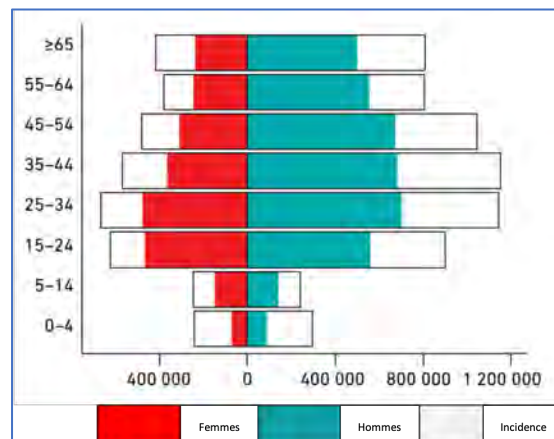
Les estimations de l'incidence de la tuberculose réparties par âge et par sexe sont illustrées à la figure 11 (mondiale), 90% des adultes et 10% des enfants (âgés de moins de 15 ans). À l'échelle mondiale, 64% des cas concernaient des hommes et des garçons et 36%, des femmes et des filles.<sup>2</sup> Les pourcentages les plus élevés de cas de tuberculose chez les hommes montrent que la tuberculose affecte davantage les hommes que les femmes et que les lacunes en matière de dépistage et de notification des cas sont plus importantes chez les hommes



**Figure 11: Distribution des cas de TB par âge et par sexe, dans le monde, 2017**

(« WHO | World Health Organization » 2018).

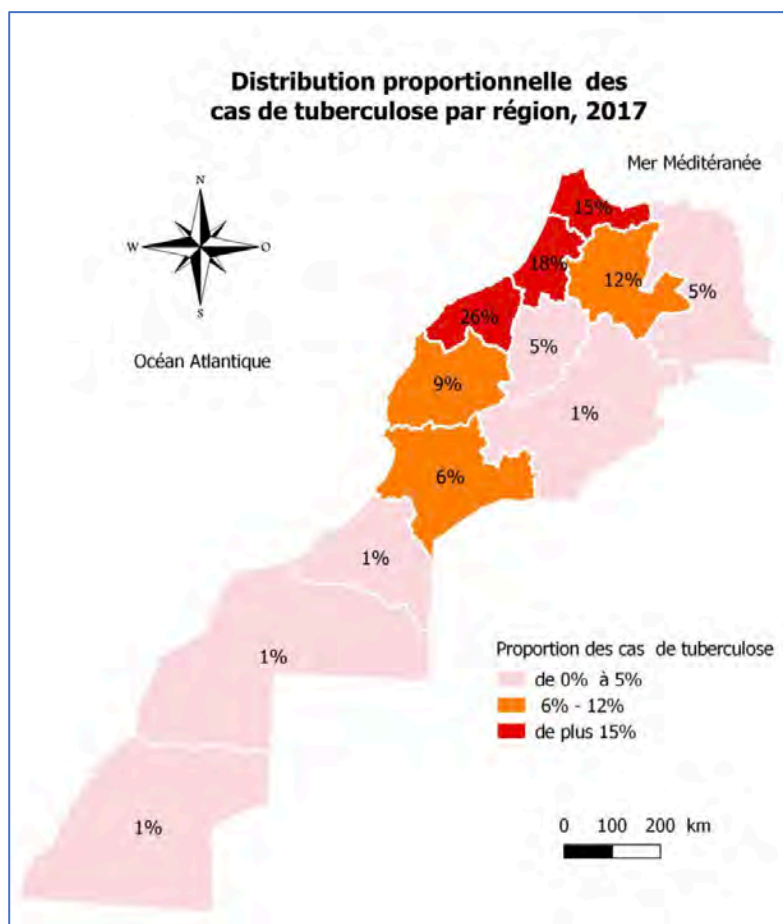
Les estimations de la mortalité par tuberculose en 2017 réparties par âge et par sexe dans le monde (figure 12), représentent 63% des décès sont des hommes et des garçons, et 37% des femmes et filles. Les enfants âgés de moins de 15 ans représentent 15 % du nombre total de décès.



**Figure 12: Distribution de la mortalité de TB par âge et par sexe, dans le monde, 2017**

▪ **La tuberculose au Maroc**

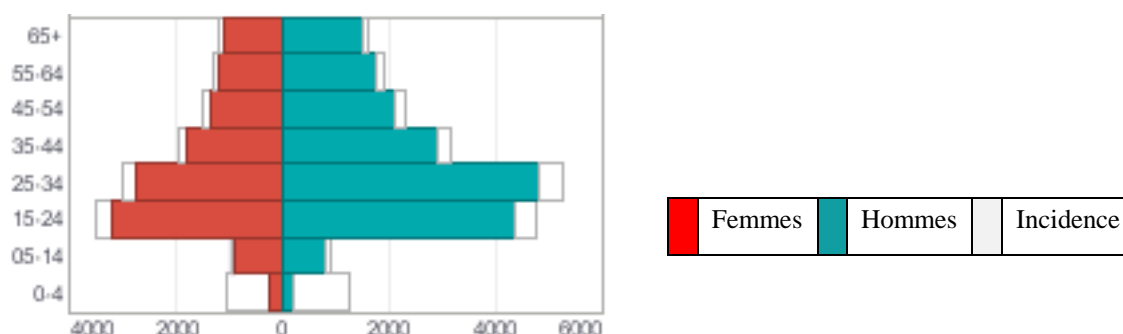
Au Maroc, en 2017 un total de 30 897 cas de tuberculose, toutes formes confondues, a été déclaré par le ministère de la Santé. Cette maladie demeure un problème de santé publique non négligeable. L’incidence de la tuberculose atteint 91/100 000 habitants, avec 36 000 nouveaux cas chaque année, dont 31 000 ont accès au traitement. Ces cas de tuberculose, sont répartis sur 6 régions enregistrant 86%, dont vit 78% de la population marocaine (figure 13).



**Figure 13: Distribution proportionnelle des cas de tuberculose par région, Maroc, 2017**

Au Maroc, comme dans plusieurs autres pays, les taux de détection de la tuberculose (TB) chez les hommes est plus élevés que chez les femmes. Les nouveaux cas enregistrés en 2017, représentent un sex-ratio H/F = 1,5 au Maroc, et ce ratio (H/F) s’établit à 1,7 au plan mondial. Les enfants moins de 15 ans représentent 11,6 % de la populations atteintes, ainsi que pour les patients âgés entre 15 ans et 44 ans le pourcentage s’élève à 64 %. Selon le ministère de

la santé au Maroc, l'âge moyen des patients atteints de la tuberculose est de 36 ans (Figure 14).



**Figure 14: Distribution des cas de TB par âge et par sexe, Maroc 2017**

Selon l’OMS, l’incidence a baissé de 1,1 % comme moyenne annuelle entre 1990 et 2015, ce pourcentage reste en dessous des attentes, tant les déterminants de la maladie sont multiples et se rapportent essentiellement aux facteurs de vulnérabilité socio-économiques des malades notamment la pauvreté, l’habitat insalubre, et la malnutrition qui jouent un rôle important dans la propagation de cette maladie, qui demeure un sérieux problème de santé publique au Maroc.

#### 1.2.5. Traitement de la Tuberculose

Le traitement de la tuberculose consiste généralement à prendre des antibiotiques pendant plusieurs mois. Bien que la tuberculose soit une maladie grave qui peut être fatale si elle n'est pas traitée, les décès sont rares si le traitement est terminé. Le traitement de la tuberculose dépend du type de la maladie (Rabahi et al. 2017) :

- **TB Pulmonaire** : Si le patient est diagnostiqué avec une tuberculose pulmonaire active, où les poumons sont atteints et les symptômes sont présents, une combinaison d’antibiotiques est prescrite dont le traitement le plus courant est l’isoniazide INH, ainsi que trois autres médicaments, la rifampicine, le pyrazinamide et l’éthambutol. La durée du traitement dépendra de l’état de santé générale et de la gravité de la tuberculose, il est important de continuer à suivre le traitement exactement comme prescrit afin d’éliminer la bactérie M.Tuberculosis définitivement. L’interruption du

traitement va provoquer une résistance aux antibiotiques qui est potentiellement grave à cause de la difficulté du traitement.

- **TB extra-pulmonaire** : Les schémas thérapeutiques sont les mêmes que ceux prescrits pour les formes pulmonaires. Une maladie extra-pulmonaire sévère, caractérisée par une atteinte neurologique, abdominale, bilatérale de la plèvre, péricardique, osseuse, articulaire ou systémique, nécessite quatre médicaments en phase intensive et parfois une durée de traitement de 9 mois (par exemple, en cas d'atteinte neurologique). En cas d'inflammation importante, la prescription de stéroïdes est recommandée. Cependant, le pronostic dépend strictement de la précocité de l'administration des médicaments antituberculeux (Sotgiu et al. 2015).

Le diagnostic et le traitement efficace des personnes atteintes de tuberculose évitent des millions de décès chaque année (environ 54 millions sur une période allant de 2000 à 2017), mais les lacunes en matière de dépistage et de traitement restent importantes et persistantes. La gestion thérapeutique actuelle des souches sensibles et résistantes aux médicaments doit encore être améliorée. Le taux de succès global du traitement est inférieur à la proportion recommandée de 85% par l'Organisation mondiale de la santé et, par conséquent, le niveau de résistance aux médicaments augmente.

En 2017, 464 633 cas de tuberculose ont été signalés chez des personnes vivant avec le VIH (51% des 920 000 nouveaux cas estimés la même année), dont 84% sous traitement antirétroviral. La plupart des lacunes dans la détection et le traitement ont été observées dans la Région africaine de l'OMS, où le fardeau de la tuberculose associée au VIH est le plus lourd.

L'objectif de l'OMS entre 2018 et 2022 est de détecter et traiter 40 millions de personnes atteintes de la tuberculose, en se basant sur les dernières données des résultats du traitement pour les nouveaux cas qui indiquent un taux de réussite global du traitement de 82% en 2016. Cela représente une réduction de 86% en 2013 et de 83% en 2015 ; dans les pays où le nombre de notifications a augmenté, la notification des résultats du traitement n'a pas suivi le même rythme.

# Objectifs du travail sur la maladie Tuberculose

- **Thématique**

La tuberculose est une maladie infectieuse causée par une bactérie appelée *Mycobacterium tuberculosis*. Cette maladie se transmet par voie aérienne. La tuberculose est dite multi-résistante (MDR), en raison de sa résistance à au moins deux des principaux antituberculeux. La tuberculose peut également être ultrarésistante (XDR), soit des souches résistantes à au moins deux médicaments de deuxième intention.

- **Objectifs**

L'objectif du projet est d'élaborer un workflow d'analyse génomique de données issues du séquenceur Illumina Miseq dans le but d'analyser la bactérie tuberculose pour différents patient marocains atteints de la tuberculose afin d'identifier les gènes de résistances et effectuer une comparaison génomique et d'avoir les caractéristiques de la souche de tuberculose marocaine.

- **Objectifs généraux**

Séquencer quatre souches de la bactérie *mycobacterium tuberculosis*

Assemblage, annotation et caractérisation de la souche de *mycobacterium tuberculosis*

Identification, analyse et comparaison des éléments génétiques communs des quatre souches séquencées isolés chez des adultes atteints de la tuberculose

L'objectif principal de mon projet de thèse est de contribuer à améliorer et à approfondir les connaissances du génome complet de l'espèce *Mycobacterium Tuberculosis* d'origine marocaine, et de caractériser les résistances pour chaque souche ainsi d'établir un workflow d'analyse génomique dédié à cette bactérie.

## 2. Cas n°2 : La maladie de Parkinson

### 2.1. Généralités

La MP (MP) doit son nom au médecin généraliste londonien James Parkinson, qui a décrit de manière vivante de nombreuses caractéristiques cliniques de la maladie en 1817 (National Collaborating Centre for Chronic Conditions (Great Britain) 2006). La MP est une maladie neurologique dégénérative chronique du système nerveux central qui affecte principalement le système moteur, elle est considérée comme le deuxième trouble neurodégénératif progressif le plus fréquent après la maladie d'Alzheimer. Elle affecte environ 1% de la population après l'âge de 50 ans et environ 2,5% de la population âgée de plus de 70 ans (L. M. L. de Lau et Breteler 2006). Le risque de développement de la MP au cours de la vie est de 2,0% chez les hommes et de 1,3% chez les femmes (Oguh et Videnovic 2012). La MP idiopathique, également appelée MP sporadique, est la forme la plus courante de la maladie. En général, elle est associée à des symptômes moteurs associés à un déficit en dopamine dans les noyaux gris centraux en raison de la dégénérescence des neurones dopaminergiques dans la *pars compacta* de la *substantia nigra* (SNpc) (Obeso et al. 2000).

### 2.2. Symptômes

#### 2.2.1. Symptômes moteurs

L'observation d'un ou plusieurs de ces symptômes est le principal moyen utilisé par les médecins pour diagnostiquer la MP (Marvanova 2016). Les quatre principaux symptômes moteurs de la MP sont les suivants :

- **Le Tremblement de repos** est le symptôme le plus caractéristique de la MP. Il s'agit d'un tremblement présent principalement au repos, il disparaît avec le mouvement, mais revient souvent lorsque le membre est maintenu dans une position comme tenir une cuillère. Les tremblements de la MP peuvent toucher presque toutes les parties du corps, mais le plus souvent les doigts, suivis des mains, de la mâchoire et des pieds - dans cet ordre. Ce tremblement commence généralement de façon unilatérale, il prédomine au niveau des extrémités des membres supérieurs. Les tracés accélérométriques montrent un pic de fréquence entre 4 et 6Hz (Stanley et al. 2010).

- **La rigidité** est définie par une résistance accrue lors de la mobilisation passive d'un membre, indépendamment de la direction et de la vitesse du mouvement, et constitue l'une des caractéristiques cardinales de diagnostic de la MP. Elle est due à une hypertonie

musculaire généralisée, continue et homogène. La rigidité concerne la quasi-totalité des muscles avec une prédominance du versant fléchisseur sur le versant extenseur (Delwaide, Sabbatino, et Delwaide 1986). La rigidité progressera plus rapidement chez certaines personnes que chez d'autres (Baradaran et al., 2013).

- **L'akinésie ou bradykinésie** est un symptôme cardinal de la MP, qui correspond à la raréfaction voir la privation du mouvement (Wilson 1925) et à la lenteur de son exécution, respectivement. En d'autres termes, les mouvements volontaires se caractérisent par un défaut d'initiation avec apparition d'un délai important entre la volonté de réaliser le mouvement et son exécution. Les activités de routines sont donc grandement perturbées (Berardelli et al. 2001)

- **L'instabilité posturale** est une manifestation tardive de la maladie, qui survient après de nombreuses années d'évolution. Elle se manifeste par des troubles de l'équilibre qui provoquent des chutes de plus en plus fréquentes. Il s'agit probablement du symptôme le plus incommodant, puisque la mobilité est réduite lors des déplacements. Si l'instabilité posturale apparaît très tôt, il y a une forte possibilité qu'elle soit symptomatique d'une autre maladie que la MP idiopathique.

### 2.2.2. Symptômes non-moteurs

Les symptômes non-moteurs associés à la MP sont : la dépression, l'anxiété, le trouble de sommeil, les troubles cognitifs, la douleur, la fatigue, la constipation, les troubles dysautonomiques, etc. Ils sont souvent négligés et pas bien pris en charge sur le plan thérapeutique.

- **La dépression et l'anxiété** font partie d'un ensemble de symptômes qui peuvent se manifester au début de la MP en précédant les symptômes moteurs. La dépression est souvent présente avant les symptômes moteurs car la dopamine joue un rôle dans la régulation de l'humeur. C'est un symptôme qui peut prédire une MP (Choudhury et Daadi 2018), mais pas le seul. La dépression et l'anxiété sont des symptômes non moteurs assez fréquents de la MP. Elles peuvent varier en gravité et s'améliorer avec le traitement dopaminergique et la psychothérapie, telle que la thérapie cognitivo-comportementale (Choudhury et Daadi 2018).

- **Le trouble de sommeil** touche les personnes atteintes de la MP. L'incapacité de s'endormir ou l'insomnie primaire est moins courante que l'incapacité de rester endormi ou



l'insomnie secondaire. D'autres personnes atteintes de la MP ont des rêves éclatants, bien que ceux-ci soient généralement dus aux effets secondaires des médicaments anti-parkinsoniens.

- **Les troubles cognitifs** : Ces troubles se manifestent chez les patients à un stade avancé de la maladie ou chez les personnes âgées atteintes de la maladie. Les problèmes de pensée, de recherche de mots et de jugement sont fréquents. Cependant, si ces symptômes apparaissent aux premiers stades de la maladie, ils peuvent être les symptômes d'un trouble apparenté (par exemple, la démence à corps de Lewy) plutôt que la MP idiopathique. De nombreuses personnes signalent des difficultés à effectuer plusieurs tâches à la fois et à organiser leurs activités quotidiennes. La confusion peut également être un effet secondaire de certains médicaments contre la MP (Choudhury et Daadi 2018).

- **La douleur** est un symptôme non moteur, complexe à détecter chez les parkinsoniens dont le seuil de perception nociceptif est plus bas que chez le sujet normal. Les patients sont davantage sensibles à la douleur. L'origine de ce symptôme reste mal comprise. Dans certaines situations, la douleur est associée à une dystonie lorsque les effets dopaminergiques (DA) s'estompent. La perte de dopamine induit une altération de la transmission nociceptive en abaissant le seuil de sensibilité mécanique et thermique (Charles et al. 2018). En outre, la douleur neuropathique cliniquement observée et d'autres sensations de douleur musculaire sont décrites par les patients parkinsoniens (Skogar et Lokk 2016). Les mécanismes physiopathologiques à l'origine de ce phénomène sont les plus probables selon lesquels la dopamine, en liaison avec d'autres monoamines telles que la noradrénaline et la 5-hydroxytryptamine (5-HT), interagissent par des voies inhibitrices et excitatrices. Les anomalies dans les voies descendantes affectent le traitement central de la douleur.

- **La fatigue** est un problème fréquent chez les personnes atteintes de la MP. Entre 30 et 70% des personnes atteintes de la MP souffrent de fatigue excessive tous les jours (Friedman et al. 2016). C'est un symptôme qui peut apparaître plusieurs années avant que des symptômes moteurs ne se manifestent. Bien qu'elle ait un impact significatif sur la qualité de vie, la compréhension de son origine ou de ses mécanismes reste difficile à déterminer. Les premières preuves suggèrent également que la fatigue dans la MP n'est pas due à un déficit en dopamine (Chong et al. 2018).

- **La constipation** est l'un des symptômes non moteurs les plus fréquents associé à un désordre au niveau du système autonome et des troubles gastro-intestinaux de la MP. Environ

50 ~ 80% de patients atteints de la MP souffrent de constipation (Ashraf et al. 1997). Il a été rapporté que la constipation peut précéder les symptômes moteurs et que les personnes atteintes de constipation présentent un risque relativement élevé de développer une MP. Avec d'autres troubles, la constipation peut prédire la survenue d'une MP. Cependant, les patients atteints de la MP ne signalent pas systématiquement le trouble de constipation, et de ce fait il n'est pas pris en charge sur le plan thérapeutique (Yu et al. 2018).

▪ **Troubles dysautonomiques** s'accroissent avec le temps, associant selon les cas : hypotension artérielle orthostatique, troubles vésico-sphinctériens (impériosités mictionnelles).

### 2.3. Les causes de la MP

Les causes de la MP n'ont toujours pas été identifiées, bien que le rôle de la génétique, de facteurs environnementaux ou d'une combinaison des deux puisse être mis en évidence. La MP peut également être causée secondairement suite à une exposition à certaines neurotoxines et des affections moins courantes telles que : un infarctus cérébral multiple et des affections dégénératives.

La plupart des cas de MP résultent probablement d'une interaction complexe de facteurs environnementaux et génétiques. Ces cas sont classés comme sporadiques et surviennent chez des personnes sans antécédents apparents de ce trouble dans leur famille. La cause de ces cas sporadiques reste incertaine.

Le rôle des facteurs génétiques dans les formes génétiques de la MP a été démontré par la découverte de locus<sup>40</sup> chromosomiques associés à la maladie et de gènes responsables des formes mendéliennes. Parmi les formes autosomiques dominantes, les gènes *SNCA* et *LRRK2* sont officiellement associés à la maladie. Le gène *SNCA* joue un rôle central dans la pathogenèse de la maladie et la mutation *G2019S* de *LRRK2* est la plus fréquente dans la population caucasienne. D'autres gènes ont récemment été associés à des formes telles que les gènes *VPS35*, *EIF4G1*, etc.

Parmi les formes autosomiques récessives, les gènes *PRKN*, *PINK1* et *DJ-1* sont identifiés et les mutations de *Parkin* représentent 50% des formes précoces de la maladie.

---

<sup>40</sup> En génétique, un locus (pluriel « locus » ou « loci ») est une position fixe (d'un gène ou d'un marqueur génétique) sur un chromosome.

D'autres études se focalisant sur l'ensemble du génome ont identifié des facteurs de susceptibilité au développement de la MP. Les variantes les plus robustes se trouvent dans les gènes *SNCA*, *MAPT* et *LRRK2*.

Des mutations hétérozygotes du gène de la *glucocérébrosidase* sont également des facteurs de risque de la MP ou peuvent être des mutations causales avec une pénétrance réduite.

L'étude de la pathogénie de ces gènes a permis de proposer des mécanismes associés à la neurodégénérescence tels que le dysfonctionnement mitochondrial, le stress oxydatif, l'altération des voies de dégradation des protéines et des organites cellulaires et la perturbation du transport des vésicules.

Le développement actuel des technologies de séquençage de nouvelle génération permet d'identifier de nouvelles variantes associées à la MP et aidera à mieux comprendre les mécanismes associés à la mort neuronale afin de développer des traitements efficaces pour la progression de la maladie (Mutez, Chartier-Harlin, et Destée 2013).

Certains facteurs environnementaux, comme une exposition importante à des pesticides ou à certains métaux lourds et des traumatismes crâniens répétés, peuvent accroître le risque de développer la MP. La plupart des personnes n'ont pas de cause environnementale claire pour le diagnostic de la MP et, étant donné que de nombreuses années peuvent s'écouler entre l'exposition à un facteur environnemental et l'apparition de symptômes de la MP, le lien est souvent difficile à établir. Cependant, il semble probable que des facteurs environnementaux influencent le développement de la MP, notamment chez les personnes présentant également une susceptibilité génétique.

#### 2.4. Traitement

La physiopathologie de la MP et le développement des modèles animaux parkinsoniens ont été la base du développement des stratégies thérapeutiques. Il n'existe aucun traitement préventif des processus de neurodégénérescence à l'origine de la maladie, mais les médicaments existants peuvent aider à contrôler et soulager essentiellement les troubles moteurs. Pour remédier à cette pathologie, trois types d'approches existent :

- **Une approche pharmacologique** de référence repose sur l'utilisation de **la lévodopa (L-dopa)**, un précurseur de la dopamine. Elle est particulièrement efficace pour atténuer les difficultés de mouvements, les tremblements et la rigidité des membres. L'efficacité de la L-

dopa diminue avec le temps (Hinz, Stein, et Uncini 2011). L'efficacité du traitement à la L-dopa pour la MP varie énormément d'un individu à l'autre, en fonction de la composition de leur microbiote. La L-Dopa est décarboxylée en dopamine active, mais si le microbiote intestinal métabolise la L-dopa avant qu'elle ne traverse la barrière hémato-encéphalique, le traitement est inefficace. Le gène et les enzymes bactériens responsables de cette biodégradation du médicament est identifié (Maini Rekdal et al. 2019).

- **Les agonistes dopaminergiques (AD)** imitent les effets de la dopamine, ce médicament est prescrit dès que le diagnostic est établi, il est associé à la L-dopa lorsque la maladie est à un stade avancé. D'autres traitements médicamenteux existent en particulier pour optimiser l'efficacité de la L-dopa (Lees, Hardy, et Revesz 2009)

- **Approche thérapeutique neurochirurgicale** (Benazzouz et al. 1993) : La stimulation cérébrale profonde consiste en la stimulation électrique à haute fréquence de différents noyaux des ganglions de la base et plus particulièrement du noyau sous-thalamique (NST) dans le contexte de la MP. La stimulation cérébrale profonde se fait grâce à l'implantation bilatérale d'électrodes dans le cerveau qui sont reliées à un stimulateur implantable sous la clavicule qui va permettre l'envoi d'impulsions électriques. Ces stimulations permettent d'améliorer de façon spectaculaire les symptômes moteurs de la maladie et en parallèle réduire les mouvements involontaires associés aux traitements médicamenteux avec des effets secondaires limités (Faggiani et Benazzouz 2017).

## 2.5. Les causes de la MP

Les causes de la MP n'ont toujours pas été identifiées, bien que le rôle de la génétique, de facteurs environnementaux ou d'une combinaison des deux puisse être mis en évidence. La MP peut également être causée secondairement suite à une exposition à certaines neurotoxines et des affections moins courantes telles que : un infarctus cérébral multiple et des affections dégénératives.

La plupart des cas de MP résultent probablement d'une interaction complexe de facteurs environnementaux et génétiques. Ces cas sont classés comme sporadiques et surviennent chez des personnes sans antécédents apparents de ce trouble dans leur famille. La cause de ces cas sporadiques reste incertaine.

Le rôle des facteurs génétiques dans la MP a été démontré par la découverte de locus<sup>41</sup> chromosomiques associés à la maladie et de gènes responsables des formes mendéliennes. Parmi les formes autosomiques dominantes, les gènes *SNCA* et *LRRK2* sont officiellement associés à la maladie. Le gène *SNCA* joue un rôle central dans la pathogenèse de la maladie et la mutation *G2019S* de *LRRK2* est la plus fréquente dans la population caucasienne. D'autres gènes ont récemment été associés à des formes telles que les gènes *VPS35*, *EIF4G1*, etc.

Parmi les formes autosomiques récessives, les gènes *PRKN*, *PINK1* et *DJ-1* sont identifiés et les mutations de *Parkin* représentent 50% des formes précoces de la maladie.

D'autres études se focalisant sur l'ensemble du génome ont identifié des facteurs de susceptibilité au développement de la MP. Les variantes les plus robustes se trouvent dans les gènes *SNCA*, *MAPT* et *LRRK2*.

Des mutations hétérozygotes du gène de la *glucocérebrosidase* sont également des facteurs de risque de la MP ou peuvent être des mutations causales avec une pénétrance réduite.

L'étude de la pathogénie de ces gènes a permis de proposer des mécanismes associés à la neurodégénérescence tels que le dysfonctionnement mitochondrial, le stress oxydatif, l'altération des voies de dégradation des protéines et des organites cellulaires et la perturbation du transport des vésicules.

Le développement actuel des technologies de séquençage de nouvelle génération permet d'identifier de nouvelles variantes associées à la MP et aidera à mieux comprendre les mécanismes associés à la mort neuronale afin de développer des traitements efficaces pour la progression de la maladie (Mutez, Chartier-Harlin, et Destée 2013).

Certains facteurs environnementaux, comme une exposition importante à des pesticides ou à certains métaux lourds et des traumatismes crâniens répétés, peuvent accroître le risque de développer la MP. La plupart des gens n'ont pas de cause environnementale claire pour le diagnostic de la MP et, étant donné que de nombreuses années peuvent s'écouler entre l'exposition à un facteur environnemental et l'apparition de symptômes de la MP, le lien est souvent difficile à établir. Cependant, il semble probable que des facteurs environnementaux

---

<sup>41</sup> En génétique, un locus (pluriel « locus » ou « loci ») est une position fixe (d'un gène ou d'un marqueur génétique) sur un chromosome.

influencent le développement de la maladie de Parkinson, notamment chez les personnes présentant également une susceptibilité génétique.

## 2.6. Traitement

La physiopathologie de la MP et le développement des modèles animaux parkinsoniens ont été la base du développement des stratégies thérapeutiques. Il n'existe aucun traitement préventif des processus de neurodégénérescence à l'origine de la maladie, mais les médicaments existants peuvent aider à contrôler et soulager essentiellement les troubles moteurs. Pour remédier à cette pathologie, trois types d'approches existent :

- **Une des approches pharmacologiques** de référence repose sur l'utilisation de **la lévodopa (L-dopa)**, un précurseur de la dopamine. Elle est particulièrement efficace pour atténuer les difficultés de mouvements, les tremblements et la rigidité des membres. L'efficacité de la L-dopa diminue avec le temps (Hinz, Stein, et Uncini 2011). L'efficacité du traitement à la L-dopa pour la MP varie énormément d'un individu à l'autre, en fonction de la composition de leur microbiote. La L-Dopa est décarboxylée en dopamine active, mais si le microbiote intestinal métabolise la L-dopa avant qu'elle ne traverse la barrière hémato-encéphalique, le traitement est inefficace. Le gène et les enzymes bactériens responsables de cette biodégradation du médicament est identifié (Maini Rekdal et al. 2019).

- **Les agonistes dopaminergiques (AD)** imitent les effets de la dopamine, ce médicament est prescrit dès que le diagnostic est établi, il est associé à la L-dopa lorsque la maladie est à un stade avancé. D'autres traitements médicamenteux existent en particulier pour optimiser l'efficacité de la L-dopa (Lees, Hardy, et Revesz 2009)

- **Approche thérapeutique neurochirurgicale** (Benazzouz et al. 1993) : La stimulation cérébrale profonde consiste en la stimulation électrique à haute fréquence de différents noyaux des ganglions de la base et plus particulièrement du noyau sous-thalamique (NST) dans le contexte de la MP. La stimulation cérébrale profonde se fait grâce à l'implantation bilatérale d'électrodes dans le cerveau qui sont reliées à un stimulateur implantable sous la clavicule qui va permettre l'envoi d'impulsions électriques. Ces stimulations permettent d'améliorer de façon spectaculaire les symptômes moteurs de la maladie et en parallèle réduire les mouvements involontaires associés aux traitements médicamenteux avec des effets secondaires limités (Faggiani et Benazzouz 2017).

## 2.7. Épigenétique de la MP

### 2.5.1. L'épigenétique et les miRNAs

L'épigenétique est l'étude des modifications héréditaires de l'expression des gènes (gènes actifs par rapport aux gènes inactifs) et ce terme définit l'ensemble des mécanismes permettant l'expression du phénotype d'une cellule, sans modifier la séquence fondamentale de l'ADN (Deichmann 2016) (Goldberg, Allis, et Bernstein 2007). Le terme épigenétique a été utilisé lors du début d'une vaste recherche axée sur la combinaison de la génétique et de la biologie du développement menée par des scientifiques réputés tel que *Conrad H. Waddington* (Waddington 2012). Ce terme inventé par Waddington en 1957 (Waddington 1957), est dérivé du mot grec «épigénèse» qui décrivait à l'origine l'influence des processus génétiques sur le développement. Au cours des années 1990, l'assimilation génétique a élucidé la base moléculaire des observations de Conrad Waddington (Waddington 2012) dans lesquelles le stress environnemental provoquait l'assimilation génétique de certaines caractéristiques phénotypiques chez les drosophiles. Depuis lors, les efforts de recherche se sont concentrés sur la découverte des mécanismes épigénétiques liés à ces types de changements. Il existe trois mécanismes de régulation épigénétique : la méthylation de l'ADN, les mécanismes de modifications covalentes et non covalentes de la chromatine et le plus intéressant pour notre étude c'est la régulation de l'expression génétique par des ARNs non codants (Goldberg, Allis, et Bernstein 2007). Les ARNs non codants représentent 98% du transcriptome chez l'Homme (Shruti, Shrey, et Vibha 2011). Ils peuvent être classés en deux groupes (R. Wu et al. 2016) :

- **ARNs de ménage** sont exprimés de manière constitutive et jouent des rôles critiques dans la biosynthèse protéique. Ce groupe contient des ARNs ribosomiques (*ARNr*), des ARNs de transfert (*ARNt*), des ARNs nucléaires (*snARN*), et des petits ARNs nucléaires (*snoARN*).
- **ARNs régulateurs** sont impliqués dans la modulation épigénétique de l'expression des gènes. Ils sont constitués des **petits ARNs non codants** (V. N. Kim 2005), et des longs ARNs non codants (> 200nt) dont leur principal rôle est la régulation de la chromatine (Quinodoz et Guttman 2014) (R. Wu et al. 2016).

Les petits ARNs non codants, dont la taille est comprise entre 18 et 30 nucléotides en moyenne, sont les acteurs clés du **mécanisme de l'inhibition des gènes** (*gene silencing*) (V. N. Kim 2005). C'est une voie de régulation dans laquelle la formation d'un double brin d'ARN interfère avec l'expression des gènes, en inhibant une séquence spécifique. Au niveau moléculaire, elle est médiée par une famille de ribo-nucléo-protéines, nommées RISC (*RNA-induced silencing complexe*) (Hammond 2015) (Pratt et MacRae 2009).

Les petits ARNs non codants sont incorporés dans le complexe RISC, et servent de guide au complexe pour cibler l'ARNm cible via une interaction de type Watson-Crick entre les nucléotides. Cette interaction aboutit à la diminution de l'expression du gène cible. Selon la nature de l'ARN incorporé dans le complexe RISC, ainsi que le type cellulaire dans lequel s'exerce la régulation, le mécanisme de régulation du gène ciblé peut s'effectuer (Pratt et MacRae 2009) au niveau :

- de la synthèse protéique à travers une répression de la traduction,
- du transcrit du gène, via la dégradation de l'ARNm,
- du génome à travers la formation d'hétérochromatine ou par l'élimination d'une partie de l'ADN.

Les petits ARNs non codants sont répartis dans trois types, qui sont présents dans toutes les espèces, y compris les mammifères :

- **les microRNAs** (*miRNAs*) subissent une série d'étapes de biogenèse qui convertissent le transcrit primaire du *miRNA* en un *miRNA* mature actif (environ 22 nucléotides) (Hammond 2015). Ils sont générés à partir de la région double brin d'un ARN précurseur en forme d'épingle (V. N. Kim 2005),
- **les petits ARNs interférants** (*siRNAs* : *small interfering RNAs*) viennent d'un long double brin d'ARN (V. N. Kim 2005),
- **les piwi ARNs** (*piRNAs*) sont principalement exprimés dans les cellules germinales, où leur rôle principal est d'inhiber l'activité des éléments transposables (Ozata et al. 2019) (Mani et Juliano 2013) (Zuo et al. 2016).

Les types de petits ARNs non codants se distinguent par leur biogénèse, leur mécanisme d'action et le type cellulaire dans lequel ils agissent (Tableau 3). Les *miRNAs* et les *siRNAs* ne sont souvent pas distinguables par leur forme mature, ils se différencient majoritairement selon leur biogénèse.



## 2.5.2. Les miRNAs, de régulateurs importants de l'expression de gènes

### a. La découverte des miRNAs

Les miRNAs sont des petits ARNs endogènes simple brin, d'une longueur comprise entre 18 et 21 nucléotides. Ils exercent principalement une régulation post-transcriptionnelle négative de l'expression des gènes en ciblant la partie 3' UTR (*untranslated region* ; région non codante) des ARN messagers (ARNm) (Bartel 2004).

La découverte des miRNAs a été faite pour la première fois en 1993 (Lee, Feinbaum, et Ambros 1993), deux groupes de recherche indépendants (Lee, Feinbaum, et Ambros 1993) (Wightman, Ha, et Ruvkun 1993) ont mis en évidence qu'un petit ARN de 21 nucléotides, nommé *lin-4*, contrôle les phases de développement chez *Caenorhabditis elegans*, par un mécanisme de régulation post-transcriptionnelle du gène *lin-14*. Ces deux groupes de chercheurs émettent l'hypothèse que le petit ARN *lin-4*, qui ne code pas pour les protéines et contient une séquence anti-sens complémentaire d'une séquence répétée située dans la partie 3' non traduite (3'UTR), inhibe l'expression du gène *lin-14* à travers une interaction anti-sens ARN-ARN. En 2000, un autre petit ARN régulateur nommé *let-7* a été mis en évidence, toujours chez *C.elegans* (Reinhart et al. 2000). Le *let-7* (figure 15) régule l'expression de gènes qui contrôlent les étapes de développement et de différenciation cellulaire, et en particulier le *lin-41*. La séquence *let-7* permet de déterminer le pattern développemental, ainsi elle est trouvée chez diverses espèces, incluant les vertébrés, les ascidies, les hémicordés, les mollusques, les annélides et les arthropodes (Pasquinelli et al. 2000).

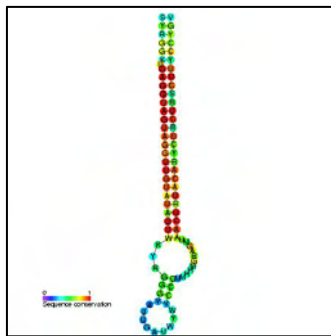


Figure 15:Précurseur miR-let-7

**Tableau 3: Les caractéristiques principales des trois types de petits ARNs non codants exprimés chez les mammifères**

	Type cellulaire	Biogénèse	Mécanisme d'action	Références
<b>miRNAs</b>	Tous	Clivage en deux étapes par les enzymes Drosha et Dicer	Inhibition de la traduction de l'ARNm	(Bartel 2004) (V. N. Kim 2005) (Tolia et Joshua-Tor 2007)
<b>siRNAs</b>	Tous	Clivage par Dicer	Clivage de l'ARNm, modification des histones et/ou de l'ADN	(V. N. Kim 2005) (Tolia et Joshua-Tor 2007)
<b>piRNAs</b>	Les cellules germinales, dans quelques types de cellules somatiques	Via la modification de transcrits, puis le modèle ping-pong pour l'amplification	La suppression de l'activité des éléments transposables et la méthylation de l'ADN.	(Tolia et Joshua-Tor 2007) (Watanabe et al. 2011)

L'année 2001 (Lagos-Quintana et al. 2001) a connu la découverte de nouveaux petits ARNs similaire à lin-4 et let-7 (N. C. Lau et al. 2001), qui ont été identifiés chez les vertèbres et chez les invertébrés, et qui régulent divers processus biologiques (Lee et Ambros 2001). Ces petits ARNs régulateurs furent officiellement identifiés et nommés miRNAs (Pasquinelli et al. 2000).

### **b. Les miRNAs régulateurs de l'expression génétique**

Les miRNAs sont capables de réguler entre 30% et 60% des gènes chez les mammifères (Chekulaeva et Filipowicz 2009) et déterminent le protéome de la cellule (Filipowicz, Bhattacharyya, et Sonenberg 2008) (Lewis, Burge, et Bartel 2005). Les miRNAs sont impliqués dans la régulation de l'ensemble des processus biologiques, ainsi que dans un contexte physiologique et pathologique. Particulièrement, le rôle des miRNAs est bien décrit dans les étapes de développement, de différenciation cellulaire, de prolifération, et d'apoptose (Vidigal et Ventura 2015), ainsi dans le développement des pathologies comme les cancers (Oliveto et al. 2017), les neuropathologies (Jakubowski et Labrie 2016).

Certains miRNAs peuvent servir de cible thérapeutique, ainsi, ils peuvent servir de biomarqueurs liés à l'apparition et la progression de certaines maladies (Lanford et al. 2010).

### **c. L'organisation génomique des miRNAs**

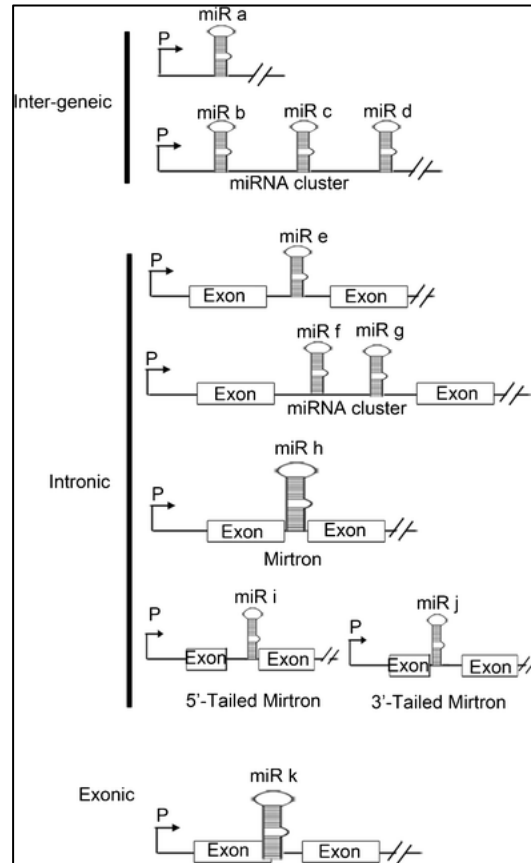
Les gènes des miRNAs ont été classés selon leur position dans le génome en trois grandes catégories (Rodriguez et al. 2004) (Figure 16)

- **Inter-génique :** Ce sont les gènes de miRNAs qui se situent dans des régions génomiques distinctes de celles connues pour être des unités de transcription fonctionnelles, et peuvent être monocistroniques<sup>42</sup> ou polycistroniques. Ils possèdent leurs propres promoteurs et autres éléments régulateurs. Cette catégorie représente 50% des gènes codants pour des miRNAs chez les mammifères (O'Carroll et Schaefer 2013).
- **Intronique :** ce sont des miRNAs localisés dans les introns de gènes référencés, codant ou non pour une protéine (Lagos-Quintana et al. 2001). Cette catégorie représente 40% des gènes (O'Carroll et Schaefer 2013). Il a été montré que la localisation de certains gènes introniques de miRNAs est conservée entre plusieurs espèces, démontrant leur rôle important au cours de l'évolution (N. C. Lau et al. 2001) (Rodriguez et al. 2004).

---

<sup>42</sup> miRNAs monocistroniques : ne possédant qu'un seul cistron et code pour une seule chaîne polypeptidique

- **Exonique** : les gènes de miRNAs se situent en chevauchant un exon non codant pour une protéine et un intron d'un gène. Cette catégorie de gènes est bien plus rare que les deux précédentes, elle représente 10% des miRNA. La biogénèse de cette classe reste peu étudiée (O'Carroll et Schaefer 2013).



**Figure 16: Représentation des différentes localisations des gènes de miRNAs (Ul Hussain 2012)**

Les gènes des miRNAs peuvent être inter-génique et possèdent leur propre promoteur et éléments de régulation, ou alors introniques, ou exoniques (Ul Hussain 2012).

Une étude illustre le mécanisme de la boucle de régulation qui se réalise lors de la co-expression d'un miRNA et de sa cible par la présence du gène du miRNA dans l'intron du gène cible. En effet, le travail décrit le rôle du *miR-26b* dans le contrôle de la différenciation neuronale à travers la régulation de l'expression de son gène hôte CTDSP2 (Dill et al. 2012). Les gènes de miRNAs chez les mammifères représentent 50% et se situent à proximité d'autres gènes codant pour des miRNAs, ces miRNAs sont regroupés en cluster et ils ont souvent

(mais pas toujours), une relation entre eux, c'est-à-dire qu'ils vont réguler les mêmes gènes ou une même famille de gènes (Lagos-Quintana et al. 2001) (N. C. Lau et al. 2001).

### 2.5.3. L'épigénétique et le Parkinson

Le changement épigénétique est un phénomène normal et naturel, mais il peut aussi être influencé par plusieurs facteurs, notamment l'âge, l'environnement, le mode de vie et un état pathologique associé à une maladie.

Les recherches ont montré le rôle de l'épigénétique dans le développement et la progression de nombreuses maladies neurodégénératives, y compris la MP. La modification épigénétique, permet les modifications de l'expression ou de la fonction des gènes sans modification de la séquence de l'ADN, notamment la méthylation de l'ADN, les modifications ultérieures de l'histone et les ARN non codants (Feng, Jankovic, et Wu 2015).

Ces études ont été axées sur la génétique, malgré le fait que les mutations pathogènes décrites n'ont été observées que chez environ 10% des patients parkinsoniens. Cette observation confirme le fait que la MP est un trouble multifactoriel. Les différences dans l'expression des miRNAs la modification de l'histone et les niveaux de méthylation de l'ADN ont été décrits, soulignant l'importance des facteurs épigénétiques dans l'étiologie de la MP. Toute altération du niveau de méthylation dans les gènes liés à la MP pourrait être liée à la pathogenèse de la maladie, probablement en raison de modifications de l'expression génique (Jakubowski et Labrie 2016).

Les biomarqueurs<sup>43</sup> de la MP sont indispensables au diagnostic précoce, ainsi qu'à la prévision de la progression de la maladie et des résultats du traitement. À cet égard, l'épigénome, qui est partiellement dynamique, est très prometteur pour le développement de biomarqueurs moléculaires de cette maladie. Les marques épigénétiques sont modifiées à la fois par la séquence de l'ADN et par les facteurs environnementaux associés à cette maladie, et de telles marques pourraient servir de prédicteur unificateur des individus à risque.

Des anomalies épigénétiques ont été détectées dans la MP et d'autres maladies neurodégénératives dépendantes de l'âge, dont certaines auraient été signalées à un stade précoce et pourraient être réversibles avec des médicaments contre cette maladie.

---

<sup>43</sup> Les biomarqueurs sont des molécules qui représentent une signature pathologique facilement identifiable et quantifiable pour une maladie ou un trouble particulier.

De nouveaux rapports indiquent que certaines différences épigénétiques observées dans le cerveau des parkinsoniens sont détectables dans des tissus plus facilement accessibles (Jakubowski et Labrie 2016).

Ces dernières années, les biomarqueurs se sont révélés essentiels pour identifier les individus à risque pour la maladie, suivre la progression de la maladie et sélectionner une approche thérapeutique adaptée à chaque cas (Y. Wu, Le, et Jankovic 2011).

La perte d'environ 70% de neurones dopaminergiques chez un patient atteint de la MP déclenche l'apparition des symptômes moteurs tels que la rigidité, le tremblement au repos, la bradykinésie et l'instabilité posturale (Graeber 2009) (Jakubowski et Labrie 2016), l'objectif de recherche des biomarqueurs de cette maladie est de fournir aux professionnels de la santé et aux chercheurs des outils permettant un diagnostic précis avant que la mort neuronale ne survienne (Graeber 2009). En plus de surveiller les individus à risque pour un diagnostic précoce, les biomarqueurs peuvent accélérer la découverte et le développement de nouvelles approches thérapeutiques.

# Objectifs du travail sur la MP

## ▪ Thématique

La MP est une maladie neurodégénérative caractérisée par la dégénérescence progressive des neurones dopaminergiques de la *SNpc*. Les traitements pharmacologiques et chirurgicaux actuels permettent d'améliorer et contrôler les symptômes moteurs ainsi que certains symptômes non-moteurs associés à la maladie, mais ils sont sans effet sur la dégénérescence neuronale et sa progression dans le temps.

## ▪ Objectifs

L'objectif du travail de thèse était de rechercher des biomarqueurs de la maladie afin d'identifier les individus à risque, ce qui permettrait de développer des stratégies thérapeutiques neuroprotectrices. L'identification des biomarqueurs pour détecter les stades précoces de la MP permettra de démarrer les traitements neuroprotecteurs avant ou juste au début de la perte neuronale. L'intérêt est majeur puisque les symptômes cliniques de la MP n'apparaissent que quand 60 à 70 des neurones dopaminergiques sont morts. Les biomarqueurs pourraient également permettre d'identifier les personnes présentant des risques de développer la maladie, de mesurer sa progression et de choisir les traitements les plus adaptés.

### • Objectifs généraux

- Générer et utiliser des modèles animaux de la MP. Deux modèles de rats ont été utilisés, le rat 6-hydroxydopamine (6-OHDA) et le rat Alpha-synucléine ( $\alpha$ -syn) obtenus par une injection stéréotaxique dans le cerveau de 6-OHDA ou  $\alpha$ -syn respectivement.
- Caractériser le degré de sévérité des déficits et de la dégénérescence des neurones dopaminergiques, en utilisant une approche biochimique de dosage des monoamines qui sont la dopamine, la noradrénaline et la sérotonine.
- Rechercher des biomarqueurs de prédisposition à la maladie dans les deux modèles animaux ce qui aiderait à déterminer leurs implications dans le pronostic du patient. Nous avons utilisé une approche biomoléculaire ciblant les microRNAs par séquençage d'ARN (RNA-Seq).
- Générer un pipeline spécifique pour l'analyse et l'identification des miRNAs.

- **Objectifs personnels**

- Maitriser l'approche in vivo des études
- Maitriser l'approche moléculaire de recherche des biomarqueurs
- Appliquer une approche génomique pour proposer les voies régulant le développement de la pathologie



# CHAPITRE 2 : Workflow d'analyse génomique pour la Maladie de Tuberculose

L'objectif de ce chapitre est de décrire les étapes d'analyse génomique pour la maladie tuberculose avec les outils utilisés en citant les résultats obtenus pour chaque étape du workflow.

La tuberculose est une maladie infectieuse causée par la bactérie *Mycobacterium tuberculosis*, transmissible par des gouttelettes en suspension dans l'air. Le risque d'infection dépend de la durée d'exposition, de l'intensité de l'exposition et du statut immunitaire de la personne exposée.

Le système immunitaire élimine les bactéries immédiatement chez plus de 80% des personnes exposées. Dans une faible proportion des personnes infectées, les bactéries sont isolées et restent dormantes mais viables. C'est ce qu'on appelle la tuberculose latente. On estime que 5 à 10% des personnes atteintes de tuberculose latente développeront une tuberculose active au cours de leur vie.

La forme la plus contagieuse de tuberculose est la tuberculose pulmonaire, en particulier les cas à frottis positif où des bacilles tuberculeux peuvent être vus lors d'un examen microscopique direct des expectorations. En l'absence de traitement, on estime que chaque personne atteinte de tuberculose active infectera en moyenne entre 10 et 15 personnes chaque année.

La tuberculose est présente dans toutes les régions du monde. En 2017, le plus grand nombre de nouveaux cas de tuberculose est survenu dans les régions de l'Asie du Sud-Est et du Pacifique occidental (62%), suivi de la région africaine (25%).

En 2017, 87% des nouveaux cas de tuberculose sont survenus dans les 30 pays fortement touchés par la tuberculose. Huit pays représentaient les deux tiers des nouveaux cas de tuberculose : Inde, Chine, Indonésie, Philippines, Pakistan, Nigéria, Bangladesh et Afrique du Sud.

La tuberculose multirésistante (TB-MDR) est une forme de tuberculose causée par une bactérie qui ne répond pas à l'isoniazide et à la rifampicine, les 2 médicaments anti-tuberculeux de première ligne les plus puissants. La tuberculose multirésistante est traitable et curable en utilisant des médicaments de deuxième intention. Cependant, les options de traitement de deuxième intention sont limitées et nécessitent une chimiothérapie extensive (jusqu'à 2 ans de traitement) avec des médicaments coûteux et toxiques. La tuberculose ultrarésistante (TB-XDR) est une forme plus grave de tuberculose multirésistante causée par une bactérie qui ne répond pas aux médicaments anti-TB de deuxième intention les plus efficaces, laissant souvent les patients sans option de traitement supplémentaire.

En 2017, la TB-MDR demeure une crise de santé publique et une menace pour la sécurité sanitaire. L'OMS estime à 558 000 le nombre de nouveaux cas de résistance à la rifampicine, le médicament de première intention le plus efficace, dont 82% étaient atteints de tuberculose multi-résistante. Le fardeau de la TB-MDR repose en grande partie sur trois pays - l'Inde, la Chine et la Fédération de Russie - qui représentent ensemble près de la moitié des cas dans le monde. Environ 8,5% des cas de tuberculose multi-résistante avaient une tuberculose ultrarésistante (TB-XDR) en 2017.

À l'échelle mondiale, seuls 55% des patients atteints de TB-MR sont actuellement traités avec succès. En 2016, l'OMS a approuvé l'utilisation d'un schéma thérapeutique court et standardisé pour les patients atteints de TB-MDR qui ne présentent pas de souches résistantes aux médicaments antituberculeux de deuxième intention. Ce traitement prend 9 à 12 mois et coûte beaucoup moins cher que le traitement conventionnel de la TB-MDR, qui peut durer jusqu'à 2 ans. Les patients atteints de TB-XDR ou présentant une résistance aux antituberculeux de deuxième intention ne peuvent pas utiliser ce traitement mais doivent être soumis à un schéma thérapeutique plus long pour la TB-MDR, auquel un des nouveaux médicaments (bedaquiline et delamanide) peut être ajouté.

L'OMS a également approuvé en 2016 un test de diagnostic rapide pour identifier rapidement ces patients. 62 pays ont commencé à utiliser des schémas thérapeutiques plus courts pour la TB-MDR. À la fin de 2017, 62 pays ont signalé avoir introduit la *bedaquiline*<sup>44</sup> et 42 pays, le *delamanide*<sup>45</sup>, dans le but d'améliorer l'efficacité des schémas thérapeutiques contre la TB-MDR.

Au début de mes travaux de thèse, aucun génome complet de *Mycobacterium Tuberculosis* isolé de patients marocains n'a été soumis dans les banques de données biologiques, ni séquencé dans un laboratoire marocain.

Dans ce contexte, quatre souches multi-résistantes de *Mycobacterium tuberculosis* ont été isolées de quatre patients (hommes) marocains au sein de l'hôpital Militaire d'Instruction

---

<sup>44</sup> bedaquiline (noms de code : TMC2074 et R2079105) est une molécule de la famille des diarylquinoléines constituant le principe actif d'un nouveau médicament antituberculeux nommé Sirturo

<sup>45</sup> delamanid est un médicament anti-tuberculeux indiqué dans les formes résistantes aux antibiotiques traditionnels, Il s'agit d'un dérivé nitro-dihydro-imidazooxazole inhibant la synthèse de l'acide mycolique, un constituant de la paroi du mycobacterium

Mohamed V de Rabat au Maroc et ont été séquencé dans le laboratoire de Biotechnologie Médicale de la faculté de Médecine et Pharmacie de Rabat – Maroc en 2016. Plusieurs objectifs ont été fixés, la priorité c'est d'avoir un génome de référence marocain afin d'identifier les caractéristiques de cette bactérie en la comparant avec différents génomes internationaux ainsi qu'un workflow d'analyse génomique adapté aux souches séquencées. Ce chapitre est présenté sous forme de trois articles (deux publiés et un soumis),

## II. Partie expérimentale (Biologique)

### 1. Culture

La culture des quatre souches isolées (MTB1\_M, MTB2\_M, MTB13\_M, MTB20\_M) est faite dans le milieu de Lowenstein-Jensen Medium, ce dernier sert à l'isolement des mycobactéries.

### 2. Extraction et dosage d'ADN

Après l'isolement et la culture des quatre souches de *Mycobacterium Tuberculosis* (MTB1\_M, MTB2\_M, MTB13\_M, MTB20\_M) depuis l'expectoration de quatre patients à l'hôpital militaire d'Instruction Mohammed V au Maroc. L'ADN génomique total a été extraite à l'aide du kit *GenoLyse (Hain lifescience)*. C'est une technique qui isole de l'ADN à partir d'une cellule en qualité suffisante (Rasmussen, Morrissey, et Hebert 2009).

L'extraction de l'ADN a été faite au sein du laboratoire de bactériologie de l'hôpital militaire d'Instruction Mohammed V.

L'ADN extraite a été dosé au sein de la salle d'extraction d'ADN du laboratoire de Biotechnologie en utilisant le spectrophotomètre "*NanoVue™ Plus*" afin d'obtenir une concentration supérieure à **50ng/μL**

### 3. Préparation de librairie

Pour la préparation de librairie, le kit "*Nextera XT DNA Library Prep V600*" (Illumina 2018b) a été choisi parce qu'il est dédié au séquençage des petits génomes (*Small Genome Sequencing*) tels que les bactéries, les virus et les archées ainsi il permet de séquencer les amplicons et les plasmides. Le diagramme suivant illustre le workflow de préparation de librairie d'ADN en utilisant le kit *Nextera XT* pour huit échantillons. Les points d'arrêt sécurisés sont marqués entre les étapes (Figure 17).



**Figure 17: Workflow de préparation de librairie avec le kit Nextera XT**

Ce protocole a été utilisé pour la préparation de 24 échantillons parmi eux les souches de *Mycobacterium tuberculosis* traitées.

#### 4. Séquençage des 4 souches MTB

##### 4.1. Séquençage par Illumina-MiSeq

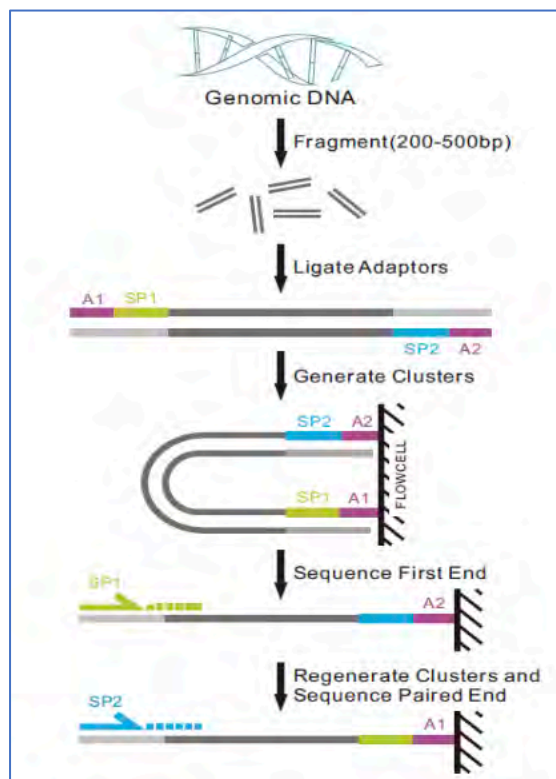
Les quatre MTB ont été séquencé avec le système de séquençage MiSeq qui permet le séquençage en simultané de plusieurs millions de fragments d'ADN. La technologie utilisée repose sur une amplification en ponts des fragments d'ADN sur support solide (*flowcell*) suivi d'un séquençage par synthèse chimique (SBS) des colonies d'ADN préalablement amplifiés. Il existe 2 supports de 1 ligne générant 15 millions de lecture en *singne read* (30 millions de lectures en *paired-end*) en version V2 ou 25 millions de reads (50 millions de

lectures en *paired-end*) en version V3. La taille de séquences générés peut aller de 50 SR, 50 PE, 100 SR, 100 PE, 150 SR, 150 PE, 250 SR ou 250 PE (Illumina 2018a).

#### 4.2. Applications et méthodes utilisées pour le séquençage de MTB

Une fois la préparation de librairie est terminée, elle est suivie de la création de la *sample sheet* (annexe1) qui contient plusieurs paramètres tels que le nom du kit utilisé lors de la préparation de librairie, le nom d'application et de workflow, et le type de librairie.

Pour nos données MTB, au niveau de la *sample sheet* la méthode de séquençage choisi est *Paired-end* (Figure 18) parce qu'elle permet de séquencer les deux extrémités d'un fragment et de générer des données de bonne qualité, et produit deux fois plus de reads qu'avec la méthode *Single-end*, l'utilisation de cette méthode de séquençage facilite les étapes d'analyse génomique et surtout la détection des variants (Nakazato, Ohta, et Bono 2013).



**Figure 18: Protocole de séquençage paired-end**

*Small genome sequencing*, c'est l'application utilisée pour le séquençage, elle est dédiée au petit génome ( $\leq 5$  Mb) tel que les bactéries, les virus, ...

Pour lancer le séquençage, le logiciel MiSeq Control Software est utilisé. Le séquençage a duré presque 44h, la première étape effectuée après la fin du séquençage, c'est la vérification de la qualité du run, dont on a obtenu 83.4 % des résultats dont leurs qualités est supérieur à Q30.

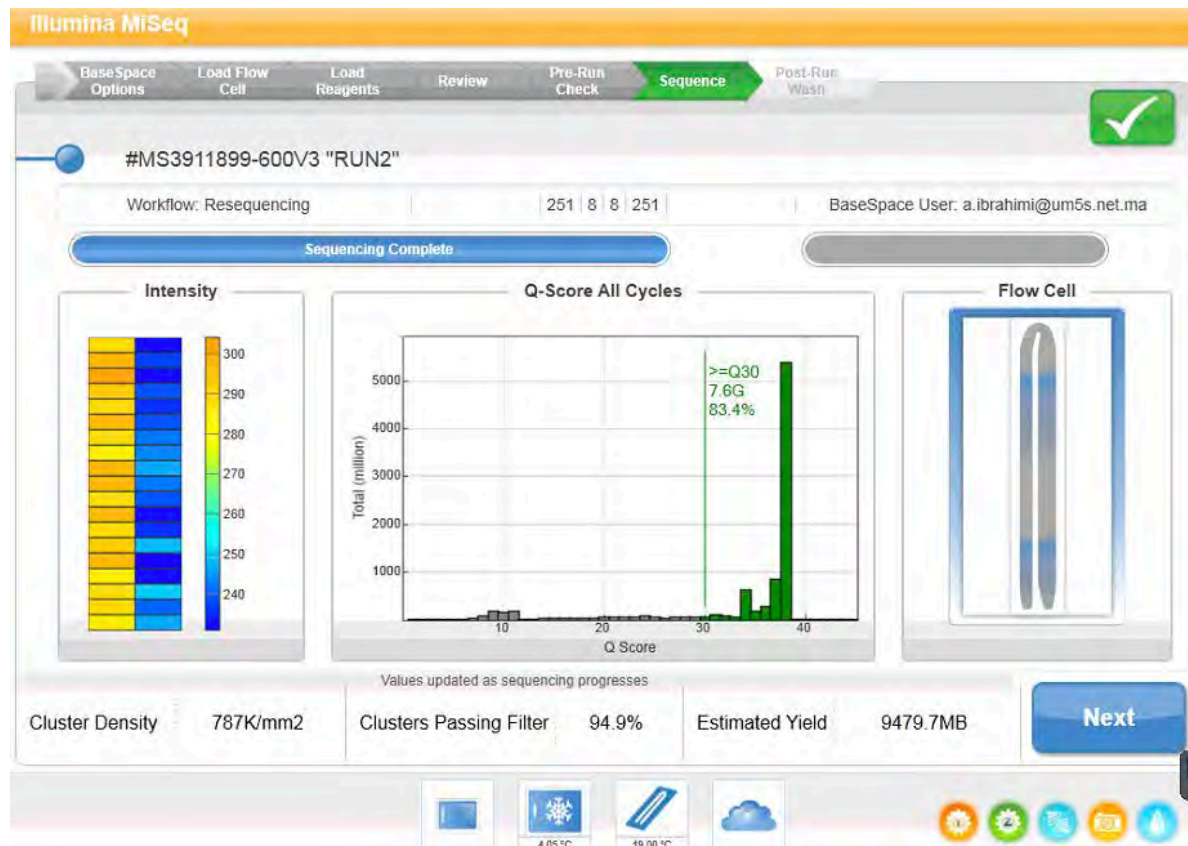


Figure 19: Résultat du séquençage sur le séquenceur Illumina MiSeq

### III. Analyse génomique des 4 MTB

L'analyse génomique permet l'identification, la mesure ou la comparaison de caractéristiques génomiques telles que la séquence d'ADN, la variation structurelle, l'expression génique ou l'annotation d'éléments régulateurs et fonctionnels à l'échelle génomique.

Les méthodes d'analyse génomique nécessitent généralement un séquençage à haut débit ou une hybridation de puces à ADN et la bio-informatique. Cette approche est devenue un outil essentiel pour le contrôle de la résistance aux antibiotiques spécifiquement celle liée à des bactéries. De nombreuses études ont mis en évidence l'intérêt du WGS (*Whole Genome Sequencing*) et l'analyse génomique en tant qu'outils de contrôle quotidien des infections et



pour certains agents pathogènes, en tant qu'outils de diagnostic principal pour détecter la résistance aux antibiotiques (Köser, Ellington, et Peacock 2014).

L'analyse génomique est l'enchaînement d'outils bio-informatiques et statistiques dédiés à des étapes prédéfinies différentes permettant de créer un workflow qui résume les étapes d'analyse génomique effectués pour 4 MTB.

## 1. Le prétraitement

Le séquençage des 4 échantillons de MTB a produit des fichiers de format FastQ (Cock et al. 2010), ce dernier est un format de fichier texte permettant de stocker les séquences nucléiques avec leurs qualités (Annexe 2). Après l'obtention de ces fichiers, nous commençons l'étape du prétraitement, cette étape se compose de deux étapes : commençant par le contrôle qualité suivi de correction et filtrage des données brutes.

### 1.1. Contrôle qualité

Le contrôle qualité est la première étape primordiale à effectuer après avoir l'obtention des données brutes du séquenceur, cette étape permet d'évaluer la qualité de ces données avant de commencer les analyses bio-informatiques, et de vérifier la conformité des données en déterminant :

- Le nombre, la qualité et la longueur des reads
- Les contaminations
- La présence d'adaptateurs
- ...

Pour effectuer le contrôle qualité de ces données, plusieurs outils bio-informatiques existent, tels que MultiQC (Ewels et al. 2016), Prinseq (Schmieder et Edwards 2011), FastQC (S. Andrews 2010), etc ...

Les données issues du séquenceur MiseQ sont fournies au format FastQ (Cock et al. 2010), donc le choix de FastQC (S. Andrews 2010) est adapté, il permet une visualisation graphique des différents métriques d'intérêt avec un code couleur spécifique.

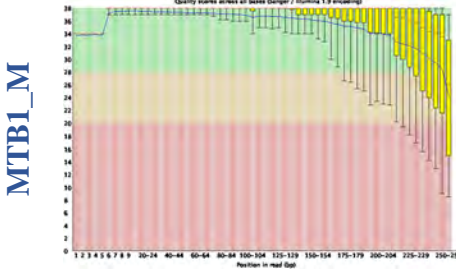
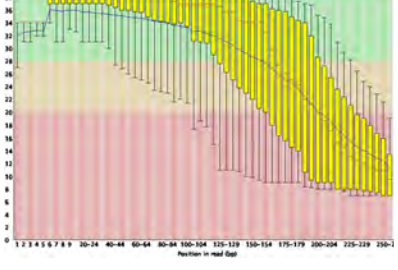
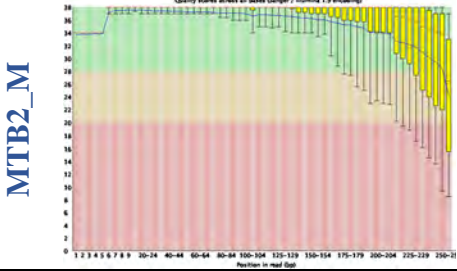
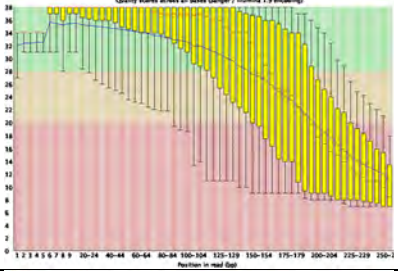
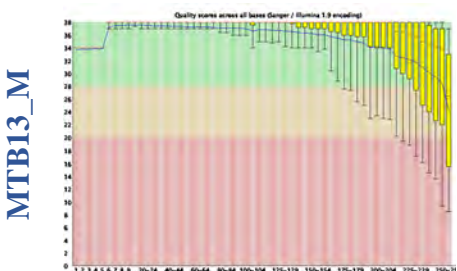
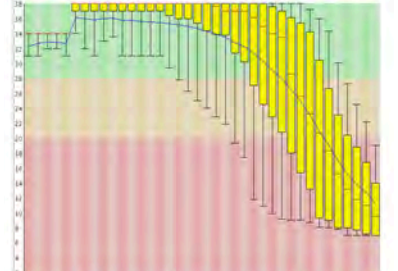
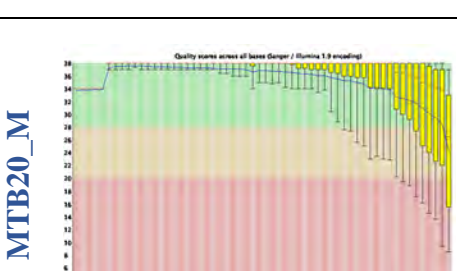
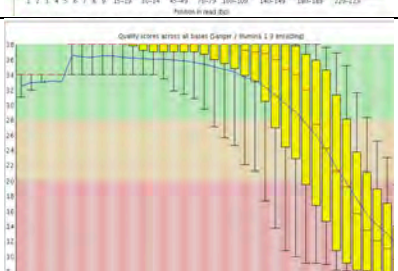
Voici les informations obtenues pour la souche MTB1 en utilisant FastQC :

- ***Per Base Sequence Quality*** : Qualité de la séquence de base, c'est le premier indicateur de la qualité des données issues du séquençage.

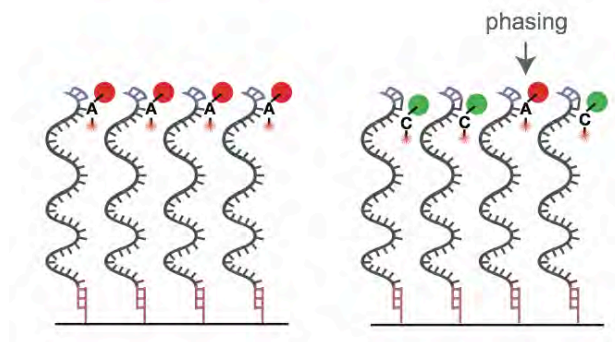
Le tableau 4 regroupe des graphiques représentant la qualité (score Phred, en ordonnée) de chaque base (en abscisse) pour tous les reads. À chaque position du read, la qualité de tous les reads est représentée sous la forme d'un boxplot. La médiane est en rouge. Le code couleur vous indique les scores de très bonne qualité en vert, bonne qualité en orange et mauvaise en rouge. Généralement, la qualité baisse en fin de reads.

La qualité du *forward* est très bonne, la médiane est supérieure à 25 et le code couleur est en vert, par contre le *reverse* est de mauvaise qualité (code couleur en rouge), cette diminution de la qualité s'explique par la méthode de séquençage utilisée par Illumina, qui repose sur la procédure de synthèse. Au cours de chaque cycle du séquençage, le séquenceur lave les produits chimiques qui incluent des variants pour les quatre nucléotides sur la flowCell (qui possède des grappes "*Cluster*" différentes avec des fragments d'ADN identiques pour chaque grappe), les nucléotides ont un bloqueur (capuchon de terminaison) de sorte qu'une seule base est ajoutée à chaque molécule d'ADN à la fois. Après la détection du signal de fluorescence couplé, le bloqueur peut être supprimé et le cycle peut recommencer. De cette façon, les fragments d'ADN de chaque groupe sont séquencés de manière synchrone en exprimant des signaux de fluorescence spécifiques. Au cours du processus de séquençage, différentes erreurs peuvent survenir.

**Tableau 4: Per Base Sequence Quality des 4 MTB avec FastQC**

	<b>Forward</b>	<b>Interprétation</b>	<b>Reverse</b>	<b>Interprétation</b>
<b>MTB1_M</b>		<b>Bonne qualité</b> : la médiane est supérieure à 25 et le quartile supérieur de 10		<b>Mauvaise qualité</b> : la médiane est inférieure à 20 à partir de la position 200
<b>MTB2_M</b>		<b>Bonne qualité</b> : la médiane est supérieure à 25 et le quartile supérieur de 10		<b>Mauvaise qualité</b> : la médiane est inférieure à 20 à partir de la position 185.
<b>MTB13_M</b>		<b>Bonne qualité</b> : la médiane est supérieure à 25 et le quartile supérieur de 10		<b>Mauvaise qualité</b> : la médiane est inférieure à 20 à partir de la position 180
<b>MTB20_M</b>		<b>Bonne qualité</b> : la médiane est supérieure à 25 et le quartile supérieur de 10		<b>Mauvaise qualité</b> : la médiane est inférieure à 20 à partir de la position 179

Donc, la diminution de la qualité de la séquence est principalement due à l'étape du séquençage appelée phasing qui signifie que le bloqueur d'un nucléotide n'est pas correctement éliminé après la détection du signal. Au cycle suivant, aucun nouveau nucléotide ne peut se lier sur ce fragment d'ADN et l'ancien nucléotide est détecté une nouvelle fois, le signal de fluorescence de cet ancien nucléotide étant (probablement) différent du signal synchrone des autres nucléotides (Figure 19).



**Figure 20: Processus de l'étape Phasing**

Alors ce fragment d'ADN aura un cycle de retard (déphasé), polluant le signal lumineux que la caméra du séquenceur doit détecter. Un effet similaire se produit si un nucléotide a un capuchon de terminaison de défaut (pré-phasing). Ces erreurs se produisent avec une faible probabilité. Mais avec le temps (avec une longueur de read croissante), ils s'additionnent et polluent de plus en plus le signal lumineux. Le signal devient de plus en plus asynchrone. Et puisque le signal lumineux est utilisé pour calculer les scores de qualité, le signal asynchrone résulte en un score de qualité de séquence décroissant.

Pour toutes les souches (MTB1\_M, MTB2\_M, MTB13\_M et MTB20\_M), la qualité du fichier *reverse* est décroissante et cela est due à un processus indésirable mais inévitable.

A partir de ce module résultant de FastQC, il est donc nécessaire de réaliser un filtrage des reads en éliminant les bases les plus mauvaises de façon à avoir des données de séquençage contenant le moins d'erreurs et d'incertitudes possibles.

- ***Per Sequence Quality Scores*** : Scores de qualité par séquence

Ce module représente le nombre de reads pour chaque score de qualité. Il est très informatif s'il y a eu un problème lors du séquençage.

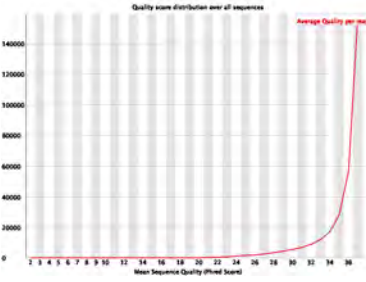
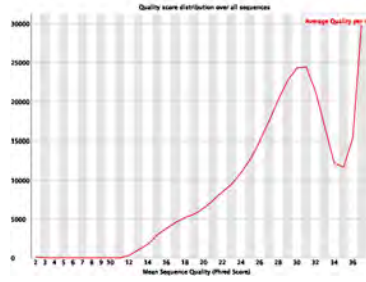
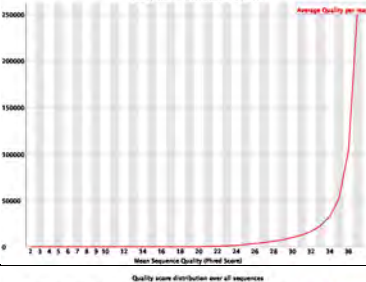
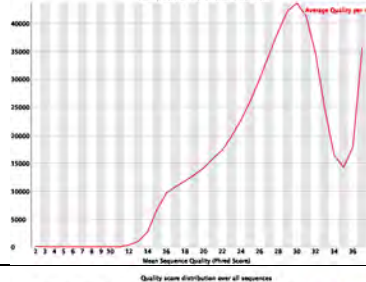
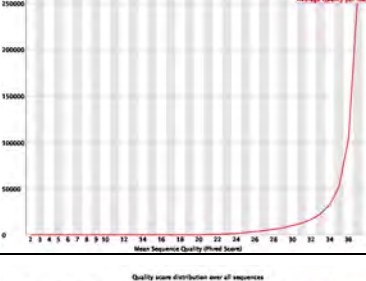
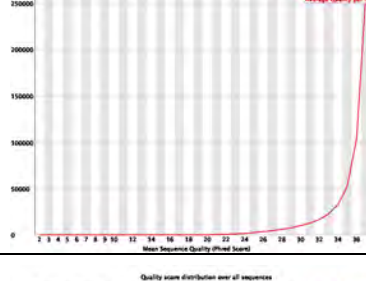
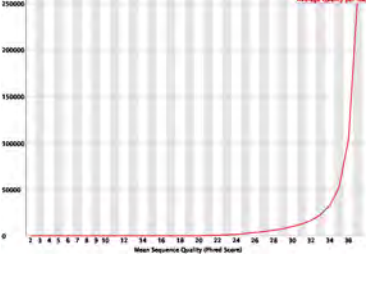
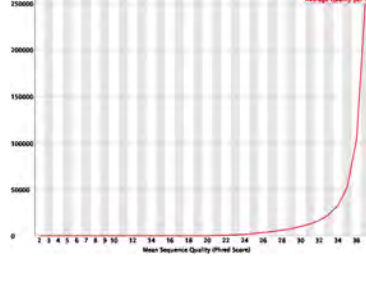
Le tableau 5 représente les graphiques du module ***Per Sequence Quality Scores*** de toutes les souches séquencées, ces graphiques représentent le nombre de séquences en fonction de leur qualité moyenne. La distribution du nombre de séquences par scores permet de voir si une s'il y a une proportion des séquences ayant une faible qualité. Toutes les données sont de bonnes qualités.

**NB :**

**Warning** : si le pic de la qualité moyenne est inférieur à Q27

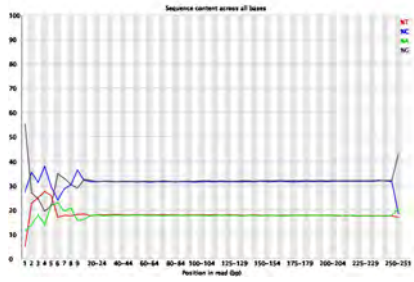
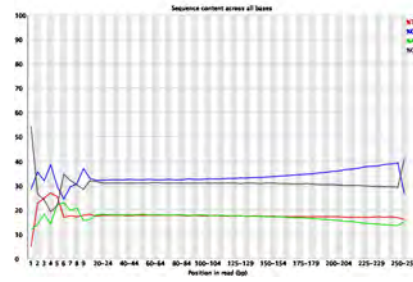
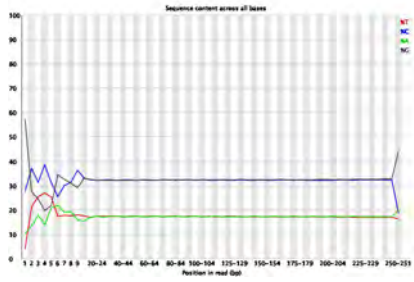
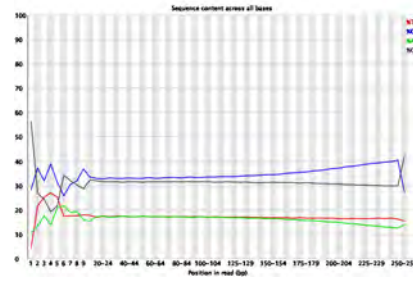
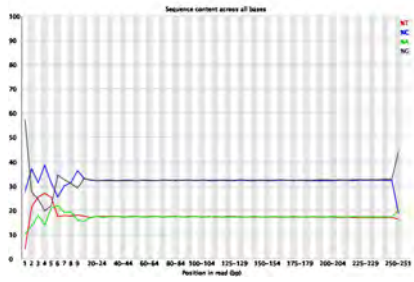
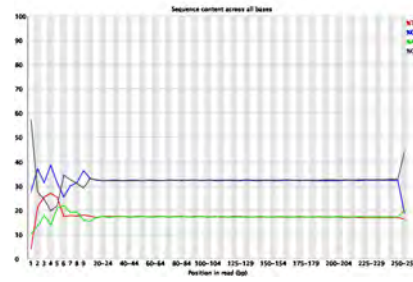
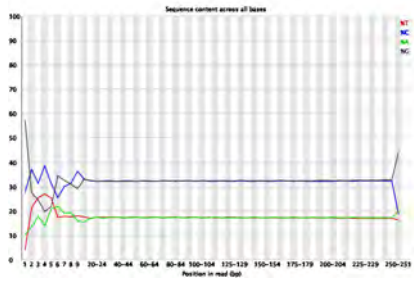
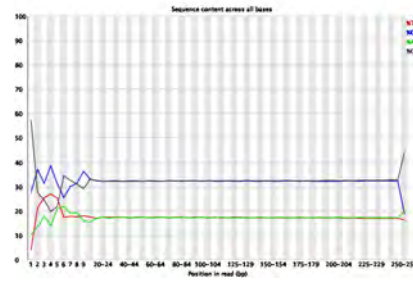
**Erreur** : si le pic de la qualité moyenne est inférieur à Q20.

Tableau 5: Per Sequence Quality Scores des 4 MTB

	Forward	Interprétation	Reverse	Interprétation
MTB1_M		<b>Bonne qualité</b> : Tous les reads ont une qualité supérieure à 28		<b>Bonne qualité</b> : la majorité des reads ont une qualité supérieure à 28
MTB2_M		<b>Bonne qualité</b> : le pic de la moyenne de qualité de tous les reads est supérieur à 28		<b>Mauvaise qualité</b> : la médiane est inférieure à 20 à partir de la position 185.
MTB13_M		<b>Bonne qualité</b> : le pic de la moyenne de qualité est supérieur à 28		<b>Mauvaise qualité</b> : la médiane est inférieure à 20
MTB20_M		<b>Bonne qualité</b> : le pic de la moyenne de qualité est supérieur à 28		<b>Mauvaise qualité</b> : la médiane est inférieure à 20

- **Per Base Sequence Content** : contenu par séquence de base : Ce graphique représente le % de chaque base (A, C, G, T) à chaque position du *read*.

**Tableau 6:Per Base Sequence Content des 4 MTB**

	Forward	Interprétation	Reverse	Interprétation
<b>MTB1_M</b>		Les pourcentages entre les bases dépassent les 20% entre la position 1 et 9 et se stabilisent après		Les pourcentages entre les bases dépassent les 20% entre la position 1 et 9 et se stabilisent jusqu'à la position 175
<b>MTB2_M</b>		Les pourcentages entre les bases dépassent les 20% entre la position 1 et 9 et se stabilisent après		Les pourcentages entre les bases dépassent les 20% entre la position 1 et 9 et se stabilisent jusqu'à la position 154
<b>MTB13_M</b>		Les pourcentages entre les bases dépassent les 20% entre la position 1 et 9 et se stabilisent après		Les pourcentages entre les bases dépassent les 20% entre la position 1 et 9 et se stabilisent jusqu'à la position 154
<b>MTB20_M</b>		Les pourcentages entre les bases dépassent les 20% entre la position 1 et 9 et se stabilisent après		Les pourcentages entre les bases dépassent les 20% entre la position 1 et 9 et se stabilisent jusqu'à la position 154

## MultiQC pour toutes les données

D'après les résultats fournis par MultiQC, il est nécessaire de réaliser un filtrage des *reads* en éliminant les bases les plus mauvaises de façon à avoir des données de séquençage contenant le moins d'erreurs et d'incertitudes possibles.

### 1.2. Filtration - Trimming

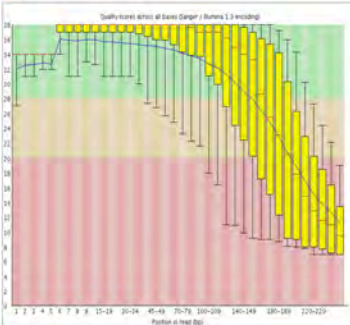
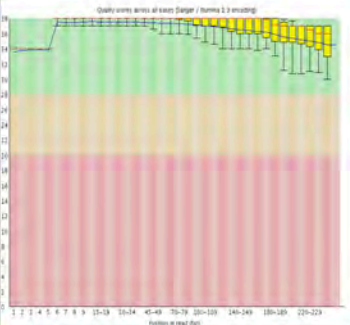
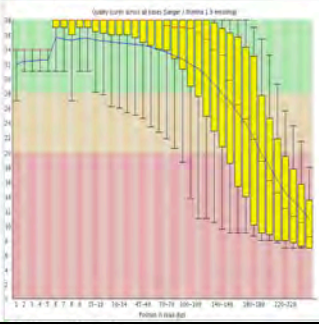
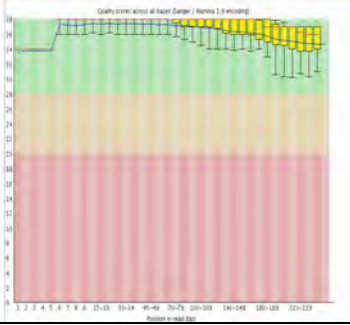
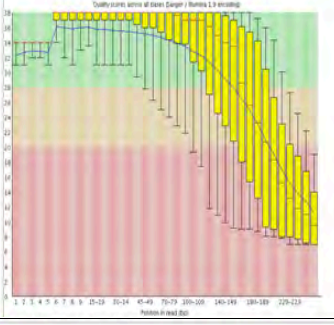
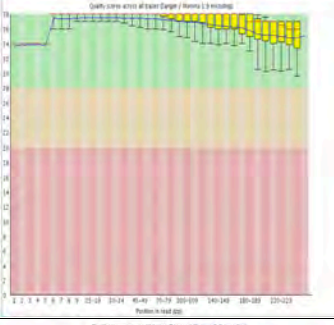
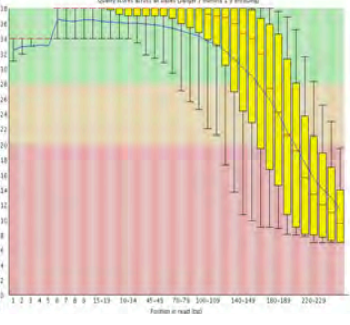
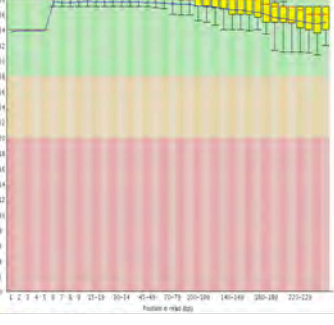
Les aléas de la chimie des plates-formes de séquençage, les erreurs de manipulation, les aberrations causées par la machine, etc., sont tous des facteurs à l'origine d'une diminution de la qualité des *reads* (Lo et Chain 2014). Il faut noter que les *reads* produits par la majorité des technologies de séquençage modernes, présentent une détérioration du score de qualité des bases vers la fin de l'extrémité 3' et vers la fin de l'extrémité 5'. Pour remédier aux soucis de qualité qui diminuent vers la fin du *read*, le programme de trimming *Sickle* (Joshi N A et Fass J N 2011) a été choisi pour couper l'extrémité 3' et l'extrémité 5' en fixant le seuil de la qualité. Ce programme éliminera également des *reads* tout en se basant sur le seuil de longueur.

Ainsi au sein d'un même *read*, toutes les bases n'ont pas la même valeur de qualité, c'est-à-dire que la probabilité que la base soit correcte n'est pas la même. Ainsi, plus la valeur du score de qualité d'une base donnée est élevée, plus la probabilité d'avoir la bonne base est forte, d'où l'utilisation du logiciel *Trimmomatic* (Bolger, Lohse, et Usadel 2014) est nécessaire dans l'étape de prétraitement, et il est plus souple et optimisé pour les données issues du séquençage d'Illumina. Ce logiciel prend en charge des données de format *\*.Fastq* pour ajuster la qualité.

Pour vérifier la qualité après le filtrage, l'utilisation du logiciel FastQC est nécessaire, ci-dessous le résultat du Reverse avant filtrage et après



Tableau 7: Per Base Sequence Quality des 4 fichiers Reverse avec FastQC après filtrage

	Reverse avant filtrage	Interprétation	Reverse après filtrage	Interprétation
MTB1_M		<b>Mauvaise qualité :</b> la médiane est inférieure à 20 à partir de la position 200		<b>Bonne qualité :</b> la médiane est supérieure à 25 et le quartile supérieur de 10
MTB2_M		<b>Mauvaise qualité :</b> la médiane est inférieure à 20 à partir de la position 185.		<b>Bonne qualité :</b> la médiane est supérieure à 25 et le quartile supérieur de 10
MTB13_M		<b>Mauvaise qualité :</b> la médiane est inférieure à 20 à partir de la position 180		<b>Bonne qualité :</b> la médiane est supérieure à 25 et le quartile supérieur de 10
MTB20_M		<b>Mauvaise qualité :</b> la médiane est inférieure à 20 à partir de la position 179		<b>Bonne qualité :</b> la médiane est supérieure à 25 et le quartile supérieur de 10

## 2. Assemblage

L'assemblage génomique consiste à aligner et/ou fusionner des fragments d'ADN ou ARN issues du séquençage de nouvelle génération afin de reconstruire la séquence originale ou s'approcher de la famille d'origine. C'est une étape qui succède au séquençage après vérification de qualité. Il existe deux types d'assemblages, l'assemblage avec référence génomique et l'assemblage de *novο* (Pop 2009a).

### 2.1. Assemblage par référence – Mapping

L'assemblage avec référence permet d'aligner les résultats issus du séquenceur par rapport à un génome de référence existant dans les banques de données biologiques internationales.

L'assemblage par référence (ou mapping) est un processus qui permet de comparer les séquences obtenues par le séquenceur avec une séquence génomique connue (référence) dans l'objectif d'identifier des zones communes, de déterminer des sites fonctionnels (site catalytique, zone d'interaction, ...) et de prédire la ou les fonctions d'une protéine (si détecter). Ce processus permet un certain nombre de prédictions.

Les outils d'alignement disponibles se basent sur deux algorithmes différents : *hash table method* (H. Li et Homer 2010) et *Burow Wheler Method* (H. Li et Durbin 2009). Deux outils ont été choisis pour effectuer le mapping : *Mauve* et *Bowtie 2*

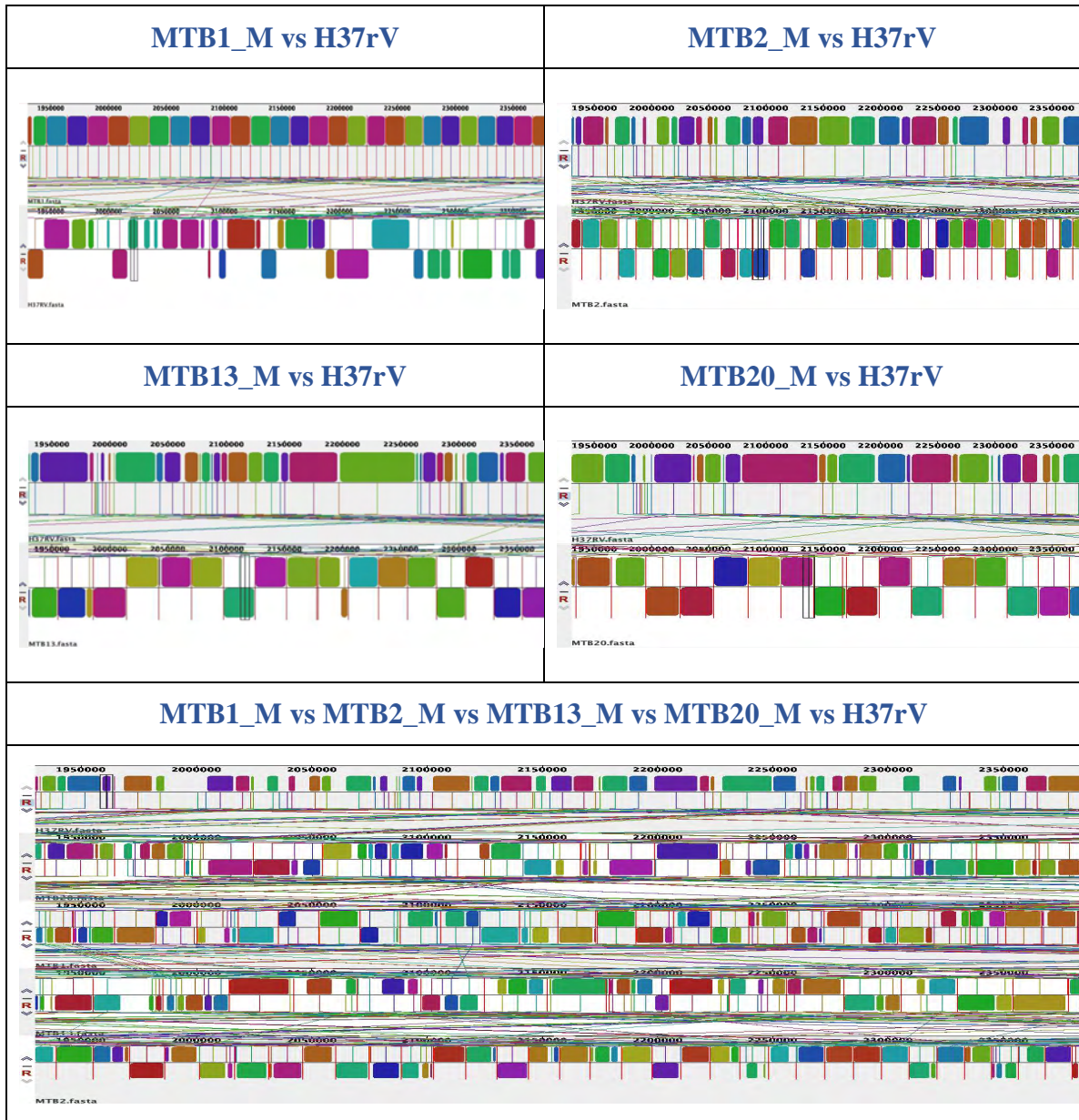
#### ▪ Alignement par Mauve

Dans notre cas, le choix du logiciel d'alignement Mauve qui se base sur la méthode BW, il permet de comparer plusieurs génomes en même temps et d'identifier les régions génomiques conservées (consensus), les réarrangements et inversion dans ces mêmes régions. En outre, il permet l'identification des substitutions, des indels et des insertions de nucléotides à l'aide de l'alignement multiple traditionnel des régions consensus (Darling et al. 2004).

Pour étudier la similarité et la colinéarité des quatre souches MTB (MTB1\_M, MTB2\_M, MTB13\_M, MTB20\_M) avec la souche de référence *Mycobacterium tuberculosis H37rV* dont le génome complet est publié sur NCBI sous le numéro **NC\_000962.3**, ce génome est téléchargé depuis la banque de données GenBank sous le format "*Fasta*".

Mauve identifie et aligne les régions de colinéarité locale, appelées blocs colinéaires locaux (LCB). Chaque bloc colinéaire local est une région homologue de séquence partagée par deux ou plusieurs des génomes étudiés.

Tableau 8: Résultat d'alignement avec Mauve avec MTB H37rV



L'alignement de MTB1\_M et le génome de référence H37rV avec Mauve (Tableau 8), a déterminé 470 LCBs. Le programme a aussi relevé 251 SNPs, concernant l'alignement avec la souche MTB2\_M, Mauve a détecté un nombre de LCBs égale à 386, 53 SNPs. Pour MTB13\_M avec H37rV, 346 LCBs et 7187 SNPs ont été détectés, et pour MTB20\_M, 294 LCBs et 2529 SNPs ont été déterminés.

En comparant toutes les souches MTB avec H37rV, Mauve détecte 11332 SNPs.

- **Alignement par Bowtie 2**

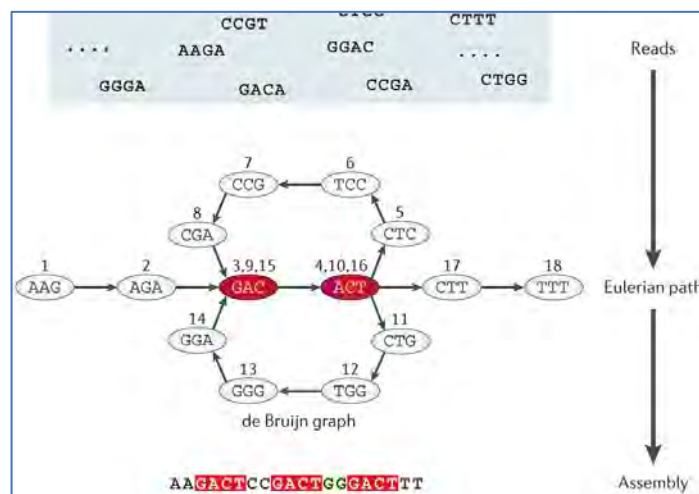
Bowtie 2 combine les atouts de l'indexation rapide avec la flexibilité et la rapidité des algorithmes de programmation dynamique à accélération matérielle afin de combiner la vitesse élevée, sensibilité et précision (Langmead et Salzberg 2012).

## 2.2. Assemblage de NOVO

Même si un génome de référence est disponible, un assemblage de novo (Miller, Koren, et Sutton 2010) doit être effectué, car il permet de récupérer les transcriptions transcrites à partir de séquences du génome manquantes dans l'assemblage du génome. Ce type d'assemblage du génome consiste à analyser les *reads* de façon à reconstruire de longues séquences via les chevauchements possibles entre *reads* et de combiner les séquences obtenues du séquenceur en séquences contiguës, appelées *contigs*.

Les séquenceurs de nouvelle génération génèrent deux types et pour chaque type de *read* un algorithme d'assemblage est dédié : Les algorithmes de type *Overlap-Layout-Consensus* (OLC) sont adaptés à l'assemblage des longs *reads* (Staden 1979), Ces algorithmes commencent par calculer tous les chevauchements possibles entre toutes les paires de *reads* fournies (Pop 2009b). Cette méthode est l'une des premières utilisées avec succès pour assembler une séquence génomique (Miller, Koren, et Sutton 2010). Le principe général est de fusionner des séquences chevauchantes, en commençant par celles ayant les chevauchements les plus significatifs, jusqu'à la formation d'une séquence unique. Par conséquent, la plupart des assembleurs disponibles capables de gérer des données typiques générées par Illumina utilisent une approche basée sur un *graphe de De Bruijn* et basée sur k-mer. Le principe de cet algorithme vise à simplifier la recherche d'un chemin

*Hamiltonien*<sup>46</sup> dans le graphe en le transformant en recherche d'un chemin *Eulérien*<sup>47</sup> (Compeau, Pevzner, et Tesler 2011). Cette méthode commence dans un premier temps, à fragmenter les séquences qui sont à assembler en sous séquences de même taille ( $k$ ), appelées *k-mer*. Deux *k-mer* vont pouvoir être assemblés si leurs séquences ne divergent que par le premier nucléotide de l'un et le dernier nucléotide de l'autre. De cette façon il est possible d'établir un lien entre chacun des *k-mer* (Compeau, Pevzner, et Tesler 2011) (Pevzner, Tang, et Waterman 2001). Ensuite, les séquences chevauchantes sont découpées en 4-mer. Les nœuds du graphe représentent les séquences communes de longueur  $k-1=3$ . Deux sous-séquences sont reliées s'il existe une concordance de séquence entre les 3 premiers nucléotides de l'un et les 3 derniers de l'autre. Il est alors possible de fournir la séquence consensus. Un des paramètres essentiels à une bonne utilisation des DBG, la taille de *k-mer*, n'est pas aussi facile à manier. La taille de *k-mer* choisie change grandement l'allure du graphe produit. Il existe une taille optimale de *k-mer* que l'on décrira comme «  $k$  » qui est un nombre impair (Zerbino et Birney 2008) forcément plus petit que la taille de des reads, le choix de la valeur de *k-mer* pourrait devenir nuisible à l'assemblage.



**Figure 21: Représentation du Graph de De Bruijn**

Le séquençage des quatre souches MTB (MTB1\_M, MTB2\_M, MTB13\_M, MTB20\_M) avec l'application "*Small genome sequencing*" de la technologie Illumina Miseq permet

<sup>46</sup> Un **chemin hamiltonien** d'un graphe orienté ou non orienté est un chemin qui passe par tous les sommets une fois et une seule.

<sup>47</sup> En mathématiques, un cycle **eulérien** se caractérise par le fait qu'il est possible de suivre le graphe à partir de n'importe quel sommet en passant une seule fois par chaque arête pour revenir au sommet de départ.

d'avoir des courtes *reads* et la méthode de DBG est mieux adaptée aux *reads* produites par cette technologie de séquençage. Parmi les outils utilisant cet algorithme le logiciel *SPAdes* (Bankevich et al. 2012) qui est un assembleur créé pour les petits génomes et surtout pour les génomes de bactéries. *SPAdes* construit un DBG à partir de *k-mers* de différentes tailles. Ainsi, une plus petite valeur de *k-mer* dans les régions à faible couverture minimise la fragmentation, alors qu'une plus grande valeur de *k-mer* dans les régions fortement couvertes réduit le nombre de *reads* chimériques. Des contigs sont alors déduits à partir du graphe. Il est recommandé pour *SPAdes* d'augmenter la longueur de *k-mer* de 21 jusqu'à ce que la longueur *k-mer* atteigne 127.

### 2.3. Le contrôle qualité de l'assemblage

L'évaluation de la qualité du résultat de l'assemblage sous format "*\*.Fasta*" obtenue par le logiciel *SPAdes*, a été effectuée par le logiciel *Quast* (Gurevich et al. 2013), qui permet de calculer plusieurs paramètres (nombre de contigs/scaffolds, taille du plus grand contig, N50, taux de GC<sup>48</sup> ...) principaux pour définir une bonne qualité d'assemblage de novo.

Nous essayons au laboratoire de produire des séquences de très haute qualité. L'assemblage de génomes complets destinés à être comparés nécessite une phase de vérification. Les critères de qualité d'une séquence génomique complète sont :

- **Nombre de contigs** : Le nombre total de contigs obtenus après l'assemblage
- **Taille du génome assemblé** : le nombre de bases assemblées
- **N50** : N50 d'un contig de taille N, signifie que 50% des bases assemblées sont contenues en contigs de longueur N ou plus (Schatz, Delcher, et Salzberg 2010).

---

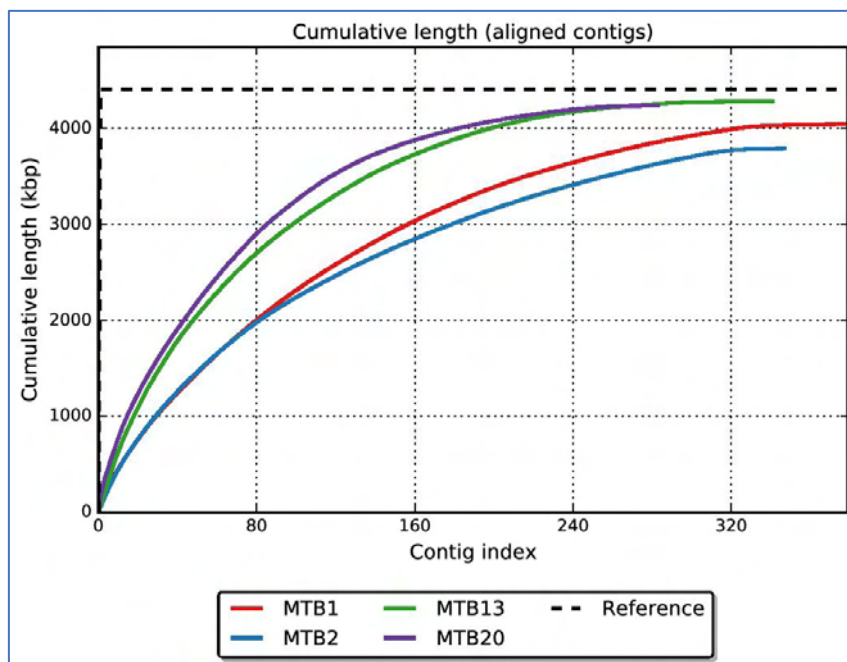
<sup>48</sup> Le taux de GC (ou % GC) d'une séquence d'ADN est défini comme la proportion de bases de cette séquence étant soit une cytosine (C), soit une guanine (G).

L'assemblage de Novo pour les souches MTB a été fait par l'assembleur *SPAdes 3.7.1*. Le résultat obtenu par *Quast* en utilisant la souche de référence *Mycobacterium Tuberculosis H37Rv* est détaillé dans le Tableau ci-dessous.

**Tableau 9: Résultat d'assemblage des 4 souches MTB**

	MTB1_M	MTB2_M	MTB13_M	MTB20_M
Fraction du génome <sup>49</sup>	91.37	85.67	96.70	95.87
Nombre de contigs	285	289	292	298
Taille du génome	4 064 236	4 689 658	4 347 307	4 358 959
%GC	65,22	65,34	65,37	65,30
N50	20 340	19 221	26 137	28 431

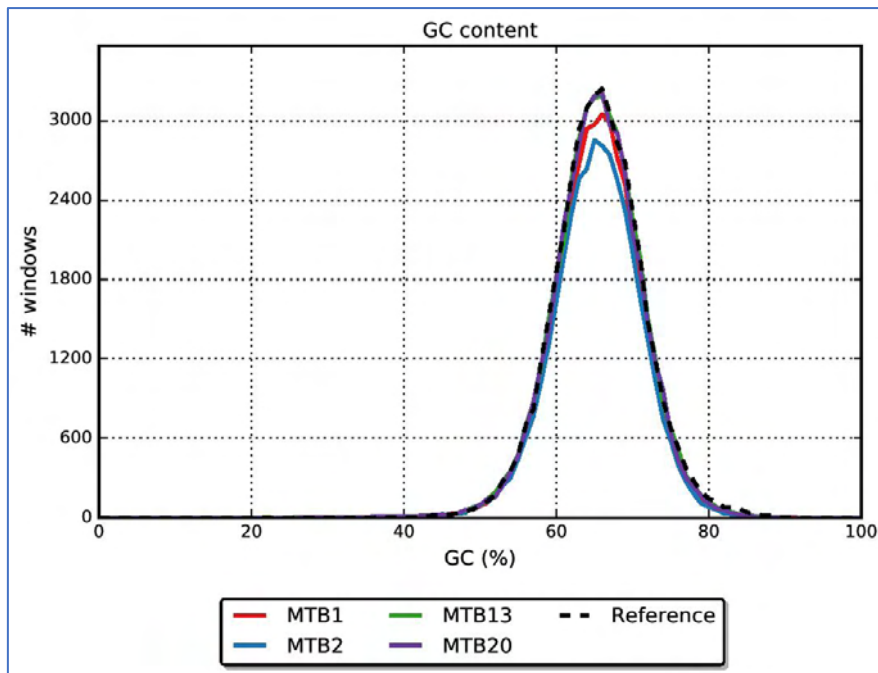
Ainsi *Quast* permet de comparer toutes les souches avec le génome de référence en représentation graphique de longueur cumulative qui montre la croissance des longueurs de contig. Sur l'axe des x, les contigs sont classés du plus grand au plus petit. L'axe des y donne la taille des x plus grands contigs de l'assemblage.



**Figure 22: Graphique de longueur cumulative (Quast)**

<sup>49</sup> La fraction du génome (%) est le pourcentage de bases alignées dans le génome de référence. Une base dans le génome de référence est alignée s'il existe au moins un contig avec au moins un alignement sur cette base.

Le graphique de contenu GC montre la distribution du contenu GC dans les contigs des 4 MTB en les comparant avec le %GC du génome de référence.



**Figure 23: Graphique du contenu GC des 4 MTB et référence (Quast)**

La valeur x est le pourcentage GC (0 à 100%).

La valeur y est le nombre de fenêtres de 100 pb ne se chevauchant pas, dont le contenu GC est égal à x%.

Pour un seul génome, la distribution est typiquement gaussienne. Cependant, pour les assemblages contenant des contaminants, la distribution GC semble être une superposition des distributions gaussiennes, donnant un tracé avec des pics multiples.

### 3. Annotation

L'annotation du génome est un processus clé pour identifier les régions codantes et non codantes d'un génome, la localisation des gènes et leurs fonctions. L'analyse de la séquence d'ADN avec des logiciels d'annotation de génome permet de rechercher et de cartographier des gènes, des exons-introns, des éléments régulateurs, des répétitions et des mutations. Les bases de données génomiques sont essentielles pour récupérer des informations sur les fonctions de nom de gènes, de produit protéique et de séquence d'ADN.



L'annotation génomique consiste à analyser la séquence nucléotidique qui constitue l'information brute pour en extraire l'information biologique nécessaire pour la compréhension du fonctionnement de la cellule et les relations entre les gènes.

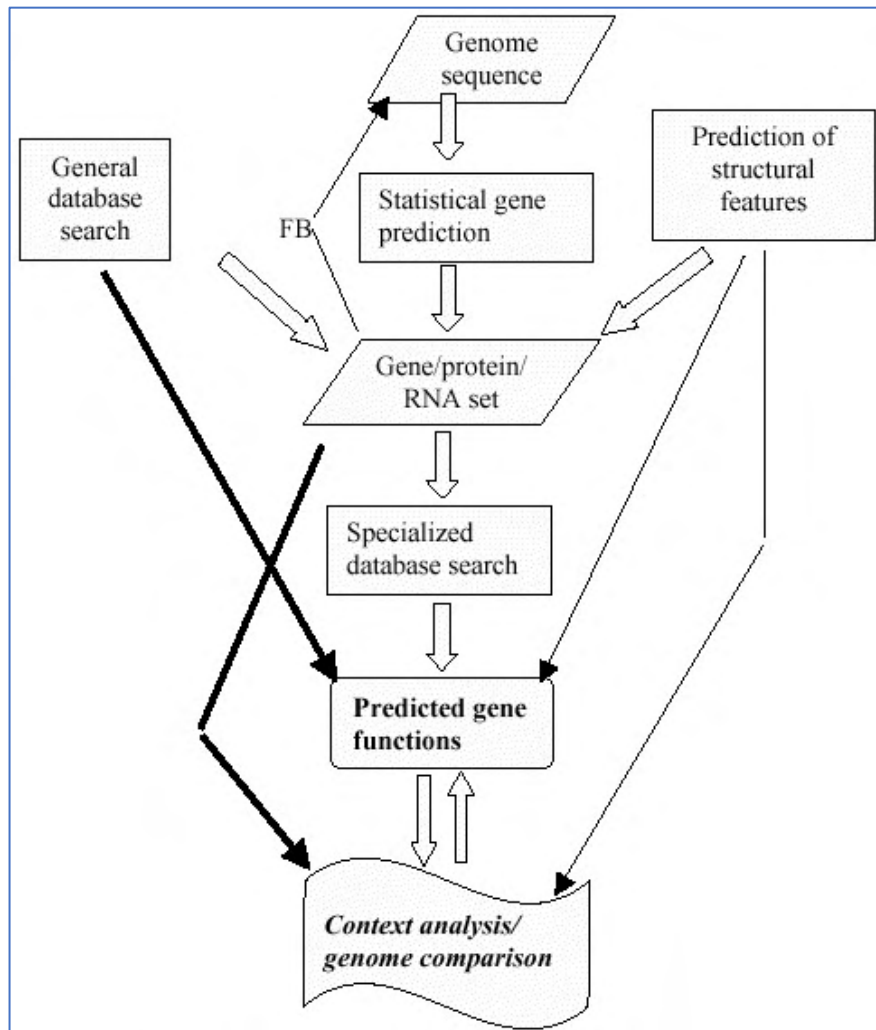
L'annotation génomique *in silico*<sup>50</sup> utilise uniquement des algorithmes recherchant des similarités (de séquence, de structure, de motifs...), permettant de prédire la fonction d'un gène. Elle aboutit au transfert « automatique » de l'information figurant dans l'étiquette d'un gène « similaire » d'un génome déjà annoté au génome en cours d'annotation (X. Zhou et Rokas 2014).

La figure 23 montre un schéma approximatif du workflow d'annotation du génome, en commençant par la séquence du génome, suivi de la prédiction statistique des gènes et la recherche de bases de données polyvalentes à des fins de similarité de séquences. La contribution des autres approches, en particulier la recherche de bases de données spécialisées, comprenant des bases de données de domaines, telles que Pfam, SMART et CDD ainsi que des bases de données orientées sur le génome, telles que COG ou KEGG et l'analyse du contexte génomique varient considérablement d'un projet à l'autre. Jusqu'à présent, ces méthodes et ressources relativement nouvelles restent la base de la recherche de base de données traditionnelle dans l'annotation du génome (Koonin et Galperin 2003).

On distingue trois types d'annotation : syntaxique ou structurale, l'annotation fonctionnelle et l'annotation relationnelle (X. Zhou et Rokas 2014).

---

<sup>50</sup> *in silico* est une locution adverbiale d'inspiration latine qui désigne une recherche ayant été effectuée par des modèles informatiques.



**Figure 24: Workflow d'annotation génomique**

FB: feedback from gene identification for correction of sequencing errors, primarily frameshifts

### 3.1. L'annotation structurale ou syntaxique

L'annotation structurale ou l'annotation syntaxique consiste à identifier les éléments génétiques présents dans le génome. Ces éléments peuvent inclure les séquences codant des protéines et des molécules d'ARN (ARNt, ARNr, ARNsn, ...), les promoteurs, les exons, les introns, les CDS, les transposons, ainsi certaines séquences répétées et séquences régulatrices de l'expression des gènes. La détection de ces régions d'intérêt se fait à l'aide de la comparaison de séquences ou de motifs et de prédictions *ab initio* aussi appelée De Novo ou intrinsèques.

### 3.2. L'annotation fonctionnelle

L'annotation fonctionnelle permet de déterminer la fonction des éléments identifiés lors de l'annotation structurale particulièrement la fonction des gènes protéiques. Cette annotation se base sur la recherche d'homologie entre la séquence inconnue considérée et un ensemble de séquences annotées, pour identifier le maximum d'informations utiles à la fonction du gène et de son produit prédit. Cependant la prédiction devra être validée expérimentalement pour en garantir la qualité d'annotation (X. Zhou et Rokas 2014)

### 3.3. L'annotation relationnelle

L'annotation relationnelle, ou contextuelle détermine les relations susceptibles d'exister entre les éléments prédits (annotation structurale) et caractérisés (annotation fonctionnelle), ces relations sont de différentes natures (Venter et al. 2001) telles que l'homologie (les protéines peuvent être regroupées en familles d'homologues), l'interaction physique (les éléments interagissent physiquement entre eux : protéine/acides nucléiques, protéine/protéine et acides nucléiques/acides nucléiques) et l'implication commune dans un processus biologique (participation à la même voie métabolique, même voie de transport ou même réseau de régulation).

### 3.4. Annotations des 4 MTB

Les 4 MTB ont été annotés par plusieurs outils :

- **Annotation avec subsystem: RAST** (*Rapid Annotation using Subsystem Technology*) est un serveur web dédié pour l'annotation des génomes procaryotes surtout pour les bactéries et les *archaea* (Aziz et al. 2008), cet outil se base sur l'annotation en utilisant le subsystem 2.0. RAST permet d'identifier les séquences codants des protéines (CDS), l'ARN ribosomal et l'ARN de transfert en utilisant FIGfams (Meyer, Overbeek, et Rodriguez 2009) et Glimmer (Arthur L. Delcher et al. 2007) (A. L. Delcher et al. 1999) qui permet de prédire les positions des gènes dans l'ADN microbien, en particulier les génomes des bactéries et les virus.
- **Annotation avec Prokka** : L'outil *Prokka* (Rapid Prokaryotic Genome annotation) est utilisé pour finaliser les résultats obtenus par RAST, c'est un logiciel d'annotation rapide de génomes procaryotes (Seemann 2014). *Prokka* combine plusieurs outils tels que *Prodigal* (Prokaryotic Dynamic Programming Gene-finding Algorithm) (Hyatt et al. 2010) pour identifier un nombre maximal de gènes, *RNAmmer* (consistent and

rapid annotation of ribosomal RNA genes) (Lagesen et al. 2007) pour prédire les ARNr, *ARAGORNE* (Laslett et Canback 2004) pour identifier les ARNt et *Infernal 1.1* (100-fold faster RNA homology searches) (Nawrocki et Eddy 2013) pour déterminer les ARN non codants en se basant sur la base de données *Rfam* (RNA families database) (Nawrocki et al. 2015). L'annotation avec *Prokka* utilise les bases de données *Pfam* (protein families database) (Finn et al. 2008) (El-Gebali et al. 2019), *TIGRFAMs* (Haft et al. 2013) et *RefSeq* de *NCBI* (O'Leary et al. 2016) (Haft et al. 2018).

- **Annotation avec TB Profiler** c'est outil en ligne rapide qui a été développé pour rapporter les profils des *Mycobacterium tuberculosis* multi-résistantes aux médicaments ainsi le type de souche à partir des séquences brutes. En utilisant une bibliothèque prédictive de 1525 mutations pour 15 médicaments anti-tuberculeux, la précision du diagnostic *in silico* était supérieure à celle de certains diagnostics commerciaux et bases de données alternatives. *TB Profiler* se base sur cette bibliothèque qui a permis de faciliter les tests de sensibilité aux médicaments basés sur les séquences génomiques (Coll et al. 2015).
- **Annotation avec MAGE** : c'est une partie de la plateforme *MicroScope* (Vallenet et al. 2009) basée sur le Web pour une annotation efficace et une analyse comparative du génome microbien (<http://www.genoscope.cns.fr/agc/microscope>). MAGE permet une annotation d'expert en tant que complément de l'annotation automatique, la plateforme permet de visualiser et parcourir les informations d'annotation sur les génomes microbiens disponibles dans le but d'améliorer la qualité de l'annotation du génome microbien, en particulier pour les génomes initialement analysés et annotés par des procédures automatiques.

#### 4. Analyse des gènes de résistance

Les tableaux ci-dessous regroupent les résultats d'annotation en se basant sur tous les outils ainsi que les gènes impliqués dans les résistances liées aux médicaments.

Tableau 10: Les résultats d'annotation des 4 MTB

Souches	Gene	Longueur	Position	Mutation	Rôle du gène	Mécanisme de résistance	Fréquence de mutation
<b>Isoniazide</b>							
MTB1_M	katG	2223	2 734 631 – 2 736 853	Ser95Thr	Catalase-peroxidase	Inhibition de la synthèse d'acide mycolique	40e96
	inhA	810	233 759 – 234 568	Arg463Leu	Enoyl ACP reductase		
MTB2_M	katG	2223	1 227 680 – 1 229 902	Glu21Gln	Catalase-peroxidase		
	fabG1	744	4 360 519 – 4 361 262	Indel: C/CTGGG	3-Oxoacyl (acyl-carrier protein)		
MTB13_M	katG	2223	1 229 902 – 1 229 902	Gly668Asp	Catalase-peroxidase		
	ahpC	588	73 818 – 74 405		Alkyl hydroperoxide reductase		
MTB20_M	inhA	744	3 677 405 – 3 678 214	Ser315Asn	Enoyl ACP reductase		
	kasA	1251	2 518 115 – 2 519 365		Ketoacyl acyl carrier protein synthase		
<b>Rifampicine</b>							
MTB1_M	rpoB	3537	23 556 – 27 092	Ala526Ser	b-subunit of RNA polymerase	Inhibition de la Synthèse de l'ARN	40e95
MTB2_M	rpoB	3537	1 531 501 – 1 535 037	Ser450Leu	b--subunit of RNA polymeras		
	rpoC	3951	1 535 082 – 1 539 032	Ala542Ala			

Souch	Gene	Longueur	Position	Mutation	Rôle du gène	Mécanisme de résistance	Fréquence de mutation
MTB13_M	rpoB	3537	3537 - 6 991	Asp103Asp	b--subunit of RNA polymerase	Inhibition de la Synthèse de l'ARN	40e95
	rpoC	3027	384 - 3 410	Gly594Glu			
MTB20_M	rpoB	3519	1 625 433 – 1 628 951	Asp103Asp	b--subunit of RNA polymerase		
	rpoC	3951	1 621 438 - 1 625 388	Gly594Glu	b--subunit of RNA polymerase		
<b>Pyrazinamide</b>							
MTB1_M	pncA	561	412 939 – 413 499	Ala23Gly	PZase	PZase est responsable de l'activation. RpsA implique en traduction PanD est responsable de synthèse du pantothenate et du CoA.	60e96
MTB2_M	pncA	561	2 647 552 – 2 648 112	Arg83Pro			
MTB13_M	pncA	561	1 268 - 1 828	Gln62Pro			
<b>Streptomycine</b>							
MTB1_M	rpsL	375	45 327 – 45 701		S12 ribosomal protein	Inhibition de la synthèse des protéines	30e60

Souches	Gene	Longueur	Position	Mutation	Rôle du gène	Mécanisme de résistance	Fréquence de mutation
MTB13_M	rpsL	375	14 933 – 15 307		S12 ribosomal protein	Inhibition de la synthèse des protéines	30e60
	gid	675	4 407 528 – 4 408 202		7-methylguanosine methyltransferase		
<b>Ethambutole</b>							
MTB1_M	embC	3288	1 213 145 - 1 216 432	Leu502Arg	Arabinosyl transferase	Inhibition de la synthèse de l'arabinogalactane	40e80
MTB2_M	embC	3285	3 321 319 - 3 324 603	Arg927Arg	Arabinosyl transferase		
	embR	1167	3 585 098 - 3 586 264	Gly386Gly	Regulator of embCAB operon		
MTB13_M	embC	3285	4 088 – 7 372	Val981Leu	Arabinosyltransferase		
MTB20_M	embB	3297	2 100 187 - 2 103 483	Met306Val	Arabinosyl transferase		
	embA	3285	2 096 906 - 2 100 190		Arabinosyl transferase		
	embC	3285	2 093 536 - 2 096 820	Val981Leu	Arabinosyl transferase		
	embR	1167	1 928 722 - 1 929 888		Regulator of embCAB operon		

Souches	Gene	Longueur	Position	Mutation	Rôle du gène	Mécanisme de résistance	Fréquence de mutation
<b>Fluoroquinolones</b>							
MTB1_M	gyrA	2517	428 836 - 431 352	Gly668Asp	DNA gyrase subunit A	Inhibition de la synthèse ADN	42e95
	gyrB	2145	431 387 - 433 531	Asp538Asn	DNA gyrase subunit B		
MTB2_M	gyrA	2517	1 147 467 - 1 149 983	Glu21Gln	DNA gyrase subunit A		
	gyrB	2145	1 145 288 - 1 147 432	Thr539Asn	DNA gyrase subunit B		
MTB13_M	gyrA	2517	22 826 – 25 342	Gly668Asp	DNA gyrase subunit A		
	gyrB	2145	20 647 – 22 791		DNA gyrase subunit B		
MTB20_M	gyrA	1787	2 409 416 - 2 411 202	Phe640Val	DNA gyrase subunit A		
	gyrB	2145	2 407 237 - 2 409 381	Asp500His	DNA gyrase subunit B		
<b>Capreomycine / Viomycine</b>							
MTB1_M	tlyA	807	2 905 940 - 2 906 746	Leu11Leu	rRNA methyltransferas	Inhibition de la synthèse protéine	40e70
MTB2_M	tlyA	807	2 920 973 - 2 921 779	Leu11Leu			



Souches	Gene	Longueur	Position	Mutation	Rôle du gène	Mécanisme de résistance	Fréquence de mutation
MTB13_M	tlyA	807	592 - 1 398	Leu11Leu		Inhibition de la synthèse protéine	40e70
MTB20_M	tlyA	807	1 202 317 - 1 203 123	Leu11Leu			
<b>Ethionamide</b>							
MTB1_M	inhA	810	233 759 - 234 568	Ser315Thr	Fatty acid enoyl acyl carrierprotein reductase A	Perturbe la biosynthèse de la paroi cellulaire	
MTB20_M	ethA	1470	2 833 125 - 2 834 594		Flavin monooxygenase		
<b>Para-aminosalicylic acid</b>							
MTB1_M	thyA	801	2 914 634 - 2 915 434	Thr202Ala	Thymidylate synthase A	Inhibition du métabolisme de l'acide folique et de la thymine	<40
MTB2_M	thyA	801	1 201 103 - 1 201 903	Thr202Ala	Thymidylate synthase A		

**Tableau 11:Récapitulatif des médicaments avec les résistances des 4 MTB**

Médicament	Gènes impliqués dans la résistance			
	MTB1_M	MTB2_M	MTB13_M	MTB20_M
Isoniazid (INH) Rifampin (RMP)	efpA, katG (S315T)	efpA, rpoB (S450L), katG (S315T)	efpA,	efpA,
aminoglycoside antibiotic gentamicin B arbekacinA tobramycin gentamicin C amikacin	aac	mtrA, aac	mtrA, aac	Aac,
fluoroquinolone antibiotic	mfpA, gyrA (S95T)	mfpA, gyrA ( (S95T)	mfpA, gyrA ( (S95T)	mfpA, gyrA ( (S95T)
Ethambutol (EMB)			embC (V981L)	embC (V981L)
Ofloxacin (OFX)			gyrA	gyrA

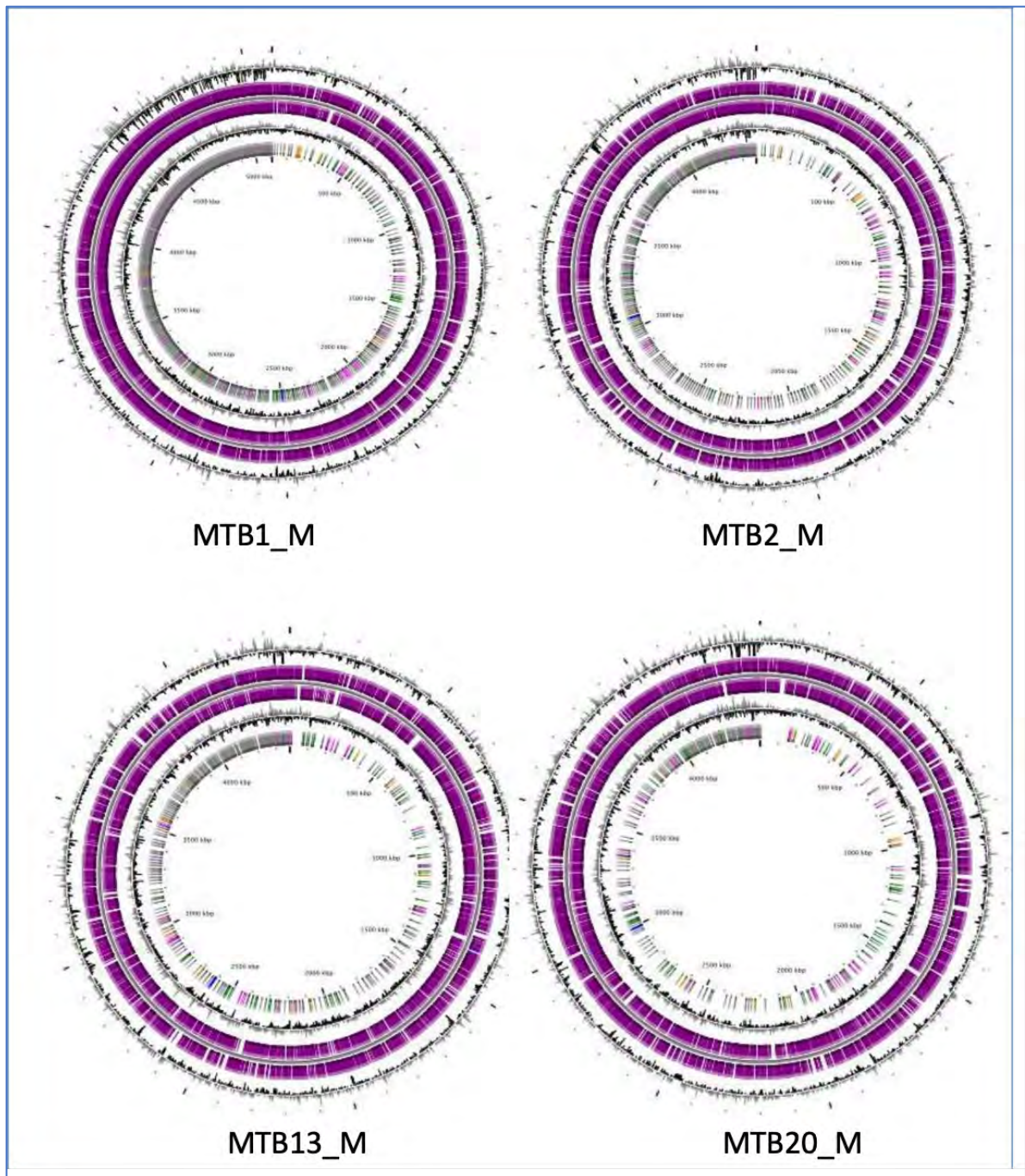
## 5. La Comparaison génomique des quatre MTB

Comparaison des gènes résistants des 4 souches avec la souche de référence H37Rv:

Une carte du génome circulaire de toutes les souches a été générée à l'aide de CGView à l'aide de contigs assemblés. Le génome circulaire des souches MTB1\_M, MTB2\_M, MTB13\_M et MTB20\_M affiche le CDS, le cadre de lecture ouvert, le contenu GC, le nombre d'ARN et le biais de GC où les 2 premiers anneaux représentent le CDS et le nombre d'ARN (ARNt, ARNr et ARNs), sur les brins avant et arrière.

- **Cartographies des quatre souches de Mycobacterium Tuberculosis**

- Couleur vert : tRNAs , -couleur bleu: ARNr



**Figure 25: Cartographie des 4 souches MTB**

Les contenus GC que *M. Tuberculosis* dépassent la moyenne du génome (65,6%) sont représentés par un tracé noir avec les pics s'étendant vers l'extérieur du cercle, tandis que les contenus GC inférieurs à la moyenne du génome s'étendent vers la marque centrale. L'intrigue la plus interne représente GC Skew. La couleur verte montre l'excédent moyen positif de G

+ C de guanine par rapport à la cytosine, tandis que la couleur violette indique l'excès moyen de G-C négatif de cytosine par rapport à la guanine.

- **KatG** : c'est un gène qui code pour la catalase-peroxydase, lorsqu'il est muté il y a une inhibition de la synthèse d'acide mycolique

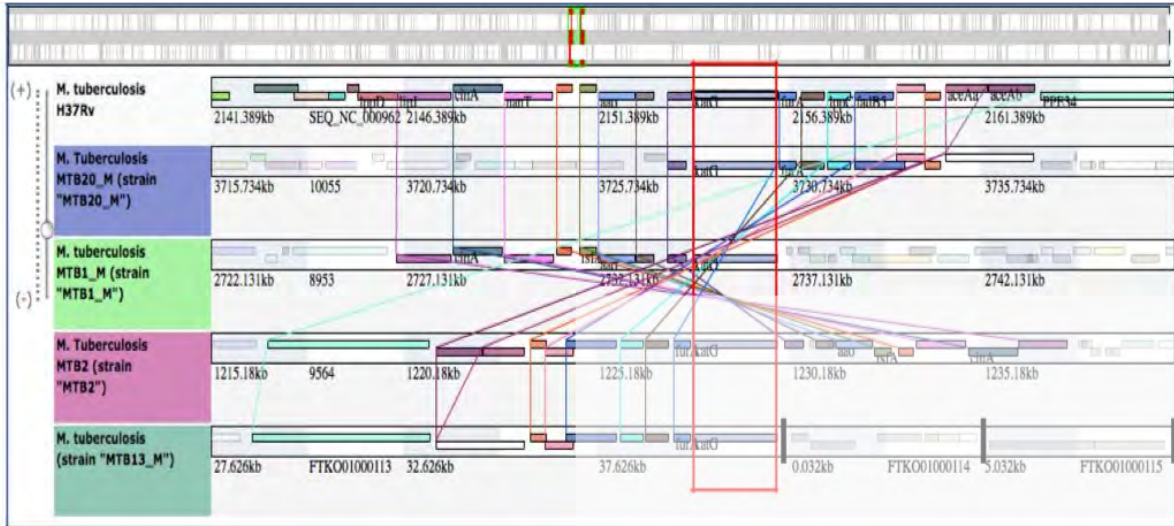


Figure 26: Position du gène *katG* sur les 4 souches et la référence H37Rv

- **RpoB** : c'est un gène qui code pour la sous unité de l'ARN polymérase, lorsqu'il est muté il y a une inhibition de la synthèse de l'ARN.



Figure 27: Position du gène *RpoB* sur les 4 souches et la référence H37Rv

- **GyrA** : c'est un gène qui code pour l'ARN méthyltransférase, lorsqu'il est muté il y a une inhibition de la synthèse des protéines.

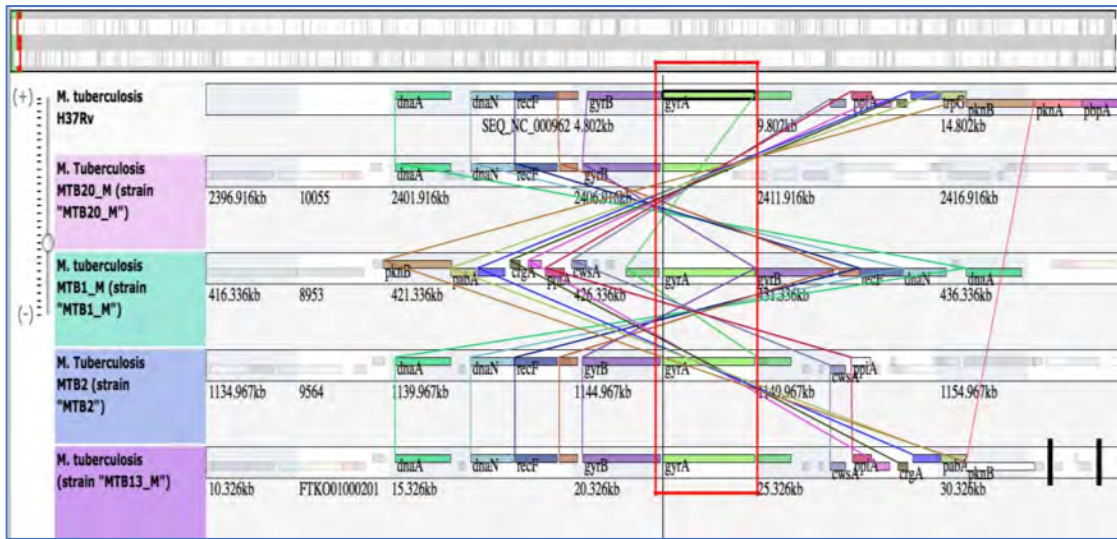


Figure 28: Position du gène *gyrA* sur les 4 souches et la référence H37Rv

- **GyrB** : c'est un gène qui code pour la sous unité B de l'ADN gyrase, lorsqu'il est muté il y a une inhibition de la synthèse de l'ADN.

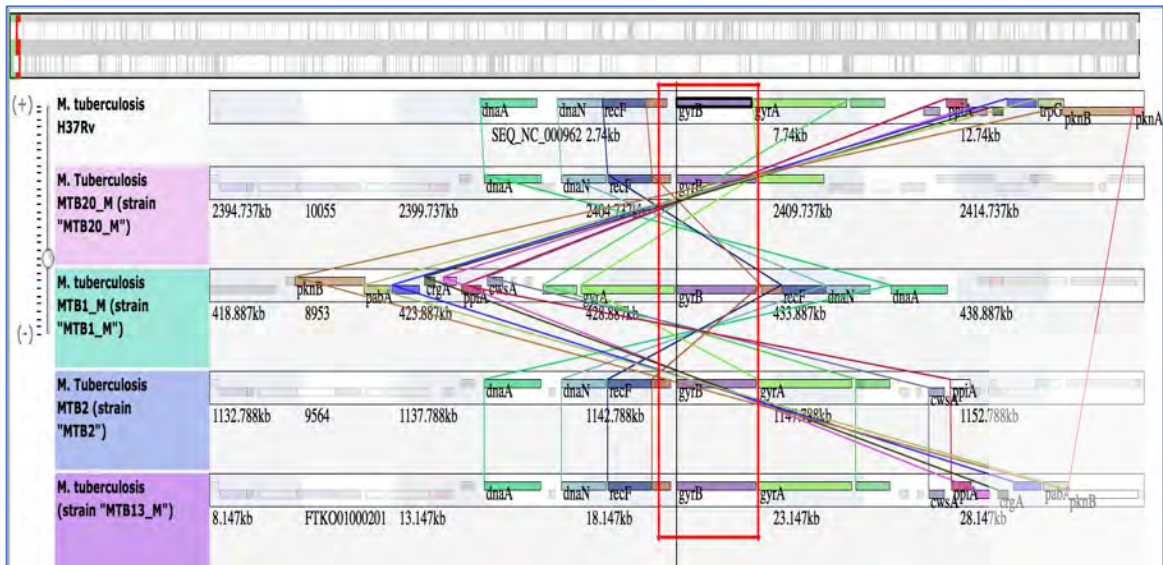


Figure 29: Position du gène *gyrB* sur les 4 souches et la référence H37Rv

- **TlyA** : c'est un gène qui code pour la sous unité B de l'ADN gyrase, lorsqu'il est muté il y a une inhibition de la synthèse de l'ADN.



Figure 30: Position du gène *tlyA* sur les 4 souches et la référence H37Rv

## 6. Analyse des gènes de virulence

La virulence est désignée par la capacité d'un agent pathogène à provoquer une maladie et, pour les mycobactéries, elle dépend de leur capacité à résider dans les cellules hôtes et à échapper aux mécanismes microbicides des macrophages. L'issue de l'infection tuberculeuse (TB) est très variable et il semble que la relation la plus étroite entre le genre *Mycobacterium* et les humains ait façonné le génome mycobactérien afin de coder des facteurs bactériens qui reflètent un programme très évolué et coordonné de stratégies d'évasion immunitaire qui interfèrent à la fois et une maladie causant une immunité adaptative même chez un hôte totalement immunocompétent. La *Mycobacterium tuberculosis* (MTB) ne possède pas de facteurs de virulence classiques, il existe de nombreux gènes associés à la virulence et des gènes de mode de vie de virulence du complexe *Mycobacterium tuberculosis* (MTBC).

Dans notre étude, l'outil VirulenceFinder 2.0 avec la version 2018-10-12 de la base de données a été utilisé pour identifier les gènes de virulence en se basant sur le facteur de virulence obtenu à partir de VFDB.

Les résultats ont montré que tous les facteurs de virulence vérifiés existaient dans toutes les souches.

**Tableau 12:Gènes de virulence pour MTB1\_M**

Gènes impliqués dans la virulence - MTB1_M					
VF Classes	Facteurs de virulence défensive, protéine de stress		Facteurs de virulence non spécifiques, enzyme, catalase-peroxydase	Facteurs de virulence non spécifiques, Métabolisme cellulaire, Enzyme, Peptidase	Facteurs de virulence offensifs, toxine, à action membranaire
Vir Gene	clpc	clpp	katP	gad, lsp	mchF
VF Label	VF0072	VF0074	XX0028	XX0023 VF0351	XX0031
Produit	classe III stress response-related ATPase, AAA+ superfamily	ATP-dependent Clp protease proteolytic subunit	catalase/hydroperoxydase HPI (I) katG	glutamate decarboxylase A, PLP-dependent gadA, Lipoprotein signal peptidase (fragment)	Putative peptide import ATP-binding protein BOV_A0347 (fragment)

**Tableau 13:Gènes de virulence pour MTB2\_M**

Gènes impliqués dans la virulence - MTB2_M					
VF Classes	Facteurs de virulence défensive, protéine de stress		Facteurs de virulence non spécifiques, enzyme, catalase-peroxydase	Facteurs de virulence non spécifiques, Métabolisme cellulaire, Enzyme, Peptidase	Facteurs de virulence offensifs, toxine, à action membranaire
Vir Gene	clpc	clpp	katP	gad	lsp
VF Label	VF0072	VF0074	XX0028	XX0023	VF0351
Produit	class III stress response-related ATPase, AAA+ superfamily	ATP-dependent Clp protease proteolytic subunit	catalase/hydroperoxydase HPI (I) katG	glutamate decarboxylase A, PLP-dependent gadA	Lipoprotein signal peptidase



**Tableau 14:Gènes de virulence pour MTB13\_M**

Gènes impliqués dans la virulence - MTB13_M					
VF Classes	Facteurs de virulence défensive, protéine de stress		Facteurs de virulence non spécifiques, enzyme, catalase-peroxydase	Facteurs de virulence non spécifiques, Métabolisme cellulaire, Enzyme, Peptidase	Facteurs de virulence offensifs, toxine, à action membranair e
Vir Gene	clpc	clpp	katP	gad	lsp
VF Label	VF0072	VF0074	XX0028	XX0023	VF0351
Product	class III stress response-related ATPase, AAA+ superfamily	ATP-dependent Clp protease proteolytic subunit	catalase/hydroperoxidase HPI (I) katG	glutamate decarboxylase A, PLP-dependent gadA	Lipoprotein signal peptidase

**Tableau 15:Gènes de virulence pour MTB20\_M**

Gènes impliqués dans la virulence - MTB20_M					
VF Classes	Facteurs de virulence défensive, protéine de stress		Facteurs de virulence non spécifiques, enzyme, catalase-peroxydase	Facteurs de virulence non spécifiques, Métabolisme cellulaire, Enzyme, Peptidase	Facteurs de virulence offensifs, toxine, à action membranair e
Vir Gene	clpc	clpp	katP	gad	mchF
VF Label	VF0072	VF0074	XX0028	XX0023	XX0031
Product	class III stress response-related ATPase, AAA+ superfamily	ATP-dependent Clp protease proteolytic subunit	catalase/hydroperoxidase HPI (I) katG	glutamate decarboxylase A, PLP-dependent gadA	

## 7. L'arbre phylogénétique

Un arbre phylogénétique est un arbre schématique qui montre les relations de parentés entre des groupes d'êtres vivants. Chacun des nœuds de l'arbre représente l'ancêtre commun de ses descendants. Dans notre cas nous avons construit un arbre phylogénétique qui contient les quatre souches de MTB avec la référence pour voir la similarité entre ces dernières.

L'outil Realphy a été utilisé, il est un pipeline en ligne qui peut déduire les arbres phylogénétiques de données de séquençage du génome entier. L'utilisateur doit seulement fournir le génome de référence au format Fasta ou Genbank (contigs ou génomes entièrement séquencés) ainsi qu'un certain nombre d'autres génomes de requête qui peut être en format Fasta ou Genbank. Les requêtes seront ensuite mappées à chacune de la référence par l'intermédiaire de bowtie2 (outils d'alignement) alors l'alignements multiples des séquences seront reconstruits d'où l'arbres phylogénétique est déduits par l'intermédiaire de PhyML.

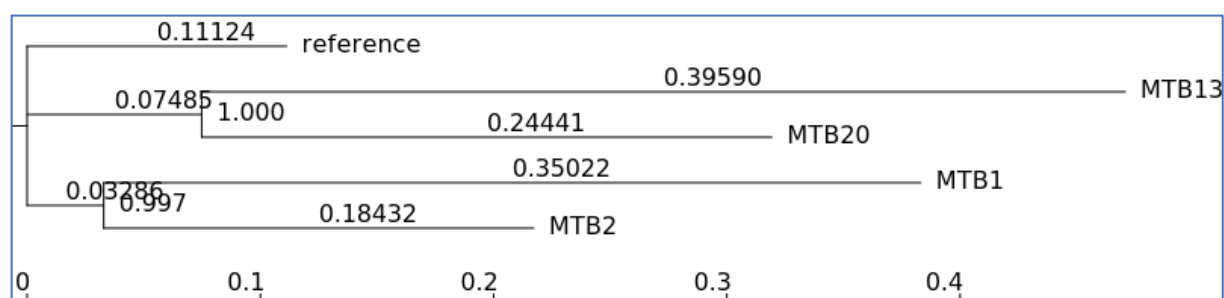


Figure 31: L'arbre phylogénétique des 4 souches MTB et la référence H37Rv

Ci-dessous le pourcentage de mapping avec la référence H37Rv :

Tableau 16:Le pourcentage de mapping des 4 souches avec la référence H37Rv

Séquences	Longueur en pb	% de mapping avec la référence
<b>H37Rv</b>	<b>4 411 532</b>	<b>100%</b>
<b>MTB13</b>	<b>4303062</b>	<b>97.54%</b>
<b>MTB20</b>	<b>4258995</b>	<b>96.54%</b>
<b>MTB1</b>	<b>4089119</b>	<b>92.69%</b>
<b>MTB2</b>	<b>3817606</b>	<b>86.53%</b>

# Discussion sur la partie Tuberculose

Les souches de *Mycobacterium tuberculosis* résistantes aux antibiotiques menacent de progresser dans la maîtrise de l'épidémie mondiale de tuberculose.

*Mycobacterium tuberculosis* est intrinsèquement résistante à de nombreux antibiotiques, ce qui limite le nombre de composés disponibles pour le traitement.

Cette résistance intrinsèque est due à un certain nombre de mécanismes, notamment une enveloppe cellulaire épaisse, cireuse et hydrophobe et à la présence d'enzymes modifiant et dégradant le médicament.

Avec l'absence de déterminants de résistance transférés horizontalement et conférée par des mutations chromosomiques, la résistance aux médicaments contre *M. tuberculosis* reste active. Ces mutations chromosomiques peuvent conférer une résistance au médicament via une modification ou une surexpression du médicament ciblé.

Bien que la diversité génétique de *M. tuberculosis* soit faible par rapport à d'autres bactéries pathogènes, il a été démontré que le fond génétique de la souche influait sur de nombreux aspects de l'évolution de la résistance aux médicaments. Le taux d'évolution de la résistance et les coûts de remise en forme des mutations de résistance aux médicaments peuvent varier en fonction du fond génétique.

La *Mycobacterium Tuberculosis* et d'autres bactéries entériques ont des gènes étendus en commun qui sont responsables de la résistance aux antibiotiques, conduisant à la prévalence croissante des phénotypes de résistance aux médicaments dans la clinique. La disponibilité des séquences du génome complet de *M. Tuberculosis* avec les profils de résistance aux antibiotiques différentiels renforce notre compréhension et l'appréciation des mécanismes de résistance aux médicaments et leur évolution au Maroc.

Au Maroc, le taux d'incidence avoisine les 87 cas par 100.000 habitants, dont la tuberculose pulmonaire représente la moitié, et affecte les jeunes de 15 à 45 ans. Chaque année près de 30.000 cas de tuberculose sont enregistrés, y compris les nouveaux cas et les cas de rechute, La propagation de cette maladie est due aux facteurs socio-économiques, notamment les conditions d'habitat, la pauvreté et la malnutrition. D'après les statistiques réalisées par le ministère de santé, montrent que 86% des cas ont été enregistrés dans six régions, totalisant une moyenne de 78% de la population du Royaume, à savoir Casablanca-Settat, Rabat-Salé-Kénitra, Tanger-Tétouan-Al Hoceima, Fès-Meknès, Marrakech-Safi et Souss-Massa.

Cette multi-résistance peut avoir des impacts en termes de traitement et de choix d'antibiotique. La propagation de cette maladie avec un taux de mortalité qui a atteint 12 cas par 100.000 patients tuberculeux en 2017, ces raisons ont motivé l'étude de la mycobacterium tuberculosis qui la bactérie responsable de la tuberculose,

Ce travail de doctorat contribue à la compréhension du profil multi-résistant de cette espèce qui a pu être isolée de différents patients atteints de la tuberculose au sein de l'Hôpital Militaire d'Instruction Mohammed V, spécifiquement le laboratoire de bactériologie.

Ce travail permet également d'étudier les différentes souches afin de mieux comprendre l'évolution et les liens phylogénétiques existants entre ces souches d'origines marocaines par l'approche de séquençage haut débit et analyse génomique comparative en utilisant plusieurs outils bio-informatiques.

Actuellement, le séquençage du génome bactérien des isolats cliniques devient de plus en plus utilisé car le coût du séquençage de l'ADN continue de diminuer. En cancérologie, en microbiologie et dans différents domaines de recherche, les technologies de séquençage de nouvelle génération ont notamment été appliquées, en effet le potentiel d'utilisation de ces technologies en microbiologie clinique reste énorme, en particulier pour la détection de facteurs de virulence et de gènes de résistance aux antibiotiques et l'épidémiologie de bactéries pathogènes.

Actuellement, de nombreux génomes de mycobacterium tuberculosis ont été séquencés entièrement ou partiellement en utilisant les NGS (plus 6000 projets) sont mentionnés sur le site GOLD (<http://www.genomesonline.org/>).

Le montage d'un projet de séquençage d'un génome est un processus long qui demande du temps pour mobiliser des ressources à la fois financières, techniques, humaines et informatiques. Ce décalage entre d'une part, les innovations et progrès des NGS et, d'autre part, la réalité de la recherche, constitue également une difficulté dans le choix de la stratégie de séquençage et d'assemblage d'un génome comme décrit dans le chapitre II.

La génomique bactérienne a apporté une meilleure compréhension des mécanismes d'évolution des génomes bactériens. L'obtention d'une culture pure est une étape essentielle pour relier le phénotype particulier d'une souche bactérienne à son contenu en gènes, notamment pour étudier spécifiquement sa résistance. Ainsi, les comparaisons génomiques ont montré que le transfert horizontal de gènes était chez les bactéries le principal moyen d'acquérir de nouvelles fonctions. La génomique bactérienne est utile aux microbiologistes, aux infectiologues et aux spécialistes en hygiène hospitalière.

Un projet de génomique comprend 6 étapes principales : le choix de l'espèce et/ou de l'échantillon à séquencer, l'extraction d'ADN, le séquençage (préparation de librairie et exécution du séquençage), l'assemblage des séquences obtenues ou le mapping, l'annotation, l'analyse des données et enfin la soumission du projet sur les banques de données. Ainsi, dans un but d'application médicale, il est particulièrement intéressant d'utiliser un génome sans effectuer la fermeture des « gap », afin de réduire les coûts jusqu'à l'obtention de l'information utile.

Dans le cas de notre étude pour *M.Tuberculosis*, aucun génome marocain complet n'était disponible dans les bases de données publiques. C'est ainsi que le séquençage et l'assemblage du génome de *M.Tuberculosis* MTB13\_M que nous avons réalisé, constitue une étape importante dans l'étude de cette espèce émergente (chapitre II).

Dans le cas de la souche *M.Tuberculosis* MTB13\_M, le séquençage a été réalisé en utilisant la technologie Illumina Miseq.

Durant le processus d'assemblage des génomes, plusieurs critères et indices sont à prendre en considération pour évaluer la qualité de l'assemblage tel que le N50 des contigs, la couverture, le taux d'erreurs d'assemblage, le nombre de contigs, et la taille totale de l'assemblage (Ekblom et Wolf 2014).

La séquence de *M.Tuberculosis* MTB13\_M a été assemblée, on a obtenu 4,347,307bp désigné comme le premier génome de référence publié dans les bases de données (chapitre II).

Dans une approche similaire nous avons effectué un projet de séquençage à haut débit (un seul run) pour 3 isolats cliniques de la souche de *M.Tuberculosis* collectés de trois autres patients de l'hôpital d'instruction militaire Mohammed V. L'annotation automatique et expert des séquences de génome *M.Tuberculosis* MTB1\_M, MTB2\_M, et MTB20\_M ont été séquencés à l'aide de la technologie Illumina Miseq.

Les quatre souches *M.Tuberculosis* ont été filtrées pour conserver les séquences possédant un score supérieur à 25. Ensuite, un assemblage de Novo permet d'obtenir 4,064,236pb, 4,689,658pb, 4,358,959pb respectivement MTB1\_M, MTB2\_M et MTB20\_M avec un pourcentage de GC de 65%.

Dans ce contexte, les nouvelles techniques de séquençages représentent un moyen indispensable pour comprendre les différents mécanismes de résistance, leur base génétique et leur dissémination intercontinentale.

Plusieurs études rapportent l'utilisation des nouvelles techniques de séquençages comme outils d'investigation pour comprendre la dissémination de souches à l'échelle d'un hôpital, d'une région, ou même d'un continent (Barzon et al. 2013).

La connaissance de la résistance à *M.Tuberculosis* est importante pour le traitement efficace des MDR et XDR.

Au Maroc, de nombreuses études ont déjà été consacrées à la prévalence de la pharmacorésistance de première ligne et aux mutations génétiques associées (Sabouni et al. 2008) (Chaoui et al. 2014). Dans une étude publiée, les auteurs ont analysé 703 souches, dont 12,8% des MDR, pour ces souches multi-résistantes, plusieurs mutations ont été détectées : Ser531Leu et Ser315Thr dans les gènes *rpoB* et *katG* associées à la résistance à la rifampicine et l'isoniazide respectivement.

L'analyse génotypique dans cette étude a révélé que parmi les souches MDR ont des mutations du gène *gyrA* ; la mutation la plus enregistrée étant Asp91Ala. Aucune mutation n'a été trouvée dans les gènes *gyrB*, *rrs* et *tlyA*.

Les analyses génomiques réalisées pendant cette thèse sur les quatre génomes *M.Tuberculosis* MTB1\_M, MTB2\_M, MTB13\_M et MTB20\_M, ces dernières ont été soumises dans les bases de données, et nous avons identifié trois mutations différentes sur le gène *rpoB* associé à la *Rifampicine* : Ala526Ser et Ser450Leu pour MTB1\_M, MTB2\_M respectivement, et Asp103Asp pour MTB13\_M et MTB20\_M ainsi des mutations sur le gène *katG* associé à l'isoniazide : Ser95Thr, Glu21Gln, Gly668Asp et Ser315Asn pour MTB1\_M, MTB2\_M, MTB13\_M et MTB20\_M respectivement.

Pour la deuxième ligne de résistance tel que le Fluoroquinolone, nous avons détecté des mutations pour deux gènes le *gyrA* et le *gyrB*, pour le premier : Gly668Asp pour MTB1\_M et MTB13\_M ainsi Glu21Gln et Phe640Val pour MTB2\_M et MTB20\_M respectivement, au niveau du gène *gyrB* : Asp538Asn, Thr539Asn, Asp500His pour MTB1\_M, MTB2\_M, et MTB20\_M respectivement. Nous avons identifié une résistance à la *Capreomycine* pour les quatre souches analysées au niveau du gène *tlyA*.

A travers les résultats obtenus par cette analyse génomique, les étapes d'analyse de données issues du séquenceur ILLUMINA Miseq sont établis dans un workflow bio-informatique dédié spécifiquement à l'analyse de *M.Tuberculosis* en se basant sur des références bibliographique qui décrivent plusieurs outils bio-informatique pour les différentes étapes d'analyse génomique.

# CHAPITRE 3 : Workflow d'analyse Transcriptomique pour la Maladie de Parkinson

L'objectif de ce chapitre est de décrire les étapes de réalisation de cette partie, commençant par la création du modèle animal parkinsonien ensuite les différentes méthodes de validation. Dans ce chapitre donne une vue détaillée sur les expériences de biologie moléculaire de l'extraction d'ARN jusqu'au séquençage, suivi par les étapes d'analyse transcriptomique pour la MP avec les outils utilisés en citant les résultats obtenus pour chaque étape du workflow.

# I. Partie expérimentale (Biologique)

## 1. Les modèles animaux de la MP

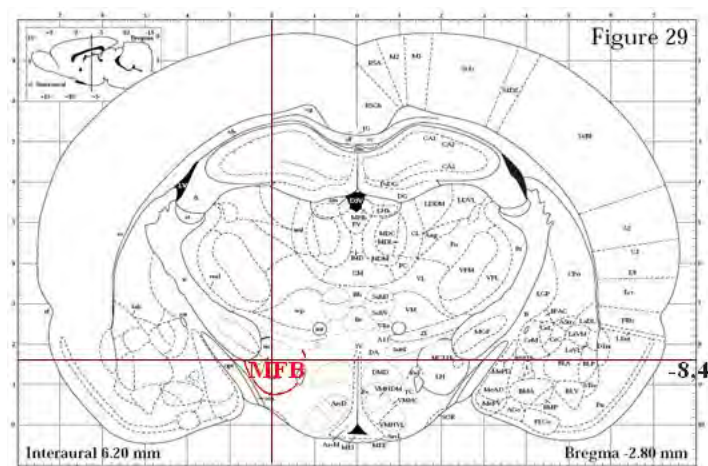
### 1.1. Animaux

Les animaux utilisés dans ce travail, font l'objet du projet qui a été soumis et accepté par le comité d'éthique de Bordeaux sous le numéro d'autorisation 2017112216001160. Pour cette étude, 25 rats adultes mâles âgés de 8 à 10 semaines environ, de la souche *Sprague Dawley* (260-280g), ont été utilisés provenant de l'élevage *Janvier (Le Genest Saint Isle, France)*. Ces animaux sont hébergés, depuis leur arrivée jusqu'à leur euthanasie, dans une pièce où la température (22°C) et le cycle nyctéméral (12H/12H - 7H-19H : ON ; 19H-7H : OFF) sont contrôlés, et la nourriture leur est accessible *ad libitum*. Toutes les considérations éthiques ont été respectées pour veiller au bien-être des animaux et minimiser leur nombre et leur souffrance. Trois groupes ont été réalisés dans cette étude : le groupe 6-OHDA (6-hydroxydopamine) (n=7) recevant une injection de 6-OHDA dans le faisceau médian du télencéphale (*MFB, median forebrain bundle*), le groupe  $\alpha$ -syn (n=8) recevant une injection de l' $\alpha$ -synucléine dans la SNpc et le groupe sham (n=10) recevant du NaCl (0.9%) dans le MFB (Ce groupe n'a pas été utilisé, nous avons utilisé le côté gauche de chaque animal comme témoin).

### 1.2. Lésion unilatérale

Les animaux ont été rendus hémiparkinsoniens par une injection stéréotaxique unilatérale de 6-OHDA ou de vecteurs viraux avec de l' $\alpha$ -syn humaine mutée dans le faisceau médian du télencéphale (MFB) selon les coordonnées stéréotaxiques de l'atlas du cerveau de rat (Paxinos G et Watson C 1996). Le MFB est un faisceau contenant les fibres dopaminergiques de la voie nigro-striée des neurones projetant de la SNpc vers le striatum. En injectant la 6-OHDA, une neurotoxine spécifique des neurones catécholaminergiques, utilisée pour détruire les neurones dopaminergique et noradrénergiques, les corps cellulaires des neurones dopaminergiques nigraux seront lésés de façon rétrograde, causant ainsi la dégénérescence caractéristique de la MP. Les neurones noradrénergiques étant protégés par un traitement spécifique.





**Figure 32: Localisation du faisceau médian du télencéphale (MFB) par ses coordonnées stéréotaxiques sur une coupe frontale de cerveau de rat.**

*Localisation du faisceau médian du télencéphale (MFB) par ses coordonnées stéréotaxiques sur une coupe frontale de cerveau de rat. Schéma adapté à partir de l'atlas stéréotaxique du cerveau de rat (Paxinos G et Watson C 1996)*

Avant l'anesthésie par injection intra péritonéale (i.p) de kétamine hydrochloride (75mg/kg) et xylazine hydrochloride (10mg/kg), le rat est traité à la désipramine (0,5%, 25mg/kg, 5ml/kg) qui inhibe la recapture de la noradrénaline (NA) pour protéger les neurones noradrénergiques de la lésion à la 6-OHDA. Le rat traité et anesthésié est placé sur un cadre stéréotaxique (Uminecanique, France). Le crâne est mis à nu puis séché pour la distinction des repères anatomiques à la surface du crâne : bregma et lambda. En se basant sur l'atlas stéréotaxique du cerveau de rat (Paxinos et Watson, 1996), le MFB droit est atteint en déplaçant la canule aux coordonnées suivantes : AP (antéro-postérieur) : -2,8 mm par rapport au bregma ; L (latéral) : 2 mm à droite de la ligne médiane ; DV (dorso-ventral) : -8,4 mm par rapport à la surface de la dure-mère. Pour le groupe Sham les mêmes coordonnées stéréotaxiques ont été utilisées, mais pour le groupe  $\alpha$ -syn deux injections sur deux sites différents du côté droit du SNpc de l' $\alpha$ -synucléine humaine mutée sous contrôle du promoteur synapsin-1 ( $2,5 \times 10^{13}$  vg/mL) dont les coordonnées sont les suivantes :

	Référence	1 <sup>e</sup> site d'injection	2 <sup>e</sup> site d'injection
AP (antéropostérieur)	Bregma	- 5,1 mm	- 5,4 mm
L (latéral)	Ligne médiane	± 2,2 mm	± 2,0 mm
DV (dorsoventral)	Surface du cerveau	- 7,8 mm.	- 7,8 mm

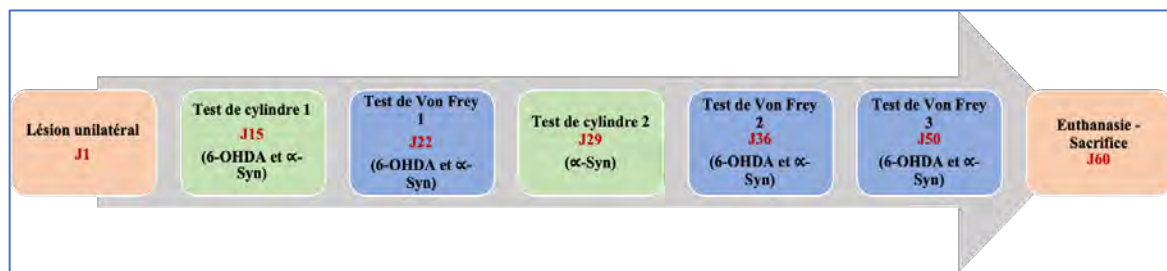
**Tableau 17: Les coordonnées des deux sites d'injection de l'Alpha-Synucleine**

La 6-OHDA (12,5 µg) est dissoute dans un volume de 2,5 µl de solution saline (NaCl 0,9%) et injectée en intracérébrale dans le MFB (droit), en utilisant une micro seringue Hamilton, ainsi que le vecteur AAV2/9 exprimant l'α-synucléine est dissout dans du NaCl 0,9 %. Il a été injecté à raison de 1µL pendant une minute par site en utilisant une micro seringue Hamilton.

Après l'injection, un délai d'attente de 5 minutes était respecté pour éviter tout reflux de la solution. Après une suture au niveau de la peau du crâne, les animaux ont reçu une injection d'antibiotique (100 µl de Borgal 7,5% en intra musculaire) et d'analgésique (80 µl de Kotefen en sous-cutané soit 2 mg/kg). Les rats ont été par la suite placés sous une lampe chauffante jusqu'à leur réveil, ensuite, ils ont été remis à l'animalerie pendant deux semaines avant le début des tests comportementaux et le sacrifice des animaux.

## 2. Études comportementales

La figure ci-dessous schématise l'ensemble de la chronologie des tests expérimentaux réalisés:



**Figure 33:Schéma chronologique du déroulement des expériences**

### 2.1. Test de Cylindre

#### • Protocole

Ce test se déroule pour la 1<sup>ère</sup> fois deux semaines après la chirurgie, il consiste à placer l'animal dans un cylindre en pléxiglace transparent (50 cm de haut, 20 cm de diamètre) nettoyé avec une solution d'acide acétique à 0,1%. Les animaux sont filmés avec une caméra placée devant le cylindre de sorte que la distance entre la caméra et le cylindre permette de bien voir les pattes de l'animal afin d'analyser visuellement les paramètres suivants : nombre de contacts de la patte gauche avec la paroi, nombre de contact de la patte droite avec la paroi et le nombre de contact des deux pattes de façon simultanée avec la paroi (Rattka et al. 2016).

Remarque : un rat normal n'ayant subi aucune lésion utilise les deux pattes de manière  $\pm$  égale. Le rat parkinsonien utilise la patte controlatérale à la lésion de façon significativement moins importante que la patte ipsilatérale.



**Figure 34:** Test de cylindre; un rat parkinsonien au cours de l'essai au cylindre (*Rattka et al. 2016*)

Le test du cylindre a été fait pour le groupe 6-OHDA deux semaines après la chirurgie et pour le groupe  $\alpha$ -Syn, le test a été fait deux semaines et quatre semaines après la lésion.

#### • Analyse statistique et résultats

Le comptage des contacts se fait en visualisant les vidéos prises pendant les tests. Les résultats sont exprimés en pourcentages d'utilisation de la patte gauche (PG) par rapport au pourcentage d'utilisation de la patte droite (PD), dite valeur contrôle. Ces valeurs sont présentées sous forme de moyennes  $\pm$  SEM. Un test de Student (*t*-test) a été utilisé pour l'évaluation statistique en utilisant le logiciel R avec la commande :

```
t.test (x, y,  
        alternative = c ("two.sided", "less", "greater"),  
        mu = 0, paired = FALSE, var.equal = FALSE,  
        conf.level = 0.95, ...)
```

- **Groupe 6-OHDA**

Deux semaines après l'injection de la 6-OHDA, l'analyse des résultats montre une diminution significative de l'utilisation de la patte gauche par rapport à la patte droite ( $p=0,002$ , tableau 18 et figure 35). Ce résultat montre la présence d'une bradykinésie du côté controlatérale à la lésion dopaminergique.

Groupe 6-OHDA	Semaine 2	
	Patte droite	Patte gauche
Moyenne	38,83	21,50
N d'échantillons	7	7
P Value	0,0021	
Significativité	**	
SD	10,13	5,86
SEM	3,83	2,21

Tableau 18: Résultat du test de student pour le test de cylindre pour le groupe 6-OHDA

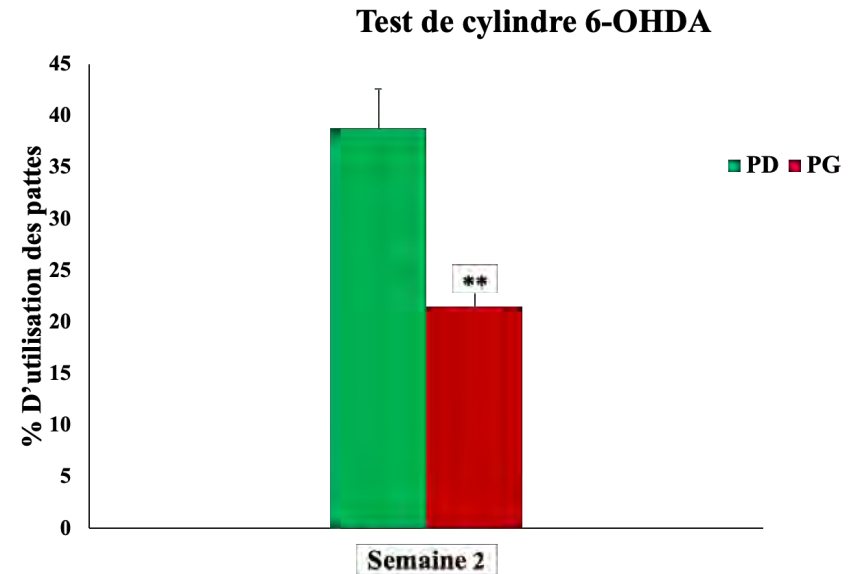


Figure 35: Taux de réussite au test de cylindre (Groupe 6-OHDA)

- **Groupe Alpha-Syn**

L'injection de l'Apha-Syn dans la SNpc provoque une diminution significative de l'utilisation de la patte gauche par rapport à la patte droite à 2 semaines ( $p=0,02$ , tableau 19, Figure 36) et à 4 semaine ( $p=0,0001$ , tableau 19, Figure 36) après la chirurgie. La comparaison des résultats des deux sessions montre le caractère progressif de cette diminution d'utilisation de la patte gauche.

Groupe Alpha-Syn	Semaine 2		Semaine 4	
	Patte droite	Patte gauche	Patte droite	Patte gauche
Moyenne	34,50	21,11	44,51	14,29
N d'échantillons	8	8	8	8
P Value	0,021		<0,0001	
Significativité	*		****	
SD	13,32	5,95	12,20	9,72
SEM	4,71	2,10	4,31	3,43

Tableau 19: Résultat du test de student pour le test de cylindre pour le groupe Alpha-Syn

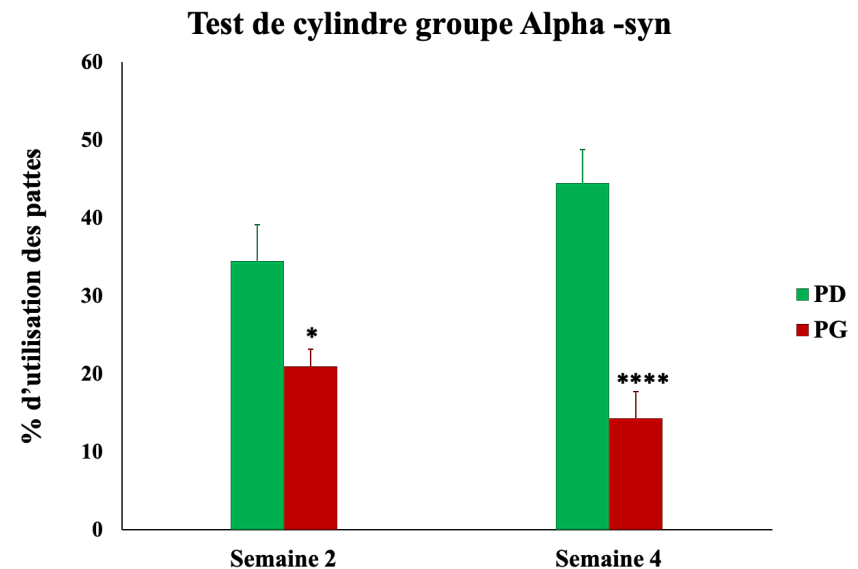


Figure 36: Taux de réussite au test de cylindre (Groupe Alpha-Syn)

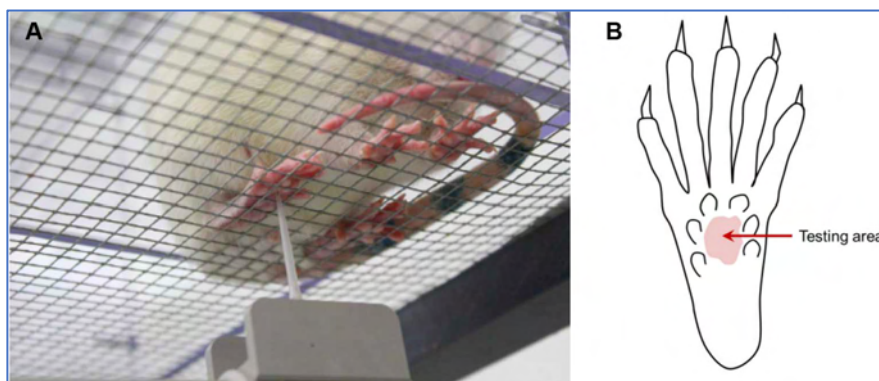
## 2.2. Test de Von Frey : Test de l'allodynie mécanique

### • Protocole

Le test électronique de Von Frey (Von Frey électronique, Bioseb) permet d'évaluer la sensibilité mécanique statique chez le rat (S. H. Kim et Chung 1991). Il peut être utilisé comme lecture comportementale pour un large éventail de modèles, notamment les modèles de douleurs inflammatoires et neuropathiques. Ce test permet d'évaluer l'allodynie mécanique chez les animaux suite à un stimulus mécanique normalement non douloureux par application d'une pression au niveau de la surface plantaire des pattes arrière ipsilatérales et controlatérales du rat, par rapport au côté opéré. Le retrait de la patte de l'animal indique le seuil de force appliquée sur la patte de l'animal, exprimé en gramme.

Les animaux ont été placés dans une cage à fond grillagé métallique du Von Frey pendant 15 minutes avant le début de l'expérience pour les habituer. Des stimulations mécaniques à l'aide d'un embout en plastique sont appliquées sous la voûte plantaire de chacune des pattes arrière de l'animal avec une force croissante, mesurée à l'aide du dynamomètre. L'expérience se réalise sur un animal en appui quadripodal en 5 sessions avec une coupure d'1 à 2 minutes entre chacune des sessions. Une moyenne des valeurs est ensuite obtenue.

Les seuils de retrait des pattes droites et gauches en fonction de la force exercée permettent de suivre l'évolution de l'allodynie (Ferrier, Marchand, et Balayssac 2016) chez les deux groupes d'animaux, et ceux-ci sont déterminés à J22, J36 et J50 après la chirurgie.



**Figure 37: Application de la pointe de Von Frey sur la patte arrière du rat (Ferrier, Marchand, et Balayssac 2016)**

- **Analyse statistique et résultats**

Les résultats du test Von Frey sont exprimés en seuil de retrait de la patte pour une force donnée (g). Ces valeurs sont présentées sous forme de moyennes  $\pm$  SEM et les comparaisons ont été réalisées en utilisant le test de student (t-test) avec le logiciel R.

```
t.test (x, y,  
        alternative = c ("two.sided", "less", "greater"),  
        mu = 0, paired = TRUE,  
        conf.level = 0.95, ...)
```

- **Groupe 6-OHDA**

La figure 38 montre les résultats de la sensibilité mécanique des deux pattes obtenus chez les rats 6-OHDA lors du test de Von Frey 22, 36 et 50 jours après l'injection stéréotaxique de la 6-OHDA. L'analyse statistique montre que le seuil de sensibilité mécanique diminue au niveau de la patte gauche controlatérale à l'injection de la 6-OHDA par rapport à la patte droite ipsilatérale ( $p < 0,0001$  pour J22 et J36 et  $p < 0,001$  à J50, Figure 38, tableau 20). Ces résultats traduisent la manifestation d'une sensibilité mécanique associée à la lésion des neurones dopaminergiques par la 6-OHDA.

Groupe 6-OHDA	Semaine 3		Semaine 5		Semaine 7	
	Patte droite	Patte gauche	Patte droite	Patte gauche	Patte droite	Patte gauche
Moyenne	144,8	108,5	152,6	109,1	157,3	110,1
N d'échantillons	5	5	5	5	5	5
P Value	<0,0001		<0,0001		<0,0001	
Significativité	****		****		****	
SD	6,14	6,40	7,90	4,29	10,32	8,89
SEM	2,74	2,86	3,53	1,92	4,61	3,98

Tableau 20: Résultat du test de student pour le test de Von Frey pour le groupe 6-OHDA

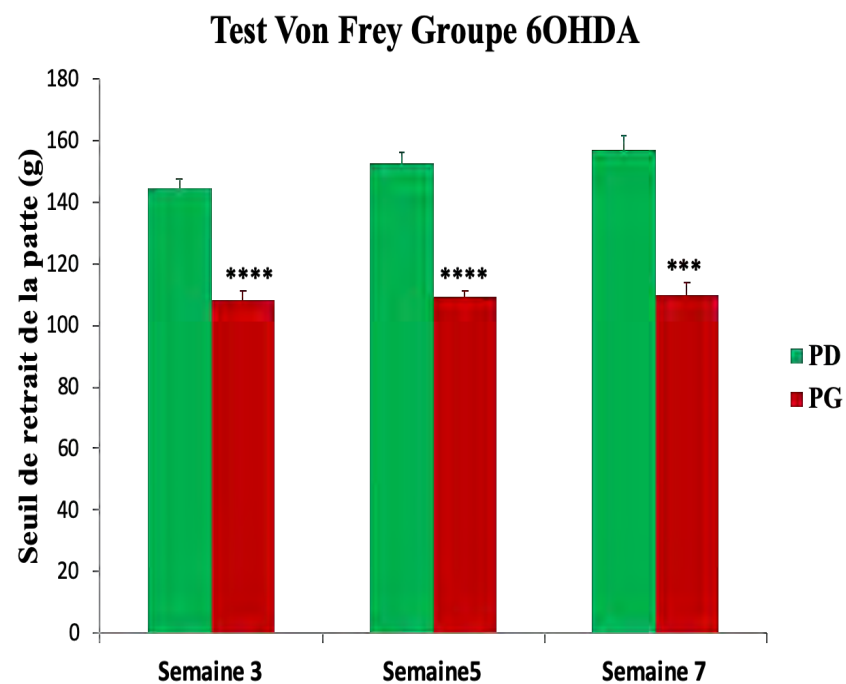


Figure 38: Taux de réussite du test Von Frey chez le groupe 6-OHDA

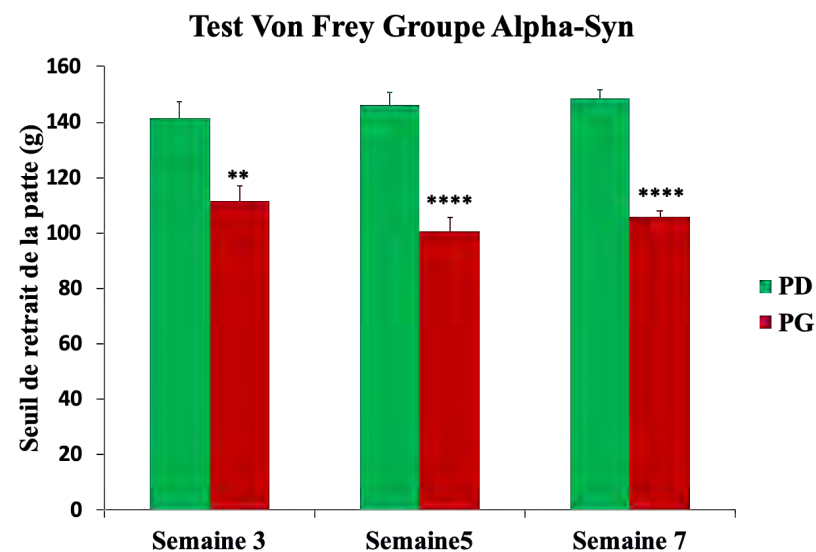


- **Groupe Alpha-Syn**

Les rats du groupe alpha-Syn ont été testés pour la sensibilisation à la douleur en utilisant un dispositif électronique von Frey : 22 jours, 36 jours et 50 jours après la chirurgie, ils ont montré une diminution progressive du seuil de retrait de la patte ipsilatérale par rapport au seuil de retrait de la patte controlatérale. Les résultats du test de student ont montré une significativité pour le test Von Frey effectué trois semaines après la chirurgie avec une p-value de 0,0026, et une p-value de <0,0001 pour la 5<sup>ème</sup> et 7<sup>ème</sup> semaine (Figure 39, Tableau 21) . Ces résultats montrent un développement progressif d'allodynie chez le groupe alpha-syn en fonction du temps.

Groupe Alpha-Syn	Semaine 3		Semaine 5		Semaine 7	
	Patte droite	Patte gauche	Patte droite	Patte gauche	Patte droite	Patte gauche
<b>Moyenne</b>	141,7	11,6	146,4	100,6	148,4	105,9
<b>N d'échantillons</b>	7	7	7	7	7	7
<b>P Value</b>	0,026		<0,0001		<0,0001	
<b>Significativité</b>	**		****		****	
<b>SD</b>	14,90	14,79	11,38	13,28	8,16	5,81
<b>SEM</b>	5,631	5,59	4,30	5,02	3,086	2,19

**Tableau 21:Résultat du test de student pour le test de Von Frey pour le groupe Alpha-Syn**



**Figure 39:Taux de réussite du test Von Frey chez le groupe alpha-Syn**

### 3. Analyse biochimique de la déplétion monoaminergique

A J60 après la chirurgie d'injection de la 6-OHDA ou de l'Alpha-Syn, les animaux sont sacrifiés par décapitation sous anesthésie générale profonde induite par injection intra-péritonéale d'un mélange de pentobarbital (Exagon) (200 mg/kg) et de Lidocaïne (20mg/kg).

Après l'arrêt cardiaque et respiratoire des animaux, le cerveau est extrait et congelé immédiatement dans de l'isopentane maintenu entre -40° et -45°C grâce à l'azote liquide.

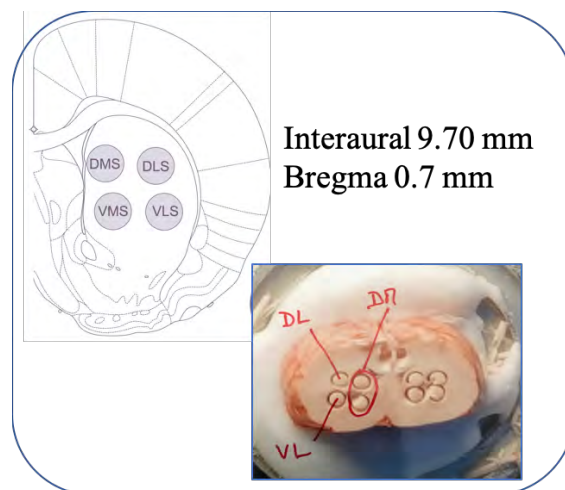
Les cerveaux extraits sont conservés à -80°C jusqu'à la découpe au cryostat qui s'effectue à -24°C et récupération des punchs de tissus.

#### 3.1. Traitement des tissus cérébraux

L'analyse biochimique a été faite sur huit structures cérébrales des deux côtés droits et gauches parmi eux : Striatum dorso-latéral (DL), Striatum ventro-latéral (VL), cortex frontal (M2).

Les cerveaux ont été découpés à l'aide d'un cryostat à -24 °C, et des prélèvements de tissus au niveau de chaque hémisphère ont été réalisés pour les différentes structures cérébrales d'intérêt à l'aide de canules en acier de 500 ou 800 µm de diamètre (Fitoussi, Dellu-Hagedorn, et De Deurwaerdère 2013).

La localisation des structures prélevées a été rapportée sur les diagrammes de sections coronales de l'atlas de Paxinos et Watson (1996) et correspond à la figure 40 ci-dessous.



**Figure 40: Les structures du cerveau prélevées pour HPLC**

Ces prélèvements ont ensuite été déposés dans des tubes Eppendorf préalablement pesés et remis dans le congélateur à -80 °C. Des photos ont été prises pour s'assurer que les échantillons ont été prélevés d'un plan similaire à chaque fois. Les tubes contenant les échantillons ont été à nouveau pesés avec précaution le jour de l'analyse biochimique (Dellu-Hagedorn, Fitoussi, et De Deurwaerdère 2017) et 100 µl d'acide perchlorique (HClO<sub>4</sub> 0,1 N, 4 ° C) ont été ajoutés.

Ensuite, les structures ont été dissociées pendant 6 secondes par un Sonicateur et centrifugés à 13 000 tours/minute pendant 30 minutes à 4 °C. Un volume de 10 à 20 µl du surnageant a été injecté dans le système de chromatographie liquide à haute pression (HPLC) (Chagraoui et al. 2019).

### 3.2. Analyse par HPLC des surnageants

La séparation des catécholamines<sup>51</sup> et indolamines<sup>52</sup> a été réalisée par chromatographie en phase liquide à haute performance (HPLC) en phase inversée. Une pompe isocratique<sup>53</sup> (modèle 116, système Gold) (« Beckman, Paris » s. d.) délivre la phase mobile à un débit de 1,2 ml/min à travers le système (pression ~ 1500 psi).

La phase mobile utilisée dans ces conditions se compose de : NaH<sub>2</sub>PO<sub>4</sub> (60mM), Na<sub>2</sub>EDTA (0,1mM), acide octyl-sulfonique (0,2mM), 7 % méthanol et ajustée à un pH de 3,9 avec de l'acide ortho-phosphorique. La colonne (Kromasyl C8, 4.6 x 150mm, 5µm, Touzard et Matignon) est précédée d'une pré-colonne (Browlee-Newgard, RP-8, 15 X 3.2 mm, 7µM). 20 µl de surnageant ont été injectés dans un injecteur manuel (vanne rhéodyne 7725) équipé d'une boucle de 200 µl (injection en boucle partielle). Le système HPLC est piloté par le logiciel Spike 2.

### 3.3. Détection électrochimique

La détection des catécholamines et les indolamines est faite par électrochimie de type coulométrique, qui se compose d'une cellule de garde (modèle 5021, ESA) dont le potentiel est à 0,1 Volt, une cellule analytique (modèle 5011, ESA) et un détecteur *Coulochem* I ou II (ESA). Des potentiels de +350 mV et -270 mV sont appliqués pour oxyder et réduire les composés, respectivement. Le traitement du signal ampérométrique se fait par le logiciel CED (« CED Spike2 : Présentation du système » s. d.). Le seuil de détection pour les différentes molécules (rapport signal/bruit = 3) est d'environ 15pg/20µl injectés.

### 3.4. Analyses statistiques

La comparaison des effets de la 6-OHDA et de l'α-syn sur les quantités tissulaires en Dopamine (DA), en Dopac (DO), en acide homovanillique (HVA), en sérotonine (5-HT), en acide 5-hydroxyindolacétique (5HIAA), et en noradrénaline (NA) ont été comparés à l'aide du test Wilcoxon, ce test non paramétrique permet de comparer deux échantillons appariés.

---

<sup>51</sup> Les **catécholamines** sont des composés organiques synthétisés à partir de la tyrosine et jouant le rôle d'hormone ou de neurotransmetteur.

<sup>52</sup> Les indolamines sont une famille de neurotransmetteurs présentant une fonction amine liée à un cycle d'indole

<sup>53</sup> Isocratique c'est l'élution au cours d'une séparation chromatographique en phase liquide, dans laquelle la composition du mélange des solvants constituant la phase éluante est gardée constante.

Les animaux sélectionnés pour l'analyse finale ont fait l'objet d'une étape de validation supplémentaire concernant l'étendue de la réduction du contenu du tissu en monoamines. Cette étape biochimique est obligatoire pour la validation des lésions qui permet l'inclusion finale des animaux dans chaque groupe.

- **Validation du modèle 6-OHDA**
  - **Striatum DL**

Les concentrations tissulaires des monoamines et leurs métabolites ont été mesurées dans différentes structures, les figures 41, 42 et 43, représentent le contenu en Dopamine, Dopac, HVA, NA, 5-HT et 5-HIAA au niveau du striatum DL.

Les résultats sont exprimés en ng / mg de tissu et chaque valeur correspond à la moyenne  $\pm$  SEM. Le pourcentage de diminution des monoamines chez les animaux du groupe 6-OHDA en comparant le coté lésé et le coté contrôle, est significatif pour la DA, le Dopac et le HVA, avec un pourcentage de diminution qui atteint 97%, 94% et 66% respectivement pour DA, Dopac et HVA.

Par contre aucune différence significative n'a été observée pour la NA et la 5-HT et la 5-HIAA. Le tableau 22, 23 et 24 résume les valeurs de concentration tissulaire des six monoamines au niveau du striatum DL, en utilisant le test de Wilcoxon-Mann-Whitney (ou test U de Mann-Whitney ou encore test de la somme des rangs de Wilcoxon) qui est un test statistique non paramétrique qui permet de tester l'hypothèse selon laquelle les médianes de chacun de deux groupes de données sont proches.

Son implémentation sous le logiciel R se fait avec le script : **wilcox.test** du package **stat**.

```
wilcox.test (x, y = NULL,  
             alternative = c ("two.sided", "less", "greater"),  
             mu = 0, paired = TRUE, exact = NULL, correct = TRUE,  
             conf.int = FALSE, conf.level = 0.95, ...)
```

Groupe 6-OHDA Striatum DL	Dopamine	
	Côté Droit	Côté Gauche
Moyenne	1725	9429
N d'échantillons	7	7
P Value	<0,0001	
Significativité	****	
SD	896,7	2511
SEM	338,90	949,14

Tableau 22: Résultat du test de Wilcoxon du groupe 6-OHDA - DL pour la dopamine

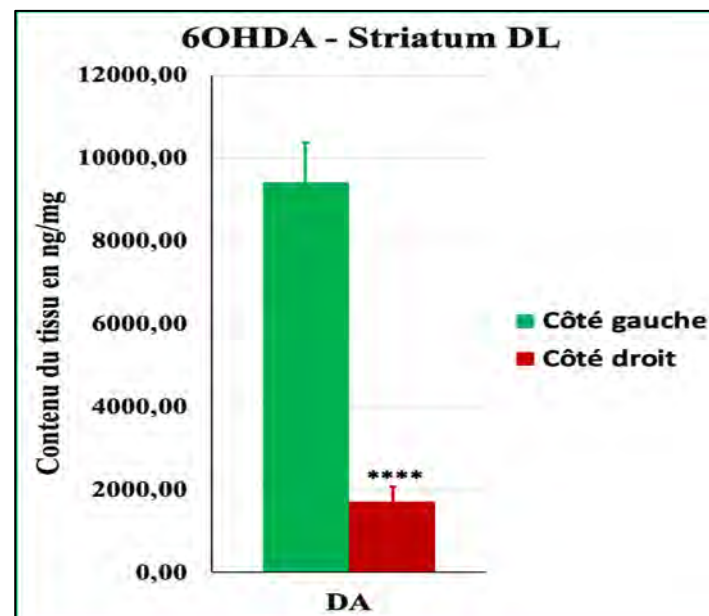


Figure 41: contenu du tissu Striatum - DL du groupe 6-OHDA en Dopamine

Groupe 6-OHDA Striatum DL	Dopac		HVA	
	Côté Droit	Côté Gauche	Côté Droit	Côté Gauche
Moyenne	238,4	1223,11	130,8	816,2
N d'échantillons	7	7	7	7
P Value	0,0005		<0,0001	
Significativité	****		****	
SD	91,86	550,58		
SEM	34,72	208,10		

Tableau 23: Résultat du test de Wilcoxon du groupe 6-OHDA - DL pour Dopac et HVA

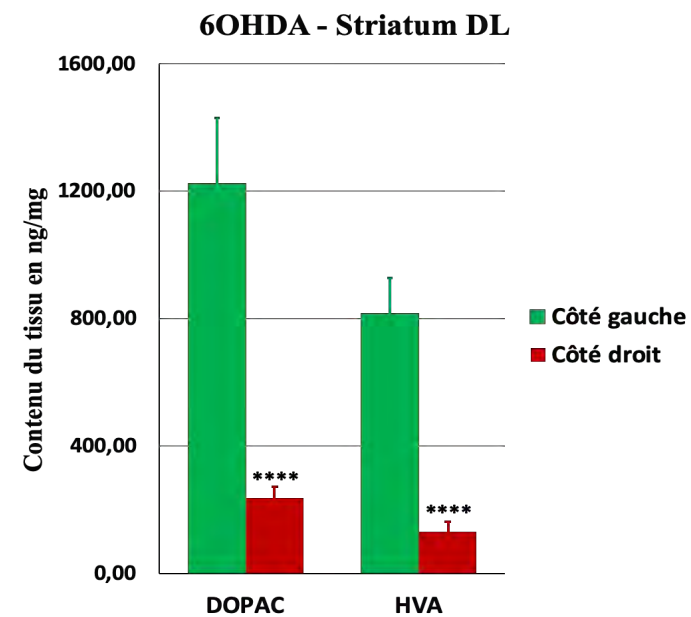


Figure 42: Contenu du tissu Striatum - DL du groupe 6-OHDA en Dopac et HVA

Groupe 6-OHDA Striatum DL	NA		5HT		5HIAA	
	Côté Droit	Côté Droit	Côté Droit	Côté Gauche	Côté Droit	Côté Gauche
Moyenne	136,6	140,4	45,87	56,30	406,8	402,5
N d'échantillons	7	7	7	7	7	7
P Value	0,85		0,2		0,95	
Significativité	ns		ns		ns	
SD	35,68	38,52	12,55	16,93	140,73	169,31
SEM	13,48	14,54	4,74	6,40	53,19	63,99

Tableau 24: Résultat du test de Wilcoxon du groupe 6-OHDA - DL pour NA, 5HIAA et 5HT

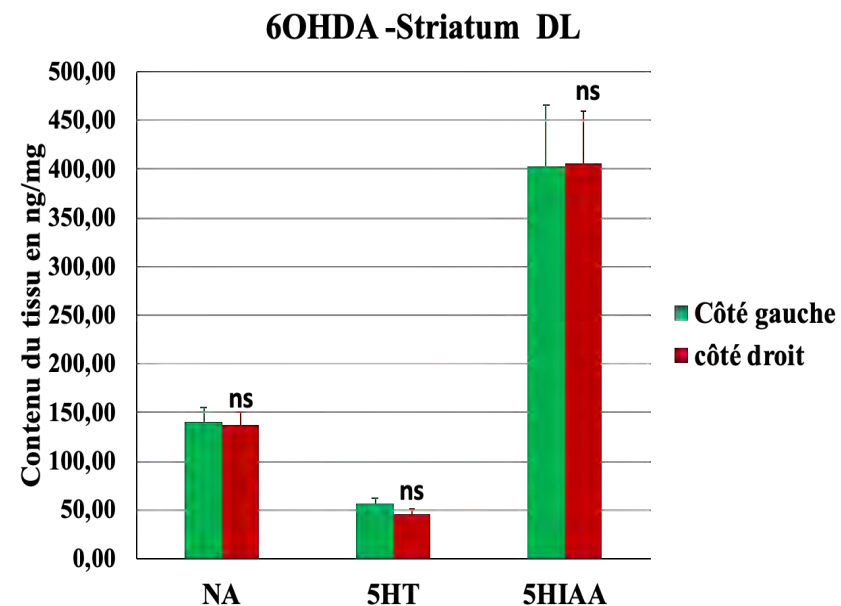


Figure 43: Contenu du tissu Striatum - DL du groupe 6-OHDA en NA, 5HT, 5HIAA

### ➤ **Striatum -VL**

Les figures 44,45 et 46, représentent le contenu en Dopamine, Dopac, HVA, NA, 5-HT et 5-HIAA au niveau du striatum VL.

Les résultats sont exprimés en ng / mg de tissu et chaque valeur correspond à la moyenne  $\pm$  SEM. Le pourcentage de diminution des monoamines chez les animaux du groupe 6-OHDA en comparant le coté lésé et le coté contrôle, est significatif pour la DA et le Dopac, avec un pourcentage de diminution qui atteint 67% et 64% respectivement pour la DA et le Dopac. Par contre aucune différence significative n'a été observée pour la HVA, NA et la 5-HT et la 5-HIAA.

Le tableau 25,26 et 27 résume les valeurs de concentration tissulaire des six monoamines au niveau du striatum VL, en utilisant le test de Wilcoxon-Mann-Whitney.



Groupe 6-OHDA Striatum VL	Dopamine	
	Côté Droit	Côté Gauche
Moyenne	1095,56	1840,79
N d'échantillons	7	7
P Value	0,0009	
Significativité	***	
SD	654,57	629,76
SEM	267,23	257,10

Tableau 25: Résultat du test de Wilcoxon du groupe 6-OHDA - VL pour Dopamine

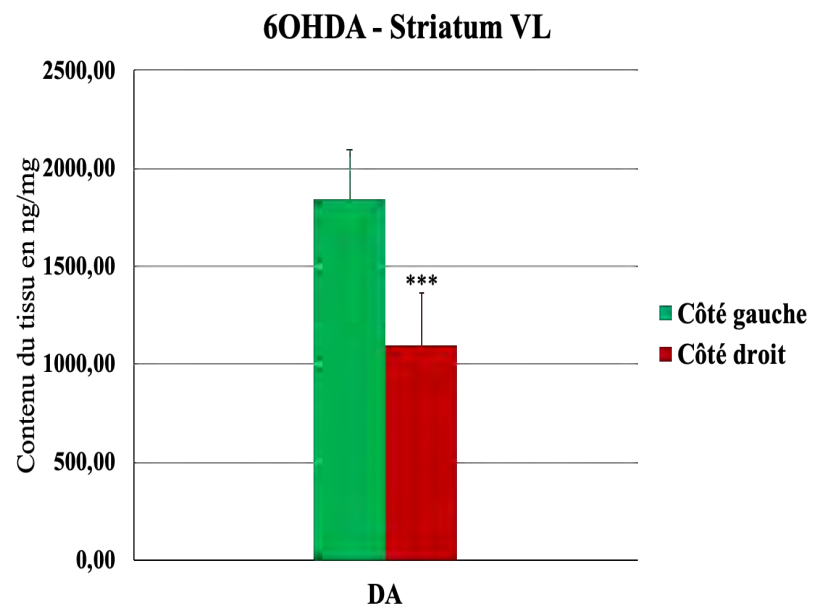


Figure 44: Contenu du tissu Striatum - VL du groupe 6-OHDA en Dopamine

Groupe 6-OHDA Striatum VL	Dopac		HVA	
	Côté Droit	Côté Gauche	Côté Droit	Côté Gauche
Moyenne	145,48	221,04	120,52	154,33
N d'échantillons	7	7	7	7
P Value	0,0488		0,05	
Significativité	*		ns	
SD	90,01	77,90	46,54	56,76
SEM	40,26	34,84	19	23,17

Tableau 26: Résultat du test de Wilcoxon du groupe 6-OHDA - VL pour Dopac et HVA

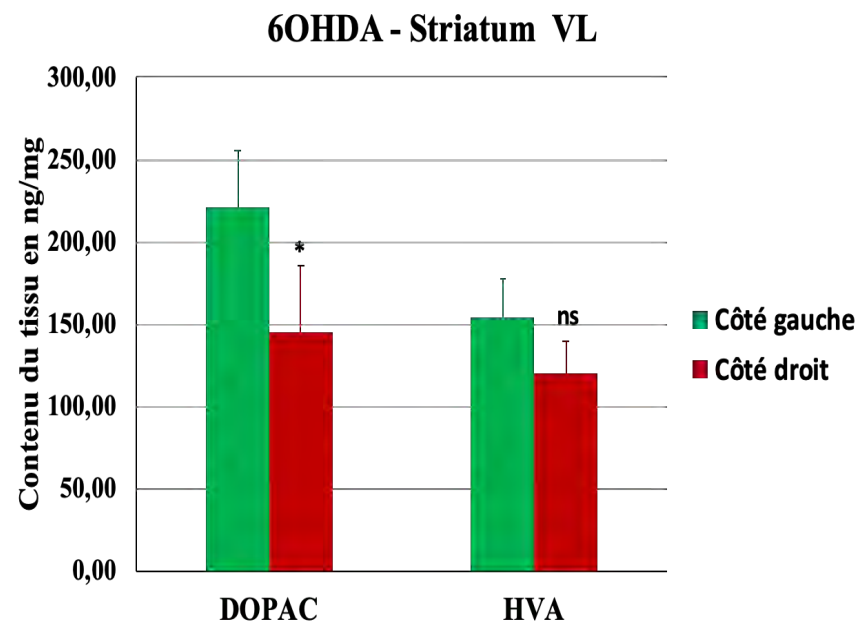


Figure 45: Contenu du tissu Striatum - VL du groupe 6-OHDA en Dopac et HVA

Groupe 6-OHDA Striatum VL	NA		5HT		5HIAA	
	Côté Droit	Côté Droit	Côté Droit	Côté Gauche	Côté Droit	Côté Gauche
Moyenne	144,61	111,50	43,68	57,44	69,31	84,02
N d'échantillons	7	7	7	7	7	7
P Value	0,43		0,19		0,29	
Significativité	ns		ns		ns	
SD	78,06	42,46	16,43	25,27	14,17	33,78
SEM	31,87	17,33	6,71	10,32	5,78	13,79

Tableau 27: sultat du test de Wilcoxon du groupe 6-OHDA - VL pour NA, 5HT et 5HIAA

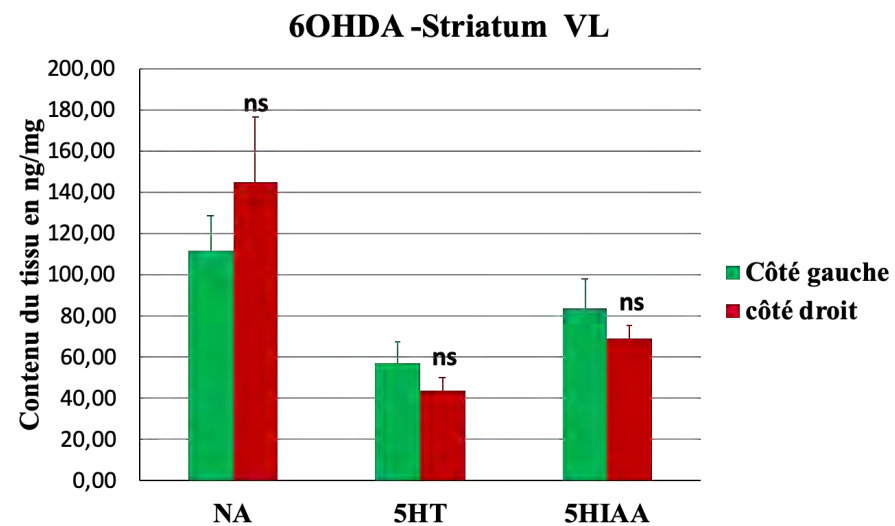


Figure 46: Contenu du tissu Striatum - VL du groupe 6-OHDA en NA, 5HT et 5HIAA

➤ **Cortex Frontal (M2)**

Les figures 47 représentent le contenu en NA et 5-HT au niveau du cortex frontal.

Les résultats sont exprimés en ng / mg de tissu et chaque valeur correspond à la moyenne  $\pm$  SEM. Le pourcentage de diminution des monoamines chez les animaux du groupe 6-OHDA en comparant le coté lésé et le coté contrôle, est significatif pour la NA et 5HT, avec un pourcentage de diminution qui atteint 47% et 39% respectivement pour la NA et 5HT.

Le tableau 28 résume les valeurs de concentration tissulaire des six monoamines au niveau du striatum VL, en utilisant le test de Wilcoxon-Mann-Whitney.

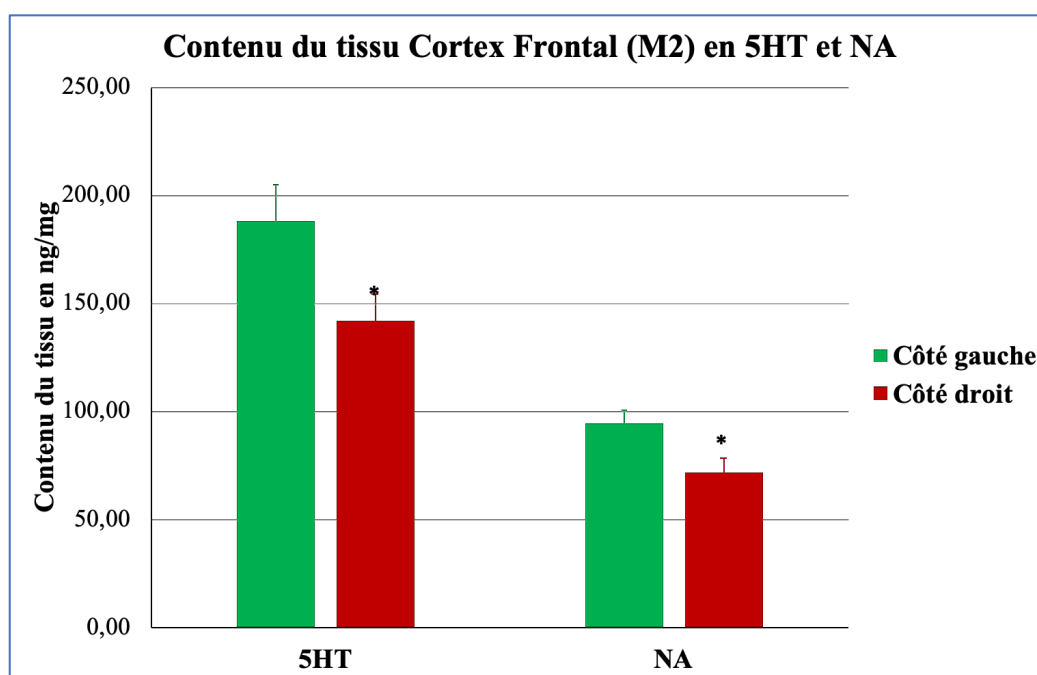


Figure 47:Contenu du tissu Cortex Frontale en 5HT et NA

Groupe 6-OHDA Cortex Frontal (M2)	5HT		NA	
	Côté Droit	Côté Gauche	Côté Droit	Côté Gauche
<b>Moyenne</b>	142,15	188,33	71,69	94,38
<b>N d'échantillons</b>	7	7	7	7
<b>P Value</b>	0,0448		0,0233	
<b>Significativité</b>	*		*	
<b>SD</b>	32,64	43,70	17,08	15,56
<b>SEM</b>	12,33	16,51	6,45	5,88

Tableau 28:Résultat du test de Wilcoxon du groupe 6-OHDA - Cortex Frontal pour 5HT et NA

- **Validation du modèle Alpha-Syn**

- **Striatum DL**

Les concentrations tissulaires des monoamines et leurs métabolites ont été mesurées dans différentes structures, les figures 48, 49 et 50, représentent le contenu en Dopamine, Dopac, HVA, NA, 5-HT et 5-HIAA au niveau du striatum D pour le groupe Alpha-Syn.

Les résultats sont exprimés en ng / mg de tissu et chaque valeur correspond à la moyenne  $\pm$  SEM. Le pourcentage de diminution des monoamines chez les animaux du groupe alpha-syn en comparant le coté lésé et le coté contrôle, est significatif pour la DA, Dopac, HVA, 5HT et 5HIAA, avec un pourcentage de diminution qui atteint 71%, 68%, 66%, 56% et 59% respectivement pour DA, Dopac, HVA, 5HT et 5HIAA. Par contre aucune différence significative n'a été observée pour la NA (Tableau 29, 30 et 31).

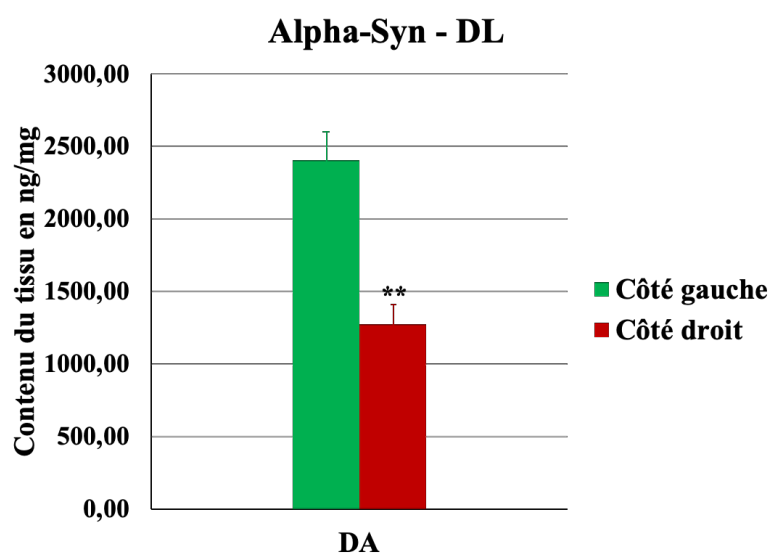


Figure 48:Contenu du tissu Striatum DL en Dopamine pour le groupe Alpha-Syn

Groupe Alpha-Syn Striatum DL	Dopamine	
	Côté Droit	Côté Gauche
Moyenne	1277,91	2407,50
N d'échantillons	8	8
P Value	0,001	
Significativité	**	
SD	387,81	549,72
SEM	137,11	194,36

Tableau 29:Résultat du test de Wilcoxon du groupe Alpha-Syn - Striatum DL pour Dopamine

Groupe Alpha-Syn Striatum DL	Dopac		HVA	
	Côté Droit	Côté Gauche	Côté Droit	Côté Gauche
Moyenne	63,03	168,32	139,28	308,60
N d'échantillons	8	8	8	8
P Value	0,004		0,001	
Significativité	*		**	
SD	48,92	107,38	36,79	84,24
SEM	17,30	37,96	13,01	29,78

Tableau 30: Résultat du test de Wilcoxon du groupe Alpha-Syn - Striatum DL pour Dopac et HVA

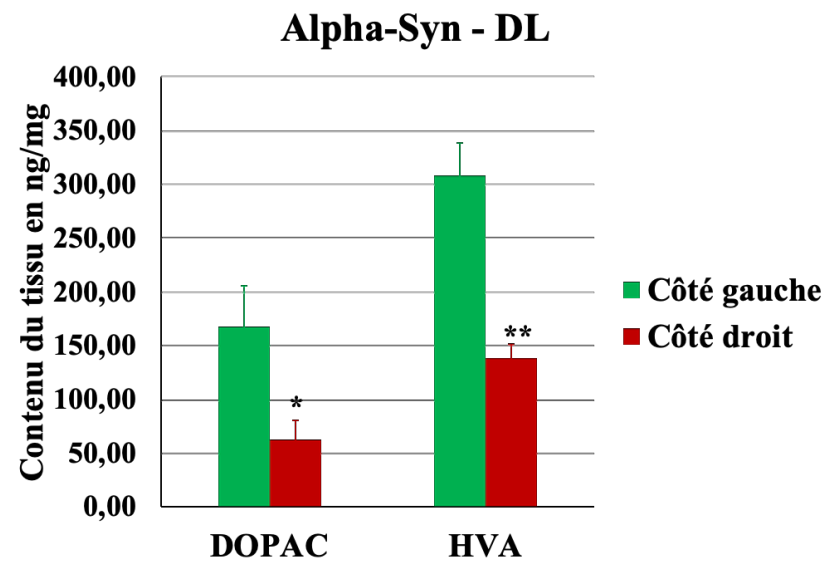
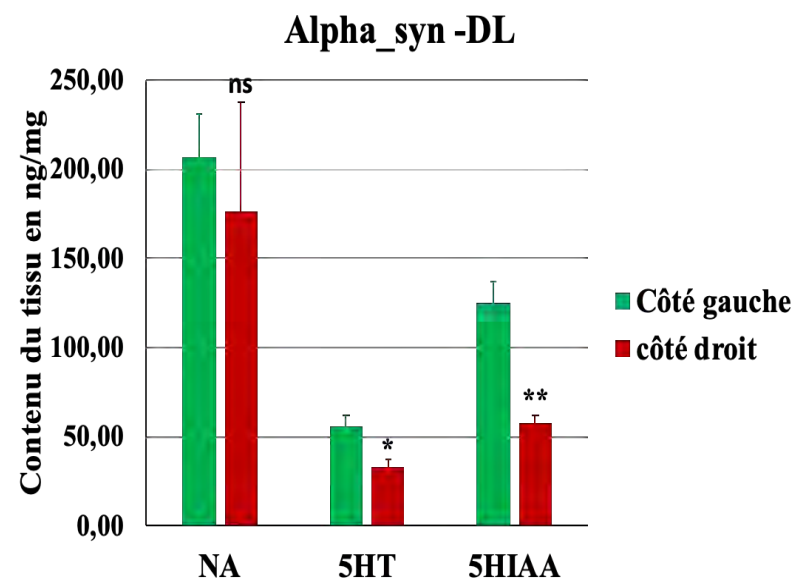


Figure 49: Contenu du tissu Striatum DL en Dopac et HVA pour le groupe Alpha-Syn

Groupe Alpha-Syn Striatum DL	NA		5HT		5HIAA	
	Côté Droit	Côté Gauche	Côté Droit	Côté Gauche	Côté Droit	Côté Gauche
Moyenne	176,62	206,85	33,30	56,28	125,42	52,06
N d'échantillons	8	8	8	8	8	8
P Value	0,67		0,01		0,001	
Significativité	Ns		*		**	
SD	172,87	69,11	11,78	15,92	13,65	34,31
SEM	61,12	24,43	4,16	5,63	4,83	12,13

**Tableau 31: Résultat du test de Wilcoxon du groupe Alpha-Syn - Striatum DL pour NA, 5HT et 5HIAA**



**Figure 50: Contenu du tissu Striatum DL en NA, 5HT et 5HIAA pour le groupe Alpha-Syn**

▪ **Striatum VL**

Les concentrations tissulaires des monoamines et leurs métabolites ont été mesurées dans différentes structures, les figures 51, 52 et 53 représentent le contenu en Dopamine, Dopac, HVA, NA, 5-HT et 5-HIAA au niveau du striatum D pour le groupe Alpha-Syn.

Les résultats sont exprimés en ng / mg de tissu et chaque valeur correspond à la moyenne  $\pm$  SEM. Le pourcentage de diminution des monoamines chez les animaux du groupe Alpha-Syn en comparant le coté lésé et le coté contrôle, est significatif pour la DA, HVA, 5HT et 5HIAA, avec un pourcentage de diminution qui atteint 56%, 40%, 60% et 62% respectivement pour DA, HVA, 5HT et 5HIAA. Par contre aucune différence significative n'a été observée pour le Dopac et la NA (Tableau 32, 33 et 34).

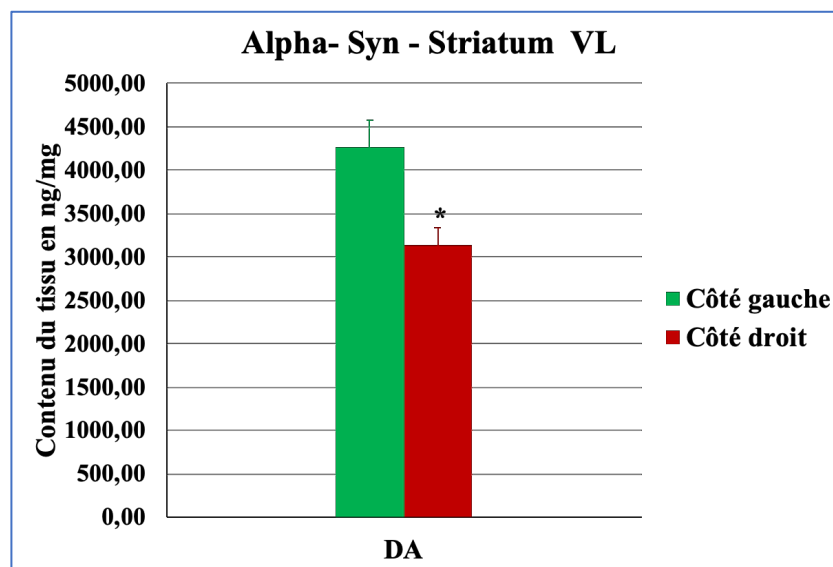


Figure 51:Contenu du tissu Striatum VL - Dopamine pour le groupe Alpha-Syn

Groupe Alpha-Syn Striatum VL	Dopamine	
	Côté Droit	Côté Gauche
Moyenne	3136,49	4271,64
N d'échantillons	8	8
P Value	0,01	
Significativité	*	
SD	590,92	861,97
SEM	208,92	304,75

Tableau 32:Résultat du test de Wilcoxon du groupe Alpha-Syn - Striatum VL pour Dopamine



Groupe Alpha-Syn Striatum VL	Dopac		HVA	
	Côté Droit	Côté Gauche	Côté Droit	Côté Gauche
Moyenne	653,64	756,15	341,19	446,82
N d'échantillons	8	8	8	8
P Value	0,54		0,02	
Significativité	ns		*	
SD	262,72	340,41	74,68	63,53

Tableau 33: Résultat du test de Wilcoxon du groupe Alpha-Syn - Striatum VL pour Dopac et HVA

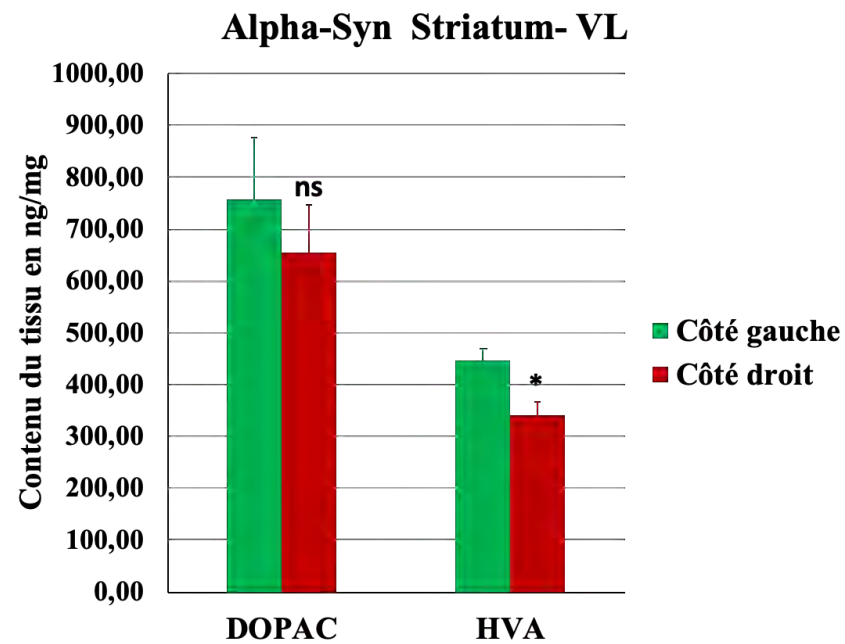


Figure 52: Contenu du tissu Striatum VL - Dopac et HVA pour le groupe Alpha-Syn

Groupe Alpha-Syn Striatum VL	NA		5HT		5HIAA	
	Côté Droit	Côté Gauche	Côté Droit	Côté Gauche	Côté Droit	Côté Gauche
Moyenne	228,83	213,34	87,88	154,51	165,60	28,84
N d'échantillons	8	8	8	8	8	8
P Value	0,5		0,02		0,03	
Significativité	ns		*		*	
SD	155,32	60,70	26,31	78,53	31,06	74,83
SEM	54,91	21,46	9,30	27,77	10,98	26,49

Tableau 34: Résultat du test de Wilcoxon du groupe Alpha-Syn - Striatum VL pour NA, 5HT et 5HIAA

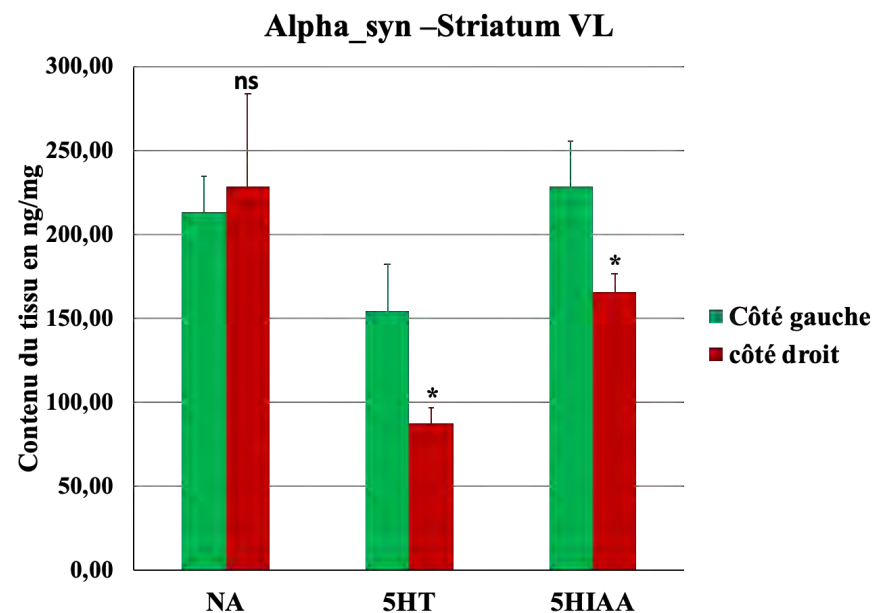


Figure 53: Contenu du tissu Striatum VL - NA, 5HT et 5HIAA pour le groupe Alpha-Syn

## II. Analyse biomoléculaire

Pour la partie d'analyse biomoléculaire, deux structures ont été utilisées : le striatum et le cortex de chaque animal pour les différents groupes.

### 1. Extraction d'ARN total

#### • Protocole expérimental

L'ARN total a été extrait en utilisant le kit *Direct-Zol RNA MicroPrep* (« Direct-Zol RNA Microprep Kits » 2019), ce protocole comprend deux parties : Préparation des échantillons et Purification de l'ARN.

- i. Préparation des échantillons : Mettre le tissu ( $\leq 5$  mg) dans 300  $\mu$ l de TRI Reagent® et homogénéiser en utilisant le FastPrep®-24<sup>54</sup>, qui lyse minutieusement et rapidement tous les tissus et cellules et permet ainsi une isolation facile et reproductible de l'ARN stable, des protéines actives et de l'ADN génomique complet. Pour éliminer les débris de particules du tissu (échantillon) homogénéisé, centrifuger et transférer le surnageant dans un tube RNase-free.
- ii. Purification de l'ARN : Ajouter un volume égal d'éthanol (95-100%) à un échantillon lysé dans du TRI Reagent®, transférer le mélange dans une colonne de colonne Zymo-Spin™ IC2 dans un tube collecteur et centrifuger à 10 000 -16 000 g pendant 30 secondes, ajouter 400  $\mu$ l d'ARN PreWash Direct-zol™ à la colonne et centrifuger, ajouter 700  $\mu$ l de tampon de lavage RNA à la colonne et centrifuger pendant 2 minutes pour éliminer complètement le tampon de lavage, transférer la colonne avec précaution dans un tube RNase-free et à la fin pour éluer l'ARN, ajouter 15  $\mu$ l d'eau DNase / RNase-free directement dans la colonne et centrifuger.

L'extraction a été faite en suivant les instructions du fabricant (« Direct-Zol RNA Microprep Kits » 2019), ce qui a permis d'enrichir les micro ARN (*miRNeasy Micro Kit, Qiagen*).

---

<sup>54</sup> L'instrument FastPrep®-24 est un homogénéisateur de table à haute vitesse offrant la vitesse et la performance ultimes pour la lyse d'échantillons biologiques. L'homogénéisation d'un maximum de 24 échantillons a lieu en moins de 40 secondes. Utilise un mouvement unique et optimisé pour perturber les cellules par le battement simultané multidirectionnel de billes de matrice de lyse spécialisées sur le matériau de l'échantillon. Développé pour les échantillons difficiles et résistants.

### • Qualité et quantité d'ARN

La quantité d'ARN purifié a été déterminée à l'aide d'un spectrophotomètre *Nanodrop-one* (Ozyme). Ce spectromètre rapporte la concentration en acides nucléiques et le rapport d'absorbance (A260/A280). La concentration est calculée en se basant sur l'équation de la Loi de Beer modifiée qui utilise la valeur d'absorbance corrigée des acides nucléiques. Ainsi, le rapport de pureté **A260/A280** est le rapport de l'absorbance corrigée à 260 nm sur l'absorbance corrigée à 280 nm. Un rapport de pureté A260/A280 au-dessus de 1,8 est normalement accepté comme " pur " pour l'ARN.

Le rapport de pureté des échantillons est compris entre 1,86 et 2,08 ce qui est considéré comme " pur ", sauf l'échantillon 70StrD dont le rapport A260/A280 est 1,72.

Nom échantillon	Volume élution (µl)	Concentration nanodrop (ng/ µl)	A260/A280
73CorD	15	32	1,95
73 CorG	15	40,4	1,94
73 StrD	15	105	1,96
73 StrG	15	81,6	1,94
75 CorD	15	108	1,93
75 CorG	15	102	1,92
75 StrD	15	65,6	1,92
75 StrG	15	34,8	1,93
70CorD	15	45	2,01
70 CorG	15	27,7	1,96
70 StrD	15	12,7	1,72
70 StrG	15	24	1,99
72CorD	15	35,9	2
72 CorG	15	31,4	2,03
72 StrD	15	13,7	2,04
72 StrG	15	24,7	2,04
76 CorD	15	45,4	1,93
76 CorG	15	54,6	2,02
76 StrD	15	49,1	1,97
76 StrG	15	46,9	1,94
78 CorD	15	36,4	1,93
78 CorG	15	61,8	2,02

78 StrD	15	28,3	1,97
78 StrG	15	42,4	1,94
80 CorD	15	18,9	1,98
80 CorG	15	26,5	2
80 StrD	15	12,8	1,83
80 StrG	15	17,4	2,06
82 CorD	15	14,8	2,08
82 CorG	15	14,8	2,08
82 StrD	15	22,9	2,03
82 StrG	15	19,1	2,05
84 CorD	15	27,2	2
84 CorG	15	63,3	1,87
84 StrD	15	63,5	1,87
84 StrG	15	20,1	2
87 CorD	15	20,1	2,03
87 CorG	15	13,9	1,93
87 StrD	15	28,6	2,01
87 StrG	15	26,1	2,06
88 CorD	15	31,9	2
88 CorG	15	11,1	1,94
88 StrD	15	33,2	2,02
88 StrG	15	29,9	2
83 CorD	15	79,6	1,86
83 CorG	15	115	1,99
83 StrD	15	88	1,98
83 StrG	15	74,4	1,97
89 CorD	15	104	1,97
89 CorG	15	101	1,97
89 StrD	15	49,6	1,93
89 StrG	15	76,8	1,97
81 CorD	15	46,5	1,93
81 CorG	15	57,3	2,01
81 StrD	15	32,1	2,06
81 StrG	15	27,9	2
85 CorD	15	58,7	1,99
85 CorG	15	92,4	1,98
85 StrD	15	54,9	2
85 StrG	15	59,2	1,86

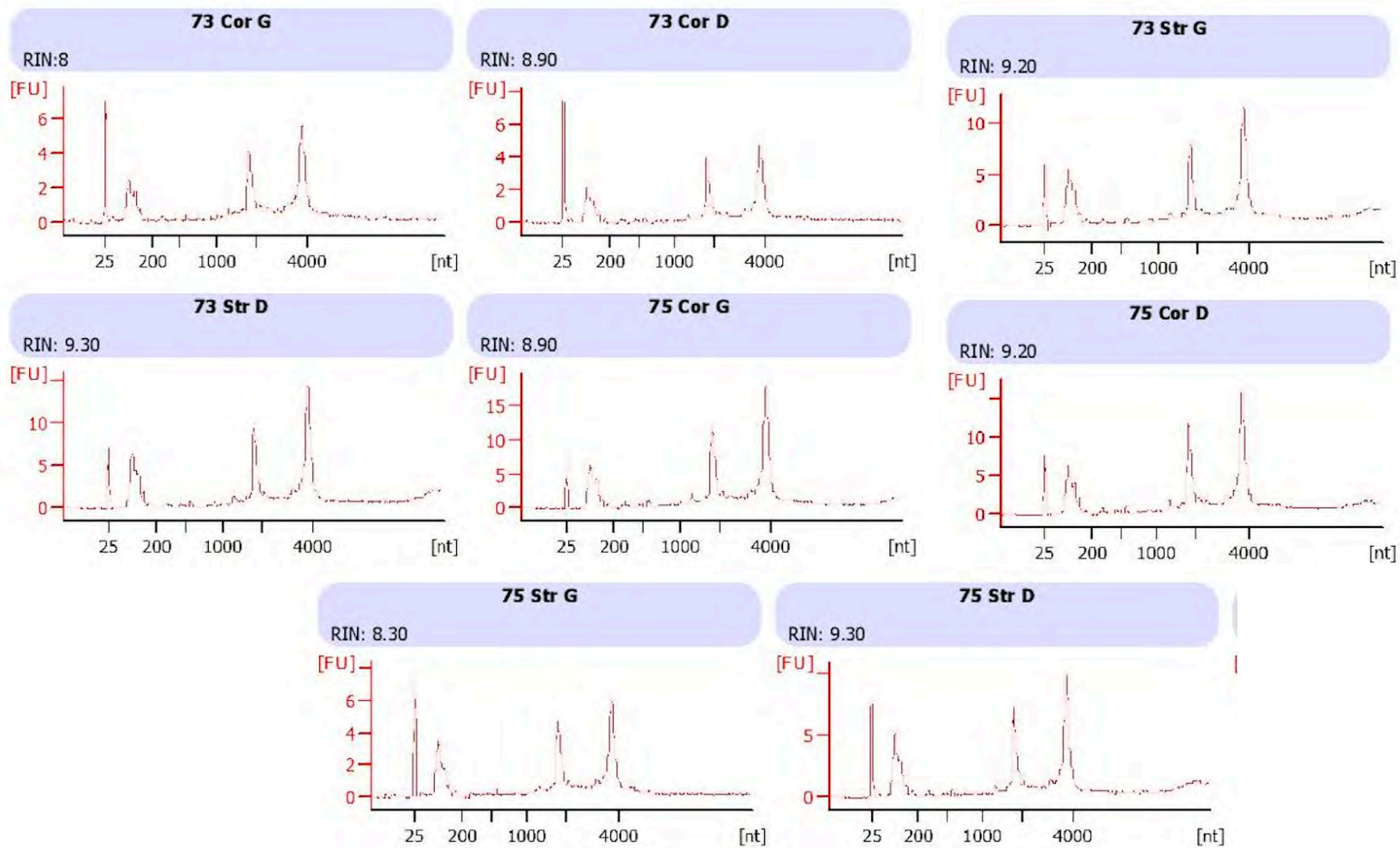
**Tableau 35:La concentration et le rapport de pureté obtenus par le Nanodrop-one**

*CorD = Cortex Droit ; CorG = Cortex Gauche ; StrD = Striatum Droit ; StrG = Striatum Gauche*

Le bioanalyzer Agilent 2100 avec une puce RNA (Agilent RNA 600 Nano kit) permet d'évaluer la qualité et l'intégrité des échantillons d'ARN total en calculant le numéro d'intégrité de l'ARN (RIN) qui est déterminé en se basant sur l'ensemble du tracé électrophorétique de l'échantillon d'ARN, y compris la présence ou non de produits de dégradation. Le RIN permet à la fois d'évaluer l'intégrité de l'ARN et comparer les échantillons (Schroeder et al., 2006).

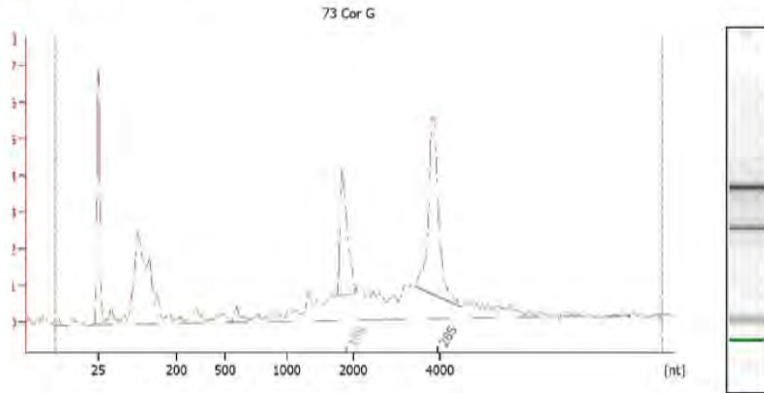
Le contrôle qualité des échantillons d'ARN se base sur trois critères fournis par le Bioanalyzer: **Ratio 28S/18S**, la valeur du **RIN** et le **profil électrophorétique**.

Les échantillons séquencés ont été choisis par rapport au pourcentage de perte de dopamine au niveau du striatum dors latéral. Nous avons sélectionné deux animaux par groupe



**Figure 54: pitulatif du résultat du Bioanalyzer pour le groupe 6OHDA**

**Electropherogram Summary**



**Overall Results for sample 1 : 73 Cor G**

RNA Area: 56,6      RNA Integrity Number (RIN): 8 (8.02.08)  
 RNA Concentration: 39 ng/µl      Result Flagging Color:    
 rRNA Ratio (28s / 18s): 1,8      Result Flagging Label: RIN:8

**Fragment table for sample 1 : 73 Cor G**

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 742	2 043	5,5	9,7
28S	3 420	4 460	9,6	16,9

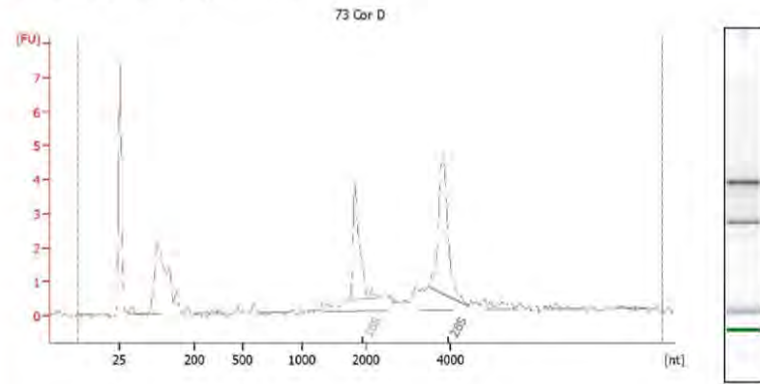
Ratio 28s/18s = 1,8 ≥ 1,5 => Bon

RIN = 8 > 7 => Bon

Groupe 60HDA

Animal 73 Cortex Gauche

**Electropherogram Summary Continued ...**



**Overall Results for sample 2 : 73 Cor D**

RNA Area: 35,9      RNA Integrity Number (RIN): 8,9 (8.02.08)  
 RNA Concentration: 25 ng/µl      Result Flagging Color:    
 rRNA Ratio (28s / 18s): 1,5      Result Flagging Label: RIN: 8,90

**Fragment table for sample 2 : 73 Cor D**

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 734	2 269	5,3	14,8
28S	3 489	4 442	7,7	21,5

Ratio 28s/18s = 1,5 ≥ 1,5 => Bon

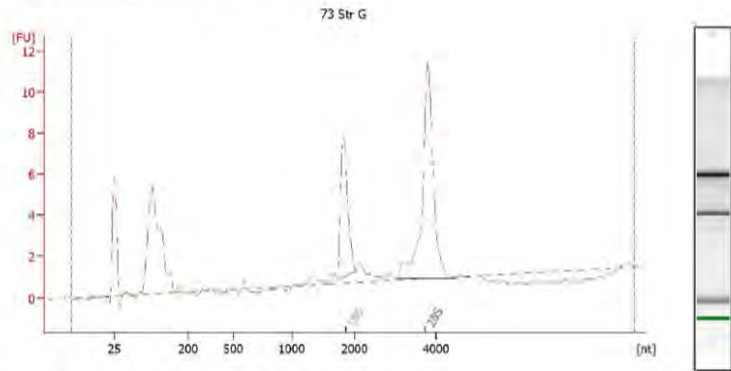
RIN = 8,90 > 7 => Bon

Groupe 60HDA

Animal 73 Cortex Droite



Electropherogram Summary Continued ...



Overall Results for sample 3 : **73 Str G**

RNA Area: 68,2 RNA Integrity Number (RIN): 9,2 (B.02.08, Anomaly Threshold(s) manually adapted)  
 RNA Concentration: 48 ng/µl  
 rRNA Ratio [28s / 18s]: 2,1 Result Flagging Color:  Result Flagging Label: RIN: 9,20

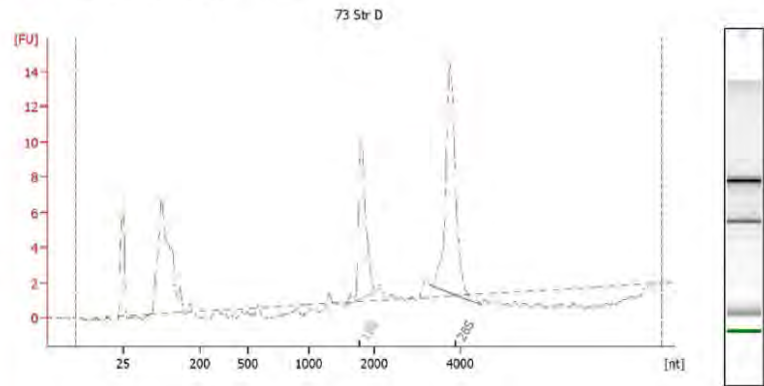
Fragment table for sample 3 : **73 Str G**

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 723	2 036	11,1	16,2
28S	2 993	4 464	23,4	34,4

Ratio 28s/18s = 2,1 ≥ 1,5 => Bon  
 RIN = 9,20 > 7 => Bon

Groupe 60HDA  
 Animal 73 Striatum Gauche

Electropherogram Summary Continued ...



Overall Results for sample 4 : **73 Str D**

RNA Area: 72,9 RNA Integrity Number (RIN): 9,3 (B.02.08)  
 RNA Concentration: 51 ng/µl Result Flagging Color:   
 rRNA Ratio [28s / 18s]: 1,8 Result Flagging Label: RIN: 9,30

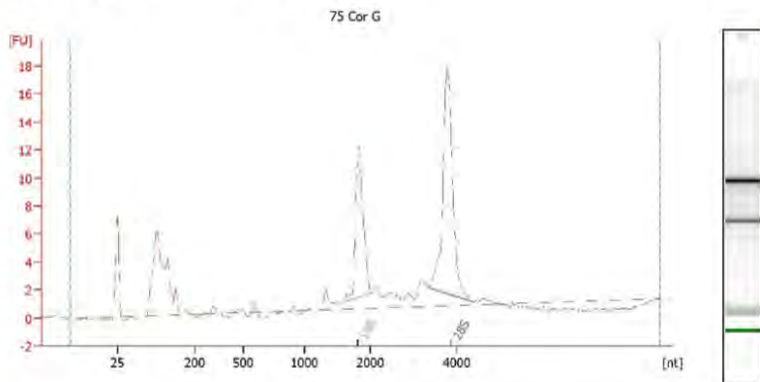
Fragment table for sample 4 : **73 Str D**

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 537	2 040	13,7	18,8
28S	3 308	4 507	24,2	33,2

Ratio 28s/18s = 1,8 ≥ 1,5 => Bon  
 RIN = 9,30 > 7 => Bon

Groupe 60HDA  
 Animal 73 Striatum Droite

Electropherogram Summary Continued ...



Overall Results for sample 5 : 75 Cor G

RNA Area: 102,7 RNA Integrity Number (RIN): 8,9 (8,02,08)  
 RNA Concentration: 72 ng/μl Result Flagging Color:    
 rRNA Ratio [28s / 18s]: 1,7 Result Flagging Label: RIN: 8,90

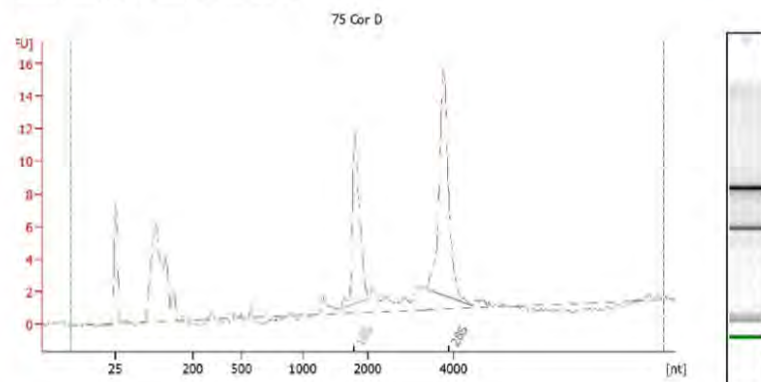
Fragment table for sample 5 : 75 Cor G

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 591	2 040	17,4	17,0
28S	3 349	4 397	29,8	29,1

Ratio 28s/18s = 1,7 ≥ 1,5 => Bon  
 RIN = 8,90 > 7 => Bon

Groupe 60HDA  
 Animal 75 Cortex Gauche

Electropherogram Summary Continued ...



Overall Results for sample 6 : 75 Cor D

RNA Area: 95,3 RNA Integrity Number (RIN): 9,2 (8,02,08)  
 RNA Concentration: 67 ng/μl Result Flagging Color:    
 rRNA Ratio [28s / 18s]: 1,5 Result Flagging Label: RIN: 9,20

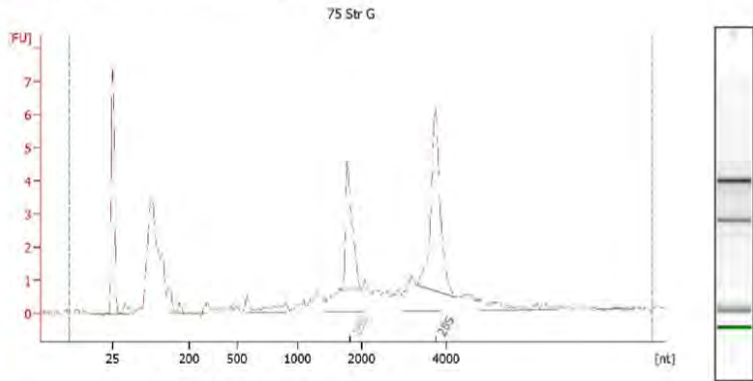
Fragment table for sample 6 : 75 Cor D

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 575	2 021	16,7	17,5
28S	3 370	4 412	25,5	26,8

Ratio 28s/18s = 1,5 ≥ 1,5 => Bon  
 RIN = 9,20 > 7 => Bon

Groupe 60HDA  
 Animal 75 Cortex Droite

Electropherogram Summary Continued ...



Overall Results for sample 7 : 75 Str G

RNA Area: 58,2      RNA Integrity Number (RIN): 8,3 (B.02.08)  
 RNA Concentration: 41 ng/µl      Result Flagging Color:   
 rRNA Ratio [28s / 18s]: 1,8      Result Flagging Label: RIN: 8,30

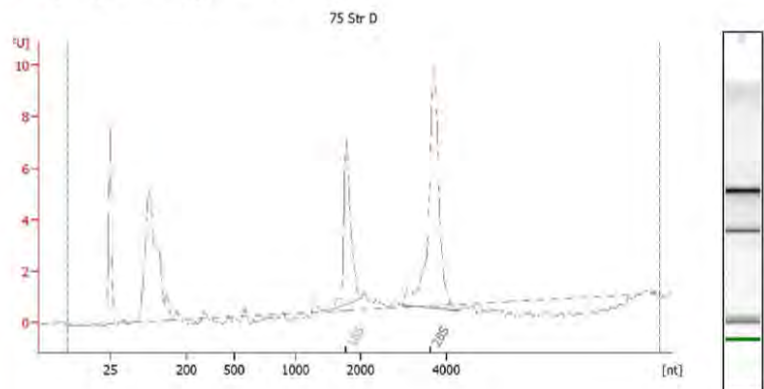
Fragment table for sample 7 : 75 Str G

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 657	1 995	5,8	9,9
28S	3 300	4 204	10,3	17,7

Ratio 28s/18s = 1,8 ≥ 1,5 => Bon  
 RIN = 8,30 > 7 => Bon

Groupe 60HDA  
 Animal 75 Striatum Gauche

Electropherogram Summary Continued ...



Overall Results for sample 8 : 75 Str D

RNA Area: 56,8      RNA Integrity Number (RIN): 9,3 (B.02.08)  
 RNA Concentration: 40 ng/µl      Result Flagging Color:   
 rRNA Ratio [28s / 18s]: 2,0      Result Flagging Label: RIN: 9,30

Fragment table for sample 8 : 75 Str D

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 570	1 984	9,8	17,3
28S	2 955	4 298	19,8	34,9

Ratio 28s/18s = 2,0 ≥ 1,5 => Bon  
 RIN = 9,30 > 7 => Bon

Groupe 60HDA  
 Animal 75 Striatum Droite

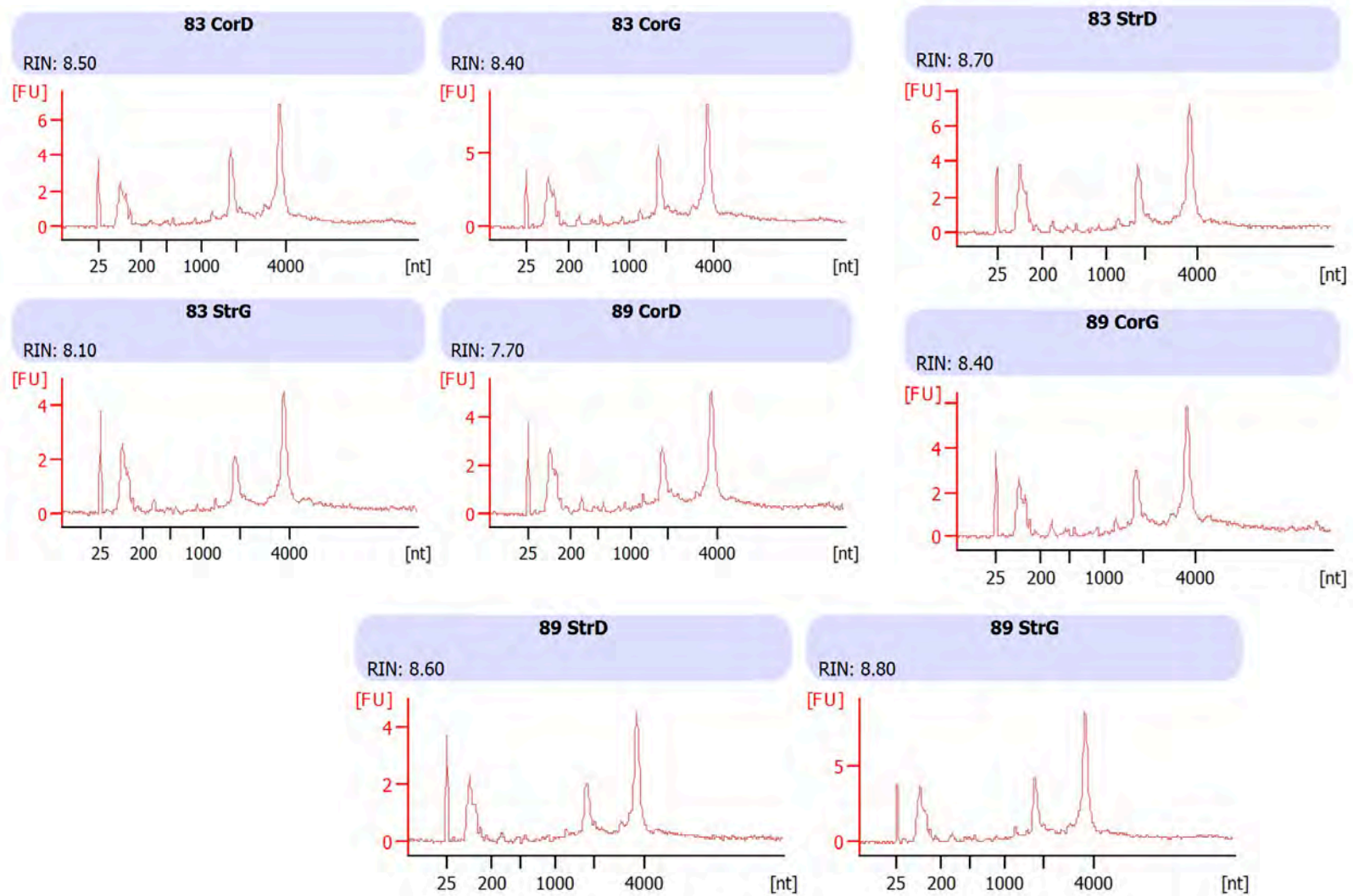
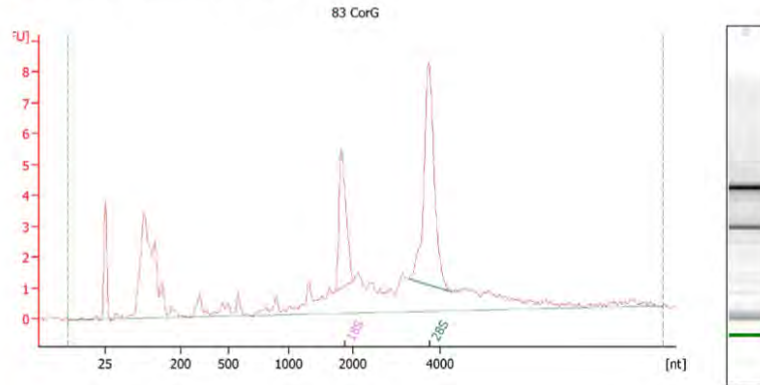


Figure 55:Récapitulatif du résultat du Bioanalyser pour le groupe Alpha\_Syn

Electropherogram Summary Continued ...



Overall Results for sample 2 : **83 CorG**

RNA Area: 77,4 RNA Integrity Number (RIN): 8,4 (B.02.08)  
 RNA Concentration: 79 ng/µl Result Flagging Color:    
 RNA Ratio [28s / 18s]: 2,0 Result Flagging Label: RIN: 8.40

Fragment table for sample 2 : **83 CorG**

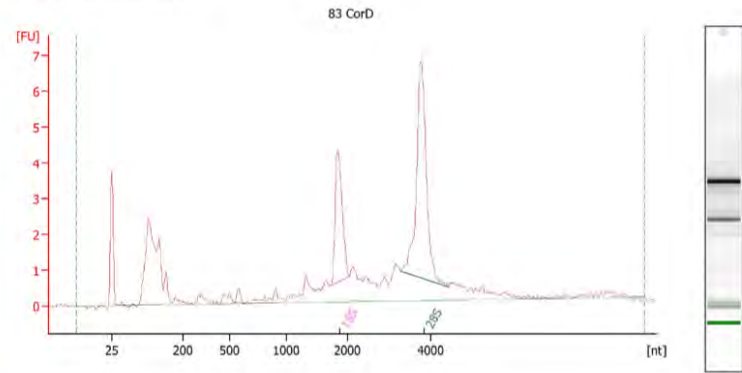
Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 721	2 014	7,3	9,5
28S	3 281	4 258	14,4	18,7

Ratio 28s/18s = 2 ≥ 1,5 => Bon

RIN = 8,40 > 7 => Bon

Groupe α-syn  
 Animal 83 Cortex Gauche

Electropherogram Summary



Overall Results for sample 1 : **83 CorD**

RNA Area: 54,1 RNA Integrity Number (RIN): 8,5 (B.02.08)  
 RNA Concentration: 55 ng/µl Result Flagging Color:    
 rRNA Ratio [28s / 18s]: 1,9 Result Flagging Label: RIN: 8.50

Fragment table for sample 1 : **83 CorD**

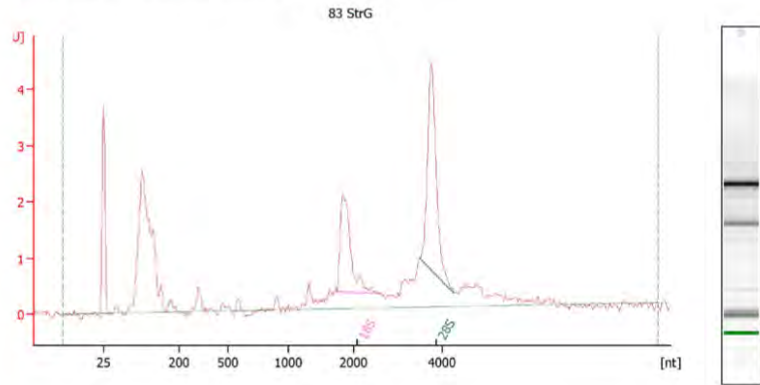
Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 728	2 027	6,3	11,6
28S	3 268	4 427	12,1	22,4

Ratio 28s/18s = 1,9 ≥ 1,5 => Bon

RIN = 8,50 > 7 => Bon

Groupe α-syn  
 Animal 83 Cortex Droite

Electropherogram Summary Continued ...



Overall Results for sample 4 : **83 StrG**

RNA Area: 36,1 RNA Integrity Number (RIN): 8,1 (B.02.08)  
 RNA Concentration: 37 ng/μl Result Flagging Color:    
 RNA Ratio [28s / 18s]: 1,5 Result Flagging Label: RIN: 8.10

Fragment table for sample 4 : **83 StrG**

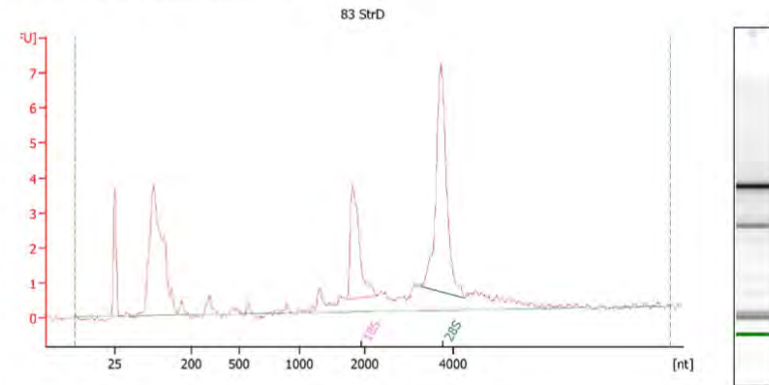
Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 737	2 510	4,1	11,4
28S	3 504	4 266	6,2	17,2

Ratio 28s/18s = 1,5 ≥ 1,5 => Bon

RIN = 8,10 > 7 => Bon

Groupe α-syn  
 Animal 83 Striatum Gauche

Electropherogram Summary Continued ...



Overall Results for sample 3 : **83 StrD**

RNA Area: 55,7 RNA Integrity Number (RIN): 8,7 (B.02.08)  
 RNA Concentration: 57 ng/μl Result Flagging Color:    
 RNA Ratio [28s / 18s]: 2,0 Result Flagging Label: RIN: 8.70

Fragment table for sample 3 : **83 StrD**

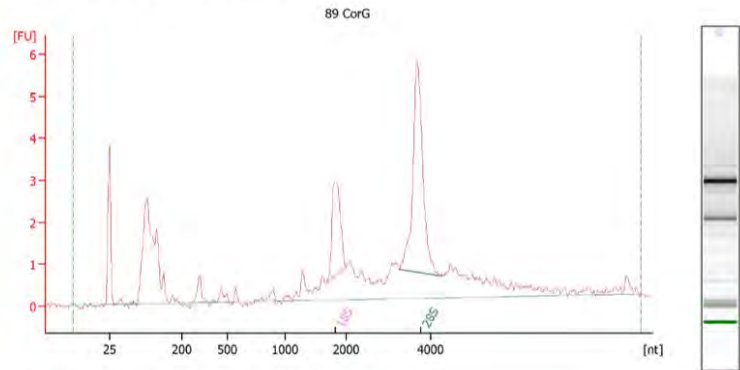
Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 717	2 265	6,7	12,0
28S	3 260	4 256	13,0	23,4

Ratio 28s/18s = 2,0 ≥ 1,5 => Bon

RIN = 8,70 > 7 => Bon

Groupe α-syn  
 Animal 83 Striatum Droite

Electropherogram Summary Continued ...



Overall Results for sample 6 : **89 CorG**

RNA Area: 52,8      RNA Integrity Number (RIN): 8.4 (B.02.08)  
 RNA Concentration: 54 ng/μl      Result Flagging Color:    
 rRNA Ratio [28s / 18s]: 2.2      Result Flagging Label: RIN: 8.40

Fragment table for sample 6 : **89 CorG**

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 657	1 983	4,5	8,6
28S	3 265	4 295	10,1	19,2

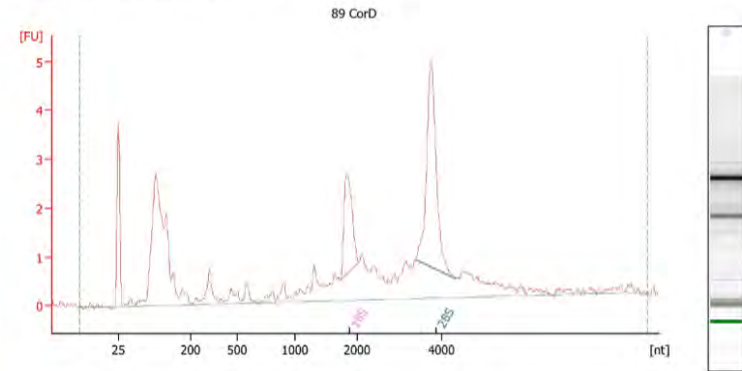
Ratio 28s/18s = 2,2 ≥ 1,5 => Bon

RIN = 8,40 > 7 => Bon

Groupe α-syn

Animal 89 Cortex Gauche

Electropherogram Summary Continued ...



Overall Results for sample 5 : **89 CorD**

RNA Area: 51,9      RNA Integrity Number (RIN): 7.7 (B.02.08)  
 RNA Concentration: 53 ng/μl      Result Flagging Color:    
 rRNA Ratio [28s / 18s]: 2.1      Result Flagging Label: RIN: 7.70

Fragment table for sample 5 : **89 CorD**

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 718	2 007	3,8	7,4
28S	3 381	4 348	8,2	15,9

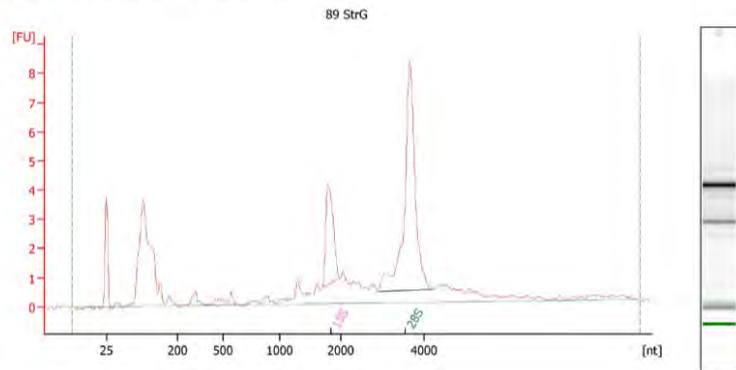
Ratio 28s/18s = 2,1 ≥ 1,5 => Bon

RIN = 7,70 > 7 => Bon

Groupe α-syn

Animal 89 Cortex Droite

Electropherogram Summary Continued ...



Overall Results for sample 8 : 89 StrG

RNA Area: 65,9 RNA Integrity Number (RIN): 8,8 (B.02.08)  
RNA Concentration: 67 ng/ul Result Flagging Color:    
rRNA Ratio (28s / 18s): 3,0 Result Flagging Label: RIN: 8.80

Fragment table for sample 8 : 89 StrG

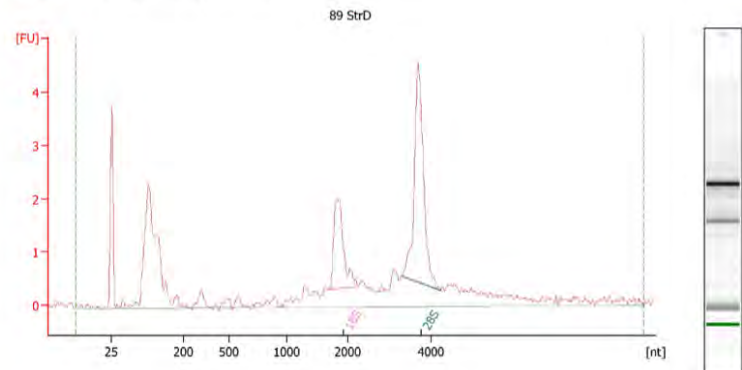
Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 688	1 979	6,0	9,0
28S	2 884	4 242	17,8	26,9

Ratio 28s/18s = 3,0  $\geq$  1,5 => Bon

RIN = 8,80 > 7 => Bon

Groupe  $\alpha$ -syn  
Animal 89 Striatum Gauche

Electropherogram Summary Continued ...



Overall Results for sample 7 : 89 StrD

RNA Area: 41,7 RNA Integrity Number (RIN): 8,6 (B.02.08)  
RNA Concentration: 42 ng/ul Result Flagging Color:    
rRNA Ratio (28s / 18s): 2,0 Result Flagging Label: RIN: 8.60

Fragment table for sample 7 : 89 StrD

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1 674	2 248	3,9	9,5
28S	3 294	4 216	7,7	18,5

Ratio 28s/18s = 2,0  $\geq$  1,5 => Bon

RIN = 8,60 > 7 => Bon

Groupe  $\alpha$ -syn  
Animal 89 Striatum Droite



## 2. Les technologies de détection et quantification des miRNAs

- **Les principales techniques de détection et quantification des miRNAs**

La détection et la quantification des miRNAs est difficile à cause de quelques caractéristiques tels que : ils sont de petite taille (<24 nucléotides), ils ne possèdent pas de caractéristique commune distinctive (les ARNm possèdent par exemple une queue poly (A) qui leur est spécifique), et leur expression varie au cours du temps et en fonction du tissu d'intérêt, en plus d'être parfois très faible (de Planell-Saguer et Rodicio 2013).

Malgré ces difficultés, il existe différentes méthodes pour détecter et quantifier les miRNAs. Les méthodes les plus couramment utilisées sont les suivantes (Tableau 36):

- **L'hybridation in situ.** Cette méthode permet de visualiser l'expression des miRNAs dans les cellules et les tissus, sur le principe de la détection de l'hybridation d'une sonde spécifique complémentaire à la séquence du miRNAs (McDougall, Dunn, et Jones 1972). Cependant cette technique ne permet que la mesure semi-quantitative du niveau d'expression.
- **RT qPCR :** La transcription inverse suivie d'une PCR quantitative, permet de faire une PCR (*Polymerase chain reaction*) quantitative à partir d'un échantillon d'ARN. Cette technique se déroule en deux phases : dans la première, correspond à la copie d'ARN en ADN complémentaire, la seconde phase correspond à quantifier les brins d'ADNc par amplification PCR quantitative et mesurer l'intensité de la fluorescence, en utilisant une amorce spécifique complémentaire du miRNA, et une amorce universelle qui reconnaît une séquence adaptatrice ajoutée lors de l'étape de transcription réverse (Schmittgen et al. 2008). Dans notre étude, le kit miScript II RT a été utilisé inclus dans le système miScript PCR qui permet la détection et la quantification sensibles et spécifiques des miRNAs. Le système miScript PCR utilise l'ARN total qui contient les miRNAs comme matériaux de départ pour la synthèse de l'ADNc. Une préparation d'ADNc unique peut être utilisée avec un tableau de PCR miScript miRNA pour profiler rapidement l'expression de miRNA matures.

Dans une réaction de transcription inverse utilisant le tampon miScript HiSpec, les miRNAs matures sont polyadénylés par la poly (A) polymérase puis convertis en ADNc par transcriptase inverse avec amorçage oligo-dT. L'ADNc est ensuite utilisé pour le profilage PCR en temps réel d'une expression miRNAs mature (à l'aide d'un tableau miScript miRNA PCR et du miScript Universal Primer) (figure 56).

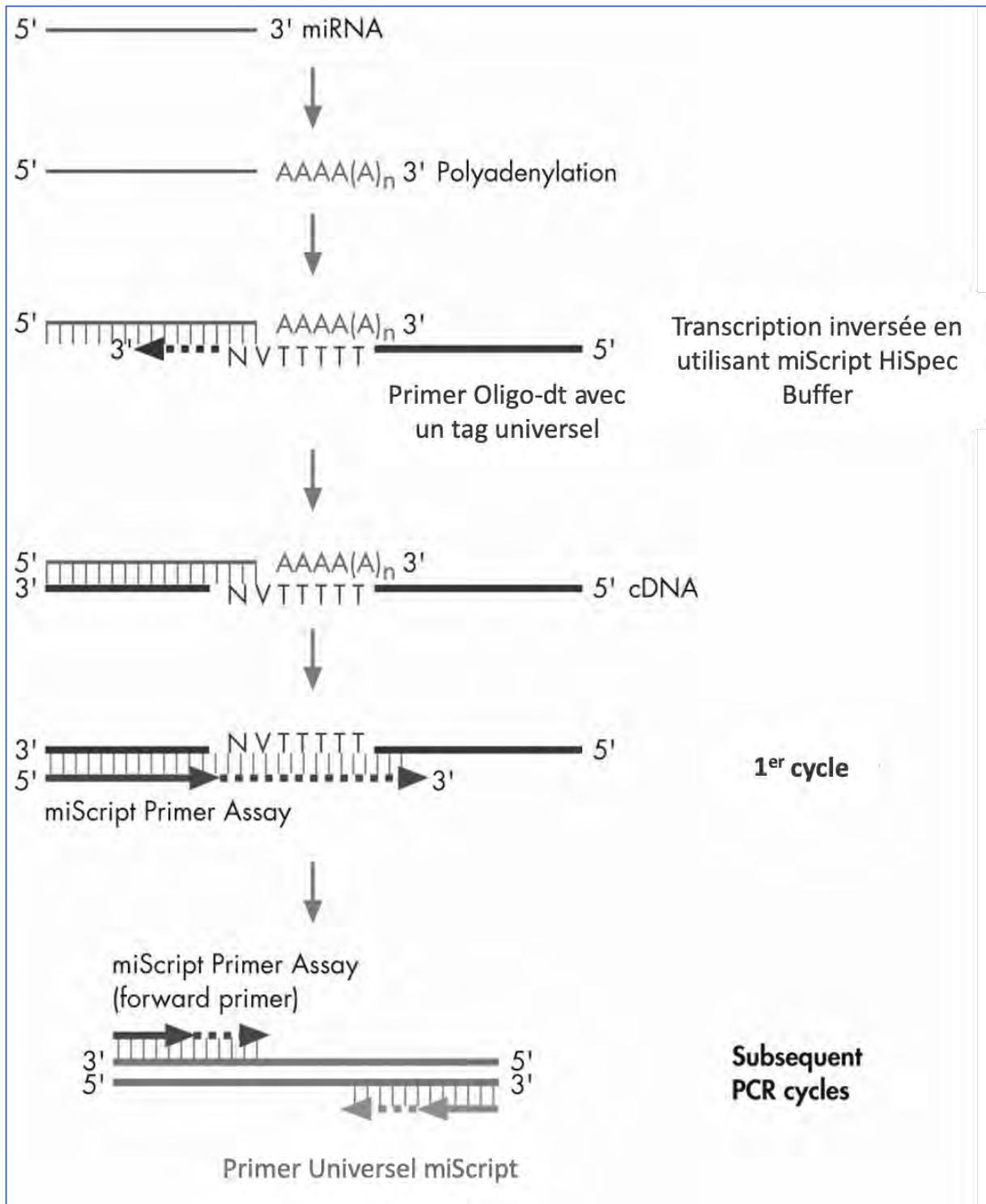


Figure 56: Conversion de miRNA matures en ADNc et détection ultérieure.

- **Les analyses de microarray.** Suivant le même principe que l'hybridation *in situ*, les miRNAs sont purifiés d'extrait cellulaires puis hybridés sur des plaques qui possèdent des sondes spécifiques complémentaires à la séquence des miRNAs d'intérêt (Duan et al. 2011) (Liu et al. 2012).
- **La PCR digitale en goutte.** Cette technique est entièrement basée sur la RT qPCR, et la différence réside dans le fait que chaque réaction de PCR est subdivisée physiquement en une multitude de réactions grâce à la formation d'une émulsion. L'analyse de l'ensemble de ces réactions, en utilisant la loi de Poisson, permet la quantification absolue des acides nucléiques (Hindson et al. 2013).
- Le séquençage d'ARN (**RNA-Seq**) permet de détecter la présence et de quantifier le niveau d'expression par une méthode de séquençage haut-débit. C'est actuellement la seule méthode qui permet de découvrir la présence de nouveaux miRNAs (Creighton, Reid, et Gunaratne 2009).

Parmi la diversité des méthodologies présentées, seul le RNA-Seq est capable à la fois de quantifier de manière absolue l'ensemble des miRNAs d'un échantillon, ainsi que d'en découvrir de nouveaux (Willenbrock et al. 2009) (Tableau 36).

Méthode	Principe	Quantification	Avantages	Inconvénients	Références
<b>Hybridation <i>In situ</i></b>	Révélation à l'aide d'une sonde spécifique	Relative et semi-quantitative	Visualisation possible du patron d'expression dans les cellules et les tissus, dans un contexte spatio-temporel, préserve l'intégrité de l'ARN, car la cellule n'est pas lysée.	Difficile de détecter plusieurs miRNAs simultanément, quantification relative, l'efficacité est dépendante de nombreux paramètres, comme la longueur et la concentration de la sonde et les conditions d'expérimentation.	(McDougall, Dunn, et Jones 1972)
<b>RT qPCR</b>	Amplification à l'aide d'amorces spécifiques	Relative, par rapport à un contrôle interne	Sensible et spécifique, rapide, détection de plusieurs miRNAs en même temps, détection des miRNAs peu exprimés.	Biais de représentativité dû à l'amplification par PCR, la spécifique dépend de la quantité de l'amorce.	(Schmittgen et al. 2008)
<b>Microarray</b>	Utilisation de sondes spécifiques	Quantification relative	Détection de la présence et de la régulation de plusieurs miRNAs simultanément, à partir de plusieurs échantillons	Spécificité de la sonde, beaucoup de bruit de fond	(Duan et al. 2011)
<b>PCR Digitale en goutte</b>	Répartition en 1 molécule par goutte, concentration de départ calculée par la loi de Poisson	Quantification absolue	Très sensible et spécifique, quantification absolue sans gène de référence, reproductibilité robuste	Actuellement moins disponible que la RT qPCR	(Hindson et al. 2013)
<b>RNA-Seq</b>	Séquençage par synthèse du brin complémentaire d'ADNc	Absolue, par rapport au nombre total de séquences	Très sensible, découverte de nouveaux miRNAs, rapide, pas de bruit de fond	Coûteux comparé aux autres techniques, souvent biais de représentativité dû à l'amplification par PCR	(Creighton, Reid, et Gunaratne 2009)

**Tableau 36: Comparaison des caractéristiques principales des différentes méthodes de détection et de quantification des miRNAs (de Planell-Saguer and Rodicio, 2013)**

### 3. Séquençage RNA-Seq

Les techniques de séquençage de l'ARN font partie des technologies de NGS, et permettent de séquencer et quantifier le transcriptôme des cellules, incluant les ARNm et les ARNs non codants et miRNAs (Z. Wang, Gerstein, et Snyder 2009). Le RNA-Seq se base sur des échantillons d'ARN, total ou fragmenté pour les longs transcrits, qui sont souvent convertis en ADN complémentaire avec des adaptateurs à l'une ou sur les deux extrémités. Selon la préparation de librairie qui précède le séquençage, chaque fragment d'ADNc est séquençé à partir de l'une ou des deux extrémités (*single-end* ou *paired-end sequencing*). La longueur des fragments séquençés varie entre 30-400pb selon la technologie utilisée. Les *reads* de séquences obtenues sont analysés selon l'objectif scientifique.

Finalement, les résultats obtenus par RNA-Seq sont validés et comptés en analyse de transcrits par des méthodes de bio-informatique (Z. Wang, Gerstein, et Snyder 2009).

Actuellement, les technologies de RNA-Seq sont devenues abordables financièrement pour les laboratoires, et permettent une quantification absolue, ainsi que la découverte de nouveaux miRNAs dans les échantillons d'intérêt. Dans ce travail de thèse, la technologie d'Illumina a été employée pour réaliser du séquençage d'ARNm et de miRNAs.

#### 3.1. Préparation de librairie avec Illumina

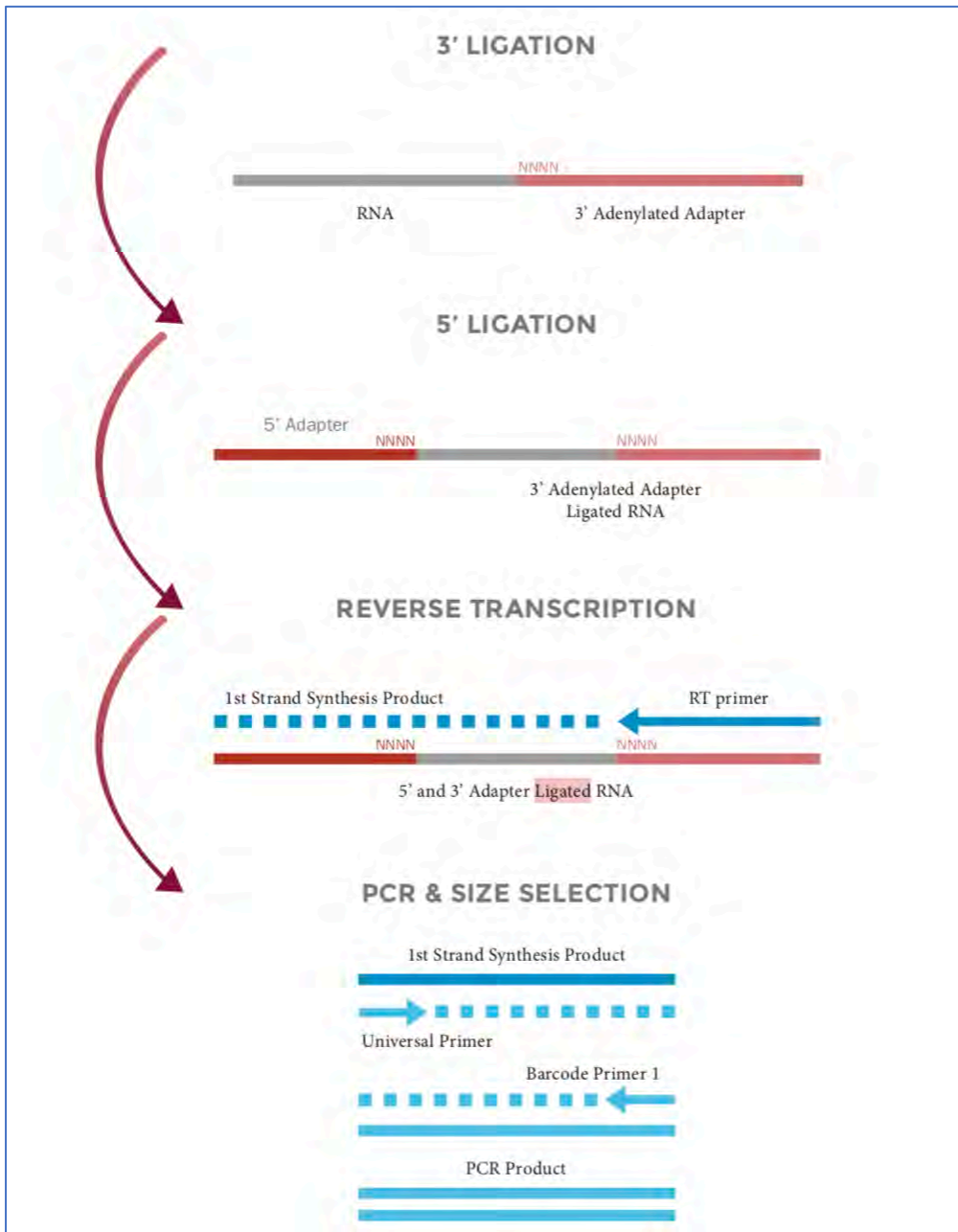
Le kit de BiooScientific : *NEXTflex Illumina Small RNA Seq Library Prep Kit* a été utilisé pour la création de librairie de miRNAs. Le protocole de préparation commence par la fixation des séquences adaptatrices (Tableau 37) aux extrémités des ARNs.

NEXTflex™ 3'4N Adenylated Adapter	rApp /NNNTGGAATTCTCGGGTGCCAAGG/ 3ddC/
NEXTflex™ 5'4N Adapter	5' GUUCAGAGUUCUACAGUCCGACGAUCNNNN
NEXTflex™ RT Primer	GCCTTGGCACCCGAGAATTCCA
NEXTflex™ Barcode Primers	CAAGCAGAAGACGGCATACTGAGATXXXXXXXXGTGACTGGAGTTCCTT GGCACCCGAGAATTCCA
NEXTflex™ Universal Primer	AATGATACGGCGACCACCGAGATCTACACGTTTCAGAGTTCTACAGT CCGA

**Tableau 37: Liste des adaptateurs et des primers fournis par le kit Bio Scientific**

Les bibliothèques ont été synthétisées à partir de 1 µg d'ARN pour chaque échantillon, en suivant les instructions du fabricant (NEXTflex Illumina Small RNA Seq Prep, BioScientific). Brevement, les adaptateurs *adenylated 3'*, suivis par les adaptateurs 5', ont été ajoutés de manière séquentielle aux brins d'ARN (figure 57).

Ensuite, les bibliothèques ont été soumises à une transcription inverse grâce à l'enzyme *M-MuLV* (Moloney Murine Leukemia Virus). Le protocole se termine par une amplification par PCR à l'aide de deux *primers* (Tableau 37) dont l'un d'eux est unique pour identifier l'échantillon car il contient un code barre unique, qui est sous forme de séquence de 6 nucléotides contenue dans la séquence du primer. La taille attendue de nos bibliothèques de miRNAs est de 150bp.



**Figure 57: Organigramme de préparation des échantillons ARN avec NEXTflex**

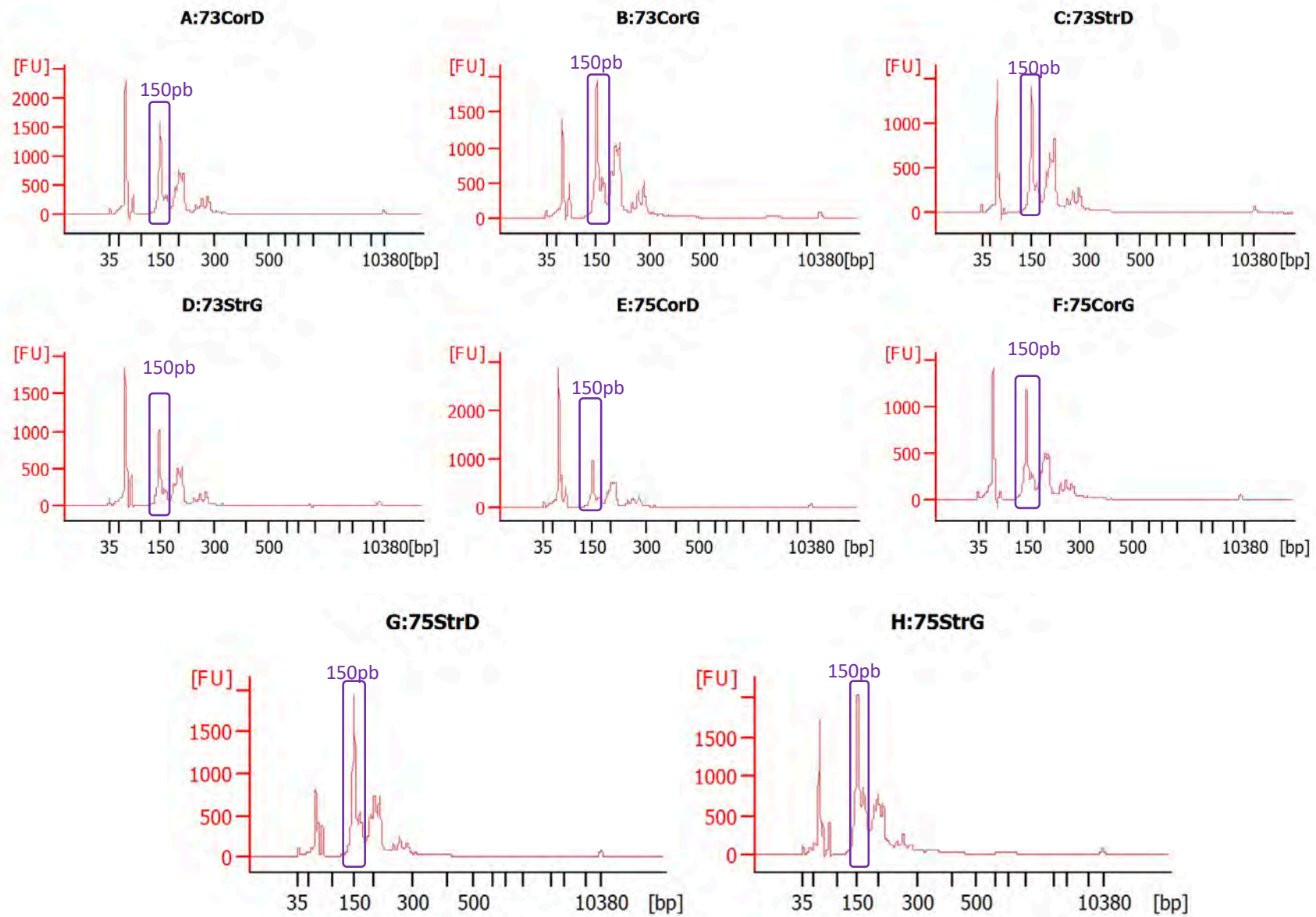
La qualité et la quantité de l'ADNc (figure 59) ont été mesurées par DNA chip (Agilent). Cette analyse montre plusieurs pics en plus de celui à 150bp correspondant à la librairie, ces pics correspondent aux contaminants habituels (ex : amorces de PCR en excès correspondant au pic à 65bp), afin d'optimiser le séquençage, une collection automatique des fragments d'ADNc de 150pb a été faite en utilisant le *Blue Pippin* (Figure 58) (« BluePippin | Sage Science » s. d.).

Ce dernier est un système automatisé d'électrophorèse pour séparer les fragments d'ADN selon une taille programmée et les collecter automatiquement (sans extraction de gel). Il permet une séparation fine, précise, sans contamination assurant la qualité des données et des résultats en aval : « reads » sur séquences utiles (Herbert 2012).



**Figure 58:BluePippine**



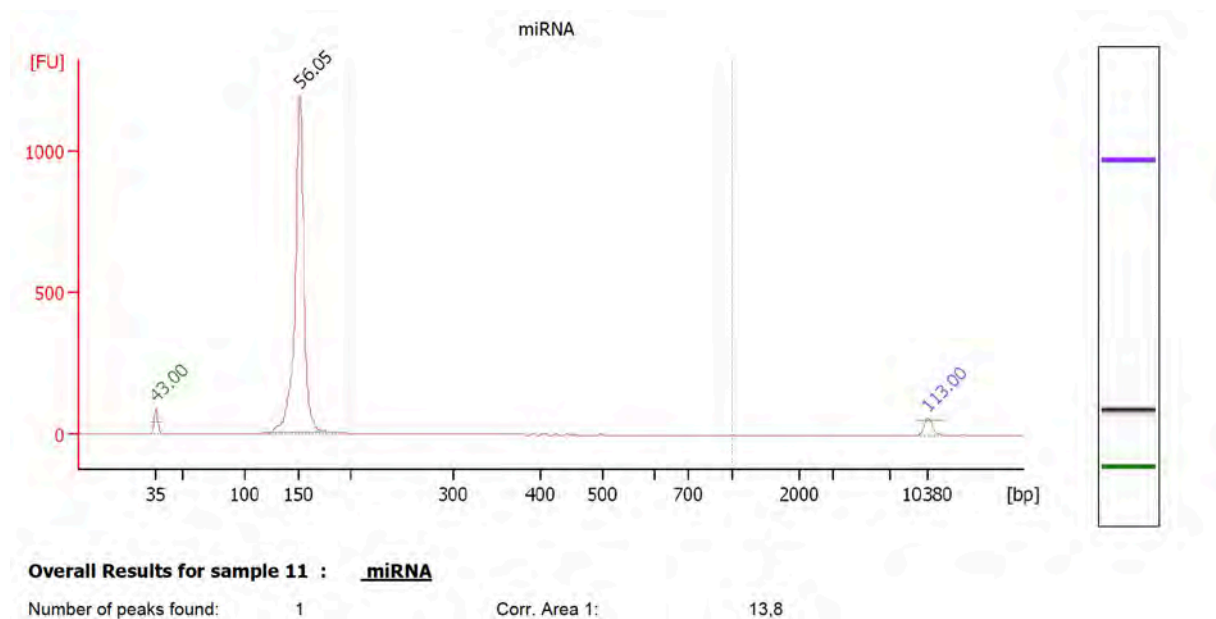


**Figure 59:Résultat du bioanalyseur Agilet de l'ADNc (librairie) pour le groupe 6-OHDA**

Les bibliothèques sont ensuite poolées de manière équimolaire et purifiées, une colonne issue du kit de QIAGEN : MinElute PCR Purification Kit a été utilisée. Pour vérifier la qualité de la bibliothèque finale préparée, une puce Agilent à ADN du bioanalyseur a été utilisée.

Dans la figure 60, aucun pic supérieur à celui d'intérêt (150pb) n'est présent suite à la sélection de la taille par le BluePippin ainsi que la purification sur colonne.

Les pics en violet (à 10380bp) et en vert (à 113bp) indiquent le marqueur de poids moléculaire.



**Figure 60: Qualité de bibliothèque pour RNAseq**

### 3.2. Exécution du séquençage RNA-Seq

Le séquençage haut-débit a été réalisé à l'aide d'un séquenceur Illumina NextSeq 500. C'est un séquenceur qui peut détecter jusqu'à 400 millions de *reads*. Pour nos échantillons, une profondeur d'au moins 8 millions de reads est nécessaire afin de pouvoir, éventuellement, détecter de nouveaux miRNAs. De plus, le séquençage s'est réalisé en single-read, c'est-à-dire à partir d'une seule extrémité d'ADNc.

## 4. Analyse transcriptomique (Bio-informatique)

### 4.1. Contrôle qualité et filtrage

Le séquençage des échantillons a produit des fichiers de format FastQ (Cock et al. 2010), la première étape primordiale à effectuer après l'obtention des données brutes du séquenceur, c'est le contrôle qualité, il a été réalisé par FastQC (S. Andrews 2010) (Voir chapitre 2).

Les reads obtenus après le séquençage possèdent à leurs extrémités les séquences des adaptateurs provenant du kit de BioScientific utilisées pour la création des bibliothèques, ces séquences ont été supprimées ainsi que les quatre nucléotides insérés entre les adaptateurs et les miRNAs en utilisant l'outil cutadapt (Martin 2011).

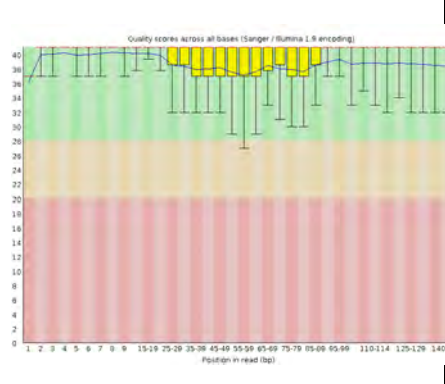
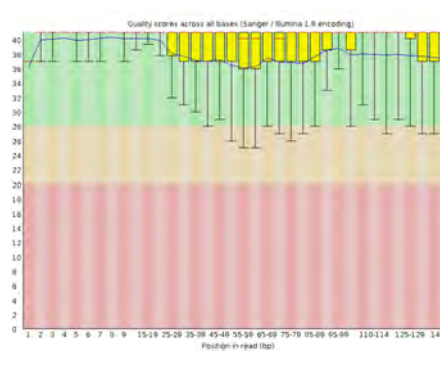
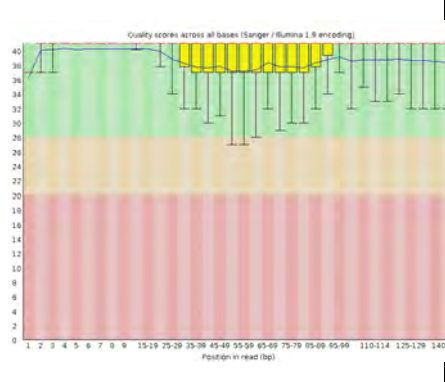
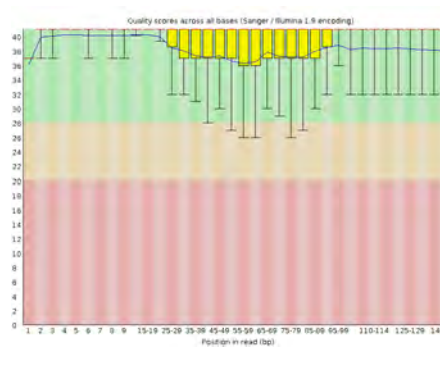
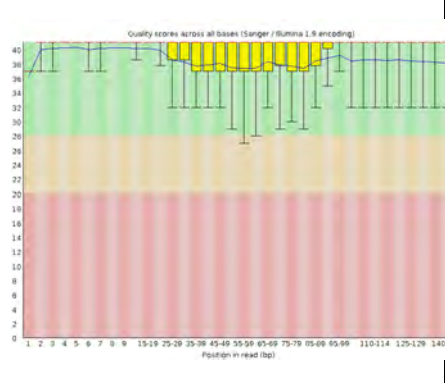
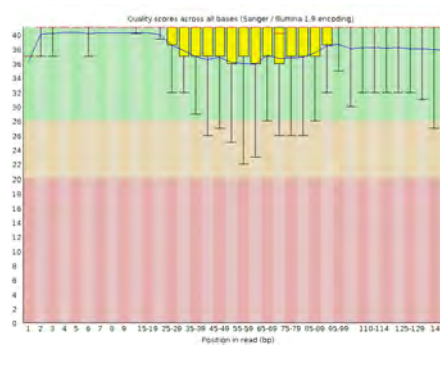
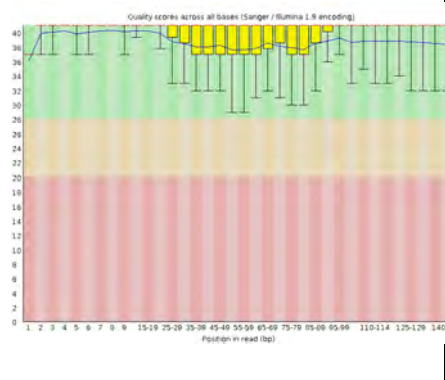
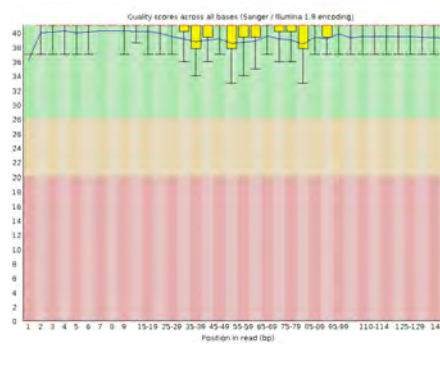
```
cutadapt -u 4 -u -4 -o trimmed.fastq reads.fastq
```

En utilisant l'option `--cut` ou son abréviation `-u`, il est possible de supprimer inconditionnellement les bases du début ou de la fin de chaque lecture. Si la longueur donnée est positive, les bases sont supprimées au début de chaque lecture. Si c'est négatif, les bases sont retirées de la fin.

Cuadapt est un outil en ligne de commande, qui prend en charge les données du séquenceur 454, Illumina et SOLiD, il propose deux algorithmes d'ajustement d'adaptateur et offre d'autres fonctionnalités utiles et il supprime les séquences d'adaptateur des données de séquençage à haut débit. Cela est nécessaire lorsque les lectures sont plus longues que la molécule séquencée, comme dans les données microRNAs.

L'outil FastQC fournit une description générale du fichier séquencé correspondant à un échantillon tel que le nombre de reads totaux, le %GC et la taille des séquences. FastQC peut évaluer la qualité des séquences à travers un score (axe des ordonnées) qu'il attribue à chaque base d'une séquence (axe des abscisses). Plus la valeur du score est élevée, plus la base est considérée de bonne qualité (Tableau 38). Cet axe des ordonnées est réparti en trois sous-groupes (vert : bonne qualité, orange : qualité raisonnable, rouge : pauvre qualité) (Voir chapitre 2).

Tableau 38: Résultat FastQC des échantillons - RNA-Seq

	Descriptions du read	Qualité du read		Descriptions du read	Qualité du read
73CorD	<p>Nombre de séquences : 10830823</p> <p>Longueur de séquence : 15 -142</p> <p>%GC = 44</p>		75CorD	<p>Nombre de séquences : 12276686</p> <p>Longueur de séquence : 15 -142</p> <p>%GC = 43</p>	
73CorG	<p>Nombre de séquences : 14207646</p> <p>Longueur de séquence : 15 -142</p> <p>%GC = 44</p>		75CorG	<p>Nombre de séquences : 9995359</p> <p>Longueur de séquence : 15 -142</p> <p>%GC = 43</p>	
73StrD	<p>Nombre de séquences : 13708425</p> <p>Longueur de séquence : 15 -142</p> <p>%GC = 44</p>		75StrD	<p>Nombre de séquences : 13220203</p> <p>Longueur de séquence : 15 -142</p> <p>%GC = 44</p>	
73StrG	<p>Nombre de séquences : 11790792</p> <p>Longueur de séquence : 15 -142</p> <p>%GC = 44</p>		75StrG	<p>Nombre de séquences : 15321652</p> <p>Longueur de séquence : 15 -142</p> <p>%GC = 44</p>	

83CorD	<p>Nombre de séquences : 33485268  Longueur de séquence : 15 -142  %GC = 44</p>		89CorD	<p>Nombre de séquences : 31849740  Longueur de séquence : 15 -142  %GC = 43</p>
83CorG	<p>Nombre de séquences : 37813213  Longueur de séquence : 15 -142  %GC = 44</p>		89CorG	<p>Nombre de séquences : 34793607  Longueur de séquence : 15 -142  %GC = 44</p>
83StrD	<p>Nombre de séquences : 39387513  Longueur de séquence : 15 -142  %GC = 44</p>		89StrD	<p>Nombre de séquences : 30977224  Longueur de séquence : 15 -142  %GC = 44</p>
83StrG	<p>Nombre de séquences : 28301734  Longueur de séquence : 15 -142  %GC = 44</p>		89StrG	<p>Nombre de séquences : 31025967  Longueur de séquence : 15 -142  %GC = 44</p>

## 4.2. Le référencement des miRNAs

- **La base de données miRBase.**

Pour regrouper les données réalisées sur miRNAs et représenter l'ensemble des miRNAs découverts à travers tout travaux de recherche, la création d'une base de données s'est révélée être indispensables. Actuellement, il existe plusieurs bases de données, mais la base de données de référence, libre d'accès, est incontestablement est miRBase (Griffiths-Jones et al. 2011) (<http://mirbase.org>).

La base de données miRBase est une base de données interrogeable contenant les séquences et les annotations miRNA publiées réparties par modèle d'étude. Les séquences publiées sont disponibles pour la recherche et la navigation, et les entrées peuvent également être récupérées par nom, mot-clé, références et annotations. Toutes les données de séquence et d'annotation sont également disponibles au téléchargement.

En septembre 2010, cette base de données contenait des informations sur 15172 microARN (Kozomara et Griffiths-Jones 2011). Ce nombre est passé à 38.589 en mars 2018 (Kozomara, Birgaoanu, et Griffiths-Jones 2019). Le registre miRBase fournit un système centralisé pour attribuer de nouveaux noms aux gènes de microARN (Griffiths-Jones 2006).

La version actuelle de miRBase est la version 22 (mars 2018), c'est la version utilisée dans cette étude.

## 4.3. La prédiction des miRNAs par miRPro

A partir des résultats obtenus par le séquençage RNA-Seq, la découverte des miRNAs est possible grâce aux algorithmes bio-informatiques. Plusieurs outils existants pour analyser et prédire les miRNAs : miRanalyzer (Hackenberg et al. 2009), miRdeep (Friedländer et al. 2008), miRExpress (W.-C. Wang et al. 2009), miRtools (Zhu et al. 2010), etc.

Chaque outil se base sur un algorithme spécifique pour un type d'espèce particulier comme miRDeep-P et miRPlant (An et al. 2014) pour les plantes, ou sur la sélectivité et l'efficacité permettant la découverte de miRNAs.

En se basant sur l'étude comparative des outils de prédictions des miRNAs (Williamson et al. 2013), le programme le plus utilisé, et plus performant dans la découverte des miRNAs est miRDeep (Friedländer et al. 2012).

Pour analyser les données de cette étude, nous avons utilisé l'outil miRPro (Shi et al. 2015) qui s'appuie sur l'algorithme principal de miRDeep, cet algorithme permet de détecter la présence de séquences de miRNAs ainsi de découvrir des nouveaux par prédiction. L'algorithme de

l'outil miRDeep s'appuie sur les caractéristiques de biogénèse des miRNAs (Friedländer et al. 2008).

L'outil miRPro adopte des approches différentes et plus précises pour réaliser l'excision des séquences adaptatrices, l'alignement des séquences sur le génome de référence, et sur la quantification des miRNAs. L'algorithme de base de miRPro permet de favoriser la découverte des variant de miRNAs, et de changement de brin dans la sélection du brin mature lors de phénomène d'"*arm switching*" (Griffiths-Jones et al. 2011). Cette fonctionnalité (Meijer, Smith, et Bushell 2014) est d'une importance cruciale dans l'analyse de la découverte de nouveaux miRNAs, l'outil miRPro permet d'annoter les séquences prédites de nouveaux miRNAs de manière homogène entre tous les échantillons étudiés. Il permet de réaliser des analyses d'expression différentielle (Shi et al. 2015). Pour toutes ces caractéristiques, le programme miRPro a été choisi pour réaliser les analyses RNA-Seq.

#### 4.3.1. Pipline d'analyse miRPro

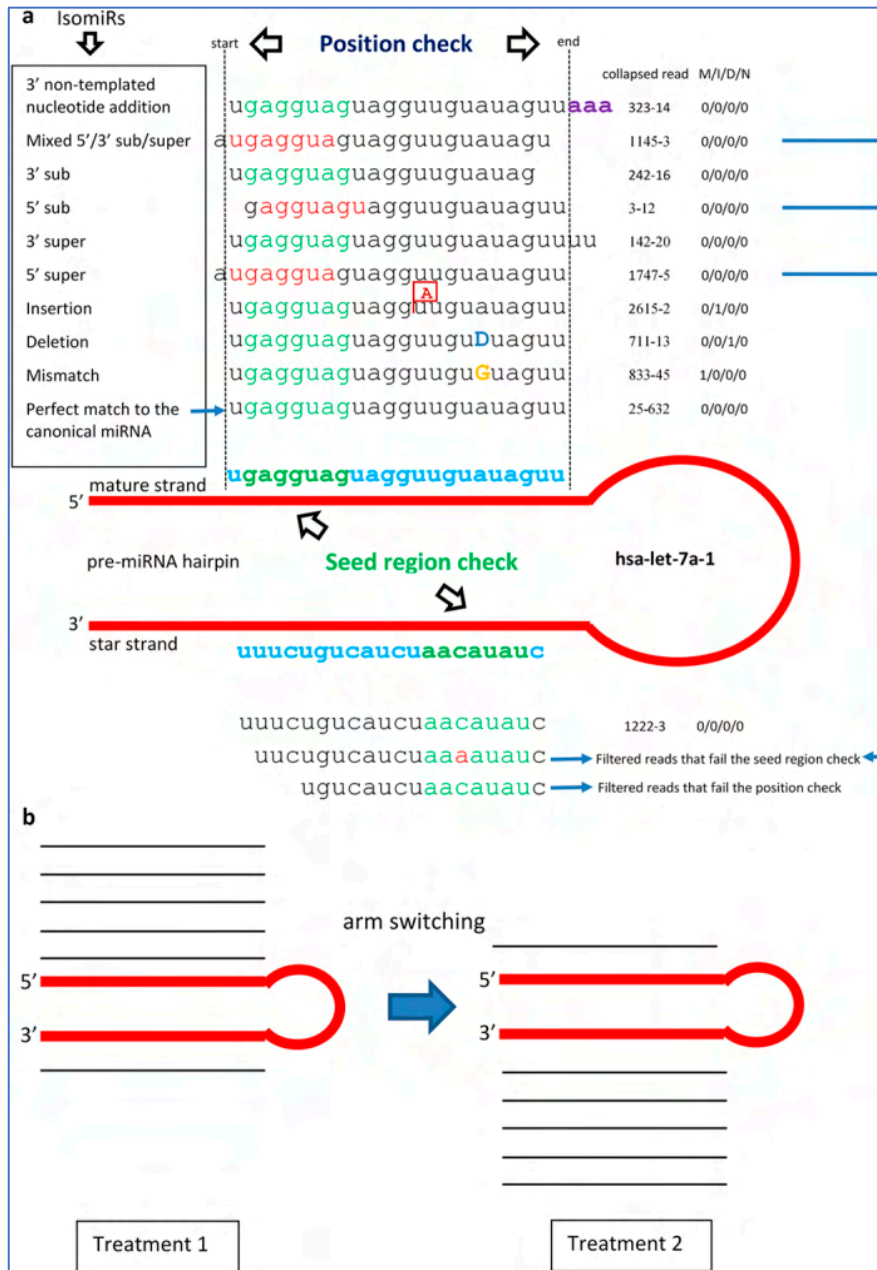
Le package mirPRO est composé d'un programme principal (c'est-à-dire, mirpro) et de plusieurs programmes composants (voir la Figure 61). Certains de ces programmes peuvent être utilisés indépendamment. Par exemple, *mirpro\_findAdapter* est un programme générique de détection de séquences d'adaptateurs pour les données RNA-Seq. *Mirpro feature\_pro* est conçu pour cataloguer les lectures cartographiées en termes d'annotation de gène; et *mirpro\_armSwitch* est un programme spécifique pour la détection de « Arm Switching ». mirPRO utilise quelques outils tiers (par exemple, Novoalign (H. Li et Homer 2010), HTSeq (Anders, Pyl, et Huber 2015), randfold (Bonnet et al. 2004) et RNAfold (Lorenz et al. 2011)), qui doivent être installés en tant qu'outils prérequis. Les programmes principaux et composants permettent un réglage avancé de leurs paramètres ajustables. Pour les utilisateurs généraux, l'ensemble du package, à l'exception de *mirpro\_armSwitch*, peut être exécuté automatiquement en lançant le programme principal, qui prend en entrée une ou plusieurs données FASTQ. *mirpro\_armSwitch* doit être appelé séparément une fois que le programme principal a généré des résultats pour différents traitements ou échantillons.

À partir des données brutes, mirPRO effectue d'abord un filtrage de la qualité des reads. Pour un alignement efficace des reads à la référence, les reads propres finales dans chaque échantillon sont ensuite réduites en termes de contenu de séquence, les nombres d'expression étant comptés. Ensuite, Novoalign (H. Li et Homer 2010) est utilisé pour aligner les reads réduites sur les pré-miRNAs (épingles à cheveux) téléchargées à partir de miRBase (Kozomara, Birgaoanu, et Griffiths-Jones 2019). Novoalign permet de signaler les alignements

selon le meilleur score d'alignement pour chaque read. Si un read ne contient qu'une seule correspondance avec le meilleur score, elle sera traitée comme une correspondance unique.

En s'appuyant à la fois sur les miRNAs matures et leurs séquences pré-miRNAs (en épingle à cheveux) téléchargées à partir de miRBase (Kozomara, Birgaoanu, et Griffiths-Jones 2019), mirPRo effectue une quantification connue des miRNAs à l'aide de l'algorithme illustré à la figure 61. Cet algorithme assure également qu'il n'y a pas de problème de sur-dénombrement pour les reads de mRNAs matures en divisant de manière égale le nombre de reads parmi leurs épingles à cheveux mappées, si les reads ont des mappages non uniques. Sur la base de données de classification de la famille pré-miRNAs téléchargées à partir de miRBase, mirPRo fournit une quantification de l'expression de chaque famille en ajoutant toutes les reads d'ARN matures quantifiées aux miRNAs pertinents appartenant à une famille donnée. mirPRo prédit de nouveaux miRNAs en utilisant l'algorithme de miRDeep2 (Friedländer et al. 2012) (Mackowiak 2011).





**Figure 61: Les principaux algorithmes de mirPro pour l'exploration et la quantification des variants de miRNAs (Shi et al. 2015).**

a) Identification IsomiR. Le précurseur humain miRNA «hsa-let-7a-1» est utilisé à titre d'illustration, et les lectures condensées ne sont pas des données réelles. mirPro autorise les erreurs de base (mismatch et indel), permet le contrôle de position et le contrôle facultatif de la région de départ dans la quantification mature de miRNAs. Les isomères annotés par mirPro comprennent des reads de miRNAs matures avec des mésappariements, des insertions, des délétions ou un mélange, avec une addition de nucléotides à l'extrémité 3' et un décalage des nucléotides (super ou sub) à leurs extrémités 5', 3' ou aux deux extrémités.

Le "D" majuscule dans la séquence alignée signifie la suppression. La colonne à l'identificateur ("XXX-YYY" ) pour les reads, où "XXX" est un nombre unique et "YYY" le nombre de reads. La colonne "M / I / D / N" représente le nombre de "mésappariements/insertions /délétions/nucléotides N" dans l'alignement.

b) Détection de Switching ARM : Les reads supplémentaires sont mappées sur le 5' du précurseur dans le traitement 1, tandis que les autres reads sont mappées sur le 3' dans le traitement 2. Cela indique que deux miRNAs différents sont générés à partir de deux brins différents du même précurseur dans deux applications différentes (exemple : différents tissus).

Pour l'analyse des variations de miRNAs, mirPro détecte des variants d'isomiRs - miRNAs différents de leurs miRNAs matures canoniques annotées dans miRBase (Kozomara, Birgaoanu, et Griffiths-Jones 2019).

mirPro fournit un catalogage de toutes les séquences propres si les informations d'annotation du génome sont fournies au format GTF, afin que les utilisateurs puissent comprendre l'évaluation globale du séquençage et de la qualité de l'alignement des miRNAs. Ensuite, le HTSeq (Anders, Pyl, et Huber 2015) est utilisé pour comptabiliser les reads cartographiées en termes de caractéristiques annotées (par exemple, gènes codant pour des protéines, miRNAs, snRNA, ARNr, et ARNnc). mirPro génère un fichier de séquence contenant des reads propres non mappées au format FASTA, afin que les utilisateurs puissent examiner les données plus en détail afin de déterminer les sources potentielles de contamination par séquençage.

Dans notre cas, après le contrôle qualité et le filtrage des données issues du séquençage RNA-Seq, l'utilisation du programme miRPro permet d'identifier et de quantifier les miRNAs exprimés dans les échantillons.

Le génome de référence *rattus norvegicus* ( tableau 39) a été téléchargé à travers l'interface de la plateforme *UCSC Genome Browser Gateway* (*UCSC Genome Browser Gateway 2018*).

<b>UCSC Genome Browser assembly ID</b>	Rn6
<b>Sequencing/Assembly provider ID</b>	RGSC Rnor_6.0
<b>Assembly date</b>	Jul. 2014
<b>Accession ID</b>	GCA_000001895.4
<b>NCBI Genome ID</b>	<b>73</b> (Rattus norvegicus)
<b>NCBI Assembly ID</b>	191871 (Rnor_6.0)
<b>NCBI BioProject ID</b>	10629

**Tableau 39: Les caractéristiques du génome de référence rattus norvegicus**

Au début, miRPro aligne les reads avec le génome de référence en utilisant l'outil Novoalign<sup>55</sup>, ensuite la découverte des miRNAs est faite en se basant sur miRDeep2 (Friedländer et al. 2012) intégré dans le programme miRPro, et optimisé (figure 62).

MiRDeep2 scanne l'ADN de 5' en 3' sur les deux brins et repère l'alignement de reads le plus important sur 70nt. Ensuite, il excise deux précurseurs potentiels à partir de cet alignement et il crée un index de ces précurseurs, puis il procède à l'alignement des reads avec les précurseurs potentiels afin de tester s'ils correspondent à une structure tige-boucle classique des miRNAs. Finalement, les structures prédictives sont effectuées par le l'outil RNAfold qui est chargé dans le programme miRPro.

La commande utilisée pour exécuter l'outil mirPro :

```
mirpro -i trimmed.fastq -d ./ -m mature.fa -p hairpin.fa -s
rno -a null -f miFam.dat --novel 1 --other mmu --other hsa
-g genome.fa --index genome.idx --gtf genes.gtf -t 1 -r 1
```

<sup>55</sup> Novoalign : Outil d'alignement pour aligner les courtes séquences sur un ensemble indexé de séquences de référence. Généralement utilisé pour aligner les reads de bouts simples et appariés d'Illumina.



Les données utilisées dans cette commande peuvent être télécharger de miRBase.

- mature.fa : Format Fasta de toutes les séquences miRNAs matures
- hairpin.fa : Format Fasta de toutes les épingles à cheveux de miRNAs.
- miFam.dat : Classification des familles des séquences en épingle à cheveux miRNAs
- genome.fa : Génome de référence (Rn6).
- genome.idx : creation du fichier index du génome de référence en utilisant la commande `novoindex genome.idx genome.fa`
- gene.gtf : Annotation des genes avec le format GTF téléchagé de Ensembl.

Les résultats obtenus par miRPro sous forme de plusieurs fichiers (tableau 40) :

filter_miRNA_mature_report	enregistre le nombre d'espèces de miRNAs matures des espèces étudiées.
filter_miRNA_precursor_report:	enregistre le nombre d'espèces précurseurs de miRNAs des espèces étudiées.
filter_miRNA_mature_XXX_report	enregistre le nombre d'espèces de miRNAs matures des espèces apparentées "XXX".
mapping_report.csv	statistiques des mappages dans tous les échantillons.
novel_mature.fa	les séquences de miRNAs matures prédites dans tous les échantillons au format fasta. L'identifiant est "XXX-novel-miR-YYY" ("XXX" est le nom de l'espèce; "YYY" est un numéro unique identique à son précurseur).
novel_precursor.fa	les séquences prédites des miRNAs précurseurs dans tous les échantillons au format fasta. L'identifiant est "XXX-novel-mir-YYY" ("XXX" est le nom de l'espèce; "YYY" est un numéro unique). Les descriptions incluent les emplacements du génome au format "chromosome: strand: start: end".
novel_precursor.str	structure des miRNAs précurseurs prédits sous forme de point.

novel_quantifier_report.csv	le rapport de quantification pour les nouveaux miRNAs.
precursor_mature_relation.csv	relation entre le pré-miRNAs et le miRNAs mature
quantifier_report.csv	le rapport de quantification pour les miRNA connus.
result_family.csv	nombre de familles de gènes miRNAs connues dans tous les échantillons.
result_mature.csv	nombre de miRNAs connus dans tous les échantillons.
result_precursor.csv	les comptes de pré-miRNAs connus dans l'échantillon.
result_novel_mature.csv	nombre de nouveaux miRNAs matures dans tous les échantillons.
result_novel_precursor.csv	comptes de nouveaux pré-miRNAs dans l'échantillon.
statistics.csv	les comptes et pourcentages des reads traitées dans tous les échantillons.
variation_report.csv	les comptes et les pourcentages des lreads de miRNAs présentant des différences de séquence par rapport aux valeurs de miRNAs correspondantes dans tous les échantillons.

**Tableau 40: Liste des fichiers obtenus par miRPro**

Finally, the differential expression analysis of miRNAs between the structures was performed using the DESeq2 program under the R software (figure 62).

The standard steps of differential expression analysis are grouped into a single function, **DESeq** (Love, Huber, et Anders 2014). The result tables are generated using the **result** function, which extracts a table of results with the changes in log2 fold change, the p-value and the adjusted p-value (figure 62).

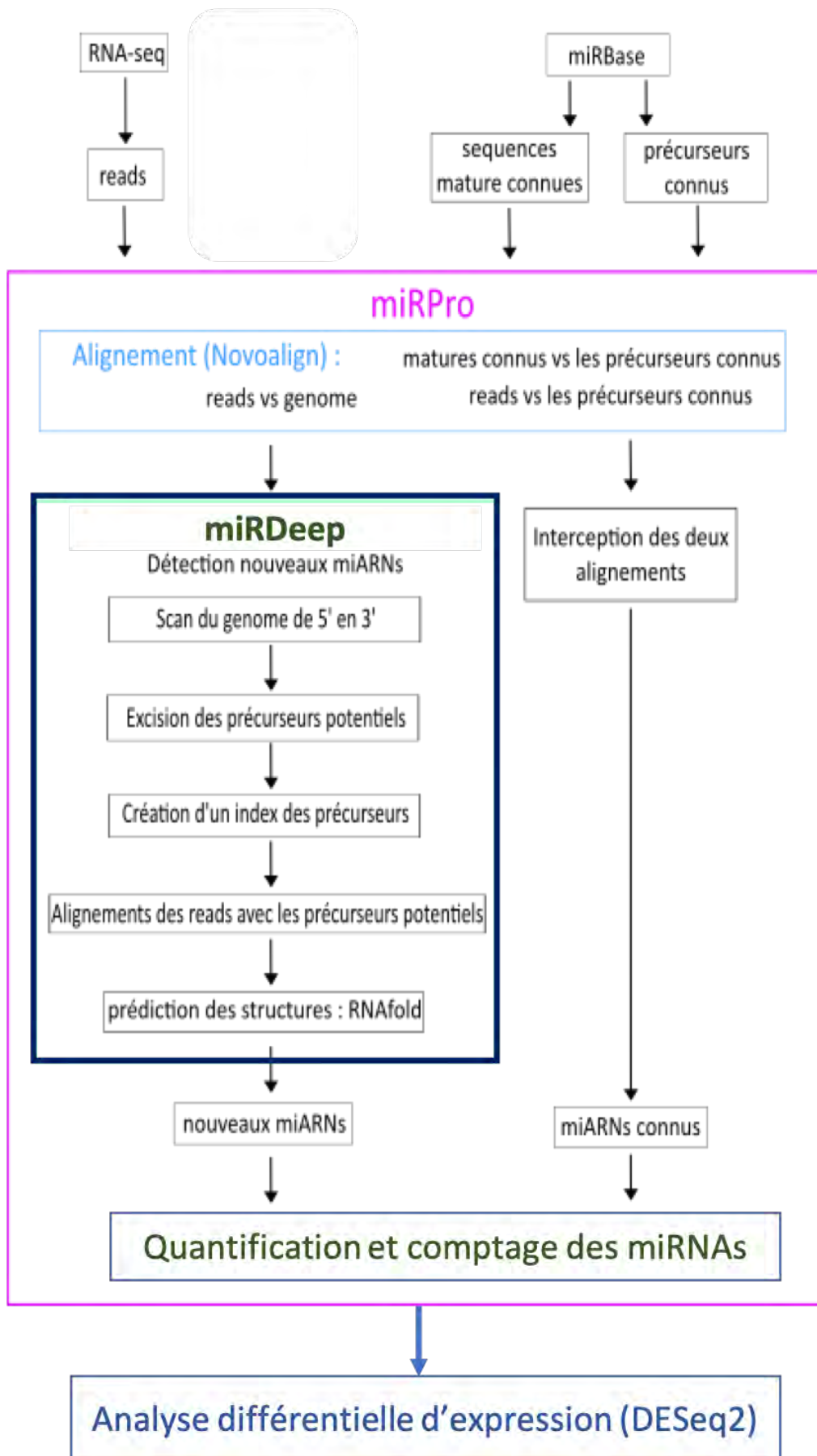


Figure 62: Schéma des étapes d'analyse bio-informatique des données issues de RNA-Seq

#### 4.3.2. Résultats miRPro

La première étape est de vérifier le nombre de miRNAs connus et référencés dans miRBase exprimés dans les différents échantillons en se basant sur les résultats de miRPro, on a obtenu **480 miRNAs** (tableau 41) et **476 miRNAs** (tableau 42) respectivement pour le groupe 6-OHDA et le groupe Alpha-Syn, communs entre les différents échantillons appartenant au même groupe.

Échantillons Groupe 6OHDA	Nombre de miRNAs	% de miRNAs matures référencés par miRBase
75CorD	542	69.0196
75CorG	545	70.3268
75StrD	552	72.549
75StrG	555	72.1569
73CorD	528	70.8497
73CorG	538	71.2418
73StrD	560	73.2026
73StrG	553	72.1569

**Tableau 41:Résultat miRPro pour le groupe 6-OHDA**

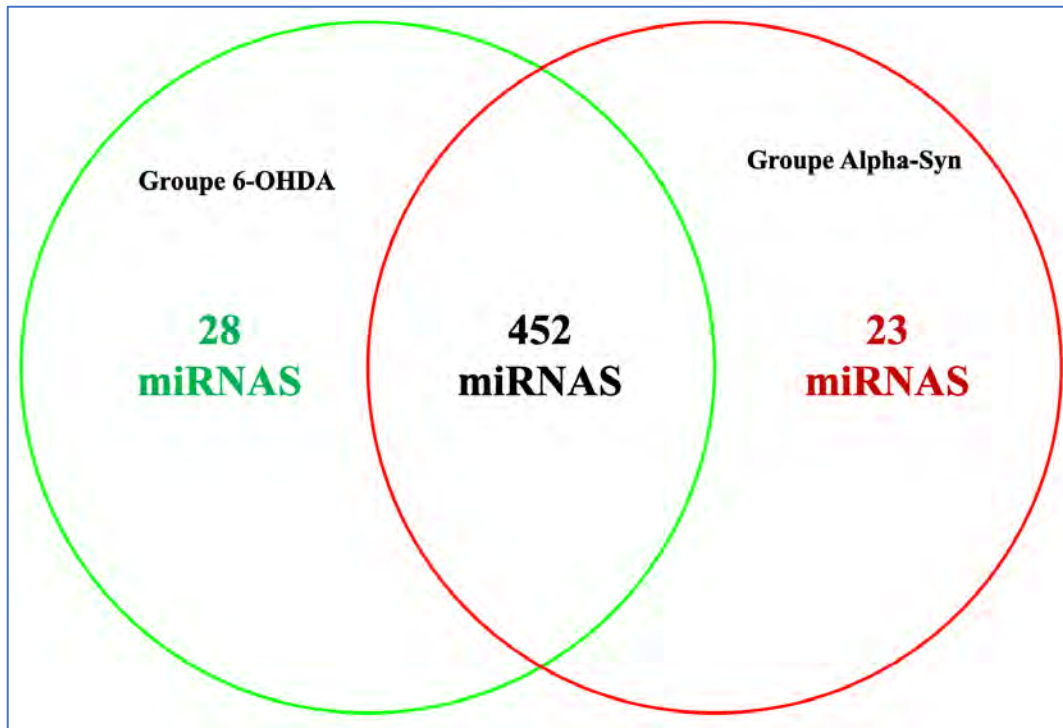
Échantillons Groupe Alpha-Syn	Nombre de miRNAs	% de miRNAs matures référencés par miRBase
83CorD	475	70.9804
83CorG	486	73.4641
83StrD	486	75.817
83StrG	486	70.1961
89CorD	476	72.8105
89CorG	480	72.0261
89StrD	500	76.4706
89StrG	465	73.8562

**Tableau 42:Résultat miRPro pour le groupe Alpha-Syn**



Pour identifier les miRNAs communs entre les deux groupes, nous avons utilisé le diagramme de Venn sous le logiciel R également appelé diagramme d'Euler-Venn, est une représentation simple des ensembles par diagrammes.

La figure 63 montre que 452 miRNAs communs entre les deux groupes étudiés.



**Figure 63:Diagramme de Venn pour les miRNAs communs entre le groupe 6-OHDA et Alpha-Syn**

- **Les miRNAs sélectionnées par pvalue**

En se basant sur DESeq2 exécuté à travers le logiciel R afin de détecter l'expression différentielle entre différents groupes (Love, Huber, et Anders 2014) et différentes structures. Afin de réaliser ces analyses, le fichier obtenu de miRPro "result\_mature.csv" qui contient le nombre de miRNAs connus dans tous les échantillons a été chargé dans le logiciel en se basant sur le script ci-dessous :

```

cts<-read.table (file="result_mature.csv", sep=";", row.names=1,
header=TRUE)

coldata<-read.table (file="mirpro_annotation.csv", sep=";",
row.names=1, header=TRUE)

dds <- DESeqDataSetFromMatrix (countData = cts,colData =
coldata,design = ~ condition)

dds$condition <- factor (dds$condition, levels = c
("treated","untreated"))

dds <- DESeq (dds)

resutats<-results (dds,cooksCutoff= FALSE, contrast=c
("condition","treated","untreated"))

```

Pour chaque miRNAs et chaque comparaison de structure, on obtient une valeur de log2 (Foldchange) et de p-value.

Dans la littérature, plusieurs chercheurs communiquent l'expression relative entre deux expériences en termes de 'Fold Change' (FC), qui donne aussi la direction de l'expression relative.

Cependant, les log-ratios sont souvent utilisés pour l'analyse et la visualisation des des "Fold change". Le logarithme de la base 2 est le plus couramment utilisé (Robinson et Smyth 2008)car il est facile à interpréter.

Afin de visualiser les niveaux d'expression de miRNAs pour le groupe 6-OHDA et le groupe Alpha-Syn (côté Droit VS côté Gauche), une comparaison entre deux conditions de traitement a été faite en se basant sur la p-value et le log2 (Fold Change) sous forme de "Volcano Plot" (Figure 63 pour le groupe 6-OHDA et la figure 64 pour le groupe Alpha-Syn). Ce graphique affiche la signification statistique de la différence par rapport à l'ampleur de la différence pour chaque miRNAs dans la comparaison, généralement par le biais du changement du log base-10 et du log base-2 respectivement. Les p-value ayant une transformation négative, plus un point de données tombe haut le long de l'axe des ordonnées, plus la p-value est petite. Elle est généralement utilisée dans le volcano plot pour inclure des indicateurs de seuil pour les p-value

ajustées afin d'indiquer quels miRNAs seraient considérés comme statistiquement exprimés de manière différentielle sur la base de la p-value ajustée. Le changement de log Fold Change sur l'axe des abscisses montre des différences plus considérables dans les valeurs extrêmes, avec des points de données plus proches de 0 représentant des miRNAs ayant des niveaux d'expression moyens similaires ou identiques. Une dispersion plus large indique que deux groupes de traitement présentent un niveau de différence plus élevé en ce qui concerne l'expression des miRNAs.

Le volcano plot généré par le logiciel R à l'aide des résultats obtenus de DESeq2, avec les seuils de Log<sub>2</sub> (Fold Change) comprise entre -0,5 et 0,5 (lignes verticales pointillées en bleu) et une signification statistique élevée (-log<sub>10</sub> de la p-value, de l'axe des ordonnées). La ligne horizontale pointillée en bleu montre p-value = 0,05. Ce graphique est coloré de manière à ce que les points dont le log<sub>2</sub> (Fold Change) soient comprise en -0,5 et 0,5 ou p-value < 0,05) apparaissent en rouge, c'est miRNAs différentiellement exprimées.

Dans notre étude, l'analyse a été focalisé sur le striatum, étant une structure qui subit directement les effets de la MP et le contenu du striatum en dopamine s'effondre rapidement.



Volcano plot des miRNAs pour le groupe Alpha-Syn

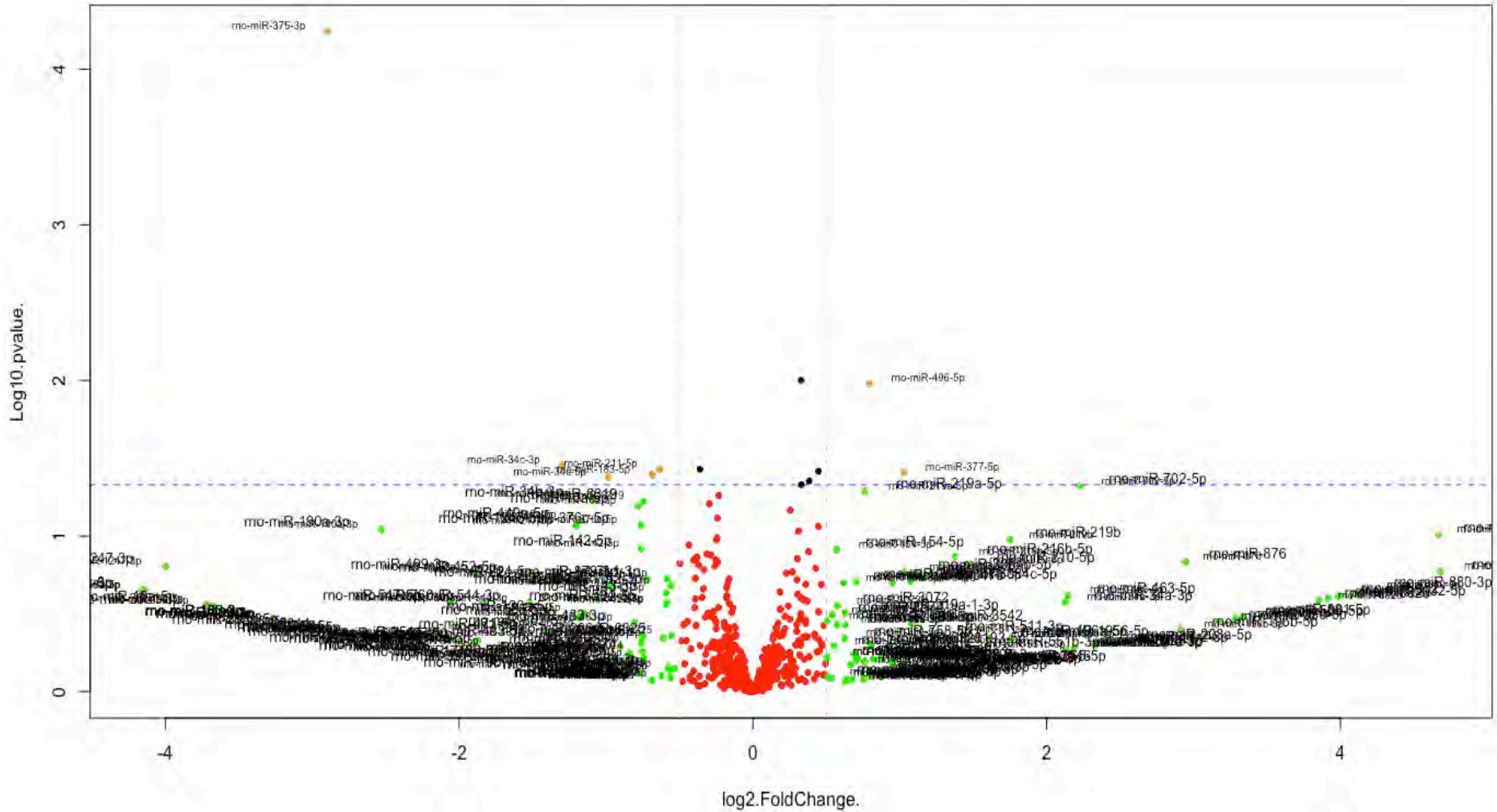


Figure 65:volcano plot groupe Alpha-Syn

- **Groupe 6-OHDA**

Les miRNAs sélectionnées à partir des résultats du volcano plot pour le groupe 6-OHDA sont générées dans les tableaux ci-dessous :

<b>Comparaison Cortex Gauche VS Droit pour le groupe 6-OHDA</b>		
Nom de miRNA	log2FoldChange	pvalue
rno-miR-219a-2-3p	-0,90516599	1,7114E-05
rno-miR-338-5p	-0,57625478	0,00691731
rno-miR-217-5p	1,712383132	0,01262068
rno-miR-7a-5p	0,513469259	0,01644107
rno-miR-338-3p	-0,67614818	0,01844175
rno-miR-139-5p	0,469090712	0,02045786
rno-miR-21-3p	-4,48983494	0,02200934
rno-miR-192-5p	-0,60297799	0,02374237
rno-miR-219a-5p	-0,77199918	0,02702871
rno-miR-374-3p	0,521582955	0,02736421
rno-miR-18a-3p	-3,13854794	0,02947415
rno-miR-181b-5p	-0,45345912	0,0327813
rno-miR-153-5p	0,641062851	0,0367054
rno-miR-143-3p	0,45094286	0,03966151
rno-miR-873-5p	0,657960871	0,03979341
rno-miR-509-3p	5,029886578	0,04405189
rno-miR-7b	0,465235524	0,04817786

**Tableau 43:Résultat de DESeq2 pour le Cortex du groupe 6OHDA**

Comparaison côté Gauche VS Droit pour le groupe 6-OHDA		
Nom de miRNA	log2FoldChange	pvalue
rno-miR-26a-3p	1,884684696	9,2904E-05
rno-miR-374-5p	0,596102037	0,00036001
rno-miR-127-5p	-0,37155006	0,00507088
rno-miR-7a-2-3p	0,56568338	0,00834454
rno-miR-421-3p	0,50145443	0,0142131
rno-miR-138-1-3p	-0,96943352	0,02266215
rno-miR-3572	1,155086776	0,02618539
rno-miR-1247-5p	1,567536037	0,02818376
rno-miR-219a-5p	-0,63802258	0,03112085
rno-miR-34b-3p	0,74157221	0,03213218
rno-miR-3068-3p	-0,42019057	0,03661708
rno-miR-122-3p	4,133418732	0,04292495
rno-miR-1306-5p	0,746896414	0,04453591
rno-miR-181b-1-3p	-0,68805449	0,04751777
rno-miR-19a-3p	1,077271567	0,0496821

**Tableau 44:Résultat de DESeq2 pour le groupe 6OHDA coté Droit VS coté Gauche**

Comparaison Striatum Gauche VS Droit pour le groupe 6-OHDA		
Nom de miRNA	log2 (FoldChange)	p-value
rno-miR-1247-5p	1,81183721	2,0296E-05
rno-miR-374-5p	0,78204721	2,1816E-05
rno-miR-7a-2-3p	0,88473649	0,0001125
rno-miR-376a-5p	-0,7228188	0,00016782
rno-miR-3068-3p	-0,687886	0,00022079
rno-let-7d-3p	0,62369021	0,00022667
rno-miR-1306-5p	1,17181647	0,00046702
rno-miR-379-5p	-0,5819823	0,0010889
rno-miR-24-2-5p	-0,639138	0,00110702
rno-miR-375-3p	1,64335568	0,00137013
rno-miR-666-3p	-0,6878272	0,00173176
rno-miR-24-3p	-0,4992104	0,00190083
rno-miR-411-5p	-0,5179919	0,00190101
rno-miR-361-5p	0,48469392	0,00324161
rno-miR-93-5p	0,48010374	0,00336481

**Tableau 45: Résultat de DESeq2 pour le striatum du groupe 6OHDA**

- **Groupe Alpha-Syn**

En se basant sur DESeq2 exécuté à travers le logiciel R afin de détecter l'expression différentielle entre différents groupes (Love, Huber, et Anders 2014) et différentes structures.

Pour chaque miRNAs et chaque comparaison de structure, on obtient une valeur de log2 (Foldchange) et de p-value.

Les miRNAs sélectionnées à partir des résultats du volcano plot pour le groupe Alpha-Syn sont générées dans les tableaux ci-dessous :



Comparaison Striatum Gauche VS Droit pour le groupe Alpha-Syn		
	log2FoldChange	pvalue
rno-miR-375-3p	4,36795265	9,8879E-41
rno-miR-34c-5p	2,21903936	3,7822E-17
rno-miR-34b-3p	2,2444119	5,0537E-07
rno-miR-1247-5p	1,41893734	1,0885E-05
rno-miR-10a-5p	1,04669884	0,00049901
rno-miR-34c-3p	2,57462235	0,00053838
rno-miR-7a-5p	0,5187257	0,00145548
rno-miR-674-5p	-0,5842344	0,00351795
rno-miR-320-3p	-0,6124978	0,00385583

**Tableau 46:Résultat de DESeq2 pour le Striatum pour le groupe Alpha-Syn**

Comparaison Cortex Gauche VS Droit pour le groupe Alpha-Syn		
	log2FoldChange	pvalue
rno-miR-219a-2-3p	-0,6194823	5,8214E-07
rno-miR-152-3p	0,74116079	7,5442E-07
rno-miR-338-5p	-0,6367316	7,9833E-07
rno-miR-143-3p	0,67698327	1,8201E-06
rno-miR-148a-3p	0,62559662	1,2506E-05
rno-miR-338-3p	-0,5188382	0,00011812
rno-miR-542-3p	0,58539075	0,00028183
rno-miR-483-3p	2,0770272	0,000304
rno-miR-379-5p	0,48818505	0,00037865
rno-miR-483-5p	6,98798372	0,00074669
rno-miR-199a-3p	0,60034317	0,00196581
rno-miR-488-3p	0,45030245	0,00271341
rno-miR-219a-5p	-0,8026508	0,00273151
rno-miR-101b-3p	0,4161405	0,00418397

**Tableau 47:Résultat de DESeq2 pour le cortex pour le groupe Alpha-Syn**

Comparaison côté Gauche VS Droit pour le groupe Alpha-Syn		
	log2FoldChange	pvalue
rno-miR-375-3p	2,89459915	6,1052E-05
rno-miR-330-3p	-0,3272582	0,00904579
rno-miR-496-5p	-0,7921842	0,00988159
rno-miR-34c-3p	1,29879325	0,03515195
rno-miR-379-5p	0,36033193	0,03596941
rno-miR-342-5p	-0,4449124	0,03632067
rno-miR-211-5p	0,63527087	0,03681035
rno-miR-377-5p	-1,0253109	0,03844583
rno-miR-183-5p	0,68365926	0,03979869
rno-miR-34c-5p	0,98385109	0,04214996
rno-miR-338-5p	-0,3828658	0,04331184
rno-miR-29b-5p	-0,3289858	0,0446555
rno-miR-702-5p	-2,2242192	0,04666171

**Tableau 48: Résultat de DESeq2 pour le groupe Alpha-Syn**

## 5. Validation des résultats par RT QPCR

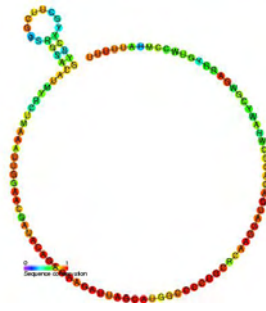
Dans notre étude de validation, on s'est focalisé sur les miRNAs existant et identifié par RNAseq su niveau du striatum qui la première structure qui subit la perte dopaminergique.

Pour confirmer l'expression de certains miRNA révélés par RNA-Seq, 15 miRNAs pour chaque groupe (tableau 45 et 46) ont été choisi en fonction de la p-value et le log2 (Fold change) obtenues par DESeq2, pour quantifier leur expression e utilisant la RT QPCR.

Une quantité de 1µg d'ARN total de chaque échantillon a été soumise à une transcription inverse en ADNc conformément aux instructions du fabricant en utilisant le kit *miScript II RT* (Qiagen) en utilisant le tampon *HiFlex* dans un *thermocycleur* (Eppendorf).

Une quantité de 5 ng d'ADNc par puits a été utilisée pour effectuer la qPCR, en utilisant des amorces spécifiques de chaque miRNA en combinaison avec une amorce universelle.

L'ARN U6 a été utilisé comme contrôle interne, c'est un petit ARN nucléaire (small nuclear RNA ou snRNA, en anglais) entrant dans la composition des petites ribonucléoprotéines nucléaires U6 (figure 66). U6 pour RT-PCR: CGCTTCACGAATTTGCGTGTCAB (L. Yang et al. 2014).



**Figure 66:snRNA U6**

La RT QPCR a été réalisé en trois exemplaire par échantillon (Triplicata), en suivant les instructions du fabricant, qui suivent ces étapes :

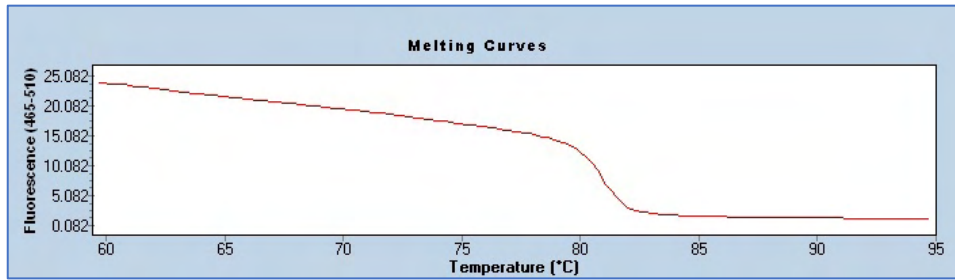
La pré-incubation : 1 cycle à 95 C° pour 30 secondes.

L'amplification : 55 cycles, le premier cycle à 95 C° pour 5 secondes, le reste à 60 C° pour 30 secondes.

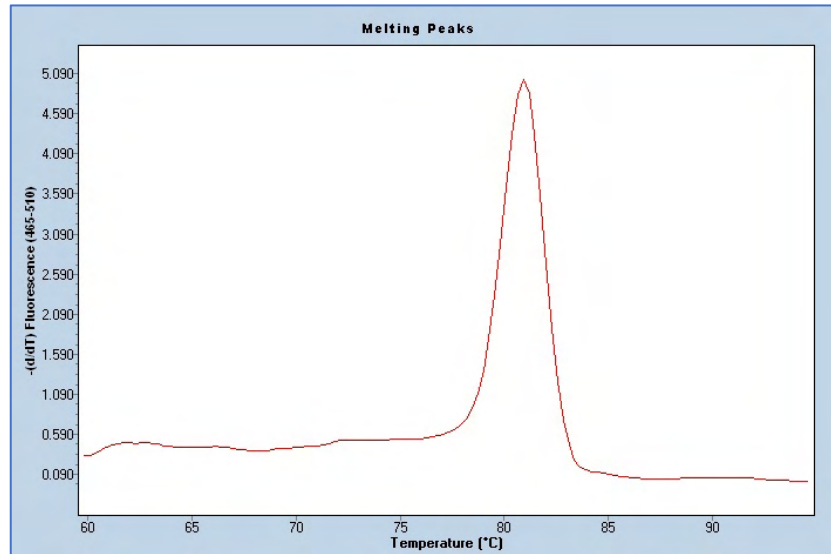
Une analyse de la courbe de dissociation du ou des produits de la PCR peut éventuellement être réalisée pour aider à vérifier leur spécificité et leur identité. L'analyse des courbes de dissociation est une étape d'analyse intégrée au logiciel des *real-time cyclers*.

Pour effectuer une analyse de courbe de dissociation, la température est augmentée très lentement d'une température basse (par exemple 65 ° C) à une température élevée (par exemple 95 ° C). A basse température, les produits de PCR sont double brin, donc *SYBR Green I* le colorant se lie à eux et la fluorescence est élevée. Cependant, à des températures élevées, les produits de PCR sont dénaturés, ce qui entraîne une diminution rapide de la fluorescence.

La fluorescence est mesurée en continu à mesure que la température augmente et que les valeurs de fluorescence sont tracées en fonction de la température. Une courbe se forme car la fluorescence diminue légèrement dans la partie inférieure de la plage de températures, mais diminue beaucoup plus rapidement à des températures plus élevées lorsque les températures de dissociation de produits de PCR non spécifiques et spécifiques sont atteintes. Les systèmes de détection calculent les premières dérivées des courbes, ce qui donne des courbes avec des pics aux Tms respectifs (figures 67 et 67). Les courbes avec des pics à une Tm inférieure à celle du produit de PCR spécifique indiquent la formation de primer-dimères, tandis que divers pics avec différents Tms ou plateaux indiquent la production de produits non spécifiques ou un frottis.



**Figure 67: Analyse de la courbe de dissociation**



**Figure 68: Courbe de dissociation des miRNA**

Analyse de la courbe de dissociation d'un produit de PCR miR-1247-3p montrant un seul pic du produit d'amplification spécifique. L'analyse de la courbe de dissociation a été réalisée à l'aide du *Rotor-Gene Q*<sup>56</sup>.

En se basant sur mirBase et mirDB<sup>57</sup> en précisant l'organisme en *Rattus norvegicus* (Norway rat) (PRJNA10629 sous NCBI) le tableau ci-dessus regroupe différentes caractéristiques des miRNAs utilisé pour la qPCR (tableau 49 pour 6-OHDA et tableau 50 pour Alpha-Syn) :

**Accession** : c'est l'accession number en anglais, il désigne un identifiant unique donné à toute séquence pour pouvoir y accéder via les banques de données biologiques.

<sup>56</sup> L'instrument Rotor-Gene Q est conçu pour effectuer un cyclage thermique en temps réel et final en utilisant la réaction en chaîne de la polymérase (PCR) et l'analyse de fusion à haute résolution dans les applications de biologie moléculaire ainsi que pour d'autres applications telles que la mesure de la concentration, analyse des protéines et cinétique enzymatique.

<sup>57</sup> miRDB est une base de données en ligne pour la prédiction de cibles miRNAs et les annotations fonctionnelles. Toutes les cibles de miRDB ont été prédites par un outil de bio-informatique, MirTarget, qui a été développé en analysant des milliers d'interactions miRNAs-cible à partir d'expériences de séquençage à haut débit

**Chromosome** : La localisation du miRNAs dans le chromosome ainsi que sa **position** et sa **séquence**.

**Cible** : nombre de cible prévues par miRDB pour le miRNAs.

**Séquences des miRNAs**

Nom du miRNA	Accession	Chromosome	Position	Séquence	Cibles
rno-mir-1247	MI0030348	Chromosome 6	134626097-134626178	ACCCGUCCCGUUCGUCCCCGGA	63
rno-mir-374-5p	MI0003552	Chromosome X	74443648-74443715	AUAUAAUACAACCUGCUAAGUG	356
rno-mir-7a-2-3p	MI0000836	Chromosome 1	140576396-140576490	UGGAAGACUAGUGAUUUUGUUGU	242
rno-mir-376a-5p	MI0003545	Chromosome 6	133873605-133873686	GGUAGAUUCUCCUUCUAUGAG	115
rno-mir-3068-3p	MI0021534	Chromosome 6	111468732-111468807	GGUGAAUUGCAGUACUCCAACA	179
rno-let-7d-3p	MI0000601	Chromosome 17	16419980-16420077	AGAGGUAGUAGGUUGCAUAGUU	15
rno-mir-1306-5p	MI0021537	Chromosome 11	86857633-86857705	CCACCUCCCCUGCAAACGUCCA	147
rno-mir-379-5p	MI0003541	Chromosome 6	133857700-133857784	UGGUAGACUAUGGAACGUAGG	110
rno-mir-24-2-5p	MI0000855	Chromosome 19	25318863-25318970	GUGCCUACUGAGCUGAAACAGU	443
rno-mir-375-3p	MI0006140	Chromosome 9	82161716-82161790	GCGACGAGCCCCUCGCACAAACC	5
rno-mir-24-3p					
rno-mir-411-5p	MI0006144	Chromosome 6	133858849-13385892	UAGUAGACCGUAUAGCGUACG	100
rno-mir-361-5p	MI0003481	Chromosome X	84708911-84708980	UUAUCAGAAUCUCCAGGGGUAC	203
rno-mir-666-3p	MI0012617	Chromosome 6	133866523-133866621	AGCGGGCACGGCUGUGAGAG	16
rno-mir-93-5p	MI0000880	Chromosome 12	19307543-19307629	CAAAGUGCUGUUCGUGCAGGUAG	252

**Tableau 49:Description des miRNAs étudiés en QPCR pour groupe 6OHDA**

Nom du miRNA	Accession	Chromosome	Position	Séquence	Cible
rno-miR-375-3p	MIMAT0005307	Chromosome 9	82161716-82161790	GCGACGAGCCCCUCGCACAAACC	11
rno-miR-34c-5p	MI0000876	Chromosome 8	55492024-55492100	AAUCACUAACCACACAGCCAGG	383
rno-miR-34b-3p	MI0000875	Chromosome 8	55492542-55492625	AGGCAGUGUAAAUAGCUGAUUGU	385
rno-miR-1247-5p	MI0030348	Chromosome 6	134626097-134626178	ACCCGUCCCCGUUCGUCCCCGGA	63
rno-miR-10a-5p	MI0000841	Chromosome 10	84165950-84166059	UACCCUGUAGAUCGAAUUUGUG	58
rno-miR-34c-3p	MI0000876	Chromosome 8	55492024-55492100	AAUCACUAACCACACAGCCAGG	119
rno-miR-7a-5p	MI0000836	Chromosome 1	140576396-140576490	UGGAAGACUAGUGAUUUUGUUGU	398
rno-miR-674-5p	MI0006159	Chromosome 3	108864462-108864538	GCACUGAGAUGGGAGUGGUGUA	332
rno-miR-320-3p	MI0000972	Chromosome 5	52134276-52134357	GCCUUCUCUCCCCGGUUCUCC	242
rno-miR-379-5p	MI0003541	Chromosome 6	133857700-133857784	UGGUAGACUAUGGAACGUAGG	110
rno-miR-23a-5p	MI0000852	Chromosome 19	25318582-25318656	GGGGUUCCUGGGGAUGGGAUUU	137
rno-miR-183-5p	MI0000928	Chromosome 4	57463569-57463678	UAUGGCACUGGUAGAAUUCACU	228
rno-miR-201-5p	MI0012608	Chromosome X	155339140-155339219	CACUCAGUAAGGCAUUGUUC	281
rno-miR-184	MI0000929	Chromosome 8	97175657-97175733	UGGACGGAGAACUGAUAAAGGGU	42
rno-miR-483-3p	MI0003485	Chromosome 1	215832088-215832160	AAGACGGGAGAAGAGAAGGGAG	33

**Tableau 50:Description des miRNAs étudiés en QPCR pour groupe Alpha-Syn**

### 5.1. Analyse statistique du résultat RT PCR

Pour obtenir une quantification comparative, nous avons mesuré le changement relatif des niveaux d'expression des miRNAs dans les échantillons. Une expérience de RT PCR exprime une expression relative, c'est-à-dire une variation de l'expression d'un miRNA entre deux échantillons.

Une expression de miRNA peut être relative à :

- Un contrôle, par ex. un miRNA de référence exprimé constant (snRNA U6)
- Un calibrateur c'est l'échantillon auquel tous les autres sont comparés. C'est l'échantillon « non traité » ou « temps zéro ». Le RQ du calibrateur donne 1 car il ne varie pas par rapport à lui-même.
- Cycle threshold : Ct Valeur à laquelle la courbe PCR croise le seuil (Figure 68)

Pour évaluer les résultats de la RT-PCR, nous avons calculé :

$$\begin{aligned}\text{Delta Ct} &= \text{Ct miRNA} - \text{Ct contrôle} \\ \text{Delta Delta Ct} &= \Delta\text{Ct traité} - \Delta\text{Ct Contrôle} \\ \text{Fold Change} &= \text{RQ} = \text{Quantification relative} = 2^{-\Delta\Delta\text{Ct}}\end{aligned}$$

Pour évaluer les résultats de la RT-PCR, le test Kruskal-Walis a été réalisé avec R sur les valeurs de  $\Delta\text{Ct}$  pour évaluer la significativité de l'expression des miRNAs et identifier les miRNAs sous-exprimé (down regulated) ou sur-exprimé (up regulated). Le niveau de signification a été fixé à  $p < 0,05$  ( Figure 69 et Figure 70).

Pour le groupe 6-OHDA :

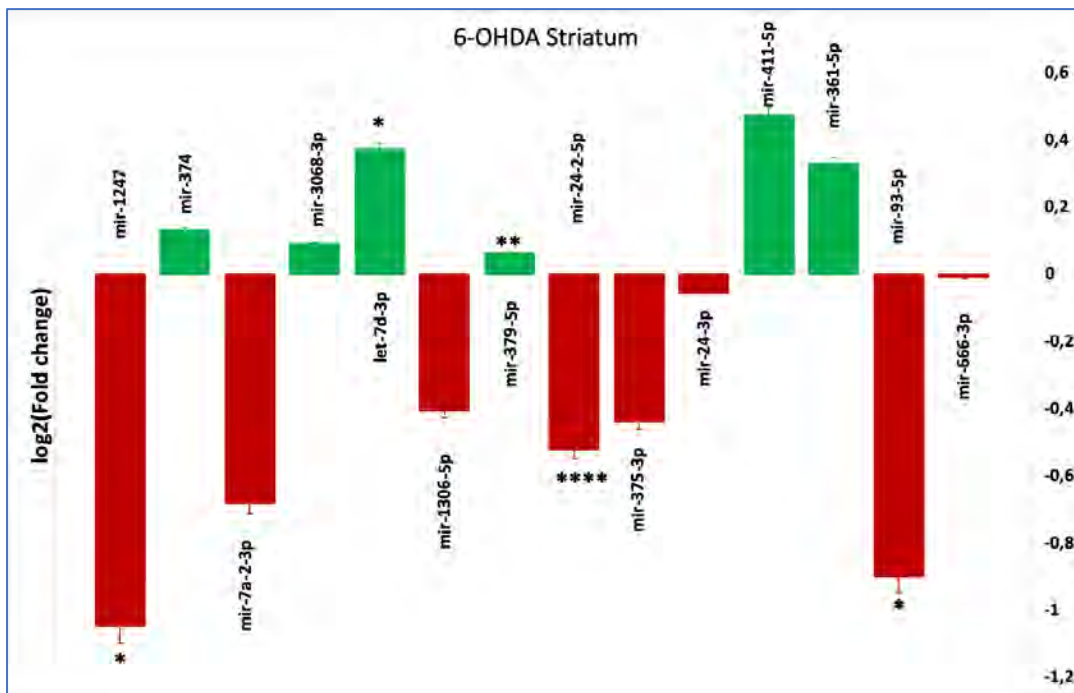


Figure 69:La différence d'expression des miRNAs dans le groupe 6-OHDA en utilisant un RTPCR

Les miRNAs sur exprimées sont en Vert et les sous exprimées en Rouge. Pour les symboles p-value : ns :  $p > 0,005$  ; \* :  $P \leq 0.05$  ; \*\* :  $P \leq 0.01$  et \*\*\*\* :  $P \leq 0.0001$

Pour le groupe Alpha-Syn :

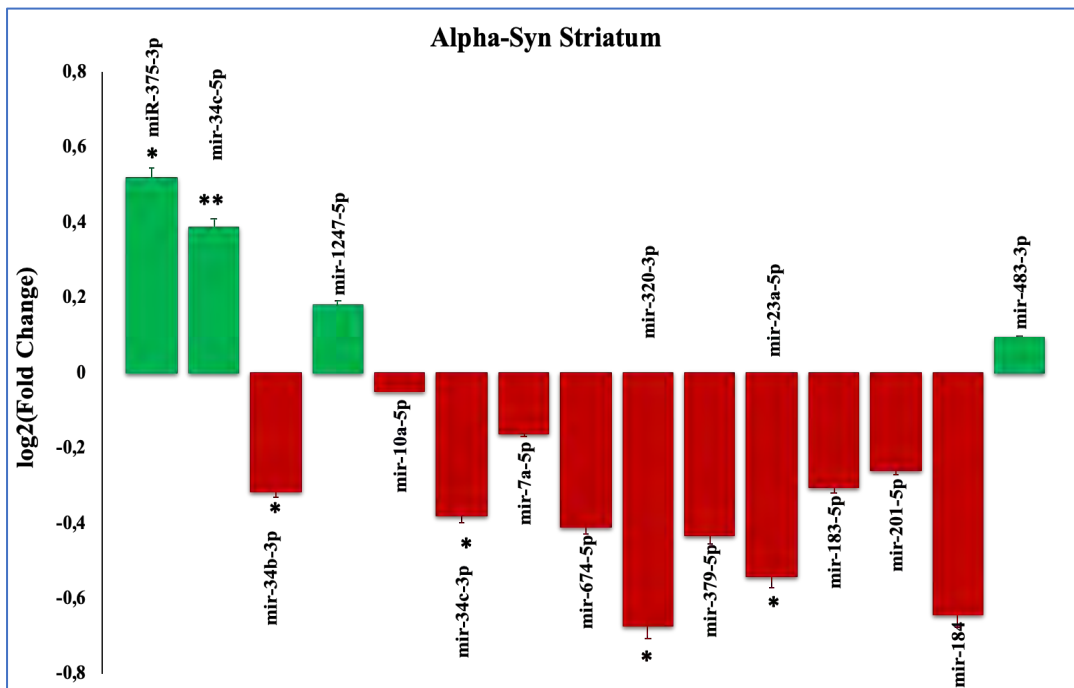


Figure 70:La différence d'expression des miRNAs dans le groupe Alpha-Syn en utilisant un RTPCR

Les miRNAs sur exprimées sont en Vert et les sous exprimées en Rouge. Pour les symboles p-value : ns :  $p > 0,005$  ; \* :  $P \leq 0.05$  ; \*\* :  $P \leq 0.01$  et \*\*\*\* :  $P \leq 0.0001$



## 5.2. Prédiction des gènes

Un miRNA interagit avec plusieurs centaines de sites répartis dans le transcriptome codant d'une cellule, et un ARNm est potentiellement ciblé par plusieurs miRNA différents, soulignant la complexité du réseau de régulation par les miRNAs. Les miRNAs régulent l'expression des gènes au niveau post-transcriptionnel par un appariement complémentaire des bases avec l'ARNm cible, conduisant à la dégradation de l'ARNm et au blocage de la traduction (Riffo-Campos, Riquelme, et Brebi-Mieville 2016). Pour une prédiction des gènes cible, les résultats obtenus par le test de Kruskal-Wallis pour évaluer la significativité de l'expression des miRNAs et identifier les miRNAs sous-exprimé (down regulated) ou surexprimé (up regulated) parmi les 15 miRNAs utilisés en RTPCR sont utilisés. Le tableau 51 contient les miRNAs sélectionnée en se basant sur la p-value < 0,05 pour faire la prédiction des gènes cibles.

Groupe 6-OHDA		Groupe Alpha-Syn	
miRNA	pvalue	miRNA	pvalue
rno-miR-1247-5p	0,0254	rno-miR-375-3p	0,01121
rno-miR-let-7d-3p	0,04313	rno-miR-34c-5p	0,005499
rno-miR-379-5p	0,006877	rno-miR-34b-3p	0,01659
rno-miR-24-2-5p	2,13E-05	rno-miR-34c-3p	0,02895
rno-miR-93-5p	0,01697	rno-miR-320-3p	0,0191
		rno-miR-23a-5p	0,02641

**Tableau 51: miRNAs sélectionnés pour les deux groupes**

Pour la prédiction des gènes cibles de ces miRNAs sélectionnées, l'outil TargetScan a été utilisé. TargetScan prédit les cibles biologiques des miRNAs en recherchant la présence de sites 8mer, 7mer et 6mer conservés qui correspondent à la région de la graine de chaque RNA (Lewis, Burge, et Bartel 2005), chez les mammifères, les prévisions sont classées en fonction de l'efficacité prévue du ciblage, calculée à l'aide des scores cumulés contextuels ++ des sites (Agarwal et al. 2015). Lors, de la recherche du nom d'un miRNA, les résultats montrent les gènes cibles en fonction du transcrit spécifique. Dans ce cas, la probabilité peut être classée par Weighted context++ score ou uniquement avec les sites conservés. Pour comprendre les résultats, il est nécessaire de comprendre chacun de ces paramètres (tableau 52).

Paramètre	Valeur	Signification
Site type	8mer > 7mer > 7mer-A1 > 6mer	Les sites correspondants dans la région de la graine (nucléotides 2 à 8 de 5' de miRNAs qui ont un appariement parfait de WC avec le 3' UTR), du plus strict au moins strict.
Context++ score	De 1 to -1	La somme de la contribution de 14 caractéristiques pour chacun des quatre types de sites, plus le score est négatif, plus le refoulement est important.
Context++ score percentile	From i to 100 - i;	Pourcentage de sites pour le miRNA dont le score de contexte est moins favorable ++.
Weighted context++ score	From 1 to -1	Les scores avec une valeur négative inférieure indiquent une prédiction plus grande de la répression.
Cumulative weighted context++ score	$C(i-1) + (1 - 2CSi) (AIRi - C(i-1))$	Ce score estime la répression totale attendue de plusieurs sites du même miRNA, pour chaque cible d'ARNm prédit.
Branch-length score	8mer: 1.8; 7mer-m8: 2.8; 7mer-A1: 3.6; 6mer: NA	Ce score est la somme des longueurs de branches phylogénétiques entre espèces contenant un site correspondant.
P <sub>CT</sub> score	Between 0 and 1	Plus le score est élevé, plus la conservation et la déstabilisation de l'ARNm attendues sont grandes.
Aggregate P <sub>CT</sub>	Value = $1 - ((1 - P_{CT}) \text{ site1} \times (1 - P_{CT}) \text{ site2})$	Pour chaque miRNA ce paramètre inclut les cibles 3'UTR conservées avec plusieurs sites manquants dans l'annotation 3'UTR humaine, mais présentes dans les annotations de souris.
Conserved sites	$\geq 0$	Nombre de sites conservés identifiés.

**Tableau 52: Les paramètres de TargetScan**

- **Groupe 6-OHDA**

Pour chaque miRNAs, nous avons sélectionné les 10 premiers gènes cibles en se basant sur le Context ++ Scor obtenu par TargetScan (tableau 53) . (Les gènes **en gras** sont déjà décrits dans la littérature comme gène impliqué dans le MP)

<b>Rno-miR-1247-5p</b>		
<b>Gène cible</b>	<b>Identifiant</b>	<b>Weighted context++ score</b>
LYRM5	ENST00000556402.1	-0,97
SLC34A1	ENST00000324417.5	-0,83
CEP152	ENST00000380950.2	-0,83
ADAM15	ENST00000360674.4	-0,81

SOX1	ENST00000330949.1	-0,81
PGF	ENST00000555567.1	-0,79
TCF23	ENST00000296096.5	-0,78
GNG10	ENST00000374293.4	-0,75
TCEB3	ENST00000418390.2	-0,75
ARFGAP3	ENST00000263245.5	-0,75
<b>Rno-miR-let-7d-3p</b>		
ZFP37	ENST00000374227.3	-1,16
ZIK1	ENST00000307468.4	-0,86
HTN1	ENST00000511674.1	-0,82
ASB11	ENST00000537676.1	-0,6
PPAPDC1B	ENST00000422581.2	-0,58
TOMM5	ENST00000544379.1	-0,58
<b>CAV1</b>	ENST00000405348.1	-0,58
TIMM8B	ENST00000504148.2	-0,58
<b>SOX1</b>	ENST00000330949.1	-0,56
<b>PTH</b>	ENST00000282091.1	-0,55
<b>Rno-miR-379-5p</b>		
EIF4G2	ENST00000526148.1	-0,61
UBE2E3	ENST00000410062.4	-0,61
FAM222B	ENST00000582266.1	-0,56
ROR1	ENST00000371079.1	-0,53
RPA3-AS1	ENST00000406829.1	-0,44
ZBTB26	ENST00000373656.3	-0,41
<b>SEMA3A</b>	ENST00000265362.4	-0,41
CYB561D1	ENST00000496961.1	-0,37
RAB30	ENST00000533486.1	-0,36
<b>SLC20A1</b>	ENST00000272542.3	-0,35
<b>Rno-miR-24-2-5p</b>		
IGIP	ENST00000333305.3	-1,07
AC007421.1	ENST00000593717.1	-0,92
KCNIP2	ENST00000348850.5	-0,89
RAP2A	ENST00000245304.4	-0,86

RNASE7	ENST00000298690.4	-0,73
STYXL1	ENST00000360591.3	-0,64
CALD1	ENST00000361388.2	-0,59
<b>FMR1</b>	ENST00000370471.3	-0,58
VWA1	ENST00000338660.5	-0,57
ACKR4	ENST00000249887.2	-0,56
<b>Rno-miR-93-5p</b>		
MARCH4	ENST00000273067.4	-0,84
PTHLH	ENST00000395872.1	-0,73
ZNF1	ENST00000371752.1	-0,68
NKIRAS1	ENST00000388759.3	-0,67
BTG3	ENST00000339775.6	-0,67
DDHD1	ENST00000323669.5	-0,67
PLEKHA3	ENST00000234453.5	-0,65
ZBTB7A	ENST00000322357.4	-0,63
FAM45A	ENST00000361432.2	-0,63
PRR15	ENST00000319694.2	-0,62

**Tableau 53:réduction des gènes par TargetScan pour le groupe 6-OHDA**

- **Groupe Alpha-Syn**

Pour chaque miRNAs, nous avons sélectionné les 10 premiers gènes cibles en se basant sur le Context ++ Scor obtenu par TargetScan (tableau 54). (Les gènes en gras sont déjà décrits dans la littérature comme gène impliqué dans le MP)

<b>Rno-miR-375-3p</b>		
<b>Gène cible</b>	<b>Identifiant</b>	<b>Weighted context++ score</b>
ELAVL4	ENST00000371824.1	-0,57
C3orf38	ENST00000486971.1	-0,5
PAX2	ENST00000370296.2	-0,41
ARL4C	ENST00000390645.2	-0,41
SLC16A2	ENST00000587091.1	-0,36
SHOX2	ENST00000490689.2	-0,35
FZD8	ENST00000374694.1	-0,33
MXI1	ENST00000332674.5	-0,32
ZFP36L2	ENST00000282388.3	-0,31
CCDC39	ENST00000273654.4	-0,29
<b>Rno-miR-34c-5p</b>		
<b>SCN2B</b>	ENST00000278947.5	-1,63
HCN3	ENST00000368358.3	-1,49
TRANK1	ENST00000428977.2	-1,4
MDM4	ENST00000391947.2	-1,24
<b>SYT1</b>	ENST00000457153.2	-1,21
<b>VAMP2</b>	ENST00000404970.3	-1,16
GREM2	ENST00000318160.4	-1,06
FAM76A	ENST00000373954.6	-0,96
NAV3	ENST00000228327.6	-0,92
PACS1	ENST00000320580.4	-0,91
<b>Rno-miR-34b-3p</b>		
UBL4A	ENST00000369660.4	-0,81
PIGP	ENST00000360525.4	-0,66
AC004899.1	ENST00000596947.1	-0,63

FAM150A	ENST00000358543.4	-0,63
KRTAP19-1	ENST00000390689.2	-0,62
TMEM47	ENST00000275954.3	-0,6
MED7	ENST00000286317.5	-0,59
TNFSF18	ENST00000404377.3	-0,57
LYSMD2	ENST00000267838.3	-0,55
TMEM210	ENST00000413619.2	-0,55
<b>Rno-miR-320-3p</b>		
IFT27	ENST00000340630.5	-0,73
STAT4	ENST00000358470.4	-0,69
TFRC	ENST00000540528.1	-0,58
RP11-160N1.10	ENST00000377548.2	-0,51
ARPP19	ENST00000566423.1	-0,51
PBX3	ENST00000342287.5	-0,51
RCN2	ENST00000394885.3	-0,48
GABRP	ENST00000519385.1	-0,47
YWHAH	ENST00000397492.1	-0,45
EREG	ENST00000244869.2	-0,45
<b>Rno-miR-23a-5p</b>		
GIPC3	ENST00000322315.5	-1,23
RP11-410N8.4	ENST00000375670.1	-1,07
MRPL52	ENST00000397505.2	-1,01
<b>GPX3</b>	ENST00000517973.1	-0,9
TMEM140	ENST00000275767.3	-0,84
RP11-362K2.2	ENST00000546977.1	-0,81
<b>APLN</b>	ENST00000307484.6	-0,73
SYNGR2	ENST00000589711.1	-0,69
EIF2D	ENST00000367114.3	-0,68
TMEM114	ENST00000568335.1	-0,65

**Tableau 54: Prédiction des gènes par TargetScan pour le groupe Alpha-Syn**

# Discussion de la partie de la MP

La MP est une maladie neurodégénérative progressive causée principalement par un manque de dopamine dans le cerveau. La dopamine est un neurotransmetteur impliqué dans le mouvement, la motivation, la mémoire et d'autres fonctions ; son niveau diminue dans le cerveau à la suite de la mort de cellules dopaminergiques. La perte de dopamine dans le cerveau du parkinsonien est une cause de déficience motrice et peut-être une des causes du déficit cognitif observé chez certains patients atteints de MP. La MP n'est généralement pas identifiée à ses débuts en raison d'une longue latence entre les premiers dommages causés aux cellules dopaminergiques et l'apparition des symptômes cliniques.

Bien que la MP soit un trouble neurodégénératif répandu, il n'existe aucun traitement disponible pour la guérir complètement. Les traitements pharmacologiques et chirurgicaux existants ne peuvent qu'améliorer les symptômes de la maladie.

La majorité des cas de MP surviennent de manière sporadique en raison de l'interaction de facteurs environnementaux et génétiques. Au moment où les symptômes se manifestent cliniquement, près de 70 à 80% des neurones dopaminergiques ont déjà dégénéré. Cela limite la durée de l'efficacité des traitements pharmacologiques. Par conséquent, il existe un fort besoin de comprendre les voies / mécanismes moléculaires impliqués dans la pathogenèse de la MP, ce qui permettrait de développer de nouvelles approches thérapeutiques neuroprotectrices. À cette fin, l'identification de biomarqueurs moléculaires tels que miRNAs impliqués dans la manifestation et la progression de la MP offre une alternative intéressante.

Notre travail met donc en évidence les différents miRNAs qui peuvent jouer un rôle important dans la pathogenèse de la MP. Par conséquent, l'identification de ces biomarqueurs moléculaires fiables permettra de distinguer la MP des autres affections apparentées, de suivre son évolution ou de donner une indication d'une réponse positive à une intervention thérapeutique. Bien que le terme "biomarqueur" soit souvent utilisé indifféremment pour décrire toute modification de l'expression des gènes ou des protéines, le groupe de travail sur les définitions des biomarqueurs du NIH le définit plus précisément comme une caractéristique objectivement mesurée et évaluée comme réponses pharmacologiques à une intervention thérapeutique » (Groupe de travail sur les définitions des biomarqueurs, 2001).

Plusieurs excellents rapports ont examiné comment utiliser et qualifier des biomarqueurs cliniques (Freeman et al. 2010). Les biomarqueurs de la MP pourraient par exemple être utilisés pour diagnostiquer la MP (marqueurs de diagnostic), prédire le risque de MP ou de progression de la maladie (marqueurs pronostiques), décrire la gravité de la maladie (marqueurs de

stadification) et appuyer le choix du traitement (marqueurs theragnostiques). Les marqueurs diagnostiques peuvent être utiles pour reconnaître la MP avant que les symptômes moteurs ne deviennent évidents ou lorsque les signes moteurs ou non moteurs ou les deux sont encore insuffisants pour définir la maladie (phase prodromique) ou même pour détecter une population asymptomatique présentant un risque de MP chez laquelle une neurodégénérescence devrait se produire (phase préclinique). Les marqueurs de diagnostic pourraient également aider à différencier la MP des autres syndromes parkinsoniens, dans la mesure où le diagnostic erroné est souvent précoce et que la confirmation du diagnostic nécessite des rapports d'autopsie. La difficulté d'identifier les critères de diagnostic précoce de la MP dépend du fait qu'aucun biomarqueur réel ne peut encore prédire l'apparition de la maladie.

Le modèle le plus fréquemment utilisé de la MP chez le rat repose sur des lésions unilatérales partielles ou totales de la SNpc (Chetrit et al. 2013) (Breit et al. 2001). Ce modèle dit hémiparkinsonien reproduit une atteinte unilatérale de lésion DAergique et des activités locomotrices, surtout pour des atteintes sévères (> 90%). Le modèle hémiparkinsonien a l'avantage de se servir du côté controlatéral à la lésion comme hémisphère contrôle. De ce fait lors de la mise en place de nos modèles nous avons choisi de réaliser des lésions 6-OHDA (6-hydroxydopamine) unilatérales, recevant une injection de 6-OHDA dans le faisceau médian du télencéphale (*MFB, median forebrain bundle*), et des lésions à l'Alpha-synucléine mutée associée à un vecteur viral dans la SNpc, afin de reproduire, de la manière la plus comparable possible, les atteintes lésionnelles du système DAergique rencontrées dans la MP, en particulier une dégénérescence plus progressive et partielle.

Par ailleurs, il a été montré que les lésions unilatérales de 6-OHDA provoquent des déficits moteurs à partir de la deuxième semaine après la lésion (Charles et al. 2018) En se basant sur le test de cylindre avec le comptage du pourcentage d'utilisation des pattes des deux côtés, nous avons obtenu une différence significative montrant la présence d'une bradykinésie du côté controlatéral à la lésion dopaminergique par rapport au côté ipsilatéral. De plus la comparaison du seuil de sensibilité mécanique entre la patte ipsilatérale (patte droite) et controlatérale (patte gauche) dans le test de Von Frey, nous avons montré une diminution significative du seuil de retrait de la patte controlatérale, ce qui signifie l'apparition d'une allodynie à partir de la troisième semaine après la chirurgie.

Pour le groupe Alpha-syn, les lésions ont conduit à des déficits moteurs progressifs en comparant les résultats du test de cylindre deux semaines ( $p = 0,002$ ) et quatre semaines ( $P < 0,0001$ ) après la lésion, ce qui montre l'apparition d'une bradykinésie du côté controlatéral à la lésion dopaminergique de manière progressive. Ainsi, pour la sensibilité mécanique utilisant



le test de von Frey: 22, 36 et 50 jours après la chirurgie, les résultats ont montré une diminution progressive du seuil de retrait de la patte controlatérale par rapport au seuil de retrait de la patte ipsilatérale. Ces résultats montrent un développement progressif de l'allodynie mécanique chez le groupe alpha-syn en fonction du temps.

Concernant la validation biochimique des modèles parkinsoniens, nos résultats indiquent une déplétion sévère de la dopamine au niveau de la partie dorsal et ventral du striatum avec un pourcentage de perte entre le coté ipsilatéral et controlatéral qui atteint 90% et 65% respectivement pour le groupe 6-OHDA et le groupe alpha-syn, sans affecter le contenu tissulaire en NA et 5-HT. Ces résultats montrent la sélectivité de la déplétion dopaminergique chez nos deux modèles animaux.

Après la caractérisation et la validation de nos deux modèles animaux, notre travail a permis d'étudier le miRnome<sup>58</sup> dans le striatum en utilisant la nouvelle technologie de séquençage dédiée pour l'ARN intitulée RNASeq. Cette technologie permet d'analyser l'expression des espèces de miRNAs ainsi que la différence d'expression des miRNAs spécifiques dans cette structure étudiée. Pour atteindre cet objectif, l'algorithme miRPro (Emamzadeh and Surguchov, 2018) a été utilisé, ce dernier se base sur l'algorithme majeur de miRDeep2 (Friedländer et al. 2012).

Malgré les améliorations apportées au programme miRPro dans la détection des miRNAs, la quantité de faux-positifs reste conséquente. Pour réduire le nombre d'erreurs dans la détection des miRNAs détectés, nous avons appliqué un filtre supplémentaire tel que la vérification du nombre de *reads* supérieur à 1 cpm (*count per million of sequences*). Néanmoins, la prédiction des miRNAs issus des analyses bio-informatiques d'échantillons de RNASeq reste insuffisante pour prouver une existence réelle. Ainsi, pour vérifier l'existence de ces miRNAs, nous avons utilisé la RTqPCR. Nos résultats montrent une différence significative de l'expression de certains miRNAs entre les différents groupes étudiés.

#### **A. Synthèse de la littérature**

Plusieurs publications ont démontré la différence d'expression de certains miRNAs

miRNA Target	Références
miR-155	(Thome et al. 2016)

---

<sup>58</sup>Le miRnome correspond à l'ensemble des espèces de miRNAs exprimées ainsi que leur quantité, lors d'un stade de développement défini ou d'une condition physiologique particulière, dans un tissu ou une structure biologique d'intérêt.

miR-16-1	(Z. Zhang et Cheng 2014)
miR-7/153	(Junn et al. 2009)
miR-133b	(J. Kim et al. 2007)
miR-133b	(N. Zhao et al. 2014)
miR-124	(A.-D. Li et al. 2019)
miR-1	(Margis, Margis, et Rieder 2011)
miR-30a	
miR-26a2	
miR-22	
miR-29a	
<b>miR-34b/c</b>	(Kabaria et al. 2015)
miR-16-2	(Margis, Margis, et Rieder 2011)
miR-126	(W. Kim et al. 2016)
<b>miR-184</b>	(Gehrke et al. 2010)
miR-16-5p/127-5p	(Hoss et al. 2016)
miR-221	(Asci et al. 2013)
miR-384-5p	(Jiang et al. 2016)
<b>miR-7</b>	(Kabaria et al. 2015) (Chaudhuri et al. 2016)
miR-127-5p	(Hoss et al. 2016)
miR-16-5p	
miR-205	(Cho et al. 2013)
<b>mir-224/320/373/379</b>	(Alvarez-Erviti et al. 2013)
miR-26b/106a	
miR-132	(D. Yang et al. 2012)

**Tableau 55: Les miRNAs impliqués dans la MP et décrits dans la littérature**

Les miRNAs **en gras** sont les miRNAs identifiés dans notre travail.

### **B. Résultats expérimentaux de notre étude**

Récapitulatif des miRNAs détectés et qui sont 4 fois plus exprimés avec une p-value < 0,05 dans le striatum des deux groupes de modèles animaux.

Les miRNAs noté en gras sont ceux que notre étude confirme par rapport à la littérature existante.

- rno-miR-1247-5p
- **rno-miR-let-7d-3p**
- **rno-miR-379-5p**
- rno-miR-24-2-5p
- rno-miR-93-5p
- rno-miR-375-3p
- **rno-miR-34c-5p**
- **rno-miR-34b-3p**
- **rno-miR-34c-3p**
- **rno-miR-320-3p**
- rno-miR-23a-5p

Les différents miRNAs jouant un rôle dans l'apparition et la progression de la MP offre une alternative intéressante. Certains miRNA tels que miR-7/153 se sont révélés être directement impliqués dans la pathogenèse de la MP, car ils inhibent l'agrégation des SNCA, tandis que plusieurs autres ont été indirectement impliqués dans cette maladie. Il existe également de nombreuses cibles prédites de divers miRNAs qui doivent être validées dans des modèles animaux ainsi que chez l'homme. Plusieurs miRNAs trouvés avec un potentiel de biomarqueur, qui pourraient aider à améliorer le diagnostic de la MP.

Les recherches scientifiques suggèrent que des altérations de l'expression de miRNAs dans le cerveau sont impliquées dans la pathogenèse de MP (Miñones-Moyano et al. 2011).

Minones-Moyano et al. a initialement signalé la diminution de l'expression de miR-34b et miR-34c dans la substance noire, ce qui a été confirmé par une recherche scientifique récente (Villar-Menéndez et al. 2014). L'inhibition de miR-34b et miR-34c dans des cellules de neuroblastome SH-SY5Y différenciées a entraîné une perte du potentiel de membrane mitochondriale et un stress oxydatif, qui sont des anomalies biochimiques associées à la MP. De plus, il a été démontré que les niveaux d'expression de la parkin et de DJ-1 étaient légèrement diminués lors

de l'inhibition de l'activité de miR-34b et de miR-34c. Compte tenu du fait que l'expression des gènes cibles devrait augmenter lorsqu'on inhibe leurs récepteurs respectifs, il est probable que miR-34b et miR-34c ne ciblent pas directement la parkin et le DJ-1. Par conséquent, l'identification des cibles physiologiques de miR-34b et miR-34c est nécessaire pour mieux comprendre le rôle de ces miRNAs dans la pathogénie de la MP. Ainsi le miR-34b et miR-34c jouent deux rôles différents dans la pathogénèse de la MP. Une expression plus faible de miR-34b ou miR-34c peut entraîner une accumulation de SNCA. D'autre part, des niveaux plus faibles de miR-34b ou miR-34c diminuent indirectement l'expression de DJ-1 et de parkin conduisant à la MP.

De même, miR-320 et le miR-379 régulent à la baisse les niveaux de LAMP-2A (lysosomal associated membrane protein 2A) qui est un gène connu comme cible thérapeutique dans la MP (Xilouri et al. 2013). De plus, miR-320 et miR-379 régulent hsc70 (heat shock cognate protein 70) qui est une protéine liée à l'agrégation de l' $\alpha$  - Synucléine et un phénomène associé à la MP (Pemberton et Melki 2012), respectivement, conduisant à une agrégation SNCA.

La LRRK2 mutée est impliquée dans la pathogénèse de la MP. Dans un modèle de MP réalisé chez la drosophile, une étude montre que le variant de LRRK2, R1441G, cible la voie let-7a / miR-184 et la bloque, ce qui conduit à une régulation à la hausse des gènes de régulation du cycle cellulaire.

D'après une étude, dans la SNpc de souris, la suppression de miR-7 a entraîné une surexpression évidente de  $\alpha$ -syn. Cliniquement, les patients atteints de MP ont des taux de miR-7 significativement plus bas dans la SNpc (McMillan et al. 2017). De ce fait, le MiR-7 joue un rôle important dans la régulation de la MP en inhibant le gène SNCA. Toutes ces données suggèrent que miR-7 régule l'expression de  $\alpha$ -syn et est donc associé à la pathogénèse de la MP. Certains miRNAs détectés dans notre étude se sont révélés être directement impliqués dans la pathogénèse de la MP, tels que miR-34b, miR-34c, miR-320 et miR-184. Nous les avons identifiés comme miRNAs d'intérêt dans le modèle Alpha-Syn ce qui suggère fortement que ce modèle est le plus proche de la pathologie chez les patients.

Détecter tout type de maladie à un stade précoce s'avère être d'une aide précieuse pour les médecins et les patients. Les miRNAs, impliqués dans divers processus de la pathogénèse de la MP, peuvent constituer un excellent biomarqueur et la promesse d'un diagnostic amélioré et efficace. La modification de plusieurs voies importantes associées à la MP et régulée par des miRNAs pourrait donner un résultat positif et améliorer l'efficacité du traitement chez les patients MP.

Il existe également de nombreuses cibles prédites de ces miRNAs qui doivent être validées dans des modèles de rats ainsi que chez l'homme. Certains miRNAs détectés dans notre étude se sont révélés être directement impliqués dans la pathogenèse de la MP.

Les résultats obtenus permettent d'avoir une réflexion sur le rôle crucial des miRNAs dans la détermination de la spécificité transcriptomique des différents modèles parkinsoniens.

# Conclusion générale

La médecine personnalisée implique une approche multidisciplinaire dans laquelle les patients, les praticiens et les chercheurs parviennent, ensemble, à une meilleure compréhension de la santé à l'échelle individuelle. Pour atteindre une telle approche, il faut utiliser la bio-informatique qui joue un rôle crucial pour offrir des services de pointe adaptés aux besoins biomédicaux. La bio-informatique permet le stockage, l'organisation, l'analyse et l'interprétation des données moléculaires biomédicales. Le projet de cette thèse avait pour objectif de proposer deux pipelines d'analyse bio-informatique appliquées à deux approches différentes, la génomique et la transcriptomique, elles même appliquées à deux pathologies : la tuberculose et la maladie de Parkinson. Le choix de ces deux pathologies revient à une augmentation importante du nombre de cas atteints, soit 36 000 nouveaux cas de tuberculose chaque année ce qui rend cette maladie un problème de santé publique au Maroc. Pour la maladie de Parkinson, elle a connu ces dernières années une augmentation de nombre de cas déclarés, la prévalence de la MP est liée à l'âge, elle concerne environ 1 à 2% des personnes âgées de plus de 65 ans et peut atteindre 4 à 5% des personnes de plus de 85 ans (Ben El Haj et al. 2016).

Face aux demandes croissantes de méthodes statistiques et d'outils bio-informatiques pour l'analyse et la gestion des énormes quantités de données "omiques" issues du séquençage de nouvelle génération, dans ce mémoire, nous avons proposé deux pipelines avec les outils utilisés et adaptés pour les deux cas étudiés.

Pour l'analyse génomique, quatre souches de *mycobacterium tuberculosis* connue comme bactérie responsable de la tuberculose, ont fait l'objet d'un travail de séquençage du génome entier, en utilisant le séquenceur Illumina MiSeq avec l'application "*Whole genome sequencing* (WGS)" en "Paired-end" et le kit "Nextera XT DNA Library Preparation". La première étape du pipeline consistait à un contrôle de qualité, suivi du filtrage des bases de mauvaise qualité qui est une étape primordiale, ce filtrage se base sur les scores de qualité attribués à chaque base. Ensuite, un mapping (alignement) afin d'estimer la couverture des souches analysées par rapport au génome de référence M.Tuberculosis H37rV, et d'extraire la partie non alignée pour l'annotation. Les résultats du mapping indiquent que le taux des reads alignés sur la séquence chromosomique de la référence

est de 92,69%, 86,53%, 97,54% et 96,54% respectivement pour MTB1\_M, MTB2\_M, MTB13\_M et MTB20\_M. Afin d'identifier les éléments génétiques (gènes de résistances, de virulence, ...) des souches, un assemblage de Novo a été effectué, suivi d'une annotation. Ce pipeline d'analyse génomique a permis d'identifier trois mutations différentes sur le gène *rpoB* associé à la *Rifampicine* appartenant à la première ligne de résistance : Ala526Ser et Ser450Leu pour MTB1\_M, MTB2\_M respectivement, et Asp103Asp pour MTB13\_M et MTB20\_M, ainsi des mutations sur le gène *katG* associé à l'isoniazide : Ser95Thr, Glu21Gln, Gly668Asp et Ser315Asn pour MTB1\_M, MTB2\_M, MTB13\_M et MTB20\_M respectivement. Pour la deuxième ligne de résistance tel que le Fluoroquinolone, nous avons détecté des mutations pour deux gènes le *gyrA* et le *gyrB*, pour le premier : Gly668Asp pour MTB1\_M et MTB13\_M ainsi Glu21Gln et Phe640Val pour MTB2\_M et MTB20\_M respectivement, au niveau du gène *gyrB* : Asp538Asn, Thr539Asn, Asp500His pour MTB1\_M, MTB2\_M, et MTB20\_M respectivement. Nous avons identifié une résistance à la *Capreomycine* pour les quatre souches analysées au niveau du gène *tlyA*.

Pour l'approche transcriptomique, la technologie RNA-Seq a été utilisée pour identifier des miRNAs liés à la MP avec l'application "single end" où chaque read est indépendant des autres. Pour atteindre cet objectif, le pipeline a commencé par un contrôle de qualité, qui représente une étape très importante puisque les reads de mauvaise qualité s'alignent difficilement sur le génome de référence, et pour cela un filtrage est primordial. Pour réaliser une cartographie du contenu des échantillons séquencés, un alignement sur le génome de référence est indispensable suivi d'un comptage pour obtenir une matrice des miRNAs avec la différence d'expression en calculant la p-value et le Fold Change. Après l'identification et la visualisation des miRNAs par prédiction bio-informatique, nous avons effectué une validation par la RT PCR dans le but de valider et quantifier les miRNAs identifiés.

Ce pipeline nous a permis d'identifier plusieurs miRNAs ayant une différence d'expression significative tels que miR-34b, miR-34c qui sont sous-exprimés et jouent deux rôles différents dans la pathogenèse de la MP. Une expression plus faible de **miR-34b** ou **miR-34c** peut entraîner une accumulation de SNCA. D'autre part, miR-34b et miR-34c ciblent directement l' $\alpha$ -syn en inhibant son expression et sa formation d'agrégation (Kabaria et al. 2015). Un SNP dans la 3'-UTR de SNCA, est capable de réduire la répression de l' $\alpha$ -syn exprimée par miR-34b (L. Zhao et Wang 2019). L'association de ce polymorphisme avec le risque de MP doit être clarifiée. L'épuisement de miR-34b ou miR-34c in vitro entraîne une réduction de la viabilité cellulaire, qui s'accompagne d'un dysfonctionnement mitochondrial d'une régulation négative

de DJ1 et de Parkin, deux protéines associées aux formes familiales de la MP (Miñones-Moyano et al. 2011).

Les résultats obtenus sont en concordance avec la littérature, ce qui renforce l'utilité du pipeline proposé.

La figure ci-dessous représente un schéma des deux pipelines :



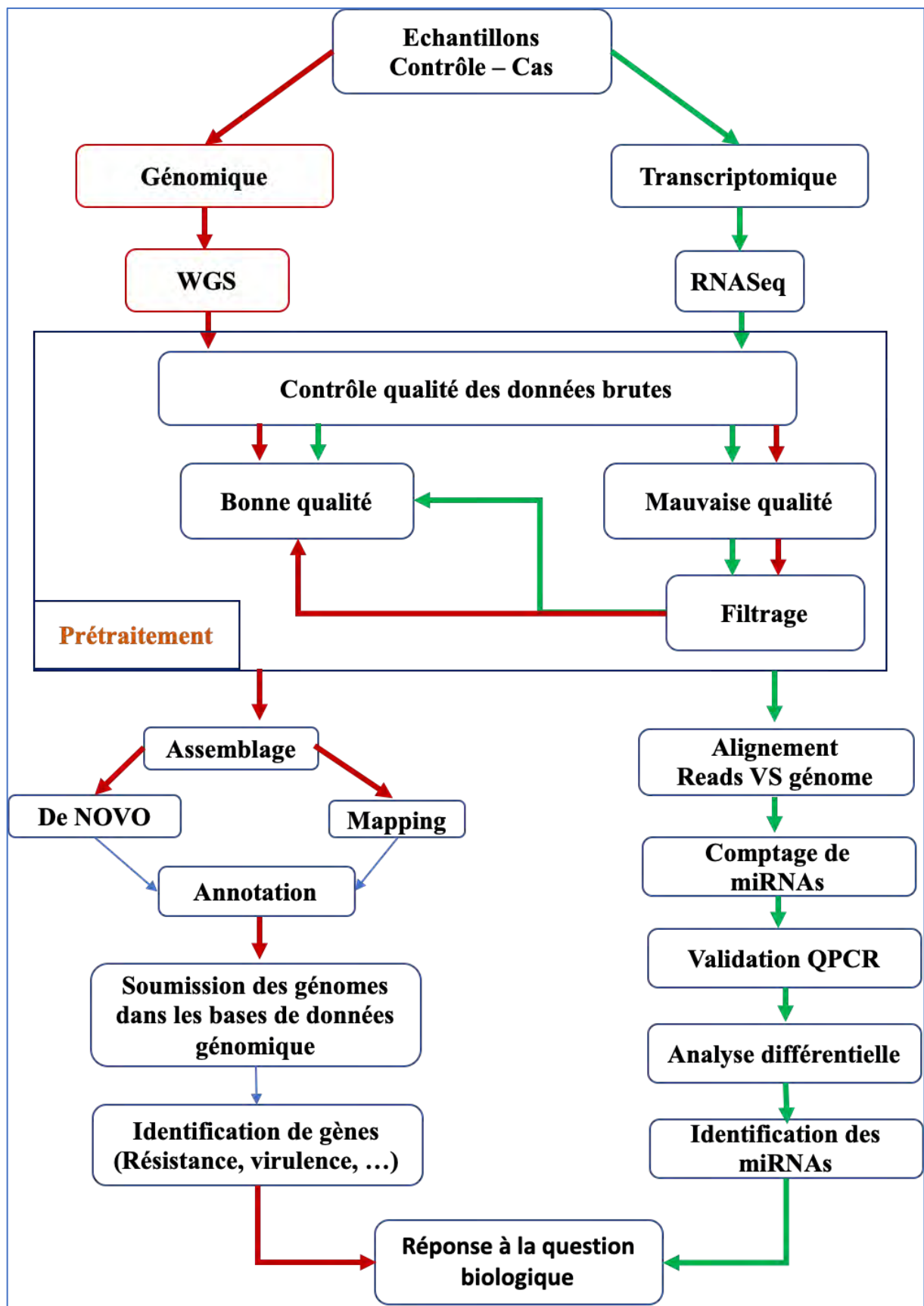


Figure 71: Pipline d'analyse génomique et transcriptomique

En conclusion, notre travail apporte une vision globale sur l'importance de l'analyse bio-informatique pour différentes pathologies, ainsi un apport biologique sur les deux pathologies étudiées.

# Résumé

## **Analyses bio-informatique de données génomiques et transcriptomiques dans deux pathologies : Tuberculose et maladie de Parkinson**

**Auteur : LAHLOU Lamiaa**

**Mots clefs :** Bio-informatique, séquençage de nouvelle génération, génomique, transcriptomique, Tuberculose, Maladie de Parkinson, Résistance, microRNAs, biomarqueurs.

Le séquençage de nouvelle génération des acides nucléiques génère une grande quantité de données brutes en un temps record. Le besoin d'analyse de ces big data a entraîné le développement de la bio-informatique qui est un champ de recherche multidisciplinaire permettant de transformer le matériel biologique issu du monde médical, en *smart data*, ce qui rend la bio-informatique incontournable pour le développement de la médecine de précision. Dans ce travail de thèse, deux approches ont été utilisées dans le contexte de deux pathologies complexes, l'approche génomique pour la tuberculose et l'approche transcriptomique pour la MP.

Des souches de *M.Tuberculosis* de quatre patients marocains atteints de la tuberculose ont été isolées, séquencées et analysées en utilisant un pipeline spécifique afin d'identifier les éléments génétiques. L'analyse génomique réalisée a permis d'identifier des mutations différentes au niveau du gène *rpoB* associé à la *Rifampicine* et du gène *katG* lié à l'*isoniazide* pour les quatre souches, ainsi que des mutations des gènes *gyrA*, *gyrB* associés à la fluoroquinolone et des mutations du gène *tlyA* associé au *Capreomycine* connues comme résistances de deuxième ligne.

Une approche transcriptomique a été utilisée pour déterminer les mécanismes moléculaires de la MP et d'identifier des biomarqueurs associés en étudiant l'expression de différents microRNAs chez deux modèles animaux de la maladie. Les résultats obtenus ont permis de valider les deux modèles sur le plan comportemental et lésionnel en se basant sur l'analyse biochimique des monoamines dans différentes structures cérébrales. Ensuite, en utilisant un pipeline d'analyse transcriptomique spécifique pour l'identification et la validation des miRNAs, nous avons identifié quelques miRNAs impliqués dans la MP, dont le miR-34b et miR-34c qui jouent deux rôles différents. Une expression plus faible de miR-34b ou miR-34c peut entraîner une accumulation de SNCA. D'autre part, miR-34b et miR-34c ciblent directement l' $\alpha$ -syn en inhibant son expression et sa formation d'agrégation. Les miRNAs identifiés vont permettre de prédire de nouveaux biomarqueurs associés à cette pathologie.

# Abstract

## **Bioinformatic analyzes of genomic and transcriptomic data in two pathologies: Tuberculosis and Parkinson's disease**

**Author:** LAHLOU Lamiaa

**Keywords:** Bioinformatics, Next Generation Sequencing, Genomics, Transcriptomics, Tuberculosis, Parkinson's Disease, Resistance, microRNAs, biomarkers.

Next-generation nucleic acid sequencing generates a large amount of raw data in record time. The need to analyze these big data has led to the development of bioinformatics, a multidisciplinary field of research that transforms biological material from the medical world into smart data, making bioinformatics essential for development of precision medicine. In this thesis work, two approaches have been used in the context of two complex pathologies, the genomic approach for tuberculosis and the transcriptomic approach for PD.

Strains of *M. Tuberculosis* from four Moroccan tuberculosis patients were isolated, sequenced and analyzed using a specific pipeline to identify genetic elements. The genomic analysis performed identified different mutations in the *rpoB* gene associated with Rifampicin and the isoniazid-related *katG* gene for the four strains, as well as mutations in the *gyrA*, *gyrB* and fluoroquinolone genes. mutations of the capreomycin-associated *tlyA* gene known as second-line resistors.

A transcriptomic approach was used to determine the molecular mechanisms of PD and to identify associated biomarkers by studying the expression of different microRNAs in two animal models of the disease. The results obtained validated both behavioral and lesional models based on the biochemical analysis of monoamines in different brain structures. Then, using a specific transcriptomic analysis pipeline for the identification and validation of miRNAs, we identified some miRNAs involved in PM, including miR-34b and miR-34c that play two different roles. Lower expression of miR-34b or miR-34c may result in SNCA accumulation. On the other hand, miR-34b and miR-34c directly target  $\alpha$ -syn by inhibiting its expression and aggregation formation. The identified miRNAs will predict new biomarkers associated with this pathology.

# ملخص

## تحليلات المعلومات الحيوية للبيانات الجينومية والنسخة في اثنين من الأمراض: مرض السل ومرض الشلل الرعاش

### المؤلف: لعلو لمياء

**الكلمات المفتاحية:** المعلوماتية الحيوية ، تسلسل الجيل القادم ، الجينومات ، السل ، مرض الشلل الرعاش ، المقاومة المؤشرات الحيوية

يولد تسلسل الحمض النووي من الجيل التالي كمية كبيرة من البيانات الخام في وقت قياسي. أدت الحاجة إلى تحليل هذه البيانات الضخمة إلى تطوير المعلوماتية الحيوية ، وهو مجال متعدد التخصصات للبحث يحول المواد البيولوجية من العالم الطبي إلى بيانات ذكية ، مما يجعل المعلومات الحيوية ضرورية لتطوير الطب الدقيق. في هذا البحث ، تم استخدام طريقتين في سياق اثنين من الأمراض المعقدة ، النهج الجيني لمرض السل والنهج النسخي لمرض الشلل الرعاش. تم عزل سلالات بكتيريا المتقطرة السلية من أربعة مرضى مغاربة بالسل وتسلسلها وتحليلها باستخدام تحليل محدد لتحديد العناصر الجينية المرتبط بالإيزونيا بالنسبة للسلالات الأربعة katG المرتبط بريفامبيسين وجين rpoB حدد التحليل الجيني إجراء طفرات مختلفة في جين المرتبط بكابريوميسين المعروف باسم مقومات الخط الثاني tlyA و طفرات الجين gyrA و gyrB. تم استخدام نهج تروسكريبتيك لتحديد الآليات الجزيئية لأمراض الشلل الرعاش وتحديد المؤشرات الحيوية المرتبطة من خلال دراسة التعبير عن الحمض النووي الريبي المختلفة في اثنين من النماذج الحيوانية للمرض. أثبتت النتائج التي تم الحصول عليها التحقق من صحة كل من النماذج السلوكية والآفات بناءً على التحليل الكيميائي الحيوي للأحاديات في هياكل حددنا بعض ، miRNAs المخ المختلفة. ثم ، باستخدام خط أنابيب تحليل تروسكريبتيك معين لتحديد والتحقق من التي تلعب دورين مختلفين. قد miR-34b و miR-34c المشاركة في مرض الشلل الرعاش ، بما في ذلك miRNAs و miR-34b من ناحية أخرى ، يستهدف كل من SNCA إلى تراكم miR-34c أو miR-34b يؤدي التعبير الأقل عن التي تم تحديدها التنبؤ المؤشرات miRNAs عن طريق تثبيط تكوين التعبير وتجميعه. سوف  $\alpha$ -syn مباشرة miR-34c الحيوية الجديدة المرتبطة بهذا المرض

# Bibliographie

Agarwal, Vikram, George W. Bell, Jin-Wu Nam, et David P. Bartel. 2015. « Predicting Effective MicroRNA Target Sites in Mammalian MRNAs ». *ELife* 4 (août). <https://doi.org/10.7554/eLife.05005>.

Alangaden, G. J., B. N. Kreiswirth, A. Aouad, M. Khetarpal, F. R. Igno, S. L. Moghazeh, E. K. Manavathu, et S. A. Lerner. 1998. « Mechanism of Resistance to Amikacin and Kanamycin in Mycobacterium Tuberculosis ». *Antimicrobial Agents and Chemotherapy* 42 (5): 1295-97.

Alvarez-Erviti, L., Y. Seow, A. H. V. Schapira, M. C. Rodriguez-Oroz, J. A. Obeso, et J. M. Cooper. 2013. « Influence of MicroRNA Deregulation on Chaperone-Mediated Autophagy and  $\alpha$ -Synuclein Pathology in Parkinson's Disease ». *Cell Death & Disease* 4 (mars): e545. <https://doi.org/10.1038/cddis.2013.73>.

Ambardar, Sheetal, Rikita Gupta, Deepika Trakroo, Rup Lal, et Jyoti Vakhlu. 2016. « High Throughput Sequencing: An Overview of Sequencing Chemistry ». *Indian Journal of Microbiology* 56 (4): 394-404. <https://doi.org/10.1007/s12088-016-0606-4>.

An, Jiyuan, John Lai, Atul Sajjanhar, Melanie L. Lehman, et Colleen C. Nelson. 2014. « MiRPlant: An Integrated Tool for Identification of Plant MiRNA from RNA Sequencing Data ». *BMC Bioinformatics* 15 (août): 275. <https://doi.org/10.1186/1471-2105-15-275>.

Anders, Simon, Paul Theodor Pyl, et Wolfgang Huber. 2015. « HTSeq—a Python framework to work with high-throughput sequencing data ». *Bioinformatics* 31 (2): 166-69. <https://doi.org/10.1093/bioinformatics/btu638>.

Andini, Nadya, et Kevin A. Nash. 2006. « Intrinsic Macrolide Resistance of the Mycobacterium Tuberculosis Complex Is Inducible ». *Antimicrobial Agents and Chemotherapy* 50 (7): 2560-62. <https://doi.org/10.1128/AAC.00264-06>.

Asci, Roberta, Fara Vallefucio, Immacolata Andolfo, Mariasole Bruno, Luigia De Falco, et Achille Iolascon. 2013. « Trasferrin Receptor 2 Gene Regulation by MicroRNA 221 in SH-SY5Y Cells Treated with MPP<sup>+</sup> as Parkinson's Disease Cellular Model ». *Neuroscience*

*Research* 77 (3): 121-27. <https://doi.org/10.1016/j.neures.2013.09.003>.

Ashraf, W., R. F. Pfeiffer, F. Park, J. Lof, et E. M. Quigley. 1997. « Constipation in Parkinson's Disease: Objective Assessment and Response to Psyllium ». *Movement Disorders: Official Journal of the Movement Disorder Society* 12 (6): 946-51. <https://doi.org/10.1002/mds.870120617>.

Aubry, Alexandra, Xiao-Su Pan, L. Mark Fisher, Vincent Jarlier, et Emmanuelle Cambau. 2004. « Mycobacterium Tuberculosis DNA Gyrase: Interaction with Quinolones and Correlation with Antimycobacterial Drug Activity ». *Antimicrobial Agents and Chemotherapy* 48 (4): 1281-88.

Aziz, Ramy K., Daniela Bartels, Aaron A. Best, Matthew DeJongh, Terrence Disz, Robert A. Edwards, Kevin Formsma, et al. 2008. « The RAST Server: Rapid Annotations Using Subsystems Technology ». *BMC Genomics* 9 (février): 75. <https://doi.org/10.1186/1471-2164-9-75>.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. « SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing ». *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 19 (5): 455-77. <https://doi.org/10.1089/cmb.2012.0021>.

Bartel, David P. 2004. « MicroRNAs: Genomics, Biogenesis, Mechanism, and Function ». *Cell* 116 (2): 281-97.

Barzon, Luisa, Enrico Lavezzo, Giulia Costanzi, Elisa Franchin, Stefano Toppo, et Giorgio Palù. 2013. « Next-Generation Sequencing Technologies in Diagnostic Virology ». *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology* 58 (2): 346-50. <https://doi.org/10.1016/j.jcv.2013.03.003>.

« Beckman, Paris ». s. d. Consulté le 10 juin 2019. <https://www.lp-chrom.com/index.php/en/beckman>.

Ben El Haj, Rafiqua, Wafaa Regragui, Rachid Tazi-Ahnini, Asmae Skalli, Naima Bouslam, Ali Benomar, Mohamed Yahyaoui, et Ahmed Bouhouche. 2016. « A Novel Homozygous p.L539F Mutation Identified in PINK1 Gene in a Moroccan Patient with Parkinsonism ». *BioMed*

*Research International* 2016. <https://doi.org/10.1155/2016/3460234>.

Benazzouz, A., C. Gross, J. Féger, T. Boraud, et B. Bioulac. 1993. « Reversal of Rigidity and Improvement in Motor Performance by Subthalamic High-Frequency Stimulation in MPTP-Treated Monkeys ». *The European Journal of Neuroscience* 5 (4): 382-89.

Berardelli, A., J. C. Rothwell, P. D. Thompson, et M. Hallett. 2001. « Pathophysiology of Bradykinesia in Parkinson's Disease ». *Brain: A Journal of Neurology* 124 (Pt 11): 2131-46. <https://doi.org/10.1093/brain/124.11.2131>.

Berglund, Eva C., Anna Kiialainen, et Ann-Christine Syvänen. 2011. « Next-Generation Sequencing Technologies and Applications for Human Genetic History and Forensics ». *Investigative Genetics* 2 (novembre): 23. <https://doi.org/10.1186/2041-2223-2-23>.

« Bioseb: Von Frey électronique ». 2011. 2011. [https://www.bioseb.com/bioseb/francais/default/item\\_id=562\\_cat\\_id=19\\_Von%20Frey%20C3%A9lectronique.php](https://www.bioseb.com/bioseb/francais/default/item_id=562_cat_id=19_Von%20Frey%20C3%A9lectronique.php).

Blanchard, J. S. 1996. « Molecular Mechanisms of Drug Resistance in Mycobacterium Tuberculosis ». *Annual Review of Biochemistry* 65: 215-39. <https://doi.org/10.1146/annurev.bi.65.070196.001243>.

« BluePippin | Sage Science ». s. d. Consulté le 29 juin 2019. <http://www.sagescience.com/products/bluepippin/>.

Bolger, Anthony M., Marc Lohse, et Bjoern Usadel. 2014. « Trimmomatic: A Flexible Trimmer for Illumina Sequence Data ». *Bioinformatics (Oxford, England)* 30 (15): 2114-20. <https://doi.org/10.1093/bioinformatics/btu170>.

Bonnet, Eric, Jan Wuyts, Pierre Rouzé, et Yves Van de Peer. 2004. « Evidence That MicroRNA Precursors, Unlike Other Non-Coding RNAs, Have Lower Folding Free Energies than Random Sequences ». *Bioinformatics (Oxford, England)* 20 (17): 2911-17. <https://doi.org/10.1093/bioinformatics/bth374>.

Breit, S., R. Bouali-Benazzouz, A. L. Benabid, et A. Benazzouz. 2001. « Unilateral Lesion of the Nigrostriatal Pathway Induces an Increase of Neuronal Activity of the Pedunculopontine Nucleus, Which Is Reversed by the Lesion of the Subthalamic Nucleus in the Rat ». *The*



*European Journal of Neuroscience* 14 (11): 1833-42. <https://doi.org/10.1046/j.0953-816x.2001.01800.x>.

Brossier, F., N. Veziris, C. Truffot-Pernot, V. Jarlier, et W. Sougakoff. 2011. « Molecular Investigation of Resistance to the Antituberculous Drug Ethionamide in Multidrug-Resistant Clinical Isolates of *Mycobacterium tuberculosis* ». *Antimicrobial Agents and Chemotherapy* 55 (1): 355-60. <https://doi.org/10.1128/AAC.01030-10>.

Campbell, Patricia J., Glenn P. Morlock, R. David Sikes, Tracy L. Dalton, Beverly Metchock, Angela M. Starks, Delaina P. Hooks, Lauren S. Cowan, Bonnie B. Plikaytis, et James E. Posey. 2011. « Molecular Detection of Mutations Associated with First- and Second-Line Drug Resistance Compared with Conventional Drug Susceptibility Testing of *Mycobacterium Tuberculosis* ». *Antimicrobial Agents and Chemotherapy* 55 (5): 2032-41. <https://doi.org/10.1128/AAC.01550-10>.

Caruccio, Nicholas. 2011. « Preparation of Next-Generation Sequencing Libraries Using Nextera™ Technology: Simultaneous DNA Fragmentation and Adaptor Tagging by in Vitro Transposition ». *Methods in Molecular Biology (Clifton, N.J.)* 733: 241-55. [https://doi.org/10.1007/978-1-61779-089-8\\_17](https://doi.org/10.1007/978-1-61779-089-8_17).

« CED Spike2: Présentation du système ». s. d. Consulté le 10 juin 2019. <http://ced.co.uk/fr/products/spkovin>.

Chagraoui, Abdeslam, Sara Whitestone, Lynn Baassiri, Julien Manem, Giuseppe Di Giovanni, et Philippe De Deurwaerdère. 2019. « Neurochemical impact of the 5-HT<sub>2C</sub> receptor agonist WAY-163909 on monoamine tissue content in the rat brain ». *Neurochemistry International* 124 (mars): 245-55. <https://doi.org/10.1016/j.neuint.2019.01.019>.

Chaoui, Imane, Naima Atalhi, Radia Sabouni, Mohammed Akrim, Mohammed Abid, Saaid Amzazi, et Mohammed ELMzibri. 2014. « Rifoligotyping assay: an alternative method for rapid detection of rifampicin resistance in *Mycobacterium tuberculosis* isolates from Morocco ». *Biotechnology & Biotechnological Equipment* 28 (6): 1095-1102. <https://doi.org/10.1080/13102818.2014.975569>.

Charles, Keri-Ann, Frédéric Naudet, Rabia Bouali-Benazzouz, Marc Landry, Philippe De Deurwaerdère, Pascal Fossat, et Abdelhamid Benazzouz. 2018. « Alteration of Nociceptive

Integration in the Spinal Cord of a Rat Model of Parkinson's Disease ». *Movement Disorders: Official Journal of the Movement Disorder Society* 33 (6): 1010-15. <https://doi.org/10.1002/mds.27377>.

Chaudhuri, Amrita Datta, Doo Chul Choi, Savan Kabaria, Alan Tran, et Eunsung Junn. 2016. « MicroRNA-7 Regulates the Function of Mitochondrial Permeability Transition Pore by Targeting VDAC1 Expression ». *The Journal of Biological Chemistry* 291 (12): 6483-93. <https://doi.org/10.1074/jbc.M115.691352>.

Chekulaeva, Marina, et Witold Filipowicz. 2009. « Mechanisms of MiRNA-Mediated Post-Transcriptional Regulation in Animal Cells ». *Current Opinion in Cell Biology* 21 (3): 452-60. <https://doi.org/10.1016/j.ceb.2009.04.009>.

Chen, Yantian, Maarten Sannaert, Scott J. Roberts, Frank P. Luyten, et Jan Schrooten. 2012. « Validation of a PicoGreen-Based DNA Quantification Integrated in an RNA Extraction Method for Two-Dimensional and Three-Dimensional Cell Cultures ». *Tissue Engineering. Part C, Methods* 18 (6): 444-52. <https://doi.org/10.1089/ten.TEC.2011.0304>.

Chetrit, Jonathan, Anne Taupignon, Lionel Froux, Stephanie Morin, Rabia Bouali-Benazzouz, Frédéric Naudet, Nabila Kadiri, Christian E. Gross, Bernard Bioulac, et Abdelhamid Benazzouz. 2013. « Inhibiting Subthalamic D5 Receptor Constitutive Activity Alleviates Abnormal Electrical Activity and Reverses Motor Impairment in a Rat Model of Parkinson's Disease ». *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 33 (37): 14840-49. <https://doi.org/10.1523/JNEUROSCI.0453-13.2013>.

Cho, Hyun Jin, Guoxiang Liu, Seok Min Jin, Loukia Parisiadou, Chengsong Xie, Jia Yu, Lixin Sun, et al. 2013. « MicroRNA-205 Regulates the Expression of Parkinson's Disease-Related Leucine-Rich Repeat Kinase 2 Protein ». *Human Molecular Genetics* 22 (3): 608-20. <https://doi.org/10.1093/hmg/dds470>.

Chong, Raymond, Lauren Albor, Chandramohan Wakade, et John Morgan. 2018. « The dimensionality of fatigue in Parkinson's disease ». *Journal of Translational Medicine* 16 (juillet). <https://doi.org/10.1186/s12967-018-1554-z>.

Choudhury, Gourav R., et Marcel M. Daadi. 2018. « Charting the Onset of Parkinson-like Motor and Non-Motor Symptoms in Nonhuman Primate Model of Parkinson's Disease ». *PloS*

*One* 13 (8): e0202770. <https://doi.org/10.1371/journal.pone.0202770>.

Chowdhury, Abhijit, Nur Hossain, Shahnewaj Bin Mannan, et Reaz Mohammad Mazumdar. 2012. « PYROSEQUENCING- PRINCIPLES AND APPLICATIONS ». *Molecular Biology* 2 (2): 12.

Cock, Peter J. A., Christopher J. Fields, Naohisa Goto, Michael L. Heuer, et Peter M. Rice. 2010. « The Sanger FASTQ File Format for Sequences with Quality Scores, and the Solexa/Illumina FASTQ Variants ». *Nucleic Acids Research* 38 (6): 1767-71. <https://doi.org/10.1093/nar/gkp1137>.

Cohen-Boulakia, Sarah, et Patrick Valduriez. 2016. « Traitement de Données Bioinformatiques Massives (Big Data) ».

Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, et al. 1998. « Deciphering the Biology of *Mycobacterium Tuberculosis* from the Complete Genome Sequence ». *Nature* 393 (6685): 537-44. <https://doi.org/10.1038/31159>.

Cole, Stewart T. 1999. « Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv ». *FEBS Letters* 452 (1): 7-10. [https://doi.org/10.1016/S0014-5793\(99\)00536-0](https://doi.org/10.1016/S0014-5793(99)00536-0).

Coll, Francesc, Ruth McNerney, Mark D. Preston, José Afonso Guerra-Assunção, Andrew Warry, Grant Hill-Cawthorne, Kim Mallard, et al. 2015. « Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences ». *Genome Medicine* 7 (1): 51. <https://doi.org/10.1186/s13073-015-0164-0>.

Compeau, Phillip E. C., Pavel A. Pevzner, et Glenn Tesler. 2011. « How to Apply de Bruijn Graphs to Genome Assembly ». *Nature Biotechnology* 29 (11): 987-91. <https://doi.org/10.1038/nbt.2023>.

Creighton, Chad J., Jeffrey G. Reid, et Preethi H. Gunaratne. 2009. « Expression Profiling of MicroRNAs by Deep Sequencing ». *Briefings in Bioinformatics* 10 (5): 490-97. <https://doi.org/10.1093/bib/bbp019>.

Damien Imbs, et Mohamed Sayed Hassan. 2017. *Bioinformatique: Travail d'étude*. Université Nice Sophia Antipolis.

Dardel, Frédéric, et François Képès. 2002. *Bioinformatique: génomique et post-génomique*.

Editions Ecole Polytechnique.

Darling, Aaron C. E., Bob Mau, Frederick R. Blattner, et Nicole T. Perna. 2004. « Mauve: Multiple Alignment of Conserved Genomic Sequence with Rearrangements ». *Genome Research* 14 (7): 1394-1403. <https://doi.org/10.1101/gr.2289704>.

Deichmann, Ute. 2016. « Epigenetics: The Origins and Evolution of a Fashionable Topic ». *Developmental Biology* 416 (1): 249-54. <https://doi.org/10.1016/j.ydbio.2016.06.005>.

Delcher, A. L., D. Harmon, S. Kasif, O. White, et S. L. Salzberg. 1999. « Improved Microbial Gene Identification with GLIMMER ». *Nucleic Acids Research* 27 (23): 4636-41.

Delcher, Arthur L., Kirsten A. Bratke, Edwin C. Powers, et Steven L. Salzberg. 2007. « Identifying bacterial genes and endosymbiont DNA with Glimmer ». *Bioinformatics (Oxford, England)* 23 (6): 673-79. <https://doi.org/10.1093/bioinformatics/btm009>.

Dellu-Hagedorn, Françoise, Aurélie Fitoussi, et Philippe De Deurwaerdère. 2017. « Correlative analysis of dopaminergic and serotonergic metabolism across the brain to study monoaminergic function and interaction ». *Journal of Neuroscience Methods* 280 (mars): 54-63. <https://doi.org/10.1016/j.jneumeth.2017.01.020>.

Delwaide, P. J., M. Sabbatino, et C. Delwaide. 1986. « Some Pathophysiological Aspects of the Parkinsonian Rigidity ». *Journal of Neural Transmission. Supplementum* 22: 129-39.

Dill, Holger, Bastian Linder, Alexander Fehr, et Utz Fischer. 2012. « Intronic MiR-26b Controls Neuronal Differentiation by Repressing Its Host Transcript, Ctdsp2 ». *Genes & Development* 26 (1): 25-30. <https://doi.org/10.1101/gad.177774.111>.

« Direct-Zol RNA Microprep Kits ». 2019. Zymo Research. 2019. <https://www.zymoresearch.com/products/direct-zol-rna-microprep-kits>.

Duan, Demin, Ke-xiao Zheng, Ye Shen, Rong Cao, Li Jiang, Zhuoxuan Lu, Xiyun Yan, et Jiong Li. 2011. « Label-Free High-Throughput MicroRNA Expression Profiling from Total RNA ». *Nucleic Acids Research* 39 (22): e154. <https://doi.org/10.1093/nar/gkr774>.

Dunn, James J., Jeffrey R. Starke, et Paula A. Revell. 2016. « Laboratory Diagnosis of Mycobacterium Tuberculosis Infection and Disease in Children ». *Journal of Clinical Microbiology* 54 (6): 1434-41. <https://doi.org/10.1128/JCM.03043-15>.

- Eid, John, Adrian Fehr, Jeremy Gray, Khai Luong, John Lyle, Geoff Otto, Paul Peluso, et al. 2009. « Real-Time DNA Sequencing from Single Polymerase Molecules ». *Science (New York, N.Y.)* 323 (5910): 133-38. <https://doi.org/10.1126/science.1162986>.
- Ekblom, Robert, et Jochen B. W. Wolf. 2014. « A Field Guide to Whole-Genome Sequencing, Assembly and Annotation ». *Evolutionary Applications* 7 (9): 1026-42. <https://doi.org/10.1111/eva.12178>.
- El-Gebali, Sara, Jaina Mistry, Alex Bateman, Sean R Eddy, Aurélien Luciani, Simon C Potter, Matloob Qureshi, et al. 2019. « The Pfam protein families database in 2019 ». *Nucleic Acids Research* 47 (Database issue): D427-32. <https://doi.org/10.1093/nar/gky995>.
- Ewels, Philip, Måns Magnusson, Sverker Lundin, et Max Käller. 2016. « MultiQC: summarize analysis results for multiple tools and samples in a single report ». *Bioinformatics* 32 (19): 3047-48. <https://doi.org/10.1093/bioinformatics/btw354>.
- Fàbrega, Anna, Sergi Madurga, Ernest Giralt, et Jordi Vila. 2009. « Mechanism of Action of and Resistance to Quinolones ». *Microbial Biotechnology* 2 (1): 40-61. <https://doi.org/10.1111/j.1751-7915.2008.00063.x>.
- Faggiani, E., et A. Benazzouz. 2017. « Deep Brain Stimulation of the Subthalamic Nucleus in Parkinson's Disease: From History to the Interaction with the Monoaminergic Systems ». *Progress in Neurobiology* 151: 139-56. <https://doi.org/10.1016/j.pneurobio.2016.07.003>.
- Fakruddin, Md, et Abhijit Chowdhury. 2012. « Pyrosequencing-An Alternative to Traditional Sanger Sequencing ». *American Journal of Biochemistry and Biotechnology* 8 (1): 14-20. <https://doi.org/10.3844/ajbbsp.2012.14.20>.
- Feng, Ya, Joseph Jankovic, et Yun-Cheng Wu. 2015. « Epigenetic mechanisms in Parkinson's disease ». *Journal of the Neurological Sciences* 349 (1): 3-9. <https://doi.org/10.1016/j.jns.2014.12.017>.
- Ferrier, Jeremy, Fabien Marchand, et David Balayssac. 2016. « Assessment of Mechanical Allodynia in Rats Using the Electronic Von Frey Test ». *BIO-PROTOCOL* 6 (18). <https://doi.org/10.21769/BioProtoc.1933>.
- Fierer, Noah, Christian L. Lauber, Kelly S. Ramirez, Jesse Zaneveld, Mark A. Bradford, et Rob

Knight. 2012. « Comparative Metagenomic, Phylogenetic and Physiological Analyses of Soil Microbial Communities across Nitrogen Gradients ». *The ISME Journal* 6 (5): 1007-17. <https://doi.org/10.1038/ismej.2011.159>.

Filipowicz, Witold, Suvendra N. Bhattacharyya, et Nahum Sonenberg. 2008. « Mechanisms of Post-Transcriptional Regulation by MicroRNAs: Are the Answers in Sight? » *Nature Reviews. Genetics* 9 (2): 102-14. <https://doi.org/10.1038/nrg2290>.

Finn, Robert D., John Tate, Jaina Mistry, Penny C. Coghill, Stephen John Sammut, Hans-Rudolf Hotz, Goran Ceric, et al. 2008. « The Pfam protein families database ». *Nucleic Acids Research* 36 (Database issue): D281-88. <https://doi.org/10.1093/nar/gkm960>.

Fitoussi, A., F. Dellu-Hagedorn, et P. De Deurwaerdère. 2013. « Monoamines tissue content analysis reveals restricted and site-specific correlations in brain regions involved in cognition ». *Neuroscience* 255 (décembre): 233-45. <https://doi.org/10.1016/j.neuroscience.2013.09.059>.

Frias-Lopez, Jorge, Yanmei Shi, Gene W. Tyson, Maureen L. Coleman, Stephan C. Schuster, Sallie W. Chisholm, et Edward F. Delong. 2008. « Microbial Community Gene Expression in Ocean Surface Waters ». *Proceedings of the National Academy of Sciences of the United States of America* 105 (10): 3805-10. <https://doi.org/10.1073/pnas.0708897105>.

Friedländer, Marc R., Wei Chen, Catherine Adamidi, Jonas Maaskola, Ralf Einspanier, Signe Knespel, et Nikolaus Rajewsky. 2008. « Discovering MicroRNAs from Deep Sequencing Data Using MiRDeep ». *Nature Biotechnology* 26 (4): 407-15. <https://doi.org/10.1038/nbt1394>.

Friedländer, Marc R., Sebastian D. Mackowiak, Na Li, Wei Chen, et Nikolaus Rajewsky. 2012. « MiRDeep2 Accurately Identifies Known and Hundreds of Novel MicroRNA Genes in Seven Animal Clades ». *Nucleic Acids Research* 40 (1): 37-52. <https://doi.org/10.1093/nar/gkr688>.

Friedman, Joseph H., James C. Beck, Kelvin L. Chou, Gracia Clark, Christopher P. Fagundes, Christopher G. Goetz, Karen Herlofson, et al. 2016. « Fatigue in Parkinson's Disease: Report from a Multidisciplinary Symposium ». *NPJ Parkinson's Disease* 2. <https://doi.org/10.1038/npjparkd.2015.25>.

Fröhlich, Eleonore. 2017. « Role of omics techniques in the toxicity testing of nanoparticles ». *Journal of Nanobiotechnology* 15 (novembre). <https://doi.org/10.1186/s12951-017-0320-3>.

Gehrke, Stephan, Yuzuru Imai, Nicholas Sokol, et Bingwei Lu. 2010. « Pathogenic LRRK2 Negatively Regulates MicroRNA-Mediated Translational Repression ». *Nature* 466 (7306): 637-41. <https://doi.org/10.1038/nature09191>.

Global Health Education. 2018. « TB Statistics | Incidence, Prevalence, High Burden ». TB Facts | TB, Tests, Drugs, Statistics. 2018. <https://www.tbfacts.org/tb-statistics/>.

Goldberg, Aaron D., C. David Allis, et Emily Bernstein. 2007. « Epigenetics: A Landscape Takes Shape ». *Cell* 128 (4): 635-38. <https://doi.org/10.1016/j.cell.2007.02.006>.

Graeber, M. B. 2009. « Biomarkers for Parkinson's Disease. » *Experimental Neurology* 216 (2): 249-53. <https://doi.org/10.1016/j.expneurol.2008.12.017>.

Griffiths-Jones, Sam. 2006. « MiRBase: The MicroRNA Sequence Database ». *Methods in Molecular Biology (Clifton, N.J.)* 342: 129-38. <https://doi.org/10.1385/1-59745-123-1:129>.

Griffiths-Jones, Sam, Jerome H. L. Hui, Antonio Marco, et Matthew Ronshaugen. 2011. « MicroRNA Evolution by Arm Switching ». *EMBO Reports* 12 (2): 172-77. <https://doi.org/10.1038/embor.2010.191>.

Gurevich, Alexey, Vladislav Saveliev, Nikolay Vyahhi, et Glenn Tesler. 2013. « QUASt: Quality Assessment Tool for Genome Assemblies ». *Bioinformatics (Oxford, England)* 29 (8): 1072-75. <https://doi.org/10.1093/bioinformatics/btt086>.

Hackenberg, Michael, Martin Sturm, David Langenberger, Juan Manuel Falcón-Pérez, et Ana M. Aransay. 2009. « MiRanalyzer: A MicroRNA Detection and Analysis Tool for next-Generation Sequencing Experiments ». *Nucleic Acids Research* 37 (Web Server issue): W68-76. <https://doi.org/10.1093/nar/gkp347>.

Haft, Daniel H, Michael DiCuccio, Azat Badretdin, Vyacheslav Brover, Vyacheslav Chetvernin, Kathleen O'Neill, Wenjun Li, et al. 2018. « RefSeq: an update on prokaryotic genome annotation and curation ». *Nucleic Acids Research* 46 (Database issue): D851-60. <https://doi.org/10.1093/nar/gkx1068>.

Haft, Daniel H., Jeremy D. Selengut, Roland A. Richter, Derek Harkins, Malay K. Basu, et Erin Beck. 2013. « TIGRFAMs and Genome Properties in 2013 ». *Nucleic Acids Research* 41 (Database issue): D387-95. <https://doi.org/10.1093/nar/gks1234>.

Hammond, Scott M. 2015. « An overview of microRNAs ». *Advanced drug delivery reviews* 87 (juin): 3-14. <https://doi.org/10.1016/j.addr.2015.05.001>.

Hasin, Yehudit, Marcus Seldin, et Aldons Lysis. 2017. « Multi-omics approaches to disease ». *Genome Biology* 18 (mai). <https://doi.org/10.1186/s13059-017-1215-1>.

Herbert, Zach. 2012. « In a Prominent Core Facility at the Dana-Farber Cancer Institute, Scientists Have Paired Pippin Prep Size Selection with Illumina's Nextera to Build Better, More Reproducible Libraries That Yield Improved Genome Assemblies. », 2.

Hindson, Christopher M., John R. Chevillet, Hilary A. Briggs, Emily N. Gallichotte, Ingrid K. Ruf, Benjamin J. Hindson, Robert L. Vessella, et Muneesh Tewari. 2013. « Absolute Quantification by Droplet Digital PCR versus Analog Real-Time PCR ». *Nature Methods* 10 (10): 1003-5. <https://doi.org/10.1038/nmeth.2633>.

Hingamp, Pascal, Nigel Grimsley, Silvia G. Acinas, Camille Clerissi, Lucie Subirana, Julie Poulain, Isabel Ferrera, et al. 2013. « Exploring Nucleo-Cytoplasmic Large DNA Viruses in Tara Oceans Microbial Metagenomes ». *The ISME Journal* 7 (9): 1678-95. <https://doi.org/10.1038/ismej.2013.59>.

Hinz, Marty, Alvin Stein, et Thomas Uncini. 2011. « Amino Acid Management of Parkinson's Disease: A Case Study ». *International Journal of General Medicine* 4 (février): 165-74. <https://doi.org/10.2147/IJGM.S16621>.

Hogeweg, Paulien. 2011. « The Roots of Bioinformatics in Theoretical Biology ». *PLoS Computational Biology* 7 (3). <https://doi.org/10.1371/journal.pcbi.1002021>.

Hoss, Andrew G., Adam Labadorf, Thomas G. Beach, Jeanne C. Latourelle, et Richard H. Myers. 2016. « MicroRNA Profiles in Parkinson's Disease Prefrontal Cortex ». *Frontiers in Aging Neuroscience* 8: 36. <https://doi.org/10.3389/fnagi.2016.00036>.

Hyatt, Doug, Gwo-Liang Chen, Philip F. Locascio, Miriam L. Land, Frank W. Larimer, et Loren J. Hauser. 2010. « Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification ». *BMC Bioinformatics* 11 (mars): 119. <https://doi.org/10.1186/1471-2105-11-119>.

Illumina. 2017. « An Introduction to Next-Generation Sequencing Technology », 16.



———. 2018a. « MiSeq Applications and Methods | A wide breadth of applications ». 2018. <https://emea.illumina.com/systems/sequencing-platforms/miseq/applications.html>.

———. 2018b. « Nextera XT DNA Library Prep Kit Reference Guide (15031942) », 30.

Jakubowski, Jennifer L., et Viviane Labrie. 2016. « Epigenetic Biomarkers for Parkinson's Disease: From Diagnostics to Therapeutics ». *Journal of Parkinson's Disease* 7 (1): 1-12. <https://doi.org/10.3233/JPD-160914>.

Jiang, Mingfang, Qiang Yun, Feng Shi, Guangming Niu, Yang Gao, Shenghui Xie, et Shengyuan Yu. 2016. « Downregulation of MiR-384-5p Attenuates Rotenone-Induced Neurotoxicity in Dopaminergic SH-SY5Y Cells through Inhibiting Endoplasmic Reticulum Stress ». *American Journal of Physiology. Cell Physiology* 310 (9): C755-763. <https://doi.org/10.1152/ajpcell.00226.2015>.

Johansen, Shanna K., Courtney E. Maus, Bonnie B. Plikaytis, et Stephen Douthwaite. 2006. « Capreomycin Binds across the Ribosomal Subunit Interface Using TlyA-Encoded 2'-O-Methylations in 16S and 23S rRNAs ». *Molecular Cell* 23 (2): 173-82. <https://doi.org/10.1016/j.molcel.2006.05.044>.

Joshi N A, et Fass J N. 2011. « Sickle: A sliding-window, adaptive, quality-based... - Google Scholar ».

[https://scholar.google.com/scholar\\_lookup?author=NA+Joshi&author=JN+Fass&title=Sickle%3A+A+sliding-window%2C+adaptive%2C+quality-based+trimming+tool+for+FastQ+files+%28Version+1.33%29+%5BSoftware%5D&publication\\_year=2011](https://scholar.google.com/scholar_lookup?author=NA+Joshi&author=JN+Fass&title=Sickle%3A+A+sliding-window%2C+adaptive%2C+quality-based+trimming+tool+for+FastQ+files+%28Version+1.33%29+%5BSoftware%5D&publication_year=2011).

Junn, Eunsung, Kang-Woo Lee, Byeong Seon Jeong, Teresa W. Chan, Joo-Young Im, et M. Maral Mouradian. 2009. « Repression of Alpha-Synuclein Expression and Toxicity by MicroRNA-7 ». *Proceedings of the National Academy of Sciences of the United States of America* 106 (31): 13052-57. <https://doi.org/10.1073/pnas.0906277106>.

Kabaria, Savan, Doo Chul Choi, Amrita Datta Chaudhuri, M. Maral Mouradian, et Eunsung Junn. 2015. « Inhibition of MiR-34b and MiR-34c Enhances  $\alpha$ -Synuclein Expression in Parkinson's Disease ». *FEBS Letters* 589 (3): 319-25. <https://doi.org/10.1016/j.febslet.2014.12.014>.

Karsch-Mizrachi, Ilene, Toshihisa Takagi, et Guy Cochrane. 2018. « The international nucleotide sequence database collaboration ». *Nucleic Acids Research* 46 (Database issue): D48-51. <https://doi.org/10.1093/nar/gkx1097>.

Kim, Jongpil, Keiichi Inoue, Jennifer Ishii, William B. Vanti, Sergey V. Voronov, Elizabeth Murchison, Gregory Hannon, et Asa Abeliovich. 2007. « A MicroRNA Feedback Circuit in Midbrain Dopamine Neurons ». *Science (New York, N.Y.)* 317 (5842): 1220-24. <https://doi.org/10.1126/science.1140481>.

Kim, S. H., et J. M. Chung. 1991. « Sympathectomy Alleviates Mechanical Allodynia in an Experimental Animal Model for Neuropathy in the Rat ». *Neuroscience Letters* 134 (1): 131-34.

Kim, V. Narry. 2005. « Small RNAs: Classification, Biogenesis, and Function ». *Molecules and Cells* 19 (1): 1-15.

Kim, Woori, Haneul Noh, Yenarae Lee, Jeha Jeon, Arthi Shanmugavadivu, Donna L. McPhie, Kwang-Soo Kim, Bruce M. Cohen, Hyemyung Seo, et Kai C. Sonntag. 2016. « MiR-126 Regulates Growth Factor Activities and Vulnerability to Toxic Insult in Neurons ». *Molecular Neurobiology* 53 (1): 95-108. <https://doi.org/10.1007/s12035-014-8989-x>.

Koonin, Eugene V., et Michael Y. Galperin. 2003. *Genome Annotation and Analysis*. Kluwer Academic. <https://www.ncbi.nlm.nih.gov/books/NBK20253/>.

Köser, Claudio U., Matthew J. Ellington, et Sharon J. Peacock. 2014. « Whole-genome sequencing to control antimicrobial resistance ». *Trends in Genetics* 30 (9): 401-7. <https://doi.org/10.1016/j.tig.2014.07.003>.

Kozomara, Ana, Maria Birgaoanu, et Sam Griffiths-Jones. 2019. « MiRBase: From MicroRNA Sequences to Function ». *Nucleic Acids Research* 47 (D1): D155-62. <https://doi.org/10.1093/nar/gky1141>.

Kozomara, Ana, et Sam Griffiths-Jones. 2011. « miRBase: integrating microRNA annotation and deep-sequencing data ». *Nucleic Acids Research* 39 (Database issue): D152-57. <https://doi.org/10.1093/nar/gkq1027>.

Kratchmarova, Irina, Dario E. Kalume, Blagoy Blagoev, Philipp E. Scherer, Alexandre V.

- Podtelejnikov, Henrik Molina, Perry E. Bickel, et al. 2002. « A Proteomic Approach for Identification of Secreted Proteins during the Differentiation of 3T3-L1 Preadipocytes to Adipocytes ». *Molecular & Cellular Proteomics: MCP* 1 (3): 213-22.
- Lagesen, Karin, Peter Hallin, Einar Andreas Rødland, Hans-Henrik Stærfeldt, Torbjørn Rognes, et David W. Ussery. 2007. « RNAmmer: consistent and rapid annotation of ribosomal RNA genes ». *Nucleic Acids Research* 35 (9): 3100-3108. <https://doi.org/10.1093/nar/gkm160>.
- Lagos-Quintana, M., R. Rauhut, W. Lendeckel, et T. Tuschl. 2001. « Identification of Novel Genes Coding for Small Expressed RNAs ». *Science (New York, N.Y.)* 294 (5543): 853-58. <https://doi.org/10.1126/science.1064921>.
- Lanford, Robert E., Elisabeth S. Hildebrandt-Eriksen, Andreas Petri, Robert Persson, Morten Lindow, Martin E. Munk, Sakari Kauppinen, et Henrik Ørum. 2010. « Therapeutic Silencing of MicroRNA-122 in Primates with Chronic Hepatitis C Virus Infection ». *Science (New York, N.Y.)* 327 (5962): 198-201. <https://doi.org/10.1126/science.1178178>.
- Langmead, Ben, et Steven L Salzberg. 2012. « Fast gapped-read alignment with Bowtie 2 ». *Nature methods* 9 (4): 357-59. <https://doi.org/10.1038/nmeth.1923>.
- Laslett, Dean, et Bjorn Canback. 2004. « ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences ». *Nucleic Acids Research* 32 (1): 11-16. <https://doi.org/10.1093/nar/gkh152>.
- Lau, Lonneke M. L. de, et Monique M. B. Breteler. 2006. « Epidemiology of Parkinson's Disease ». *The Lancet. Neurology* 5 (6): 525-35. [https://doi.org/10.1016/S1474-4422\(06\)70471-9](https://doi.org/10.1016/S1474-4422(06)70471-9).
- Lau, N. C., L. P. Lim, E. G. Weinstein, et D. P. Bartel. 2001. « An Abundant Class of Tiny RNAs with Probable Regulatory Roles in *Caenorhabditis Elegans* ». *Science (New York, N.Y.)* 294 (5543): 858-62. <https://doi.org/10.1126/science.1065062>.
- Lee, R. C., et V. Ambros. 2001. « An Extensive Class of Small RNAs in *Caenorhabditis Elegans* ». *Science (New York, N.Y.)* 294 (5543): 862-64. <https://doi.org/10.1126/science.1065329>.
- Lee, R. C., R. L. Feinbaum, et V. Ambros. 1993. « The *C. Elegans* Heterochronic Gene *Lin-4*

Encodes Small RNAs with Antisense Complementarity to Lin-14 ». *Cell* 75 (5): 843-54.

Lees, Andrew J., John Hardy, et Tamas Revesz. 2009. « Parkinson's Disease ». *The Lancet* 373 (9680): 2055-66. [https://doi.org/10.1016/S0140-6736\(09\)60492-X](https://doi.org/10.1016/S0140-6736(09)60492-X).

Lesniewski, Ryan A., Sunit Jain, Karthik Anantharaman, Patrick D. Schloss, et Gregory J. Dick. 2012. « The Metatranscriptome of a Deep-Sea Hydrothermal Plume Is Dominated by Water Column Methanotrophs and Lithotrophs ». *The ISME Journal* 6 (12): 2257-68. <https://doi.org/10.1038/ismej.2012.63>.

Lewis, Benjamin P., Christopher B. Burge, et David P. Bartel. 2005. « Conserved Seed Pairing, Often Flanked by Adenosines, Indicates That Thousands of Human Genes Are MicroRNA Targets ». *Cell* 120 (1): 15-20. <https://doi.org/10.1016/j.cell.2004.12.035>.

Li, Ang-Di, Lei Tong, Nan Xu, Yao Ye, Peng-Yin Nie, Zhen-Yu Wang, et Li-Li Ji. 2019. « MiR-124 Regulates Cerebromicrovascular Function in APP/PS1 Transgenic Mice via C1ql3 ». *Brain Research Bulletin* 153 (septembre): 214-22. <https://doi.org/10.1016/j.brainresbull.2019.09.002>.

Li, Heng, et Richard Durbin. 2009. « Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform ». *Bioinformatics (Oxford, England)* 25 (14): 1754-60. <https://doi.org/10.1093/bioinformatics/btp324>.

Li, Heng, et Nils Homer. 2010. « A survey of sequence alignment algorithms for next-generation sequencing ». *Briefings in Bioinformatics* 11 (5): 473-83. <https://doi.org/10.1093/bib/bbq015>.

Liebler, Daniel C. 2002. « Proteomic Approaches to Characterize Protein Modifications: New Tools to Study the Effects of Environmental Exposures ». *Environmental Health Perspectives* 110 Suppl 1 (février): 3-9. <https://doi.org/10.1289/ehp.02110s113>.

Liu, Lin, Yinhu Li, Siliang Li, Ni Hu, Yimin He, Ray Pong, Danni Lin, Lihua Lu, et Maggie Law. 2012. « Comparison of Next-Generation Sequencing Systems ». Research article. *BioMed Research International*. 2012. <https://doi.org/10.1155/2012/251364>.

Lo, Chien-Chi, et Patrick S. G. Chain. 2014. « Rapid evaluation and quality control of next generation sequencing data with FaQCs ». *BMC Bioinformatics* 15 (1): 366.

<https://doi.org/10.1186/s12859-014-0366-2>.

Lorenz, Ronny, Stephan H. Bernhart, Christian Höner Zu Siederdisen, Hakim Tafer, Christoph Flamm, Peter F. Stadler, et Ivo L. Hofacker. 2011. « ViennaRNA Package 2.0 ». *Algorithms for Molecular Biology: AMB* 6 (novembre): 26. <https://doi.org/10.1186/1748-7188-6-26>.

Love, Michael I., Wolfgang Huber, et Simon Anders. 2014. « Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2 ». *Genome Biology* 15 (12): 550. <https://doi.org/10.1186/s13059-014-0550-8>.

Mackowiak, Sebastian D. 2011. « Identification of Novel and Known MiRNAs in Deep-Sequencing Data with MiRDeep2 ». *Current Protocols in Bioinformatics* Chapter 12 (décembre): Unit 12.10. <https://doi.org/10.1002/0471250953.bi1210s36>.

Maini Rekdal, Vayu, Elizabeth N. Bess, Jordan E. Bisanz, Peter J. Turnbaugh, et Emily P. Balskus. 2019. « Discovery and Inhibition of an Interspecies Gut Bacterial Pathway for Levodopa Metabolism ». *Science (New York, N.Y.)* 364 (6445). <https://doi.org/10.1126/science.aau6323>.

Makarova, Kira S., et Eugene V. Koonin. 2007. « Evolutionary Genomics of Lactic Acid Bacteria ». *Journal of Bacteriology* 189 (4): 1199-1208. <https://doi.org/10.1128/JB.01351-06>.

Mani, Sneha Ramesh, et Celina E. Juliano. 2013. « Untangling the Web: The Diverse Functions of the PIWI/PiRNA Pathway ». *Molecular Reproduction and Development* 80 (8): 632-64. <https://doi.org/10.1002/mrd.22195>.

Mardis, Elaine R. 2008. « The Impact of Next-Generation Sequencing Technology on Genetics ». *Trends in Genetics* 24 (3): 133-41. <https://doi.org/10.1016/j.tig.2007.12.007>.

Margis, Regina, Rogério Margis, et Carlos R. M. Rieder. 2011. « Identification of Blood MicroRNAs Associated to Parkinson's Disease ». *Journal of Biotechnology* 152 (3): 96-101. <https://doi.org/10.1016/j.jbiotec.2011.01.023>.

Martin, Marcel. 2011. « Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads ». *EMBnet.Journal* 17 (1): 10-12. <https://doi.org/10.14806/ej.17.1.200>.

Maruri, Fernanda, Timothy R. Sterling, Anne W. Kaiga, Amondrea Blackman, Yuri F. van der Heijden, Claudine Mayer, Emmanuelle Cambau, et Alexandra Aubry. 2012. « A Systematic

Review of Gyrase Mutations Associated with Fluoroquinolone-Resistant Mycobacterium Tuberculosis and a Proposed Gyrase Numbering System ». *The Journal of Antimicrobial Chemotherapy* 67 (4): 819-31. <https://doi.org/10.1093/jac/dkr566>.

Marvanova, Marketa. 2016. « Introduction to Parkinson disease (PD) and its complications ». *The Mental Health Clinician* 6 (5): 229-35. <https://doi.org/10.9740/mhc.2016.09.229>.

Marzorati, Massimo, Lois Maignien, An Verhelst, Gabriela Luta, Robert Sinnott, Frederiek Maarten Kerckhof, Nico Boon, Tom Van de Wiele, et Sam Possemiers. 2013. « Barcoded Pyrosequencing Analysis of the Microbial Community in a Simulator of the Human Gastrointestinal Tract Showed a Colon Region-Specific Microbiota Modulation for Two Plant-Derived Polysaccharide Blends ». *Antonie Van Leeuwenhoek* 103 (2): 409-20. <https://doi.org/10.1007/s10482-012-9821-0>.

Mascher, Martin, Shuangye Wu, Paul St Amand, Nils Stein, et Jesse Poland. 2013. « Application of Genotyping-by-Sequencing on Semiconductor Sequencing Platforms: A Comparison of Genetic and Reference-Based Marker Ordering in Barley ». *PLoS One* 8 (10): e76925. <https://doi.org/10.1371/journal.pone.0076925>.

Mathys, Vanessa, René Wintjens, Philippe Lefevre, Julie Bertout, Amit Singhal, Mehdi Kiass, Natalia Kurepina, et al. 2009. « Molecular Genetics of Para-Aminosalicylic Acid Resistance in Clinical Isolates and Spontaneous Mutants of Mycobacterium Tuberculosis ». *Antimicrobial Agents and Chemotherapy* 53 (5): 2100-2109. <https://doi.org/10.1128/AAC.01197-08>.

Matthieu FALQUE. 2011. « Evolution des méthodes de genotypage ».

Maxam, A. M., et W. Gilbert. 1977. « A New Method for Sequencing DNA ». *Proceedings of the National Academy of Sciences of the United States of America* 74 (2): 560-64.

McDougall, J. K., A. R. Dunn, et K. W. Jones. 1972. « In Situ Hybridization of Adenovirus RNA and DNA ». *Nature* 236 (5346): 346-48. <https://doi.org/10.1038/236346a0>.

McMillan, Kirsty J., Tracey K. Murray, Nora Bengoa-Vergniory, Oscar Cordero-Llana, Jane Cooper, Amy Buckley, Richard Wade-Martins, et al. 2017. « Loss of MicroRNA-7 Regulation Leads to  $\alpha$ -Synuclein Accumulation and Dopaminergic Neuronal Loss In Vivo ». *Molecular Therapy: The Journal of the American Society of Gene Therapy* 25 (10): 2404-14. <https://doi.org/10.1016/j.ymthe.2017.08.017>.

Meijer, Hedda A., Ewan M. Smith, et Martin Bushell. 2014. « Regulation of MiRNA Strand Selection: Follow the Leader? » *Biochemical Society Transactions* 42 (4): 1135-40. <https://doi.org/10.1042/BST20140142>.

Meldrum, Cliff, Maria A. Doyle, et Richard W. Tothill. 2011. « Next-Generation Sequencing for Cancer Diagnostics: A Practical Perspective ». *The Clinical Biochemist. Reviews* 32 (4): 177-95.

Metzker, Michael L. 2010. « Sequencing Technologies — the next Generation ». *Nature Reviews Genetics* 11 (1): 31-46. <https://doi.org/10.1038/nrg2626>.

Meyer, Folker, Ross Overbeek, et Alex Rodriguez. 2009. « FIGfams: yet another set of protein families ». *Nucleic Acids Research* 37 (20): 6643-54. <https://doi.org/10.1093/nar/gkp698>.

Miller, Jason R., Sergey Koren, et Granger Sutton. 2010. « Assembly Algorithms for Next-Generation Sequencing Data ». *Genomics* 95 (6): 315-27. <https://doi.org/10.1016/j.ygeno.2010.03.001>.

Miñones-Moyano, Elena, Sílvia Porta, Georgia Escaramís, Raquel Rabionet, Susana Iraola, Birgit Kagerbauer, Yolanda Espinosa-Parrilla, Isidre Ferrer, Xavier Estivill, et Eulàlia Martí. 2011. « MicroRNA Profiling of Parkinson's Disease Brains Identifies Early Downregulation of MiR-34b/c Which Modulate Mitochondrial Function ». *Human Molecular Genetics* 20 (15): 3067-78. <https://doi.org/10.1093/hmg/ddr210>.

Mitchison D.A. 1979. « Basic Mechanisms of Chemotherapy ». [https://journal.chestnet.org/article/S0012-3692\(16\)30810-8/fulltext](https://journal.chestnet.org/article/S0012-3692(16)30810-8/fulltext).

Mutez, E., M. -C. Chartier-Harlin, et A. Destée. 2013. « Génétique de la maladie de Parkinson ». *Pratique Neurologique - FMC* 4 (1): 3-10. <https://doi.org/10.1016/j.praneu.2012.12.002>.

Nakazato, Takeru, Tazro Ohta, et Hidemasa Bono. 2013. « Experimental Design-Based Functional Mining and Characterization of High-Throughput Sequencing Data in the Sequence Read Archive ». *PloS One* 8 (10): e77910. <https://doi.org/10.1371/journal.pone.0077910>.

Narasimhan, Padmanesan, James Wood, Chandini Raina MacIntyre, et Dilip Mathai. 2013. « Risk Factors for Tuberculosis ». *Pulmonary Medicine* 2013.

<https://doi.org/10.1155/2013/828939>.

National Collaborating Centre for Chronic Conditions (Great Britain). 2006. *Parkinson's Disease: National Clinical Guidelines for Diagnosis and Management in Primary and Secondary Care*. London: Royal College of Physicians.

Nawrocki, Eric P., Sarah W. Burge, Alex Bateman, Jennifer Daub, Ruth Y. Eberhardt, Sean R. Eddy, Evan W. Floden, et al. 2015. « Rfam 12.0: updates to the RNA families database ». *Nucleic Acids Research* 43 (Database issue): D130-37. <https://doi.org/10.1093/nar/gku1063>.

Nawrocki, Eric P., et Sean R. Eddy. 2013. « Infernal 1.1: 100-fold faster RNA homology searches ». *Bioinformatics* 29 (22): 2933-35. <https://doi.org/10.1093/bioinformatics/btt509>.

Obeso, J. A., M. C. Rodríguez-Oroz, M. Rodríguez, J. L. Lanciego, J. Artieda, N. Gonzalo, et C. W. Olanow. 2000. « Pathophysiology of the Basal Ganglia in Parkinson's Disease ». *Trends in Neurosciences* 23 (10 Suppl): S8-19.

O'Carroll, Dónal, et Anne Schaefer. 2013. « General Principles of MiRNA Biogenesis and Regulation in the Brain ». *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology* 38 (1): 39-54. <https://doi.org/10.1038/npp.2012.87>.

Oguh, Odinachi, et Aleksandar Videnovic. 2012. « Inpatient Management of Parkinson Disease: Current Challenges and Future Directions ». *The Neurohospitalist* 2 (1): 28-35. <https://doi.org/10.1177/1941874411427734>.

O'Leary, Nuala A., Mathew W. Wright, J. Rodney Brister, Stacy Ciuffo, Diana Haddad, Rich McVeigh, Bhanu Rajput, et al. 2016. « Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation ». *Nucleic Acids Research* 44 (Database issue): D733-45. <https://doi.org/10.1093/nar/gkv1189>.

Oliveto, Stefania, Marilena Mancino, Nicola Manfrini, et Stefano Biffo. 2017. « Role of MicroRNAs in Translation Regulation and Cancer ». *World Journal of Biological Chemistry* 8 (1): 45-56. <https://doi.org/10.4331/wjbc.v8.i1.45>.

Ozata, Deniz M., Ildar Gainetdinov, Ansgar Zoch, Dónal O'Carroll, et Phillip D. Zamore. 2019. « PIWI-Interacting RNAs: Small RNAs with Big Functions ». *Nature Reviews. Genetics* 20 (2): 89-108. <https://doi.org/10.1038/s41576-018-0073-3>.



Palomino, Juan Carlos, et Anandi Martin. 2014. « Drug Resistance Mechanisms in Mycobacterium tuberculosis ». *Antibiotics* 3 (3): 317-40. <https://doi.org/10.3390/antibiotics3030317>.

Pasquinelli, A. E., B. J. Reinhart, F. Slack, M. Q. Martindale, M. I. Kuroda, B. Maller, D. C. Hayward, et al. 2000. « Conservation of the Sequence and Temporal Expression of Let-7 Heterochronic Regulatory RNA ». *Nature* 408 (6808): 86-89. <https://doi.org/10.1038/35040556>.

Paxinos G, et Watson C. 1996. *The Rat Brain in Stereotaxic Coordinates - 6th Edition*. Academic Press, San Diego. <https://www.elsevier.com/books/the-rat-brain-in-stereotaxic-coordinates/paxinos/978-0-12-374121-9>.

Pemberton, Samantha, et Ronald Melki. 2012. « The interaction of Hsc70 protein with fibrillar  $\alpha$ -Synuclein and its therapeutic potential in Parkinson's disease ». *Communicative & Integrative Biology* 5 (1): 94-95.

Penaud, Stéphanie. 2007. « Analyse de la séquence génomique et Etude de l'adaptation à l'acidité de *L. delbrueckii* ssp. *bulgaricus* ATCC11842 », 268.

Pevzner, P. A., H. Tang, et M. S. Waterman. 2001. « An Eulerian Path Approach to DNA Fragment Assembly ». *Proceedings of the National Academy of Sciences of the United States of America* 98 (17): 9748-53. <https://doi.org/10.1073/pnas.171285098>.

Planell-Saguer, Mariàngels de, et Maria Celina Rodicio. 2013. « Detection Methods for MicroRNAs in Clinic Practice ». *Clinical Biochemistry* 46 (10-11): 869-78. <https://doi.org/10.1016/j.clinbiochem.2013.02.017>.

Pop, Mihai. 2009a. « Genome Assembly Reborn: Recent Computational Challenges ». *Briefings in Bioinformatics* 10 (4): 354-66. <https://doi.org/10.1093/bib/bbp026>.

———. 2009b. « Genome assembly reborn: recent computational challenges ». *Briefings in Bioinformatics* 10 (4): 354-66. <https://doi.org/10.1093/bib/bbp026>.

Pratt, Ashley J., et Ian J. MacRae. 2009. « The RNA-Induced Silencing Complex: A Versatile Gene-Silencing Machine ». *The Journal of Biological Chemistry* 284 (27): 17897-901. <https://doi.org/10.1074/jbc.R900012200>.

Quinodoz, Sofia, et Mitchell Guttman. 2014. « Long Noncoding RNAs: An Emerging Link between Gene Regulation and Nuclear Organization ». *Trends in Cell Biology* 24 (11): 651-63. <https://doi.org/10.1016/j.tcb.2014.08.009>.

Rabahi, Marcelo Fouad, José Laerte Rodrigues da Silva, Anna Carolina Galvão Ferreira, Daniela Graner Schuwartz Tannus-Silva, et Marcus Barreto Conde. 2017. « Tuberculosis treatment ». *Jornal Brasileiro de Pneumologia* 43 (6): 472-86. <https://doi.org/10.1590/S1806-37562016000000388>.

Ranganathan, Shoba. 2009. « Towards a career in bioinformatics ». *BMC Bioinformatics* 10 (Suppl 15): S1. <https://doi.org/10.1186/1471-2105-10-S15-S1>.

Rasmussen, Rosalee S., Michael T. Morrissey, et Paul D. N. Hebert. 2009. « DNA Barcoding of Commercially Important Salmon and Trout Species (*Oncorhynchus* and *Salmo*) from North America ». *Journal of Agricultural and Food Chemistry* 57 (18): 8379-85. <https://doi.org/10.1021/jf901618z>.

Rattka, Marta, Felix Fluri, Miloš Krstić, Esther Asan, et Jens Volkmann. 2016. « A Novel Approach to Assess Motor Outcome of Deep Brain Stimulation Effects in the Hemiparkinsonian Rat: Staircase and Cylinder Test ». *Journal of Visualized Experiments : JoVE*, n° 111 (mai). <https://doi.org/10.3791/53951>.

Rawat, Richa, Adrian Whitty, et Peter J. Tonge. 2003. « The isoniazid-NAD adduct is a slow, tight-binding inhibitor of InhA, the Mycobacterium tuberculosis enoyl reductase: Adduct affinity and drug resistance ». *Proceedings of the National Academy of Sciences of the United States of America* 100 (24): 13881-86. <https://doi.org/10.1073/pnas.2235848100>.

Reinhart, B. J., F. J. Slack, M. Basson, A. E. Pasquinelli, J. C. Bettinger, A. E. Rougvie, H. R. Horvitz, et G. Ruvkun. 2000. « The 21-Nucleotide Let-7 RNA Regulates Developmental Timing in *Caenorhabditis Elegans* ». *Nature* 403 (6772): 901-6. <https://doi.org/10.1038/35002607>.

Reuter, Jason A., Damek V. Spacek, et Michael P. Snyder. 2015. « High-Throughput Sequencing Technologies ». *Molecular Cell* 58 (4): 586-97. <https://doi.org/10.1016/j.molcel.2015.05.004>.

Riffo-Campos, Ángela L., Ismael Riquelme, et Priscilla Brebi-Mieville. 2016. « Tools for

Sequence-Based miRNA Target Prediction: What to Choose? » *International Journal of Molecular Sciences* 17 (12). <https://doi.org/10.3390/ijms17121987>.

Rizzo, Jason M., et Michael J. Buck. 2012. « Key Principles and Clinical Applications of “next-Generation” DNA Sequencing ». *Cancer Prevention Research (Philadelphia, Pa.)* 5 (7): 887-900. <https://doi.org/10.1158/1940-6207.CAPR-11-0432>.

Robinson, Mark D., et Gordon K. Smyth. 2008. « Small-Sample Estimation of Negative Binomial Dispersion, with Applications to SAGE Data ». *Biostatistics* 9 (2): 321-32. <https://doi.org/10.1093/biostatistics/kxm030>.

Rodriguez, Antony, Sam Griffiths-Jones, Jennifer L. Ashurst, et Allan Bradley. 2004. « Identification of Mammalian MicroRNA Host Genes and Transcription Units ». *Genome Research* 14 (10A): 1902-10. <https://doi.org/10.1101/gr.2722704>.

Rusk, Nicole. 2009. « Cheap Third-Generation Sequencing ». *Nature Methods* 6 (avril): 244. <https://doi.org/10.1038/nmeth0409-244a>.

Russell PJ. 2010. *iGenetics: A Molecular Approach*. 3<sup>e</sup> éd.

S. Andrews. 2010. « Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data ». 2010. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

Sá, Pablo H. C. G. de, Luis C. Guimarães, Diego A. das Graças, Adonney A. de Oliveira Veras, Debmalya Barh, Vasco Azevedo, Artur L. da Costa da Silva, et Rommel T. J. Ramos. 2018. « Chapter 11 - Next-Generation Sequencing and Data Analysis: Strategies, Tools, Pipelines and Protocols ». In *Omic Technologies and Bio-Engineering*, édité par Debmalya Barh et Vasco Azevedo, 191-207. Academic Press. <https://doi.org/10.1016/B978-0-12-804659-3.00011-7>.

Sabouni, Radia, Moussa Kourout, Imane Chaoui, Annemarie Jordaan, Mohammed Akrim, Thomas C. Victor, Karim Maltouf Filali, Mohammed El Mzibri, Ouafae Lahlou, et Rajae El Aouad. 2008. « Molecular Analysis of Multidrug Resistant Mycobacterium Tuberculosis Isolates from Morocco ». *Annals of Microbiology* 58 (4): 749. <https://doi.org/10.1007/BF03175585>.

Sanger, F., S. Nicklen, et A. R. Coulson. 1977. « DNA Sequencing with Chain-Terminating

Inhibitors ». *Proceedings of the National Academy of Sciences of the United States of America* 74 (12): 5463-67.

Schadt, Eric E., Steve Turner, et Andrew Kasarskis. 2010a. « A Window into Third-Generation Sequencing ». *Human Molecular Genetics* 19 (R2): R227-240. <https://doi.org/10.1093/hmg/ddq416>.

———. 2010b. « A Window into Third-Generation Sequencing ». *Human Molecular Genetics* 19 (R2): R227-240. <https://doi.org/10.1093/hmg/ddq416>.

Schatz, Michael C., Arthur L. Delcher, et Steven L. Salzberg. 2010. « Assembly of Large Genomes Using Second-Generation Sequencing ». *Genome Research* 20 (9): 1165-73. <https://doi.org/10.1101/gr.101360.109>.

Schmieder, Robert, et Robert Edwards. 2011. « Quality Control and Preprocessing of Metagenomic Datasets ». *Bioinformatics (Oxford, England)* 27 (6): 863-64. <https://doi.org/10.1093/bioinformatics/btr026>.

Schmittgen, Thomas D., Eun Joo Lee, Jinmai Jiang, Anasuya Sarkar, Liuqing Yang, Terry S. Elton, et Caifu Chen. 2008. « Real-Time PCR Quantification of Precursor and Mature MicroRNA ». *Methods (San Diego, Calif.)* 44 (1): 31-38. <https://doi.org/10.1016/j.ymeth.2007.09.006>.

Seemann, Torsten. 2014. « Prokka: Rapid Prokaryotic Genome Annotation ». *Bioinformatics (Oxford, England)* 30 (14): 2068-69. <https://doi.org/10.1093/bioinformatics/btu153>.

Sensoy, Gulnar, Arzu Karli, Asuman Birinci, Pelin Uras, Meltem Bilgici, et Filiz Karagoz. 2016. « Multidrug-Resistant Disseminated Tuberculosis in a 9-Month-Old Infant ». *Eastern Mediterranean Health Journal* 22 (1): 59-61. <https://doi.org/10.26719/2016.22.1.59>.

Shao, Keke, Weifeng Ding, Feng Wang, Haiquan Li, Da Ma, et Huimin Wang. 2011. « Emulsion PCR: A High Efficient Way of PCR Amplification of Random DNA Libraries in Aptamer Selection ». *PloS One* 6 (9): e24910. <https://doi.org/10.1371/journal.pone.0024910>.

Shi, Jieming, Min Dong, Lei Li, Lin Liu, Agustin Luz-Madrigal, Panagiotis A. Tsonis, Katia Del Rio-Tsonis, et Chun Liang. 2015. « MirPRo-a Novel Standalone Program for Differential Expression and Variation Analysis of MiRNAs ». *Scientific Reports* 5 (octobre): 14617.

<https://doi.org/10.1038/srep14617>.

Shruti, K., K. Shrey, et R. Vibha. 2011. « Micro RNAs: Tiny Sequences with Enormous Potential ». *Biochemical and Biophysical Research Communications* 407 (3): 445-49. <https://doi.org/10.1016/j.bbrc.2011.03.058>.

Skogar, Orjan, et Johan Lökk. 2016. « Pain management in patients with Parkinson's disease: challenges and solutions ». *Journal of Multidisciplinary Healthcare* 9 (septembre): 469-79. <https://doi.org/10.2147/JMDH.S105857>.

Smith, Issar. 2003. « Mycobacterium tuberculosis Pathogenesis and Molecular Determinants of Virulence ». *Clinical Microbiology Reviews* 16 (3): 463-96. <https://doi.org/10.1128/CMR.16.3.463-496.2003>.

Smith, L. M., J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. Kent, et L. E. Hood. 1986. « Fluorescence Detection in Automated DNA Sequence Analysis ». *Nature* 321 (6071): 674-79. <https://doi.org/10.1038/321674a0>.

Sotgiu, Giovanni, Rosella Centis, Lia D'ambrosio, et Giovanni Battista Migliori. 2015. « Tuberculosis Treatment and Drug Regimens ». *Cold Spring Harbor Perspectives in Medicine* 5 (5). <https://doi.org/10.1101/cshperspect.a017822>.

Staden, R. 1979. « A strategy of DNA sequencing employing computer programs. » *Nucleic Acids Research* 6 (7): 2601-10.

Stanley, Robin E., Gregor Blaha, Robert L. Grodzicki, Michael D. Strickler, et Thomas A. Steitz. 2010. « The structures of the anti-tuberculosis antibiotics viomycin and capreomycin bound to the 70S ribosome ». *Nature structural & molecular biology* 17 (3): 289-93. <https://doi.org/10.1038/nsmb.1755>.

Steve M.Blevins, et Michael S.Bronze. 2010. « Robert Koch and the “golden Age” of Bacteriology ». <https://doi.org/10.1016/j.ijid.2009.12.003>.

Sun, Zhaogang, Jianyuan Zhang, Xuxia Zhang, Sumin Wang, Ying Zhang, et Chuanyou Li. 2008. « Comparison of GyrA Gene Mutations between Laboratory-Selected Ofloxacin-Resistant Mycobacterium Tuberculosis Strains and Clinical Isolates ». *International Journal of Antimicrobial Agents* 31 (2): 115-21. <https://doi.org/10.1016/j.ijantimicag.2007.10.014>.

Telenti, A., W. J. Philipp, S. Sreevatsan, C. Bernasconi, K. E. Stockbauer, B. Wiele, J. M. Musser, et W. R. Jacobs. 1997. « The Emb Operon, a Gene Cluster of Mycobacterium Tuberculosis Involved in Resistance to Ethambutol ». *Nature Medicine* 3 (5): 567-70.

Thome, Aaron D., Ashley S. Harms, Laura A. Volpicelli-Daley, et David G. Standaert. 2016. « MicroRNA-155 Regulates Alpha-Synuclein-Induced Inflammatory Responses in Models of Parkinson Disease ». *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 36 (8): 2383-90. <https://doi.org/10.1523/JNEUROSCI.3900-15.2016>.

Tolia, Niraj H., et Leemor Joshua-Tor. 2007. « Slicer and the Argonautes ». *Nature Chemical Biology* 3 (1): 36-43. <https://doi.org/10.1038/nchembio848>.

UCSC Genome Browser Gateway. 2018. « UCSC Genome Browser Gateway ». 2018. <https://genome.ucsc.edu/cgi-bin/hgGateway>.

Ul Hussain, Mahboob. 2012. « Micro-RNAs (MiRNAs): Genomic Organisation, Biogenesis and Mode of Action ». *Cell and Tissue Research* 349 (2): 405-13. <https://doi.org/10.1007/s00441-012-1438-0>.

Vallenet, D., S. Engelen, D. Mornico, S. Cruveiller, L. Fleury, A. Lajus, Z. Rouy, et al. 2009. « MicroScope: a platform for microbial genome annotation and comparative genomics ». *Database: The Journal of Biological Databases and Curation* 2009. <https://doi.org/10.1093/database/bap021>.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, et al. 2001. « The Sequence of the Human Genome ». *Science (New York, N.Y.)* 291 (5507): 1304-51. <https://doi.org/10.1126/science.1058040>.

Vidigal, Joana A., et Andrea Ventura. 2015. « The Biological Functions of MiRNAs: Lessons from in Vivo Studies ». *Trends in Cell Biology* 25 (3): 137-47. <https://doi.org/10.1016/j.tcb.2014.11.004>.

Villar-Menéndez, Izaskun, Sílvia Porta, Sandra P. Buira, Thais Pereira-Veiga, Sara Díaz-Sánchez, José Luis Albasanz, Isidre Ferrer, Mairena Martín, et Marta Barrachina. 2014. « Increased Striatal Adenosine A2A Receptor Levels Is an Early Event in Parkinson's Disease-Related Pathology and It Is Potentially Regulated by MiR-34b ». *Neurobiology of Disease* 69 (septembre): 206-14. <https://doi.org/10.1016/j.nbd.2014.05.030>.

Vincent, V. 1995. « Taxonomie des mycobactéries ». *Revue Française des Laboratoires* 1995 (273): 27-31. [https://doi.org/10.1016/S0338-9898\(95\)80195-2](https://doi.org/10.1016/S0338-9898(95)80195-2).

Waddington, C. H. 1957. « The Strategy of the Genes. A Discussion of Some Aspects of Theoretical Biology. With an Appendix by H. Kacser. » *The Strategy of the Genes. A Discussion of Some Aspects of Theoretical Biology. With an Appendix by H. Kacser.* <https://www.cabdirect.org/cabdirect/abstract/19580101706>.

———. 2012. « The Epigenotype. 1942 ». *International Journal of Epidemiology* 41 (1): 10-13. <https://doi.org/10.1093/ije/dyr184>.

Wang, Wei-Chi, Feng-Mao Lin, Wen-Chi Chang, Kuan-Yu Lin, Hsien-Da Huang, et Na-Sheng Lin. 2009. « MiRExpress: Analyzing High-Throughput Sequencing Data for Profiling MicroRNA Expression ». *BMC Bioinformatics* 10 (octobre): 328. <https://doi.org/10.1186/1471-2105-10-328>.

Wang, Zhong, Mark Gerstein, et Michael Snyder. 2009. « RNA-Seq: a revolutionary tool for transcriptomics ». *Nature reviews. Genetics* 10 (1): 57-63. <https://doi.org/10.1038/nrg2484>.

Wanger, Audrey, Violeta Chavez, Richard S. P. Huang, Amer Wahed, Jeffrey K. Actor, et Amitava Dasgupta. 2017. « Chapter 12 - Overview of Molecular Diagnostics Principles ». In *Microbiology and Molecular Diagnosis in Pathology*, édité par Audrey Wanger, Violeta Chavez, Richard S. P. Huang, Amer Wahed, Jeffrey K. Actor, et Amitava Dasgupta, 233-57. Elsevier. <https://doi.org/10.1016/B978-0-12-805351-5.00012-0>.

Watanabe, Toshiaki, Shin-ichi Tomizawa, Kohzoh Mitsuya, Yasushi Totoki, Yasuhiro Yamamoto, Satomi Kuramochi-Miyagawa, Naoko Iida, et al. 2011. « Role for PiRNAs and Noncoding RNA in de Novo DNA Methylation of the Imprinted Mouse *Rasgrf1* Locus ». *Science (New York, N.Y.)* 332 (6031): 848-52. <https://doi.org/10.1126/science.1203919>.

« WHO | World Health Organization ». 2018. WHO. 2018. <http://www.who.int/tb/areas-of-work/children/en/>.

Wightman, B., I. Ha, et G. Ruvkun. 1993. « Posttranscriptional Regulation of the Heterochronic Gene *Lin-14* by *Lin-4* Mediates Temporal Pattern Formation in *C. Elegans* ». *Cell* 75 (5): 855-62.

Willenbrock, Hanni, Jesper Salomon, Rolf Søkilde, Kim Bundvig Barken, Thomas Nøhr Hansen, Finn Cilius Nielsen, Søren Møller, et Thomas Litman. 2009. « Quantitative MiRNA Expression Analysis: Comparing Microarrays with next-Generation Sequencing ». *RNA (New York, N.Y.)* 15 (11): 2028-34. <https://doi.org/10.1261/rna.1699809>.

Williamson, Vernell, Albert Kim, Bin Xie, G. Omari McMichael, Yuan Gao, et Vladimir Vladimirov. 2013. « Detecting MiRNAs in Deep-Sequencing Data: A Software Performance Comparison and Evaluation ». *Briefings in Bioinformatics* 14 (1): 36-45. <https://doi.org/10.1093/bib/bbs010>.

Wilson. 1925. « THE Croonian Lectures ON SOME DISORDERS OF MOTILITY AND OF MUSCLE TONE, WITH SPECIAL REFERENCE TO THE CORPUS STRIATUM. » *The Lancet*, Originally published as Volume 2, Issue 5318, 206 (5318): 215-19. [https://doi.org/10.1016/S0140-6736\(00\)46763-2](https://doi.org/10.1016/S0140-6736(00)46763-2).

Worboys, Michael. 1990. « Robert Koch: a life in medicine and bacteriology ». *Medical History* 34 (3): 347-48.

World Health Organisation. 2018. « Tuberculosis (TB) ». 2018. <https://www.who.int/news-room/fact-sheets/detail/tuberculosis>.

Wu, Ruifang, Yuwen Su, Haijing Wu, Yong Dai, Ming Zhao, et Qianjin Lu. 2016. « Characters, Functions and Clinical Perspectives of Long Non-Coding RNAs ». *Molecular Genetics and Genomics: MGG* 291 (3): 1013-33. <https://doi.org/10.1007/s00438-016-1179-y>.

Wu, Yuncheng, Weidong Le, et Joseph Jankovic. 2011. « Preclinical Biomarkers of Parkinson Disease ». *Archives of Neurology* 68 (1): 22-30. <https://doi.org/10.1001/archneurol.2010.321>.

Xilouri, Maria, Oeystein Roed Brekk, Deniz Kirik, et Leonidas Stefanis. 2013. « LAMP2A as a Therapeutic Target in Parkinson Disease ». *Autophagy* 9 (12): 2166-68. <https://doi.org/10.4161/auto.26451>.

Xiong, Jin. 2006. *Essential Bioinformatics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511806087>.

Yang, Dehua, Ting Li, Yi Wang, Yuanjia Tang, Huijuan Cui, Yu Tang, Xiaojie Zhang, Degui Chen, Nan Shen, et Weidong Le. 2012. « MiR-132 Regulates the Differentiation of Dopamine



Neurons by Directly Targeting Nurr1 Expression ». *Journal of Cell Science* 125 (Pt 7): 1673-82. <https://doi.org/10.1242/jcs.086421>.

Yang, Li-hong, Si-lu Wang, Li-li Tang, Biao Liu, Wen-le Ye, Ling-ling Wang, Zhang-yang Wang, Meng-tao Zhou, et Bi-cheng Chen. 2014. « Universal Stem-Loop Primer Method for Screening and Quantification of MicroRNA ». *PLoS ONE* 9 (12). <https://doi.org/10.1371/journal.pone.0115293>.

Yu, Qiu-Jin, Shu-Yang Yu, Li-Jun Zuo, Teng-Hong Lian, Yang Hu, Rui-Dan Wang, Ying-Shan Piao, et al. 2018. « Parkinson disease with constipation: clinical features and relevant factors ». *Scientific Reports* 8 (janvier). <https://doi.org/10.1038/s41598-017-16790-8>.

Zaubrecher, M. Analise, R. David Sikes, Beverly Metchock, Thomas M. Shinnick, et James E. Posey. 2009. « Overexpression of the Chromosomally Encoded Aminoglycoside Acetyltransferase Eis Confers Kanamycin Resistance in Mycobacterium Tuberculosis ». *Proceedings of the National Academy of Sciences of the United States of America* 106 (47): 20004-9. <https://doi.org/10.1073/pnas.0907925106>.

Zerbino, Daniel R., et Ewan Birney. 2008. « Velvet: Algorithms for de Novo Short Read Assembly Using de Bruijn Graphs ». *Genome Research* 18 (5): 821-29. <https://doi.org/10.1101/gr.074492.107>.

Zhang, Y., B. Heym, B. Allen, D. Young, et S. Cole. 1992. « The Catalase-Peroxidase Gene and Isoniazid Resistance of Mycobacterium Tuberculosis ». *Nature* 358 (6387): 591-93. <https://doi.org/10.1038/358591a0>.

Zhang, Zhelin, et Yan Cheng. 2014. « MiR-16-1 Promotes the Aberrant  $\alpha$ -Synuclein Accumulation in Parkinson Disease via Targeting Heat Shock Protein 70 ». *TheScientificWorldJournal* 2014: 938348. <https://doi.org/10.1155/2014/938348>.

Zhao, Fei, Xu-De Wang, Luke N. Erber, Ming Luo, Ai-zhen Guo, Shan-shan Yang, Jing Gu, et al. 2014. « Binding Pocket Alterations in Dihydrofolate Synthase Confer Resistance to Para-Aminosalicylic Acid in Clinical Isolates of Mycobacterium Tuberculosis ». *Antimicrobial Agents and Chemotherapy* 58 (3): 1479-87. <https://doi.org/10.1128/AAC.01775-13>.

Zhao, Liang, et Zhiqin Wang. 2019. « MicroRNAs: Game Changers in the Regulation of  $\alpha$ -Synuclein in Parkinson's Disease ». *Parkinson's Disease* 2019 (mai).

<https://doi.org/10.1155/2019/1743183>.

Zhao, Na, Lirong Jin, Guoqiang Fei, Zhiyong Zheng, et Chunjiu Zhong. 2014. « Serum MicroRNA-133b Is Associated with Low Ceruloplasmin Levels in Parkinson's Disease ». *Parkinsonism & Related Disorders* 20 (11): 1177-80. <https://doi.org/10.1016/j.parkreldis.2014.08.016>.

Zhou, Xiaofan, et Antonis Rokas. 2014. « Prevention, Diagnosis and Treatment of High-Throughput Sequencing Data Pathologies ». *Molecular Ecology* 23 (7): 1679-1700. <https://doi.org/10.1111/mec.12680>.

Zhou, Yang, Cheng Chen, Hui Jiang, Hong-Qiu Pan, Li-Mei Zhu, et Wei Lu. 2019. « High admission rates and heavy inpatient service costs of urban tuberculosis patients in eastern China ». *BMC Health Services Research* 19 (1): 47. <https://doi.org/10.1186/s12913-019-3892-9>.

Zhu, Erle, Fangqing Zhao, Gang Xu, Huabin Hou, Linglin Zhou, Xiaokun Li, Zhongsheng Sun, et Jinyu Wu. 2010. « MirTools: MicroRNA Profiling and Discovery Based on High-Throughput Sequencing ». *Nucleic Acids Research* 38 (Web Server issue): W392-397. <https://doi.org/10.1093/nar/gkq393>.

Zuo, Lingjun, Zhiren Wang, Yunlong Tan, Xiangning Chen, et Xingguang Luo. 2016. « PiRNAs and Their Functions in the Brain ». *International Journal of Human Genetics* 16 (1-2): 53-60.

# Annexe 1

## Lancement du Séquençage et création de SampleSheet

La plateforme MiSeq est le premier système qui offre un écosystème de séquençage entièrement intégré. Le laboratoire MedBiotech a fait la mise en marche du système la première fois en février 2016.



Figure 72: Séquenceur Illumina MiSeq

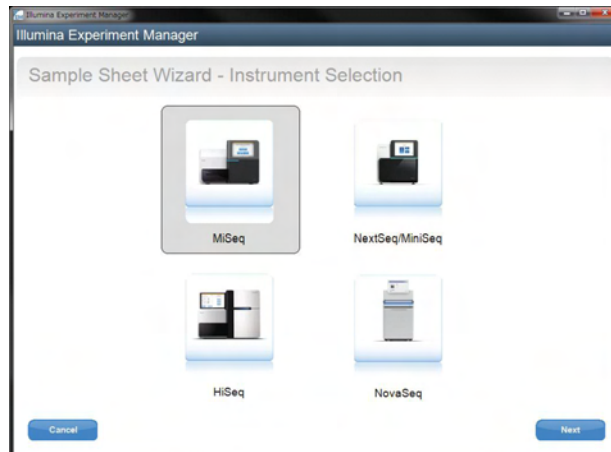
Pour créer une SampleSheet, le logiciel *Illumina Experiment Manager* est utilisé (installé au niveau du séquenceur), à travers l'application intitulée 'small genome sequencing' puis on passe à l'étape d'indexation.

1. Ouvrez le logiciel Illumina Experiment Manager (v1.11.0 utilisée à titre d'exemple).
2. Dans le l'écran principal, choisissez Create Sample Sheet.



Figure 73: Le logiciel Illumina Experiment Manager

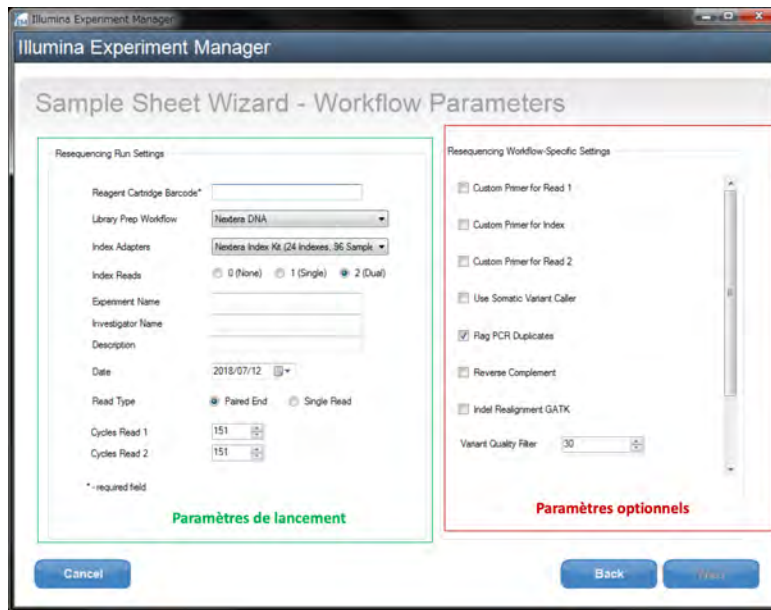
3. Sélectionnez le séquenceur (MiSeq utilisé à titre d'exemple), puis sélectionnez Suivant.



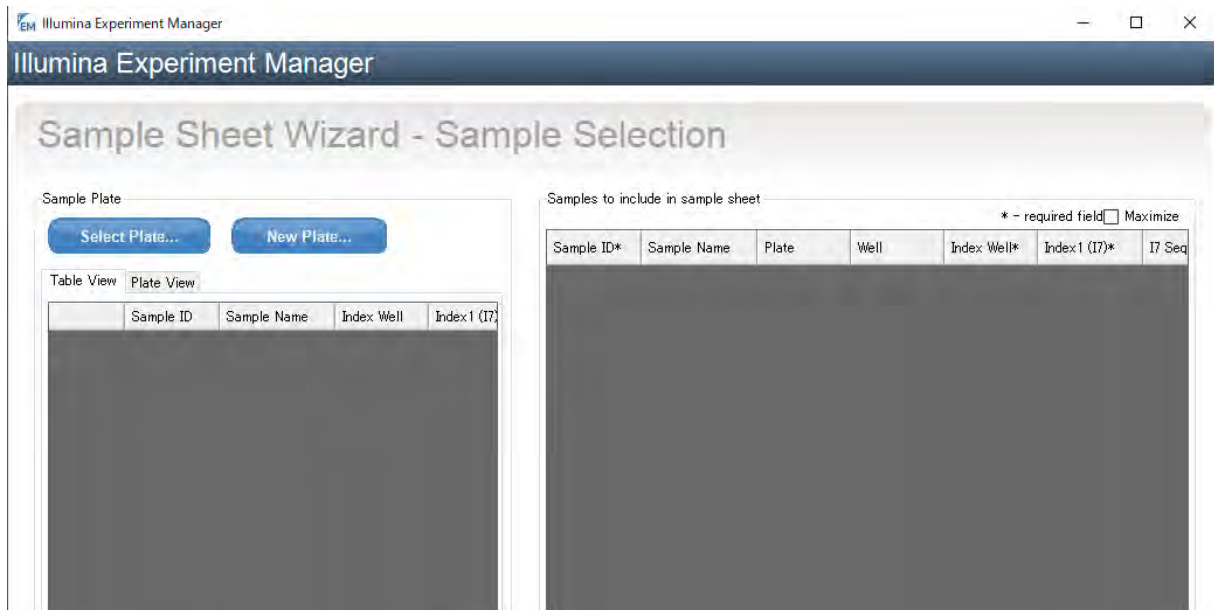
4. Sélectionnez l'application appropriée (small genome sequencing ensuite resequencing utilisé uniquement à titre d'exemple), puis sélectionnez Suivant.



5. Sur la page Paramètres de flux de travail, choisissez Nextera XT pour le kit de préparation de librairie. Entrez les autres informations requises (aucune amorce personnalisée pour les adaptateurs sparQ). Sélectionnez Suivant à la sélection de l'échantillon.



6. Sélectionnez, ajouter une ligne vide pour ajouter des lignes et entrez des informations sur l'échantillon. Les numéros de code-barres de l'adaptateur sparQ correspondent aux numéros Index1 (I7). Par exemple, sélectionnez A004 dans le champ Index1 (I7) pour sparQ Adapter Barcode 4; sélectionnez A012 dans le champ Index1 (I7) pour le code à barres sparQ Adapter 12. Entrez les autres informations requises et sélectionnez Terminer pour compléter la feuille d'échantillons.



# Annexe 2

## La description du fichier FastQ

Le format FASTQ est un format de fichier texte permettant de stocker à la fois des séquences biologiques (uniquement des séquences nucléiques) et les scores de qualité associés. La séquence et le score sont chacune codées avec un seul caractère ASCII.

```
1 @ERR000589.41 EAS139_45:5:1:2:111/1
2 CTTTCCTCCCTGCTTTCCTGGCCCCACCATTTCCAGGGAACATCTTGTCAT
3 +
4 3IIIIIIIIIIII>1IIIF9BG08E00I%IG+&?(4)%00646.C1#&(
5 @ERR000589.42 EAS139_45:5:1:2:1293/1
6 AGTTGTTAAAATCCAAGCCAATTAAGATAGTCTTATCTTTTTAAAAGAAAT
7 +
8 IIIIIIGII.AIIII=?I9G-/II=+I=4?761BA2C9I+5A711+&>1$/I
```

Chaque séquence est codée sur 4 lignes :

1. un @, suivi du nom de la séquence ;
2. la séquence elle-même suit ;
3. un + (avec éventuellement le nom de la séquence, encore une fois) ;
4. la qualité de la séquence.

La qualité de la séquence suit un codage particulier, où chaque caractère représente un nombre.

En général, l'association est la suivante :

!	“	#	\$	%	&	‘	(	)	*	+	,	-	.	/	0	1	2	3	4
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

5	6	7	8	9	:	;	<	=	>	?	@	A	B	C	D	E	F	G	H
20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39

I
40

Chaque nombre représente la probabilité  $p$  de se tromper sur la lecture d'une base. Le code représente la valeur  $-10 \log_{10} (p)$ . Par exemple, le caractère C code le nombre 34. Il représente donc une probabilité d'erreur d'environ  $4 \cdot 10^{-4}$ . Les codes les plus à droite représentent donc les meilleures qualités.

## Annexe 3

### Liste des miRNAs utilisés pour la RT QPCR

Nom du miRNA	Séquence
rno-mir-1247	GTC-CCG-TTC-GTC-CCC-GGA
rno-mir-374-5p	GCA-TAT-AAT-ACA-ACC-TGC-TAA-GTG
rno-mir-7a-2-3p	CTG-TAC-AGC-CTC-CTA-GCT-TTC
rno-mir-376a-5p	GCGGTAGATTCTCCTTCTATGAG
rno-mir-3068-3p	GGT-GAA-TTG-CAG-TAC-TCC-AAC-A
rno-let-7d-3p	CTA-TAC-GAC-CTG-CTG-CCT-TTC-T
rno-mir-1306-5p	CCACCTCCCCTGCAAACGTCCA
rno-mir-379-5p	TGG-TAG-ACT-ATG-GAA-CGT-AGG
rno-mir-24-2-5p	GTGCCTACTGAGCTGAAACAGT
rno-mir-375-3p	TGTTTCGTTTCGGCTCGCGTGA
rno-mir-24-3p	AGGCACGCGGTGAATGCC
rno-mir-411-5p	TAG-TAG-ACC-GTA-TAG-CGT-ACG
rno-mir-361-5p	TTA-TCA-GAA-TCT-CCA-GGG-GTA-C
rno-mir-666-3p	TGC-AGC-GTG-ATC-GCC-TGC-T
rno-mir-93-5p	CAAAGTGCTGTTCGTGCAGGTAG
rno-miR-375-3p	TGTTTCGTTTCGGCTCGCGTGA
rno-miR-34c-5p	AGG-CAG-TGT-AGT-TAG-CTG-ATT-GC
rno-miR-34b-3p	GAATCACTAACTCCACTGCCATC
rno-miR-10a-5p	TACCCTGTAGATCCGAATTTGTG
rno-miR-34c-3p	ACT-AAC-CAC-ACG-GCC-AGG
rno-miR-7a-5p	GCTGGAAGACTAGTGATTTTGTGT
rno-miR-674-5p	CAC-AGC-TCC-CAT-CTC-AGA-ACA-A
rno-miR-320-3p	CTG-GGT-TGA-GAG-GGC-GA
rno-miR-379-5p	TGG-TAG-ACT-ATG-GAA-CGT-AGG
rno-miR-23a-5p	GGG-TTC-CTG-GGG-ATG-GGA-TTT
rno-miR-183-5p	GGT-ATG-GCA-CTG-GTA-GAA-TTC-ACT



rno-miR-201-5p	GTA-CTC-AGT-AAG-GCA-TTG-TTC-TT
rno-miR-184	GGA-CGG-AGA-ACT-GAT-AAG-GGT
rno-miR-483-3p	CACTCCTCCCCTCCCGTCTTGT