

Année 2022/2023

N°

Thèse

Pour le

DOCTORAT EN MEDECINE

Diplôme d'État

par

Antoine CIVIT

Né le 10 novembre 1992 à Nancy, 54

Utilisation du séquençage long-fragment dans les maladies à expansion de triplets

Présentée et soutenue publiquement le 26 octobre 2023 date devant un jury composé de :

Président du Jury : Professeur Annick TOUTAIN, Génétique Médicale, PU, Faculté de Médecine - Tours

Membres du Jury :

Docteur Paul GUEGUEN, Génétique Médicale, PH, CHU – Tours

Docteur Anne TALLET, Biologie Médicale, PH, CHU – Tours

Madame Dr Marie-Laure WINTER VUILLAUME, Génétique Médicale, MCU-PH, Faculté de Médecine – Tours

Directeur de thèse : Docteur Paul GUEGUEN, Génétique Médicale - Tours

RÉSUMÉ EN FRANÇAIS

Introduction. Les maladies génétiques à expansion nucléotidique sont un groupe de maladies génétiques causées par l'expansion anormale d'une séquence de nucléotides dans un gène. La séquence est ainsi anormalement longue ayant pour conséquences diverses pathologies. Le diagnostic et la compréhension moléculaire de ces maladies se sont avérés difficiles en raison des limites des techniques conventionnelles de séquençage. Les avancées technologiques continuent de fournir de nouvelles méthodes et outils pour le diagnostic de ces maladies complexes. Ainsi le séquençage long-fragment semble être une technologie d'avenir dans le diagnostic et la caractérisation des maladies de nucléotides.

Matériels et méthodes. Cette thèse explore l'utilisation du séquençage long-fragment « Nanopore » associé à la technique CRISPR/Cas9 comme méthode innovante pour étudier les maladies à expansion de triplets et plus particulièrement le syndrome de l'X fragile, la maladie de Steinert et l'ataxie de Friedreich. Nous avons donc évalué la faisabilité du séquençage long-fragment « Nanopore » associé à la technique CRISPR/Cas9 pour détecter et quantifier les expansions de triplets dans ces trois maladies.

Résultats. Nos résultats montrent une technologie de séquençage long-fragment intéressante. Cependant la technologie CRISPR/Cas9 est difficile à utiliser. D'autres méthodes sont à expérimenter.

Conclusion. D'autres propositions méthodologiques associées au séquençage long-fragment paraissent plus performantes. Ainsi il semblerait que l'utilisation de ce type de séquençage ouvre la voie à un diagnostic plus précis et à une meilleure compréhension des mécanismes sous-jacents des maladies à expansion nucléotidique.

RÉSUMÉ EN ANGLAIS

Introduction. Repeat expansion diseases are a group of genetic diseases caused by the abnormal expansion of a nucleotide sequence in a gene. The sequence is thus abnormally long, resulting in various pathologies. Diagnosis and molecular understanding of these diseases have proven difficult due to the limitations of conventional sequencing techniques. Technological advances continue to provide new methods and tools for the diagnosis of these complex diseases. Long-fragment sequencing therefore appears to be a technology of the future in the diagnosis and characterization of expansion repeat diseases.

Materials and methods. This thesis explores the use of "Nanopore" long-fragment sequencing associated with the CRISPR/Cas9 technique as an innovative method to study repeat expansion diseases and more particularly fragile X syndrome, Friedreich's ataxia and the Steinert's disease. We therefore evaluated the feasibility of long-read sequencing associated with the CRISPR/Cas9 technology to detect and quantify triplet expansions in these three diseases.

Results. Our results show an interesting long-fragment sequencing technology. However, CRISPR/Cas9 technology is difficult to use. Other methods are worth experimenting with.

Conclusion. Other methodological proposals associated with long-fragment sequencing appear more efficient. It would appear that the use of this type of sequencing opens the way to more precise diagnosis and a better understanding of the underlying mechanisms of nucleotide expansion diseases.

MOTS-CLÉS

Maladies à expansion de triplets ; Microsatellites ; Courtes répétitions en tandem ;
Séquençage long-fragment ; CRISPR/Cas9 ;

KEYWORDS

Repeat expansion diseases ; Short tandem repeat ; Long-read Sequencing ; CRISPR/Cas9 ;

UNIVERSITE DE TOURS
FACULTE DE MEDECINE DE TOURS

DOYEN

Pr Patrice DIOT

VICE-DOYEN

Pr Henri MARRET

ASSESSEURS

Pr Denis ANGOULVANT, *Pédagogie*

Pr Mathias BUCHLER, *Relations internationales*

Pr Theodora BEJAN-ANGOULVANT, *Moyens – relations avec l'Université*

Pr Clarisse DIBAO-DINA, *Médecine générale*

Pr François MAILLOT, *Formation Médicale Continue*

Pr Patrick VOUREC'H, *Recherche*

RESPONSABLE ADMINISTRATIVE

Mme Carole ACCOLAS

DOYENS HONORAIRES

Pr Emile ARON (†) – 1962-1966

Directeur de l'Ecole de Médecine - 1947-1962

Pr Georges DESBUQUOIS (†) – 1966-1972

Pr André GOUAZE (†) – 1972-1994

Pr Jean-Claude ROLLAND – 1994-2004

Pr Dominique PERROTIN – 2004-2014

PROFESSEURS EMERITES

Pr Daniel ALISON

Pr Gilles BODY

Pr Philippe COLOMBAT

Pr Etienne DANQUECHIN-DORVAL

Pr Luc FAVARD

Pr Bernard FOUQUET

Pr Yves GRUEL

Pr Gérard LORETTE

Pr Loïc VAILLANT

PROFESSEURS HONORAIRES

P. ANTHONIOZ – P. ARBEILLE – A. AUDURIER – A. AUTRET – C. BARTHELEMY – J.L. BAULIEU – C. BERGER – JC. BESNARD – P. BEUTTER – C. BONNARD – P. BONNET – P. BOUGNOUX – P. BURDIN – L. CASTELLANI – J. CHANDENIER – A. CHANTEPIE – B. CHARBONNIER – P. CHOUTET – T. CONSTANS – C. COUET – L. DE LA LANDE DE CALAN – P. DUMONT – J.P. FAUCHIER – F. FETISSOF – J. FUSCIARDI – P. GAILLARD – G. GINIES – D. GOGA – A. GOUDEAU – J.L. GUILMOT – O. HAILLOT – N. HUTEN – M. JAN – J.P. LAMAGNERE – F. LAMISSE – Y. LANSON – O. LE FLOCH – Y. LEBRANCHU – E. LECA – P. LECOMTE – AM. LEHR-DRYLEWICZ – E. LEMARIE – G. LEROY – M. MARCHAND – C. MAURAGE – C. MERCIER – J. MOLINE – C. MORAIN – J.P. MUH – J. MURAT – H. NIVET – D. PERROTIN – L. POURCELOT – R. QUENTIN – P. RAYNAUD – D. RICHARD-LENOBLE – A. ROBIER – J.C. ROLLAND – P. ROSSET – D. ROYERE – A. SAINDELLE – E. SALIBA – J.J. SANTINI – D. SAUVAGE – D. SIRINELLI – J. WEILL

PROFESSEURS DES UNIVERSITES - PRATICIENS HOSPITALIERS

ANDRES Christian.....	Biochimie et biologie moléculaire
ANGOULVANT Denis	Cardiologie
APETOH Lionel	Immunologie
AUPART Michel.....	Chirurgie thoracique et cardiovasculaire
BABUTY Dominique	Cardiologie
BACLE Guillaume.....	Chirurgie orthopédique et traumatologique
BAKHOS David.....	Oto-rhino-laryngologie
BALLON Nicolas.....	Psychiatrie ; addictologie
BARBIER François.....	Médecine intensive et réanimation
BARILLOT Isabelle	Cancérologie ; radiothérapie
BARON Christophe	Immunologie
BEJAN-ANGOULVANT Théodora	Pharmacologie clinique
BERHOUEZ Julien	Chirurgie orthopédique et traumatologique
BERNARD Anne	Cardiologie
BERNARD Louis	Maladies infectieuses et maladies tropicales
BLANCHARD-LAUMONNIER Emmanuelle	Biologie cellulaire
BLASCO Hélène.....	Biochimie et biologie moléculaire
BONNET-BRILHAULT Frédérique	Physiologie
BOURGUIGNON Thierry	Chirurgie thoracique et cardiovasculaire
BRILHAULT Jean.....	Chirurgie orthopédique et traumatologique
BRUNEREAU Laurent	Radiologie et imagerie médicale
BRUYERE Franck.....	Urologie
BUCHLER Matthias.....	Néphrologie
CAILLE Agnès	Biostat., informatique médical et technologies de communication
CALAIS Gilles.....	Cancérologie, radiothérapie
CAMUS Vincent.....	Psychiatrie d'adultes
CORCIA Philippe.....	Neurologie
COTTIER Jean-Philippe	Radiologie et imagerie médicale
DEQUIN Pierre-François.....	Thérapeutique
DESMIDT Thomas	Psychiatrie
DESOUBEUX Guillaume.....	Parasitologie et mycologie
DESTRIEUX Christophe	Anatomie
DI GUISTO Caroline	Gynécologie obstétrique
DIOT Patrice.....	Pneumologie
DU BOUEXIC de PINIEUX Gonzague	Anatomie & cytologie pathologiques
DUCLUZEAU Pierre-Henri.....	Endocrinologie, diabétologie, et nutrition
EHRMANN Stephan	Médecine intensive – réanimation
EL HAGE Wissam.....	Psychiatrie adultes
ELKRIEF Laure.....	Hépatologie – gastroentérologie
ESPITALIER Fabien.....	Anesthésiologie et réanimation, médecine d'urgence
FAUCHIER Laurent	Cardiologie
FOUGERE Bertrand	Gériatrie
FRANCOIS Patrick.....	Neurochirurgie
FROMONT-HANKARD Gaëlle	Anatomie & cytologie pathologiques
GATAULT Philippe.....	Néphrologie
GAUDY-GRAFFIN Catherine.....	Bactériologie-virologie, hygiène hospitalière
GOUPILLE Philippe	Rhumatologie
GUERIF Fabrice	Biologie et médecine du développement et de la reproduction
GUILLON Antoine.....	Médecine intensive – réanimation
GUILLON-GRAMMATICO Leslie.....	Epidémiologie, économie de la santé et prévention
GUYETANT Serge	Anatomie et cytologie pathologiques
GYAN Emmanuel.....	Hématologie, transfusion
HALIMI Jean-Michel.....	Thérapeutique
HANKARD Régis.....	Pédiatrie
HERAULT Olivier	Hématologie, transfusion
HERBRETEAU Denis	Radiologie et imagerie médicale
HOURIOUX Christophe.....	Biologie cellulaire
IVANES Fabrice	Physiologie
LABARTHE François	Pédiatrie
LAFFON Marc	Anesthésiologie et réanimation chirurgicale, médecine d'urgence
LARDY Hubert.....	Chirurgie infantile
LARIBI Saïd.....	Médecine d'urgence
LARTIGUE Marie-Frédérique	Bactériologie-virologie
LAURE Boris.....	Chirurgie maxillo-faciale et stomatologie
LECOMTE Thierry.....	Gastroentérologie, hépatologie

LEGRAS Antoine.....	Chirurgie thoracique
LESCANNE Emmanuel.....	Oto-rhino-laryngologie
LEVESQUE Éric.....	Anesthésiologie et réanimation chirurgicale, médecine d'urgence
LINASSIER Claude	Cancérologie, radiothérapie
MACHET Laurent	Dermato-vénéréologie
MAILLOT François	Médecine interne
MARCHAND-ADAM Sylvain	Pneumologie
MARRET Henri	Gynécologie-obstétrique
MARUANI Annabel	Dermatologie-vénéréologie
MEREGHETTI Laurent.....	Bactériologie-virologie ; hygiène hospitalière
MITANCHEZ Delphine	Pédiatrie
MOREL Baptiste	Radiologie pédiatrique
MORINIERE Sylvain.....	Oto-rhino-laryngologie
MOUSSATA Driffa	Gastro-entérologie
MULLEMAN Denis.....	Rhumatologie
ODENT Thierry.....	Chirurgie infantile
OUAISSI Mehdi	Chirurgie digestive
OULDAMER Lobna	Gynécologie-obstétrique
PAINTAUD Gilles	Pharmacologie fondamentale, pharmacologie clinique
PATAT Frédéric	Biophysique et médecine nucléaire
PERROTIN Franck	Gynécologie-obstétrique
PISELLA Pierre-Jean.....	Ophtalmologie
PLANTIER Laurent.....	Physiologie
REMERAND Francis	Anesthésiologie et réanimation, médecine d'urgence
ROINGEARD Philippe.....	Biologie cellulaire
RUSCH Emmanuel.....	Epidémiologie, économie de la santé et prévention
SAINT-MARTIN Pauline.....	Médecine légale et droit de la santé
SALAME Ephrem.....	Chirurgie digestive
SAMIMI Mahtab	Dermatologie-vénéréologie
SANTIAGO-RIBEIRO Maria	Biophysique et médecine nucléaire
SAUTENET-BIGOT Bénédicte	Thérapeutique
THOMAS-CASTELNAU Pierre	Pédiatrie
TOUTAIN Annick.....	Génétique
VELUT Stéphane.....	Anatomie
VOURC'H Patrick.....	Biochimie et biologie moléculaire
WATIER Hervé	Immunologie
ZEMMOURA Ilyess	Neurochirurgie

PROFESSEUR DES UNIVERSITES DE MEDECINE GENERALE

DIBAO-DINA Clarisse
LEBEAU Jean-Pierre

PROFESSEURS ASSOCIES

MALLET Donatien

PROFESSEUR CERTIFIE DU 2ND DEGRE

MC CARTHY Catherine.....

MAITRES DE CONFERENCES DES UNIVERSITES - PRATICIENS HOSPITALIERS

AUDEMARD-VERGER Alexandra	Médecine interne
BISSON Arnaud	Cardiologie (CHRO)
BRUNAUT Paul	Psychiatrie d'adultes, addictologie
CARVAJAL-ALLEGRIA Guillermo	Rhumatologie (au 01/10/2021)
CLEMENTY Nicolas	Cardiologie
DOMELIER Anne-Sophie	Bactériologie-virologie, hygiène hospitalière
DUFOUR Diane	Biophysique et médecine nucléaire
FOUQUET-BERGEMER Anne-Marie	Anatomie et cytologie pathologiques
GARGOT Thomas	Pédopsychiatrie
GOUILLEUX Valérie	Immunologie
HOARAU Cyrille	Immunologie
KERVAREC Thibault	Anatomie et cytologie pathologiques
LE GUELLEC Chantal	Pharmacologie fondamentale, pharmacologie clinique
LEDUCQ Sophie	Dermatologie
LEFORT Bruno	Pédiatrie
LEJEUNE Julien	Hématologie, transfusion
LEMAIGNEN Adrien	Maladies infectieuses
MACHET Marie-Christine	Anatomie et cytologie pathologiques
MOUMNEH Thomas	Médecine d'urgence
PARE Arnaud	Chirurgie maxillo-faciale et stomatologie
PIVER Éric	Biochimie et biologie moléculaire
ROUMY Jérôme	Biophysique et médecine nucléaire
STANDLEY-MIQUELESTORENA Elodie	Anatomie et cytologie pathologiques
STEFIC Karl	Bactériologie
TERNANT David	Pharmacologie fondamentale, pharmacologie clinique
VAYNE Caroline	Hématologie, transfusion
VUILLAUME-WINTER Marie-Laure	Génétique

MAITRES DE CONFERENCES DES UNIVERSITES

AGUILLON-HERNANDEZ Nadia	Neurosciences
NICOLOU Antonine	Philosophie – histoire des sciences et des techniques
PATIENT Romuald	Biologie cellulaire
RENOUX-JACQUET Cécile	Médecine Générale

MAITRES DE CONFERENCES ASSOCIES

AUMARECHAL Alain	Médecine Générale
BARBEAU Ludivine	Médecine Générale
CHAMANT Christelle	Médecine Générale
ETTORI Isabelle	Médecine Générale
MOLINA Valérie	Médecine Générale
PAUTRAT Maxime	Médecine Générale
PHILIPPE Laurence	Médecine Générale
RUIZ Christophe	Médecine Générale
SAMKO Boris	Médecine Générale

CHERCHEURS INSERM - CNRS - INRAE

BECKER Jérôme.....	Chargé de Recherche Inserm – UMR Inserm 1253
BOUAKAZ Ayache	Directeur de Recherche Inserm – UMR Inserm 1253
BOUTIN Hervé.....	Directeur de Recherche Inserm – UMR Inserm 1253
BRIARD Benoit.....	Chargé de Recherche Inserm – UMR Inserm 1100
CHALON Sylvie	Directrice de Recherche Inserm – UMR Inserm 1253
DE ROCQUIGNY Hugues	Chargé de Recherche Inserm – UMR Inserm 1259
ESCOFFRE Jean-Michel.....	Chargé de Recherche Inserm – UMR Inserm 1253
GILOT Philippe.....	Chargé de Recherche Inrae – UMR Inrae 1282
GOMOT Marie.....	Chargée de Recherche Inserm – UMR Inserm 1253
GOUILLEUX Fabrice	Directeur de Recherche CNRS – EA 7501 - ERL CNRS 7001
GUEGUINOU Maxime.....	Chargé de Recherche Inserm – UMR Inserm 1069
HEUZE-VOURCH Nathalie.....	Directrice de Recherche Inserm – UMR Inserm 1100
KORKMAZ Brice.....	Chargé de Recherche Inserm – UMR Inserm 1100
LATINUS Marianne.....	Chargée de Recherche Inserm – UMR Inserm 1253
LAUMONNIER Frédéric	Directeur de Recherche Inserm – UMR Inserm 1253
LE MERRER Julie	Directrice de Recherche CNRS – UMR Inserm 1253
MAMMANO Fabrizio.....	Directeur de Recherche Inserm – UMR Inserm 1259
MEUNIER Jean-Christophe	Chargé de Recherche Inserm – UMR Inserm 1259
PAGET Christophe	Directeur de Recherche Inserm – UMR Inserm 1100
RAOUL William	Chargé de Recherche Inserm – UMR Inserm 1069
SECHER Thomas.....	Chargé de Recherche Inserm – UMR Inserm 1100
SI TAHAR Mustapha	Directeur de Recherche Inserm – UMR Inserm 1100
SUREAU Camille	Directrice de Recherche émérite CNRS – UMR Inserm 1259
TANTI Arnaud	Chargé de Recherche Inserm – UMR Inserm 1253
WARDAK Claire.....	Chargée de Recherche Inserm – UMR Inserm 1253

CHARGES D'ENSEIGNEMENT

Pour l'éthique médicale

BIRMELE Béatrice.....Praticien Hospitalier

Pour la médecine manuelle et l'ostéopathie médicale

LAMANDE Marc.....Praticien Hospitalier

Pour l'orthophonie

BATAILLE Magalie.....Orthophoniste
CLOUTOUR Nathalie.....Orthophoniste
CORBINEAU Mathilde.....Orthophoniste
EL AKIKI Carole.....Orthophoniste
HARIVEL OUALLI Ingrid.....Orthophoniste
IMBERT Mélanie.....Orthophoniste
SIZARET Eva.....Orthophoniste

Pour l'orthoptie

BOULNOIS Sandrine.....Orthoptiste

SERMENT D'HIPPOCRATE

En présence des enseignants et enseignantes
de cette Faculté,
de mes chers condisciples
et selon la tradition d'Hippocrate,
je promets et je jure d'être fidèle aux lois de l'honneur
et de la probité dans l'exercice de la Médecine.

Je donnerai mes soins gratuits aux indigents,
et n'exigerai jamais un salaire au-dessus de mon travail.

Admis(e) dans l'intérieur des maisons, mes yeux
ne verront pas ce qui s'y passe, ma langue taira
les secrets qui me seront confiés et mon état ne servira pas
à corrompre les mœurs ni à favoriser le crime.

Respectueux(euse) et reconnaissant(e) envers mes Maîtres,
je rendrai à leurs enfants
l'instruction que j'ai reçue de leurs parents.

Que les hommes et les femmes m'accordent leur estime
si je suis fidèle à mes promesses.
Que je sois couvert(e) d'opprobre
et méprisé(e) de mes confrères et consœurs
si j'y manque.

Remerciements

À la Présidente du jury,

Madame le Professeur Annick TOUTAIN,

Professeur des Universités, Praticien Hospitalier, Génétique médicale

Vous avez su m'accueillir à Tours et m'encadrer depuis le début de mon internat. Vous me faites l'honneur de juger ce travail et de présider ce jury de thèse. Je vous prie d'accepter l'expression de mon plus profond respect.

Aux différents membres du jury,

À Madame le Docteur Anne TALLET,

Praticien Hospitalier, Biologie médicale

Je vous remercie d'avoir accepté de juger mon travail et de prendre part à ce jury.

À Madame le Docteur Marie-Laure WINTER VUILLAUME,

Maître de conférences des universités, Praticien Hospitalier, Génétique médicale

Je te remercie de m'avoir encadré pendant ces années et m'avoir donner le goût de la génétique moléculaire.

À mon directeur de thèse, Monsieur le Docteur Paul GUEGUEN,

Praticien Hospitalier, Génétique médicale

Je te remercie très sincèrement de m'avoir proposé d'encadrer ce travail de thèse, tout en me laissant la liberté du sujet et de la manière de le traiter. Tu as été présent, patient et d'un grand soutien durant ce travail.

À Monsieur le Professeur Patrick VOURC'H,
je vous remercie d'avoir été disponible, à l'écoute ces années passées, et pour les nombreux
encouragements que vous m'avez prodigués.

À Monsieur le Docteur Médéric JEANNE,
je te remercie d'avoir pris le temps de transmettre tes réflexions et ton savoir, de discuter des
dossiers, de répondre à toutes mes questions.

À Madame le Docteur Stéphanie ARPIN,
je te remercie pour ta bonne humeur et ta disponibilité.

À Madame le Docteur Noémie CELTON,
je te remercie de m'avoir accueilli pendant 6 mois au laboratoire de cytogénétique et du temps
que tu as pu m'accorder.

À Madame le Docteur Nathalie LE DU,
je te remercie pour les différents échanges et ton ouverture d'esprit.

À Madame Isabelle MORTEMOSQUE et à Monsieur le Docteur Edouard COTTEREAU,
je vous remercie pour les quelques moments passés ensemble.

À mes différents cointernes, et tout particulièrement à Solène,
je vous remercie d'avoir été à mes côtés.

Au service de génétique tout entier,
je vous remercie pour le temps accordé et votre convivialité.

À mes parents, pour leur soutien indéfectible,

À mes grands-mères,

À mon frère,

À ma soeur,

À Anna,

Au reste de ma famille,

À mes amis,

TABLE DES MATIÈRES

<u>INTRODUCTION</u>	p.15
<u>REVUE DE LA LITTÉRATURE</u>	p.17
<u>I - Maladies à expansion de nucléotides</u>	p.17
A -Pathologies associées au gène FMR1.....	p.20
B - Ataxie de Friedreich.....	p.23
C - Myotonie de Steinert.....	p.25
<u>II-Techniques actuelles de diagnostic</u>	p.28
A - Southern Blot.....	p.28
B - PCR.....	p.30
<u>III - Nouvelles technologies</u>	p.32
A - Edition génomique et CRISPR/Cas9.....	p.32
B - Séquençage long-fragment.....	p.35
<u>MATÉRIEL ET MÉTHODES</u>	p.40
<u>RÉSULTATS</u>	p.45
<u>DISCUSSION</u>	p.53
<u>CONCLUSION</u>	p.57
<u>ANNEXES</u>	p.58
<u>BIBLIOGRAPHIE</u>	p.59

Les maladies génétiques à expansion nucléotidique sont causées par un nombre de répétitions anormalement élevé d'une petite séquence de nucléotides (triplets ou autres) au sein d'un gène. Ces expansions de répétitions ont pour conséquence des dysfonctionnements moléculaires, comme des anomalies de transcription, d'épissage du pré-ARNm ou de traduction de l'ARNm affectant le niveau d'expression et/ou la fonction de la protéine codée. Les conséquences de ces expansions nucléotidiques sont nombreuses, à savoir des atteintes du système nerveux, des troubles cardiaques ou musculaires par exemple (1).

Le diagnostic clinique des maladies à expansion nucléotidique est difficile du fait de la nature même des expansions nucléotidiques, qui peuvent varier en taille entre les individus et même au sein des cellules d'un même individu. Une évaluation clinique précise et minutieuse, comprenant entre autres des examens morphologiques, neurologiques, et des examens fonctionnels musculaires est souvent nécessaire avant de conduire des analyses moléculaires. Le recours à des examens d'imagerie médicale est parfois requis.

Le diagnostic clinique est confirmé par un diagnostic moléculaire. Historiquement, la technique utilisée pour ces diagnostics était le Southern Blot (1). Cette technique longue à réaliser et lourde à mettre en place n'est plus beaucoup employée dans les laboratoires de génétique moléculaire. Elle présente toutefois un avantage important, celui de pouvoir donner une taille aux expansions de nucléotides, qui sont parfois de très grandes tailles. Cette approche est parfois intéressante pour étudier les corrélations génotypes-phénotypes. Les techniques les plus pratiquées à ce jour sont basées sur la PCR, à savoir les PCR d'exclusion, la TP-PCR et les PCR de type AmpliPhi.

Aujourd'hui, la PCR est souvent le premier test effectué pour détecter une expansion de nucléotides. Cette technologie est fiable, rapide à mettre en place et adaptée aux expansions de tailles modérées. La PCR se décline sous différentes techniques qui présentent chacune des avantages et des inconvénients. Globalement ces techniques de PCR ne détectent pas toutes les tailles d'expansion nucléotidique, en particulier les grandes expansions et ne permettent pas de déterminer la taille de l'expansion. Elles sont utiles pour exclure un diagnostic positif. Par ailleurs, le nombre de répétition des motifs nucléotidiques peut varier considérablement au sein d'une population (de plusieurs dizaines à plusieurs milliers de répétitions pour une même maladie) et même chez un individu donné (mosaïcisme). Il est donc souvent nécessaire

de compléter la PCR avec un Southern Blot. Pour les expansions de triplets particulièrement grandes, supérieures à 100 triplets répétés qui sont hors de portée pour la PCR, le Southern Blot est souvent utilisé. Cette technique nécessite toutefois une grande quantité d'ADN (5-10 µg) comparativement à la PCR. La résolution du Southern Blot peut parfois être insuffisante pour distinguer des expansions de taille quasi équivalente. Enfin, le Southern Blot est en général moins sensible pour les expansions de petite taille (< inférieur à 100 répétitions) que la PCR.

Le développement de nouvelles technologies en biologie moléculaire a permis d'améliorer le diagnostic moléculaire de ces maladies à expansions nucléotidiques. Le séquençage de long-fragments d'ADN en molécules uniques (Long-Read Sequencing) est une de ces approches développée ces dernières années. Elle permet l'étude de fragments d'ADN beaucoup plus longs par rapport aux techniques de séquençage haut débit de deuxième génération NGS (Next-Generation Sequencing). Le système CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR Associated protein 9) est une autre technologie de biologie moléculaire qui a été développée récemment. Cette méthode d'édition du génome permet des modifications très précises de l'ADN qui ont de nombreuses applications notamment en médecine. Cette technique permet par exemple l'enrichissement des échantillons d'ADN en certains gènes, l'enrichissement étant le processus par lequel la proportion d'un gène ou d'une région génomique, dans un échantillon, est augmentée par rapport au génome total.

L'association du séquençage long-fragment et de la technique CRISPR/Cas9 pourrait offrir des possibilités très intéressantes d'analyses dans les domaines de la recherche et du diagnostic moléculaires de maladie génétiques, en particulier pour les maladies génétiques complexes comme les maladies à expansion de triplets.

Ma thèse a pour objectif d'explorer l'utilisation du séquençage long-fragment « Nanopore » associé au système CRISPR/Cas9 comme méthode innovante pour le diagnostic moléculaire des maladies à expansion de triplets et plus particulièrement le syndrome de l'X fragile, la maladie de Steinert et l'ataxie de Friedreich. Nous avons donc évalué la faisabilité de cette technologie pour détecter et quantifier les expansions de triplets responsables de ces maladies.

REVUE DE LA LITTÉRATURE

I - Maladies génétiques et séquences répétées

1 - Physiologie

Le génome humain est composé à plus d'un tiers de séquences répétées. Ces séquences sont répétées en tandem et sont classées selon la taille du motif répété en séquences satellites, minisatellites ou microsatellites. Les microsatellites, répétitions en tandem de séquences de courtes tailles, représentent entre 1-3 % du génome humain (2). Ces microsatellites sont des marqueurs polymorphes dispersés le long du génome humain. Ils sont utilisés depuis longtemps comme outils pour conduire des études de liaison génétique.

La conservation de ces séquences répétées dans les espèces eucaryotes (3) a suggéré qu'elles pourraient être impliquées dans des fonctions biologiques importantes. Ces séquences répétées ont, en effet, un impact sur les niveaux de transcription de gènes situés à proximité ou de gènes dans lesquels elles sont situées, ceci via une modification de structure de la chromatine. Les séquences répétées facilitent par exemple le repliement tridimensionnel du génome et modulent les niveaux de méthylation de l'ADN. Certaines séquences répétées peuvent être méthylées, ce qui conduit à une diminution voire à une absence d'expression des gènes (4). Une séquence répétée un grand nombre de fois dans le promoteur d'un gène peut avoir pour conséquence une augmentation de l'expression de celui-ci, via une liaison aux de transcription plus efficace. Des structures secondaires stables formées par certaines séquences répétées peuvent au contraire inhiber l'expression génique, ceci en empêchant le fonctionnement de la transcription. Ces mécanismes ont été décrits en lien avec les diverses séquences répétées, dont les microsatellites (séquences courtes répétées en tandem). Les microsatellites peuvent donc stimuler ou inhiber la transcription des gènes en fonction de leur séquence et de leur localisation dans ou à proximité des gènes en question (2).

Les séquences courtes répétées en tandem sont également impliquées dans la biologie des ARN. Ces séquences dans les ARN modifient l'épissage des pré-ARNm en ARNm mature via la formation de structures tridimensionnelles complexes. Elles régulent aussi les niveaux de

traduction des ARNm en protéines. Certaines d'entre elles, quand elles sont localisées dans les régions 5'UTR ou dans les régions codantes, influent sur l'efficacité de la traduction (5).

Les séquences courtes répétées en tandem agissent aussi au niveau protéique. Les triplets dans la séquence codante correspondent à des codons répétés conduisant à des répétitions d'acides aminés dans les protéines codées. Ces répétitions dans les protéines modifient la structure protéique et peuvent ainsi faciliter la formation de complexes multiprotéiques (5).

2 - Facteurs contribuant aux mécanismes pathogènes des maladies à expansion de nucléotides

Les séquences répétées en tandem ont été découvertes il y a une trentaine d'années pour être directement à l'origine de maladies génétiques humaines. Dans la majorité de ces maladies, des troubles neurologiques sont présents. La découverte de l'expansion de ces séquences répétées a par ailleurs transformé notre compréhension de l'hérédité mendélienne. L'instabilité de certaines des séquences répétées dans le génome humain conduit en effet à des expansions nucléotidiques qui peuvent varier en taille entre différentes générations dans une même famille, entraînant de génération en génération une aggravation des manifestations cliniques et l'apparition de la maladie à des âges plus précoces. Ce mécanisme appelé anticipation clinique est fréquent. L'importance de cette anticipation et l'influence du parent qui transmet la maladie varie selon la maladie (6).

Les expansions nucléotidiques pathologiques proviennent de séquences répétées polymorphes de tailles normales (c'est-à-dire non pathologiques). Le niveau de variabilité de ces tailles normales diffère selon la séquence répétée et donc de la maladie. Plus les séquences répétées sont longues plus elles sont susceptibles de s'étendre (expansion) quand elles sont transmises d'une génération à l'autre, et donc de faire apparaître les maladies. Il est donc possible que des mutations par expansion apparaissent dans des familles qui n'avaient pas d'antécédent de telles maladies. Ainsi, l'absence d'antécédent dans une famille ne devrait pas dissuader un clinicien à prescrire un diagnostic moléculaire recherchant une expansion nucléotidique (6).

L'une des caractéristiques les plus marquantes des maladies liées à l'expansion de séquences répétées est la grande variabilité des symptômes pour une même maladie. Cette hétérogénéité des symptômes est principalement influencée par la taille des séquences répétées.

La physiopathologie d'une maladie causée par des séquences répétées nucléotidiques est influencée par divers éléments. La composition en nucléotides de la séquence répétée, telle que GAA, CAG, CTG par exemple, participe à la conformation intramoléculaire de l'ADN au niveau de ces répétitions nucléotidiques. La taille de la séquence répétée pathologique dans une maladie donnée est un autre facteur clé, déterminant à la fois la sévérité de la maladie et sa présentation clinique. Le nombre de répétition peut varier de quelques dizaines à plus d'un millier parfois. Les séquences pathologiques de plus courte taille sont généralement localisées dans des régions du gène qui codent les protéines. Ainsi, des modifications mineures dans le motif de répétition peuvent avoir des conséquences importantes sur la fonction de la protéine codée. En comparaison, les grandes expansions se situent souvent dans des introns ou des régions intergéniques. Les liens entre la taille précises de ces grandes répétitions et les phénotypes observés (symptômes, sévérité) de la maladie sont plus difficiles à établir. La localisation des séquences répétées au sein du gène a bien sûr aussi un impact direct sur les mécanismes moléculaires impliqués dans la maladie. Les séquences répétées situées à proximité ou dans un promoteur peuvent fortement modifier l'expression génique, celles localisées dans les introns peuvent perturber l'épissage des pré-ARNm en piégeant les protéines impliquées dans l'épissage et/ou en modifiant l'épissage alternatif. Les séquences répétées situées dans les régions 3'UTR peuvent aussi modifier la stabilité des ARNm. Enfin, les séquences répétées présentes dans les séquences codantes influent sur la traduction en protéine. Depuis quelques temps on sait également que des séquences répétées situées dans des régions non traduites des ARN (pré-ARNm), comme les introns ou les régions 5'UTR peuvent être le support d'une traduction dite RAN (Repeat Associated Non-AUG translation), produisant des protéines potentiellement toxiques (ref ?). Le degré d'instabilité méiotique pour une répétition donnée détermine le degré d'anticipation d'une maladie à expansion de triplets. Le degré d'instabilité méiotique dépend de nombreux facteurs, notamment le nombre de répétitions, le fait que la répétition soit continue ou interrompue, la séquence de la répétition et le positionnement de la répétition par rapport aux nucléosomes et à d'autres facteurs chromosomiques. De plus, une instabilité somatique dans les maladies à expansion de

répétitions est clairement présente. Plus la répétition est longue, plus la probabilité d'instabilité somatique est grande (6).

A - Pathologies associées au gène FMR1 (7, 8)

1 - Syndrome X fragile (Martin-Bell Syndrome) (FXS)

Le syndrome de l'X fragile est la cause la plus fréquente de retard mental monogénique. Il atteint surtout les hommes. Il se caractérise par un trouble du neurodéveloppement dès les premières années de vie, une déficience intellectuelle (modérée à sévère), associés à des troubles du comportement et des troubles du spectre de l'autisme (TSA). Une dysmorphie faciale peut être présente mais reste discrète. Les caractéristiques physiques cranio-faciales sont difficilement reconnaissables chez l'enfant en bas âge et deviennent plus évidentes avec l'âge. Classiquement, la croissance staturale est normale et un périmètre crânien important est observé. Une macro-orchidie est identifiée chez pratiquement tous les hommes après la puberté. Les signes cliniques ne sont ni spécifiques, ni constants et, notamment chez le jeune enfant, la déficience intellectuelle est parfois isolée. Une microcéphalie et un micro-génitalisme sont des critères d'exclusion d'un syndrome de l'X fragile. L'atteinte chez les filles est en général plus modérée avec des troubles cognitifs de sévérité très variable.

Le syndrome de l'X fragile est causé dans la grande majorité des cas par une expansions de triplets CGG localisés en 5'UTR du gène *FMR1* qui aboutit à une perte de la transcription et donc de la traduction de la protéine FMRP. Dans de très rares cas, le syndrome est causé par une mutation ponctuelle ou une délétion du gène *FMR1*.

2 - Insuffisance ovarienne prématurée liée au gène FMR1 (FXPOI)

Le FXPOI est défini comme un hypogonadisme hypergonadotrope avant l'âge de 40 ans. Les femmes atteintes présentent une insuffisance ovarienne prématurée et donc une ménopause plus précocément que la population générale. Elles souffrent d'infertilité et de symptômes de type ménopausique (symptômes vasomoteurs, changements d'humeur, sécheresse vaginale). Certaines femmes continuent d'avoir une ovulation et des règles irrégulières pendant des années après le diagnostic. D'autres sont complètement aménorrhéiques. Les femmes atteintes

présentent un risque de séquelles de santé à long terme dues à un environnement hypoestrogénique et un risque accru de développer une maladie thyroïdienne.

3 - Ataxie – tremblements liés au gène *FMRI* (FXTAS)

Il s'agit d'un syndrome neurologique d'apparition tardive qui est retrouvé chez certains hommes et femmes. Le diagnostic est porté sur des critères cliniques, neuroradiologiques, moléculaires, voire neuropathologiques. Le FXTAS est caractérisé par une ataxie cérébelleuse progressive d'apparition tardive, d'un tremblement intentionnel et d'un déficit cognitif. L'âge d'apparition des symptômes se situe généralement entre 60 et 65 ans mais varie considérablement selon les individus. La progression des symptômes est également très variable selon les individus. La déficience intellectuelle commence généralement par une déficience des fonctions exécutives et s'étend à d'autres domaines. Près de la moitié des personnes atteintes de FXTAS répondent aux critères de démence. Les troubles psychiatriques sont courants.

4 - Pathologie moléculaire et corrélation génotype-phénotype

Le mode de transmission est lié au chromosome X.

Le gène *FMRI* code pour la protéine FMRP qui se lie à certains ARNm pour les véhiculer du noyau au cytoplasme. Dans les neurones, les complexes FMRP-ARNm sont transportés jusqu'à la synapse.

Le gène *FMRI* comporte, dans sa région 5'UTR, une répétition de triplets CGG dont le nombre est variable. On distingue différentes classes d'allèles :

- Allèles normaux : 6 – 44 répétitions
- Allèles de la zone intermédiaire : 45-54 répétitions. Les allèles compris entre 50 et 54 peuvent être instables lors de leur transmission : ils ne présentent pas de risque d'expansion vers une mutation complète en une seule génération mais ils peuvent passer dans l'intervalle des prémutations (≥ 55 répétitions CGG)

- Allèles prémutés $\geq 55 - 200$ répétitions. Les allèles prémutés sont instables lors de la transmission. Le risque de passage à une mutation complète est corrélé au nombre de répétitions.
- Allèles mutés > 200 répétitions avec méthylation anormale. Les allèles mutés sont instables lors de la transmission.

5 - Anticipation

Le FXS est un trouble qui peut montrer une anticipation dans certaines familles. Généralement, l'anticipation se produit lorsque des individus moins gravement atteints avec une prémutation ou une mutation mosaïque transmettent des allèles *FMRI* instables à leurs enfants. Cependant, l'anticipation observée dans les familles dont les membres sont atteints du FXS n'est pas classique, comme c'est le cas par exemple dans la dystrophie myotonique de type 1, où la maladie elle-même s'aggrave et avec un âge plus précoce au fil des générations. De nombreuses familles transmettent des allèles *FMRI* prémutés pendant des générations avec peu ou pas de symptômes cliniques jusqu'à ce qu'une mutation complète soit produite, aboutissant à un diagnostic de FXS. La forme d'anticipation du FXS est l'augmentation du nombre d'individus atteints du FXS au fil des générations.

6 - Diagnostic moléculaire

L'analyse génotypique pour le syndrome de l'X fragile reposait sur l'utilisation de deux techniques, PCR et Southern blot, utilisées de façon séquentielle. La méthode diagnostique de référence est la méthode de Southern avec double digestion enzymatique, méthode adaptée pour l'identification des expansions importantes et la détermination du statut de méthylation du gène. On utilise la PCR afin d'exclure un diagnostic positif.

En première intention, on effectue une PCR standard afin de détecter les allèles normaux et prémutés. Si aucun allèle (ou un allèle chez la fille) n'est détecté par la PCR, la recherche d'une mutation complète par Southern Blot est recommandée. En cas de détection d'une prémutation, la présence d'une mutation complète doit être recherchée (Southern, ou PCR long-range). Aujourd'hui, l'utilisation de la PCR Amplidex (Asuragen) tend à remplacer la technique de Southern Blot.

En cas de détection d'une prémutation de grande taille ou d'une mutation complète de petite taille (200 +/- 30 répétitions), l'analyse de la méthylation est indispensable.

B - Ataxie de Friedreich (9, 10)

L'ataxie de Friedreich est l'ataxie héréditaire la plus courante. Elle affecte le système nerveux et se manifeste par des difficultés à la marche, une faiblesse musculaire, une perte de sensation et de proprioception et de troubles de la parole. Les symptômes apparaissent généralement dans l'enfance. Les symptômes s'aggravent avec le temps. Les personnes atteintes finissent par avoir besoin d'aides à la mobilité, perdent la vision et l'audition et développent d'autres complications médicales (diabète, scoliose...). La cause de décès la plus fréquente chez les patients atteints d'AF est la cardiomyopathie hypertrophique.

L'ataxie de Friedreich est une maladie autosomique récessive.

Le diagnostic d'ataxie de Friedreich est établi chez un patient par la détection de variants pathogènes bialléliques dans *FXN*. Une expansion répétée du trinuécléotide GAA, dans le premier intron du gène, est la principale cause de cette maladie. Ces répétitions conduisent à une transcription réduite du gène *FXN*, le désactivant et réduisant la production de frataxine. Cette protéine est impliquée dans la régulation mitochondriale de l'homéostasie du fer et de la production d'ATP.

La grande majorité des patients atteints par cette maladie sont homozygotes pour la mutation d'expansion GAA. 5% des patients atteints sont des hétérozygotes composites. Ces patients présentent une expansion répétée de GAA sur un allèle et une mutation ponctuelle (faux-sens, non-sens, intronique ou exonique) sur l'autre allèle.

On distingue quatre catégories d'allèles pour la séquence répétée GAA dans l'intron 1 de *FXN* :

- Allèles normaux : 5 à 33 répétitions GAA.
- Allèles prémutés : 34 à 65 répétitions GAA.
- Allèles limites : 44-100 répétitions GAA.
- Allèles à pénétrance totale : 100 à environ 1 300 répétitions GAA.

La technique employée au laboratoire pour détecter l'expansion de GAA est la TP-PCR. La PCR long-range permet de déterminer la taille de l'amplification des répétitions GAA anormalement étendue dans l'intron 1 de FXN, cependant elle ne permet pas de repérer les variations ponctuelles. Ainsi, si un seul allèle anormalement développé est identifié, une analyse de séquence de FXN est ensuite effectuée, suivie d'une analyse de délétion/duplication si aucun variant inactivant pathogène n'est trouvé lors du séquençage.

Il n'est pas possible de prédire avec précision le résultat clinique spécifique chez un individu en fonction du génotype. Des expansions plus importantes de GAA, en particulier sur l'allèle le plus petit, sont associées à une apparition plus précoce de la maladie, à une progression plus rapide de la faiblesse musculaire, à une fréquence plus élevée de cardiomyopathie et à une aréflexie dans les membres supérieurs. En général, les répétitions entre 34 et 100 ne provoquent que très peu de symptômes surtout si elles sont interrompues par des répétitions non-GAA. Les interruptions stabilisent la répétition contre l'expansion. Cependant, les répétitions ininterrompues sont considérées comme des pré-mutations et peuvent atteindre plus de 300 répétitions en une seule génération.

La pénétrance est complète chez les patients présentant des expansions répétées bialléliques de GAA à pénétrance totale et chez les hétérozygotes composites pour une expansion de GAA à pénétrance totale dans un allèle et un variant ponctuel pathogène dans l'autre allèle. Cependant, en raison de la grande variabilité de la taille des expansions, l'âge d'apparition de la maladie peut varier de façon très importante (de 5 à 50 ans) même au sein d'une même fratrie.

La taille des allèles à l'extrémité inférieure de la plage des allèles pathogènes n'a pas été clairement définie dans l'ataxie de Friedreich. Il est possible qu'une pénétrance incomplète soit associée à des allèles limites.

L'ataxie de Friedreich (FRDA) est héritée de manière autosomique récessive, par conséquent aucune anticipation n'est observée car la maladie n'est généralement pas observée sur plus d'une génération.

C - Myotonie de Steinert (11, 12)

La dystrophie myotonique de type 1 (Myotonie de Steinert, DM1) est un trouble multisystémique qui affecte les muscles squelettiques et lisses ainsi que les yeux, le cœur, le système endocrinien et le système nerveux central. Les symptômes cliniques couvrent un spectre phénotypique allant de léger à sévère. Trois phénotypes qui se chevauchent quelque peu ont été décrits : léger, classique et congénital.

La DM1 légère est caractérisée par des cataractes qui débutent généralement après 40 ans, une faiblesse musculaire et une légère myotonie. Ces patients ont généralement une espérance de vie normale.

La DM1 classique se manifeste généralement à partir de la seconde décennie par une myotonie, une faiblesse musculaire, surtout distale, une atrophie musculaire et des cataractes. La myotonie est plus prononcée après le repos et s'améliore avec l'activité musculaire. La maladie provoque une altération des mouvements fins impliquant et des difficultés à la marche. Dans la forme classique, on peut également observer un « visage myopathique » caractéristique en raison de la faiblesse et de l'atrophie des muscles faciaux. Des anomalies de conduction cardiaque sont souvent présentes. La durée de vie des patients atteints est réduite par rapport à la population générale.

La DM1 congénitale est caractérisée par une hypotonie et une faiblesse généralisée sévère à la naissance, souvent accompagnée d'une insuffisance respiratoire et d'un décès prématuré. Les manifestations prénatales peuvent inclure une réduction des mouvements fœtaux et un hydramnios, ainsi que des pieds varus équinés et une ventriculomégalie. Une déficience intellectuelle est courante.

La myotonie de Steinert se transmet de manière autosomique dominante. Elle est causée par une expansion de répétitions de triplets CTG dans la région non traduite en 3' du gène *DMPK* sur le chromosome 19q13.3, ce qui entraîne un gain de fonction de la protéine traduite. *DMPK* possède 14 exons couvrant environ 13 kb d'ADN génomique. *DMPK* code pour la myotonine-protéine kinase.

On distingue plusieurs classes d'allèles :

- Les allèles normaux ont entre 5 à 34 répétitions de triplets CTG. Les patients présentant ces allèles ne sont pas atteints par la myotonie de Steinert.
- Les allèles avec 35 à 49 répétitions CTG sont des allèles prémutés. Les individus présentant des expansions de CTG dans la plage de prémutation ne présentent pas de symptômes. La transmission de l'allèle est instable.
- Les allèles pathologiques ont plus de 49 répétitions. Les expansions de répétition anormales peuvent atteindre plusieurs milliers, en particulier chez les individus atteints de DM1 congénitale.

La physiopathologie de la maladie est liée au nombre de répétitions CTG. Un nombre de répétitions faible mais anormal (50-99) est associé à un phénotype léger. Les allèles pathologiques sont instables et prédisposent à de nouvelles expansions, ce qui conduit au phénomène génétique d'anticipation lorsqu'ils sont transmis d'une génération à l'autre. Les symptômes se manifestent plus tôt à chaque génération successive en raison de ce phénomène. La longueur de répétition du triplet CTG est également instable lors des mitoses chez les individus atteints de DM1. Cette instabilité conduit très souvent à un mosaïcisme somatique pour l'ampleur de l'expansion des triplets CTG. Ce mosaïcisme somatique est important entre les différents organes, ce qui explique en partie la grande variabilité phénotypique.

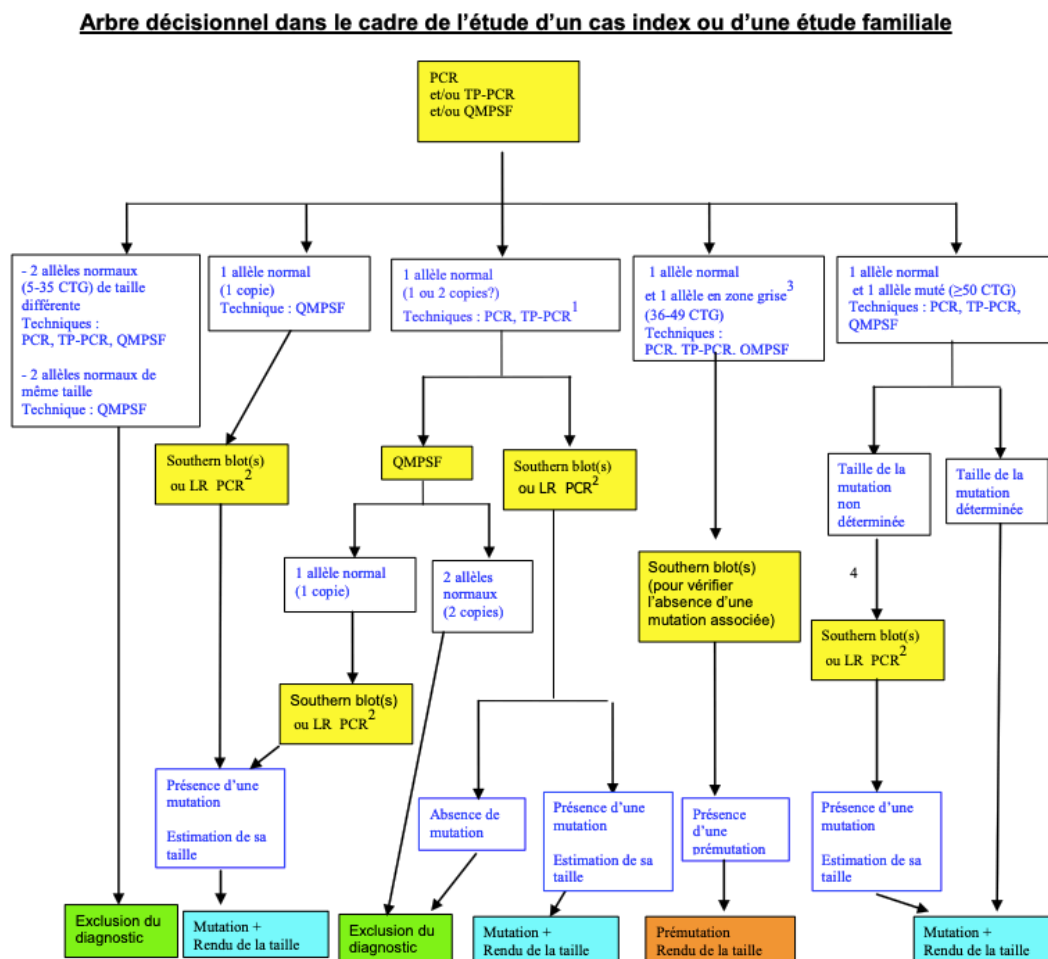
La pénétrance est élevée : près de 100 % à 50 ans.

Le diagnostic positif repose sur la mise en évidence de l'expansion de la répétition CTG. Un diagnostic de DM1 est exclu si 2 allèles de taille normale sont détectés. Différentes techniques sont possibles en fonction du contexte (Voir figure) :

- Diagnostic d'exclusion : PCR (Polymerase Chain reaction) : Les techniques utilisées dans le diagnostic d'exclusion permettent de quantifier des allèles avec moins de 100 triplets CTG
 - PCR fluorescente.
 - Quantitative multiplex PCR of short fluorescent fragments.

- TP-PCR (Triplet Repeat Primed PCR) : La TP-PCR permet également de confirmer le diagnostic de myotonie de Steinert dans les cas d'un nombre de répétitions compris entre 50 et 100 CTG.
- Confirmation diagnostique : LR PCR et Southern Blot : Ces techniques permettent d'estimer le nombre de répétitions en cas d'allèle muté.
- LR-PCR (Long-range PCR)
- Southern Blot

Figure 1 : Arbre décisionnel pour le diagnostic de la myotonie de Steinert.



Recommandations ANPGM.

II - Techniques diagnostiques actuelles (12)

A - Southern Blot

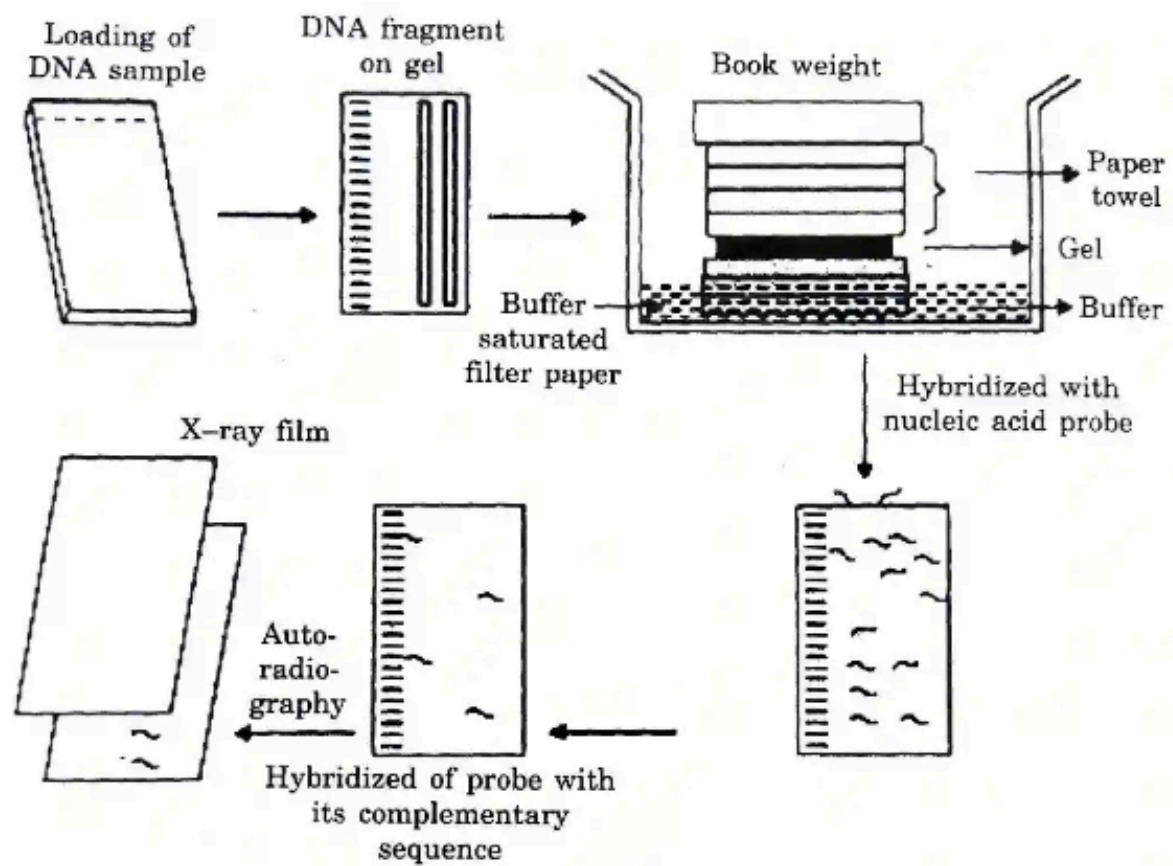
Le Southern blot est une méthode de biologie moléculaire permettant de visualiser n'importe quelle fraction du génome avec une sonde hybridant spécifiquement la séquence à analyser. Elle a été mise au point en 1975 par Edwin Southern, professeur de biologie moléculaire. Cette méthode est particulièrement lourde et n'est plus utilisée que pour analyser des grands fragments.

Principe et méthode :

- 1 - Digestion enzymatique : L'ADN est fragmenté en utilisant une ou plusieurs enzymes de restriction appropriées. L'enzyme coupe l'ADN sur un site spécifique générant des fragments d'ADN.
- 2 - Séparation des fragments d'ADN par électrophorèse sur gel d'agarose.
- 3 - Traitement des gels par la soude pour dénaturer l'ADN et obtenir de l'ADN simple brin.
- 4 - Transfert des fragments d'ADN par capillarité sur une membrane en nylon.
- 5 - Préhybridation afin de saturer la membrane.
- 6 - Hybridation avec la sonde marquée.
- 7 - Lavage de la membrane.
- 8 - Visualisation par chimioluminescence : révélation du fragment, qui donne un motif de bandes.

Figure 2 : Illustration du Southern Blot (technique utilisant la radioactivité)

(<https://www.biotechfront.com/2021/04/southern-blotting-basic-principle-steps.html>)



B - Polymerase chain reaction (PCR)

La PCR est une technique d'amplification moléculaire ciblée *in vitro*. Elle permet d'obtenir, à partir d'un échantillon d'importantes quantités d'un fragment d'ADN spécifique et de longueur définie.

On effectue une succession de réactions de réplication d'une matrice double brin d'ADN. Chaque réaction met en oeuvre deux amorces oligonucléotidiques. Les amorces encadrent la séquence à amplifier. On utilise les produits de chaque étape de synthèse comme matrice pour les étapes suivantes afin d'obtenir une amplification exponentielle.

La technique est utilisée massivement depuis la commercialisation d'une ADN polymérase résistante aux températures élevées (la Taq polymérase), qui permet une automatisation de la technique.

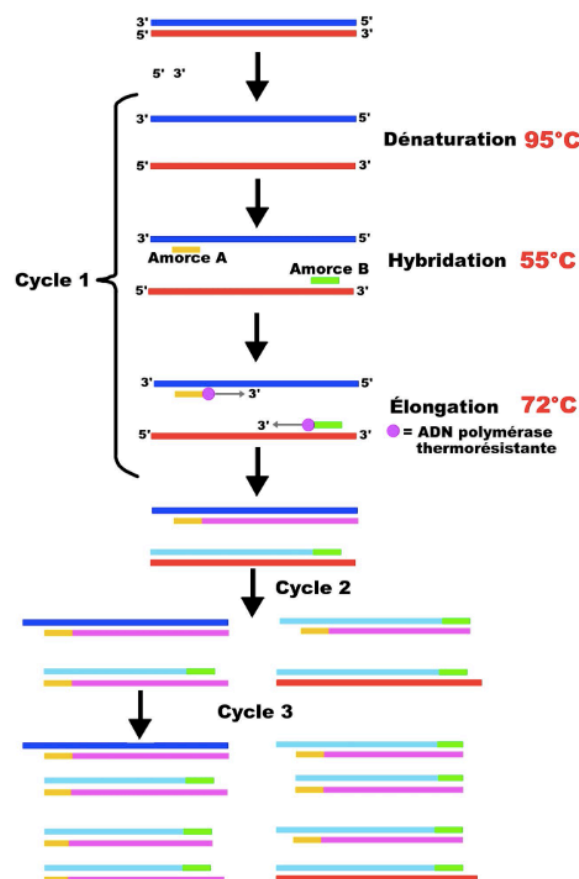


Figure 3 : Principe de la PCR (<https://microbiologiemedicale.fr/biologie-moleculaire-amplification-genique-pcr/>)

1 - Triple Repeat Primed PCR (TP-PCR)

La TP-PCR est une variante de la PCR classique. Elle permet de détecter les allèles avec expansion par fluorescence. Cependant cette variante de PCR ne peut mesurer la taille des allèles. Dans cette variante on utilise, en plus des amorces flanquantes la région génomique à étudier, 2 amorces inverses l'une par rapport à l'autre qui sont complémentaires de l'expansion de triplet en cause. À cette paire d'amorce est accroché un marqueur fluorescent.

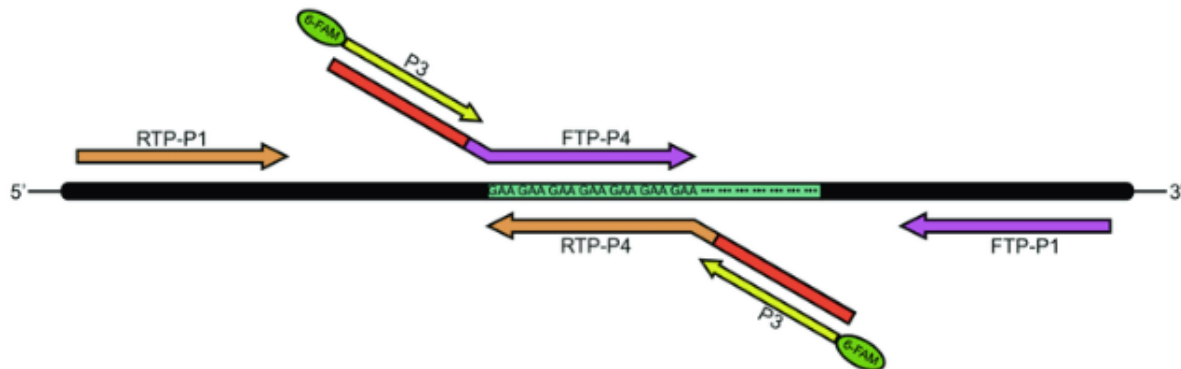


Figure 4 : Schéma représentant le principe de la TP-PCR

Nethisinghe S et al. Interruptions of the FXN GAA Repeat Tract Delay the Age at Onset of Friedreich's Ataxia in a Location Dependent Manner. International Journal of Molecular Sciences 2021;22:7507. <https://doi.org/10.3390/ijms22147507>.

2 - PCR long-range

Dans une PCR classique, il est difficile d'amplifier de très longs fragments (> 2 kb), surtout s'il contiennent beaucoup de G et de C. L'amélioration des conditions expérimentales et l'utilisation d'autres polymérases (autre que la Taq polymérase) ont permis de mettre au point la PCR long-range qui permet d'amplifier des fragments dont la longueur peut atteindre plusieurs dizaines de milliers de kb.

Ajustements techniques :

- incubation à pH plus élevé que dans une PCR normale.
- ajout de glycérol et de DMSO.
- diminution de l'étape de dénaturation et augmentation de l'étape d'élongation.

- utilisation de deux polymérases : une polymérase dénuée d'activité exonucléasique et une polymérase à haute fidélité qui possède une activité exonucléasique permettant de corriger les erreurs.

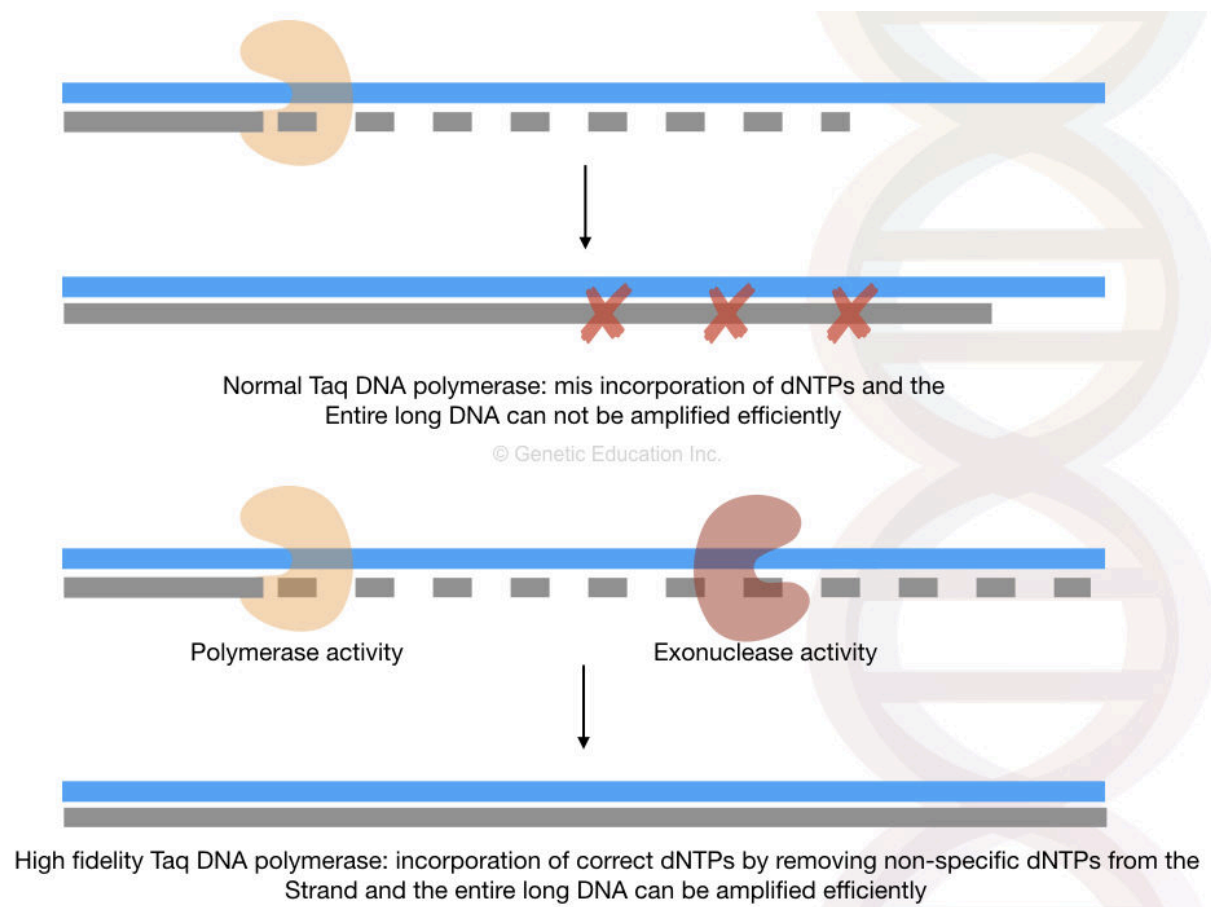


Figure 5 : Principe de la PCR long-range (<https://geneticeeducation.co.in/wp-content/uploads/2019/03/vjbj.001-e1553247863263.jpeg>)

III - Nouvelles technologies

A - Edition génomique et système CRSIPR/Cas9 (13)

L'édition génomique fait référence à la modification d'un segment spécifique du génome en substituant sa séquence par une autre. Éditer le génome était un défi pour le génie génétique, bien qu'il était envisageable d'exploiter un mécanisme naturel de réparation de l'ADN, appelé « recombinaison homologue ». Ce mécanisme permet la réparation de l'ADN endommagé, utilisant une autre molécule d'ADN similaire comme modèle. Pour éditer une séquence

spécifique, il est nécessaire de couper l'ADN cible tout en fournissant un ADN modèle similaire, mais en incorporant la séquence désirée à intégrer. La machinerie cellulaire interne réparera alors la coupure en introduisant la séquence voulue. La recombinaison homologue offre donc la possibilité d'insérer n'importe quelle séquence nucléotidique à n'importe quel endroit du génome à condition de réussir à couper l'ADN précisément à l'endroit voulu. Historiquement, deux méthodes permettaient de couper le génome de manière spécifique, bien qu'elles soient très limitées : l'utilisation de méganucléases, qui ne permettent pas de couper l'ADN de manière complètement dirigée et la production d'enzymes artificielles, un processus complexe et long.

Le système CRISPR/Cas9 a été découvert grâce à l'étude des mécanismes de défense des bactéries. La plupart des bactéries contiennent dans leur génome des séquences dites CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), composées de répétitions palindromiques et de séquences intercalaires issues du génome de leurs pathogènes. Ce système agit comme une mémoire immunitaire, permettant aux bactéries de résister à des pathogènes spécifiques grâce à l'action des protéines Cas (CRISPR-associated, associées aux CRISPR) qui permettent l'introduction de morceaux de séquence de pathogènes dans cette région génomique, quand la cellule est infectée. Certaines protéines Cas fabriquent des ARN particuliers, les ARN CRISPR (ARNcr) à partir des séquences intercalaires. Ces ARN sont ensuite associés à une protéine Cas (la protéine Cas9). L'association ARN/protéine Cas est capable de reconnaître les ADN qui portent la même séquence que l'ARN, et de les couper. Les ARNcr programment Cas9 pour lier et cliver l'ADN sur des sites identiques (ou très similaires) à la séquence d'ARNcr. Lorsque Cas9 est chargé d'un ARNcr identique à la séquence cible prévue, ainsi que d'un ARN CRISPR trans-activant (tracrARN - un ARN structural requis pour l'activité catalytique), il forme un complexe ribonucléoprotéique (RNP). Ce RNP recherche dans l'ADN génomique la région cible à l'aide de la séquence crRNA. Ainsi, l'ARN-guide peut s'apparier au brin d'ADN, formant une double hélice ADN/ARN. Cette association stabilise l'interaction avec la protéine Cas9, permettant à celle-ci de sectionner l'ADN en un point spécifique. De plus, les protéines Cas9, issues de diverses bactéries, nécessitent une séquence spécifique adjacente à la cible qui s'appelle motif adjacent au protospacer (PAM, signifiant "protospacer adjacent motif", un motif situé à côté de la séquence intercalaire, motif NGG) immédiatement à côté de l'extrémité 3' de la cible. Lorsque la protéine Cas reconnaît son PAM dans le génome, la séquence adjacente peut

s'apparier avec le site de reconnaissance de 20 pb de l'ARNcr. Si l'appariement de bases réussit, la protéine Cas créera une cassure double brin dans l'ADN génomique. Cas9 coupe les deux brins de la séquence cible 3 pb en amont du PAM. La production d'une extrémité nouvellement exposée sur un site défini constitue la base de la sélectivité de la méthode d'enrichissement médiée par Cas. Il existe une possibilité qu'un ARNcr guide la protéine Cas9 vers des emplacements non ciblés et que des cassures double brin se produisent. Il est important d'obtenir un recrutement efficace des cibles et une génération de cassure double brin au locus prévu, mais une plus grande efficacité sur la cible se traduit souvent par une fréquence hors cible plus élevée. Cette nécessité limite la capacité du système CRISPR/Cas9 à couper uniquement les ADN possédant un PAM proche de la séquence cible. Néanmoins, le PAM étant une séquence généralement courte, elle est fréquemment retrouvée dans les séquences d'ADN.

La découverte du système CRISPR/Cas9 a non seulement approfondi la compréhension de l'immunité bactérienne, mais a aussi ouvert la voie à des applications biotechnologiques. Cas9, en étant programmable, peut cibler et couper une région génomique spécifique, permettant ainsi l'édition génomique et donc par extension, la capacité de cibler et de couper n'importe quelle région du génome pour enrichir un échantillon sans amplification et le séquencer sans modification des marqueurs épigénétiques.

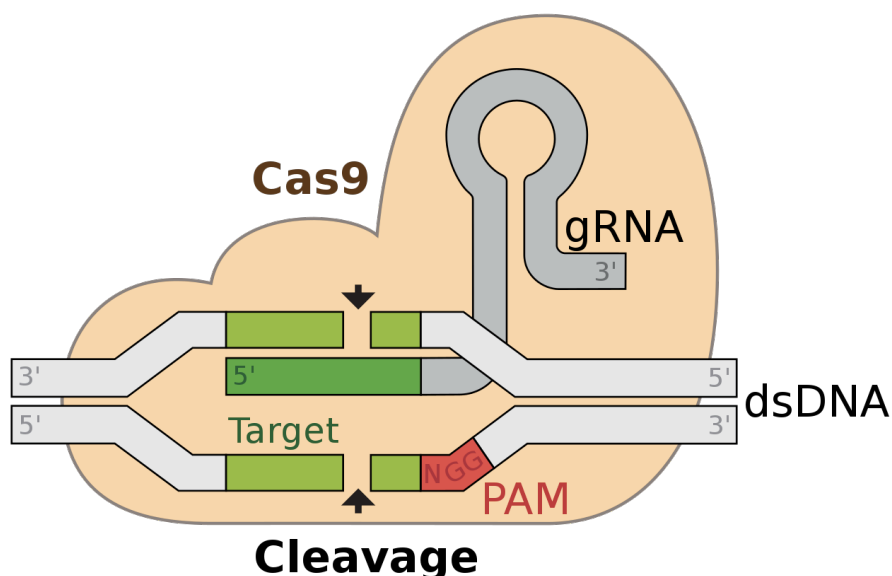


Figure 6 : Schéma explicatif de l'enrichissement d'une région génomique (https://fr.wikipedia.org/wiki/Clustered_Regularly_Interspaced_Short_Palindromic_Repeats#/media/Fichier:GRNA-Cas9.svg)

B - Nanopores et séquençage long-fragment (14, 15)

Un nanopore désigne un canal nanométrique au travers d'une membrane. Il existe de nombreux types de nanopores. Ils peuvent contribuer au passage à travers une surface de différents fluides dont des solutions chargées en acides nucléiques ou en protéines. Ils peuvent être biologiques ou synthétiques.

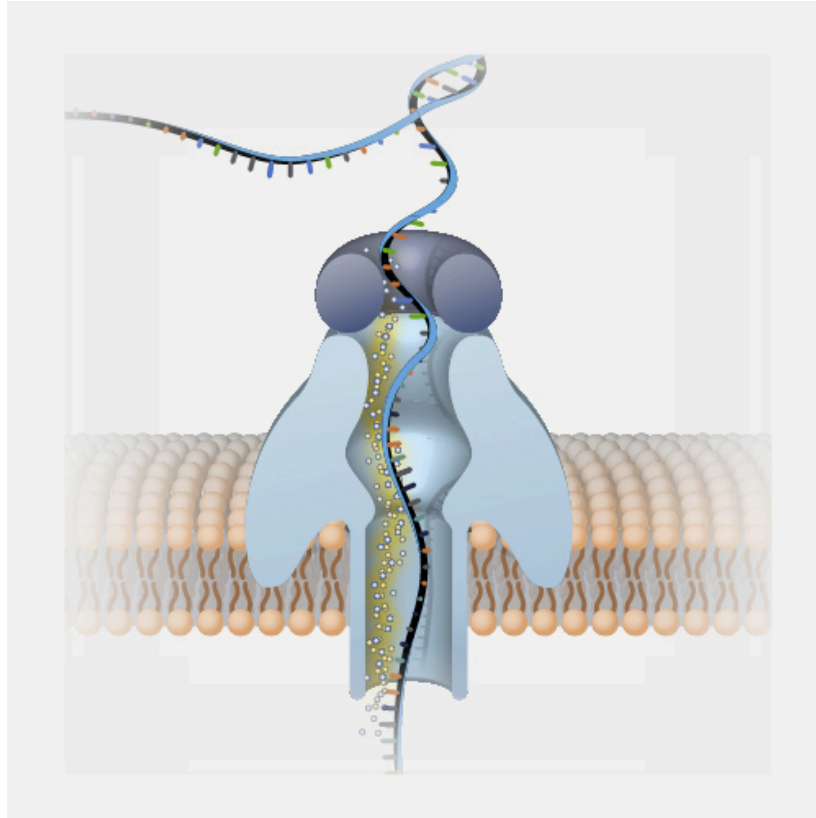


Figure 7 : Représentation d'un nanopore traversant une membrane cellulaire au travers duquel passe un brin d'acide nucléique. (<https://www.genome.gov/genetics-glossary/Nanopore-DNA-Sequencing>)

L'utilisation d'un nanopore pour le séquençage de l'ADN est basée sur l'idée qu'il fonctionne comme une sonde locale, forçant l'ADN à le traverser pour une lecture base par base. Les pionniers de cette technique ont démontré qu'en appliquant une différence de potentiel de chaque côté d'un nanopore biologique (en l'occurrence, l'alpha-hémolysine de *Staphylococcus aureus*), il était possible de faire passer des macromolécules (ADN et ARN simple brin) à travers cette cavité. La présence d'une macromolécule dans le pore modifie le courant ionique, permettant ainsi la détection du passage de chaque molécule.

Afin d'identifier chaque paire de bases traversant le nanopore, divers défis techniques ont dû être surmontés et d'importantes améliorations du système ont été nécessaires. Des ajustements ont été apportés pour réduire le bruit de mesure et augmenter la résolution temporelle (aptitude du nanopore à mesurer séparément les bases se succédant rapidement), notamment par la réduction de la surface des membranes et l'intégration de l'électrode de mesure. La vitesse de translocation (vitesse de passage de l'ADN dans le nanopore) élevée a été contrôlée en plaçant des enzymes, telles que les hélicases ou l'ADN polymérase, en amont du pore. Des modifications ont également été apportées au pore biologique pour augmenter le temps d'interaction entre l'ADN et les acides aminés dans le pore. La membrane lipidique originale a été substituée par une membrane polymérique plus stable. La différence de courant ionique induite par chaque paire de bases et leur temps de passage dans le pore ont été analysés pour distinguer chaque paire.

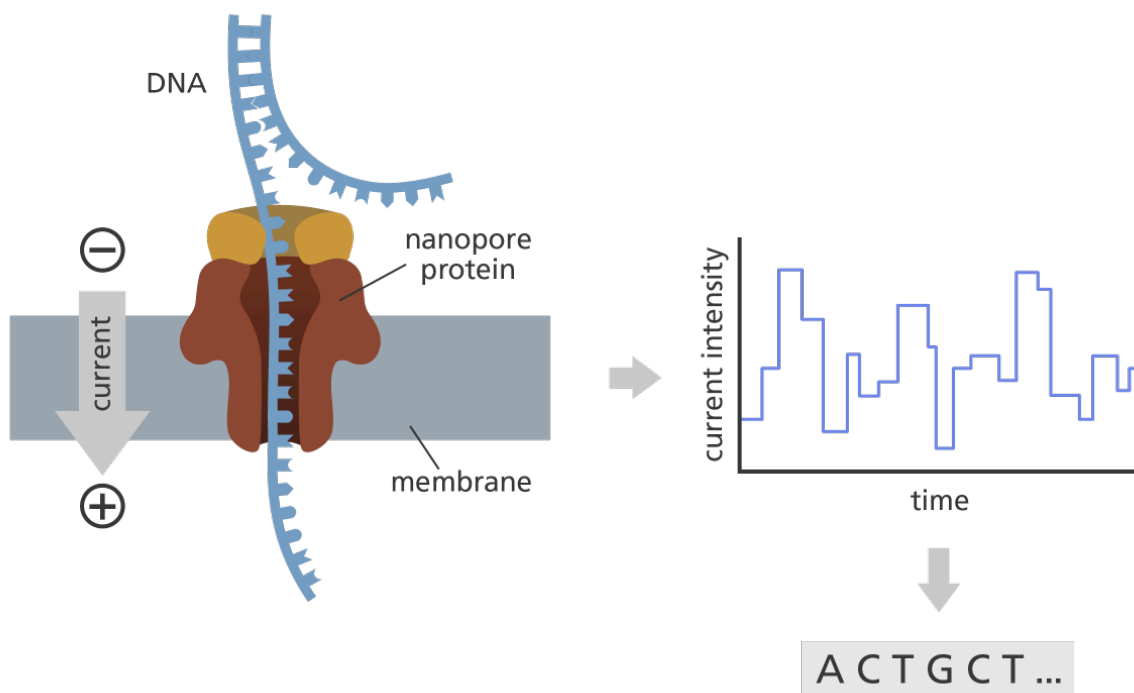


Figure 8 : Principe du séquençage par nanopores (<https://www.yourgenome.org/facts/what-is-oxford-nanopore-technology-ont-sequencing/>)

Suite à ces améliorations, la technique commercialisée par Oxford Nanopore Technologies, basée sur l'utilisation de nanopores biologiques et la détection électrique, permet de séquencer de longs fragments individuels de 10 à 1 000 kb. Malgré un taux d'erreur brut initial élevé,

l'augmentation du nombre de lecture d'un nucléotide donné (augmentation de la couverture) peut améliorer la précision jusqu'à 99,5 %. La vitesse de lecture par nanopore varie entre 250 et 450 bases par seconde. Bien que la parallélisation (nombre de réactions de séquençage dans le même temps) soit inférieure à celle des techniques de séquençage actuelles, la technologie présente des avantages significatifs.

Les technologies de séquençage à haut débit ont révolutionné le domaine de la génomique humaine, permettant aux chercheurs d'étudier et de comprendre plus facilement les processus biologiques et leur impact. Grâce à ces technologies, les chercheurs peuvent analyser des génomes entiers ou des régions d'intérêt spécifiques, avec des informations fonctionnelles supplémentaires obtenues grâce à la caractérisation et à la quantification des transcrits et des isoformes d'ARN. Ensemble, ces capacités ont fourni un aperçu sans précédent de la diversité génétique humaine et de ses implications sur la santé et la maladie. Bien qu'elles offrent des progrès significatifs en termes de vitesse et de résolution par rapport à l'ancienne technique de séquençage dite de Sanger, il existe encore un certain nombre de limitations inhérentes aux plates-formes traditionnelles de séquençage à lecture courte et à haut débit.

À ce jour, la majorité des grands génomes séquencés ont utilisé des technologies de lecture courte, qui nécessitent que l'ADN soit fragmenté (fragments de l'ordre de généralement 150 à 300 pb) avant le séquençage. Bien que ces technologies permettent une analyse du génome rapide et relativement rentable par rapport à la méthode de Sanger, leurs limitations inhérentes en matière de longueur de lecture empêchent l'analyse de larges régions génomiques correspondant à des régions répétitives et à de grands variants de structure. C'est pour ces raisons qu'une partie importante du génome humain reste difficile à séquencer. L'étape requise de fragmentation de l'ADN perd les informations de position relative, ce qui signifie que lors de l'assemblage du génome, ces petits fragments doivent être reconstitués par chevauchement avec d'autres fragments courts. Pour les régions répétées, cela peut être particulièrement difficile et entraîner le raccourcissement de régions de répétitions potentiellement longues en des longueurs beaucoup plus courtes. De la même manière, la présence de certaines variations de structure telles que des délétions, insertions, duplications, inversions et translocations peuvent être manquées lors de l'utilisation seule de lectures de séquençage courtes. Il est bien établi que certaines expansions anormales de répétition et variants de structure sont associées aux maladies humaines, ce qui rend leur caractérisation de routine nécessaire. Contrairement

aux technologies de séquençage à lecture courte, le séquençage basé sur les nanopores traite toute la longueur du fragment d'ADN présenté au pore. Des fragments complets de milliers de kb sont régulièrement traités et des longueurs de lecture ultra-longues supérieures à 4 Mb ont été obtenues (Advancing human genomics with nanopore sequencing, White Paper, Oxford Nanopore Technologies). Des lectures aussi longues sont plus susceptibles de couvrir des régions entières de séquences d'ADN répétées et de variations de structure. En conséquence, le séquençage des nanopores fournit une vision plus complète de la variation génétique. Les lectures de séquençage directes et longues fournies par la technologie des nanopores permettent également la mise en évidence de variants et de bases modifiées.

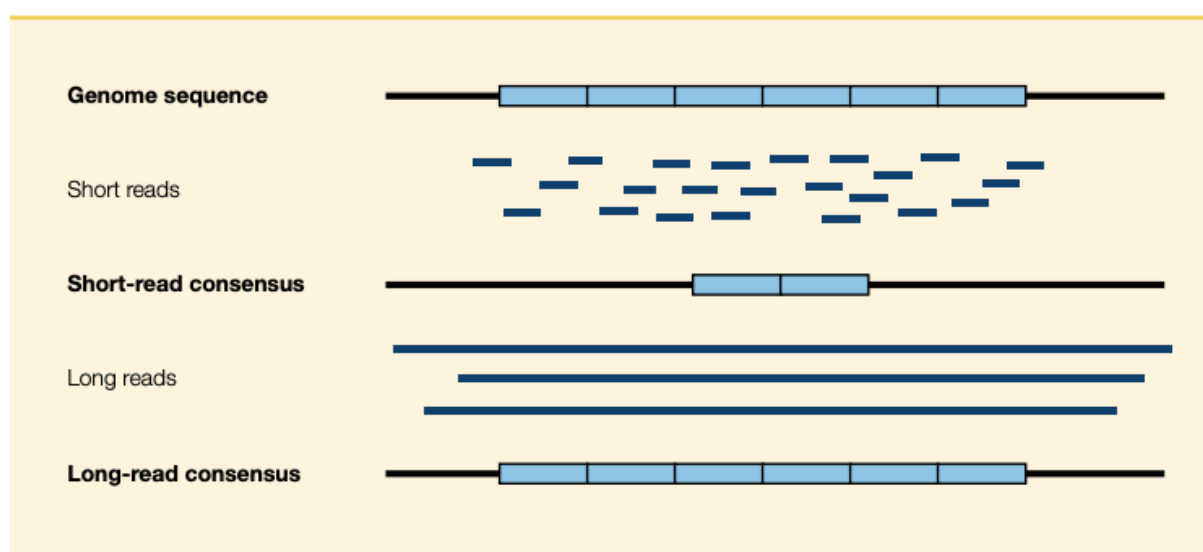


Figure 9 : Schéma présentant l'avantage théorique du séquençage long-fragment par rapport au séquençage haut-débit conventionnel pour les régions génomiques contenant des expansions de triplets. (Advancing human genomics with nanopore sequencing, White Paper, Oxford Nanopore Technologies).

Pour examiner des locus génomiques spécifiques, l'utilisation d'une stratégie de séquençage ciblé s'avère nécessaire. Cette stratégie augmente la couverture et facilite l'interprétation des données. Une diversité de méthodes de séquençage ciblé est accessible grâce au séquençage Nanopore, lequel a déjà prouvé son efficacité en utilisant des techniques d'enrichissement ciblé telle que la PCR. Actuellement, l'aptitude de CRISPR/Cas9 à enrichir de longues molécules d'ADN ciblées est également à l'étude. Les premières observations sont encourageantes pour ces méthodes, qui peuvent offrir des processus plus rapides et permettre, par le séquençage direct de la molécule visée, également l'analyse des marqueurs

épigénétiques. Une caractéristique distinctive de la technologie Nanopore est la faculté d'analyse en temps réel pour séquencer ou écarter sélectivement les molécules d'ADN lors de leur passage dans le pore. En inversant le courant sur un pore spécifique, les molécules ne répondant pas aux critères sont rejetées, laissant le pore libre pour une autre molécule d'ADN. Cette méthode d'enrichissement en temps réel en est encore à ses prémices.

La signification des modifications épigénétiques de certaines bases (cytosine, adénosine) dans l'expression et la fonction des gènes est de plus en plus reconnue. L'amplification nécessaire dans la technologie traditionnelle de séquençage à lecture courte efface ces modifications, rendant leur détection impossible sans des méthodes additionnelles de traitement d'échantillons, qui sont souvent longues et inefficaces. Le séquençage par nanopores associé à la technologie CRISPR/Cas9, ne nécessitant pas d'amplification ou de synthèse de brins, permet la détection simultanée de la base et de sa modification durant le même cycle de séquençage. Actuellement, le séquençage par nanopores détecte plusieurs bases modifiées dans l'ADN et l'ARN, y compris la pseudouridine, la N6-méthyladénosine (m6A), la 5-méthylcytosine (5mC) et la 7-méthylguanosine (m7G).

MATÉRIEL ET MÉTHODE :

Nous avons utilisé le séquençage long-fragment associé à la technologie CRISPR/Cas9 afin de mesurer l'expansion du triplet CTG dans le gène DMPK chez des patients atteints de la myotonie de Steinert. L'utilisation de la technologie CRISPR/Cas9 permet de couper le génome de part et d'autre de la région d'intérêt.

I - Matériel

Pour la mise au point de ce protocole de séquençage, nous avons utilisé 5µl d'ADN, extrait manuellement, de patients atteints de la myotonie de Steinert (> 50 CTG répétés).

Nous avons utilisé des ARN guide conçus avec chopchop (<http://chopchop.cbu.uib.no>, mode d'emploi en annexe) et designés par IDT (ARNcr).

Séquences d'ADN cibles		
AGCTCCTTAGTTGCATACACTGG	chr19:45788326	-
CCGGCCCACAACGCAAACCGCGG	chr19:45769675	+

Le kit de séquençage est le Ligation sequencing gDNA - Cas9 enrichment Kit (SQK-LSK109). La flow-cell utilisée pour le séquençage est la ONT SpotON Flow Cell Flongle. Le kit d'amorçage de la flow cell est le EXP-FLP002. Le séquenceur utilisé est le MinION Mk1C.

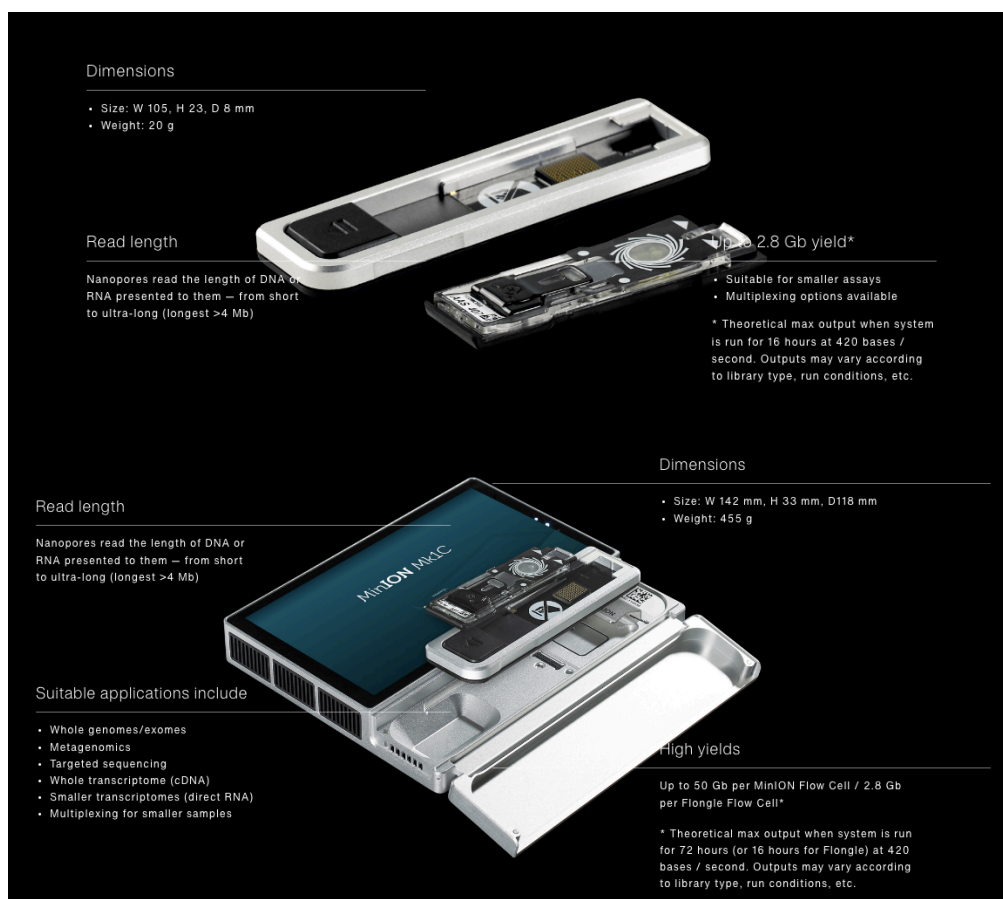


Figure 10 : Brochure montrant le séquenceur utilisé et la flow-cell (nanoporetech.com)

	Flongle	MinION & MinION Mk1C	GridION	PromethION 24	PromethION 48
Read length	Fragment length = read length. Longest read now >4 Mb				
Run time	1 min – 16 hrs	1 min – 72 hrs	1 min – 72 hrs	1 min – 72 hrs	1 min – 72 hrs
Number of flow cells per device	1	1	5	24	48
DNA sequencing yield per flow cell*	Up to 2.8 Gb	Up to 50 Gb	Up to 50 Gb	Up to 290 Gb	Up to 290 Gb
DNA sequencing yield per device*	Up to 2.8 Gb	Up to 50 Gb	Up to 250 Gb	Up to 7 Tb	Up to 14 Tb

Figure 11 : Capacités de séquençage des différents outils utilisés (Advancing human genomics with nanopore sequencing, White Paper, Oxford Nanopore Technologies)

II - Méthode : Préparation de la bibliothèque ciblée Cas9 pour le séquençage

Le protocole opératoire comporte plusieurs étapes que l'on va énumérer et illustrer brièvement.

1 - Après extraction de l'ADN, les extrémités 5' de l'ADN du patient sont déphosphorylées pour réduire la liaison des adaptateurs de séquençage aux brins ciblés.

2 – Le complexe ribonucléoprotéique formé par la molécule Cas9 associée aux ARNcr et aux tracrARNest ajouté à l'ADN génomique du patient, puis se lie à cet ADN et clive la région d'intérêt (Region Of Interest, ROI).

3 - Le clivage de l'ADNdb par le système CRISPR-Cas9 produit des extrémités franches avec des phosphates 5' (non déphosphorylées).

4 - Tout l'ADN de l'échantillon est polyadénylé, ce qui prépare les extrémités phosphates 5' pour la liaison à l'adaptateur de séquençage.

5 - Les adaptateurs de séquençage sont principalement liés aux côtés coupés par Cas9. La préparation est nettoyée pour éliminer les adaptateurs en excès à l'aide de billes AMPure XP et remise en suspension dans le tampon de séquençage. Il est à noter que les régions non ciblées ne sont pas extraites de l'échantillon. Elles ne passeront tout simplement pas à travers les nanopores car elles ne sont pas censées porter d'adaptateur de séquençage. La préparation de bibliothèque ultérieure est ajoutée à la Flow Cell pour le séquençage.

6 – L'analyse des données de séquençage a été effectuée à l'aide du logiciel MinKNOW (v19.2.2). Les « reads » ont été alignées sur le génome humain de référence (Hg38) à l'aide de Minimap2.

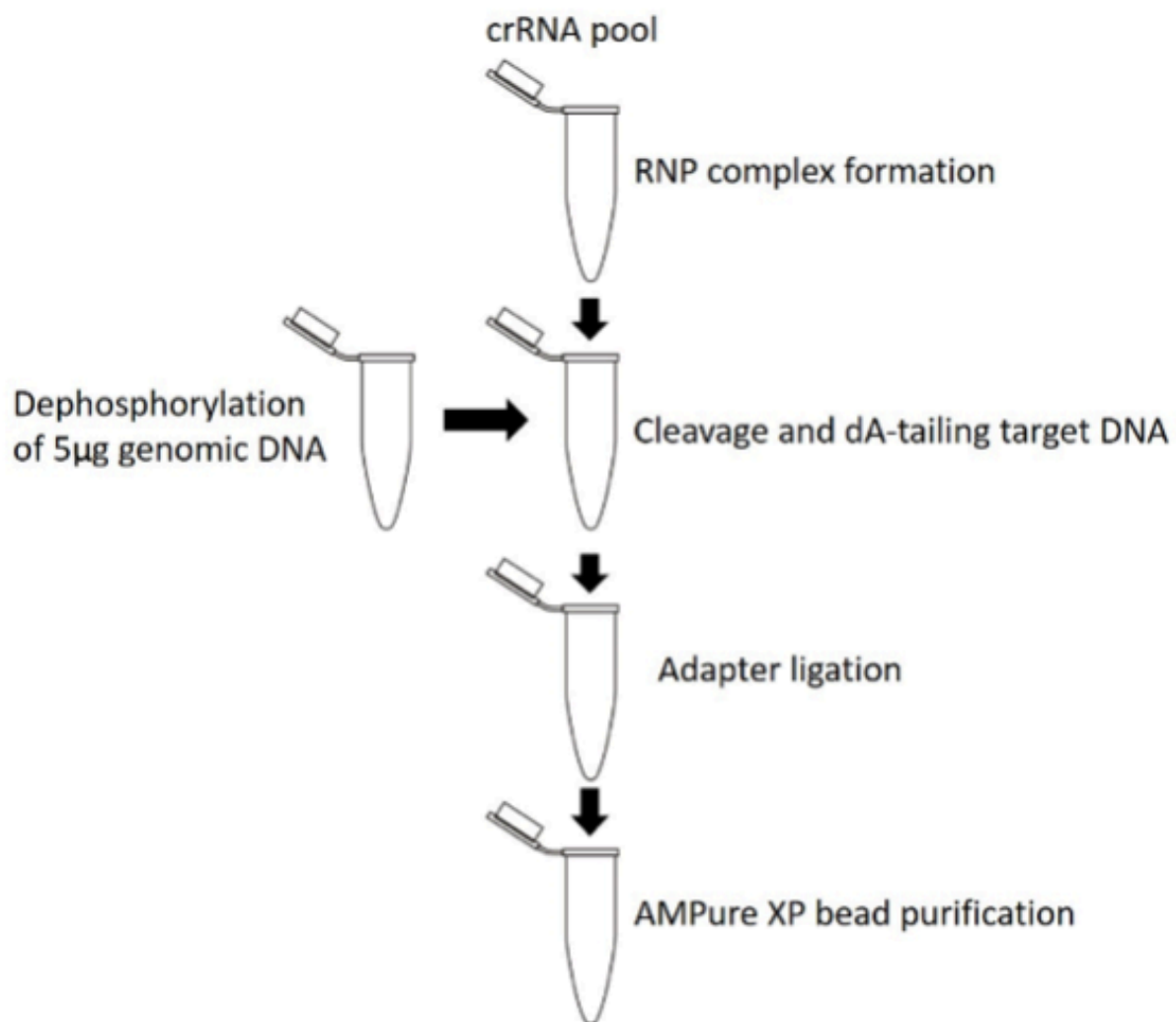


Figure 12 : Représentation globale des étapes de manipulation en vue du séquençage par nanopores

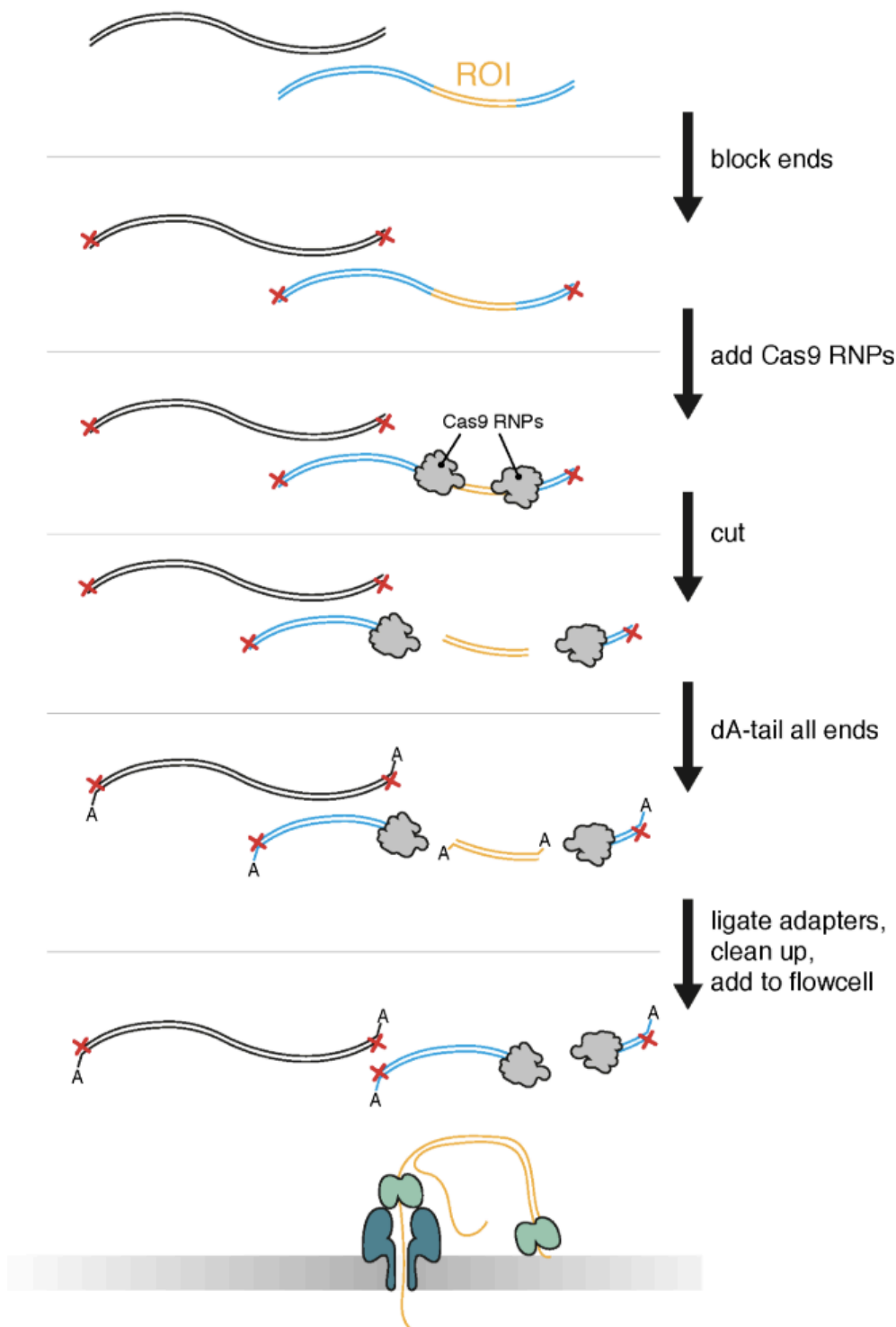


Figure 13 : Explication de l'effet de la technologie CRISPR/Cas9 dans l'échantillon.
(Nanoporetech.com)

RÉSULTATS :

L'objectif de cette thèse était de démontrer la faisabilité du séquençage long-fragment Nanopore associé à la technologie CRISPR/Cas9 dans le diagnostic des maladies à expansion de triplets. Dans un second temps, nous voulions prouver que cette technique permettait de conserver les paramètres épigénétiques des régions génomiques sélectionnées cette technique était moins fastidieuse que le Southern Blot.

Afin de pouvoir utiliser les données de séquençage, le développement d'un pipeline bio-informatique est important. En effet après séquençage, des étapes de traitement des données de séquençage sont essentielles.

Le séquençage Nanopore par MINion permet d'obtenir des fichiers fastq. Les fichiers fastq sont des fichiers textes qui permettent de stocker des séquences d'ADN. Dans notre mise au point, il fallait alors aligner ces séquences sur un génome de référence afin de pouvoir s'assurer que les séquences obtenues correspondaient aux séquences ciblées par la technologie CRISPR/Cas9. L'entreprise Nanopore met à disposition des solutions bio-informatiques gratuites. Cependant leur installation et leur complexité en fait des logiciels difficiles à utiliser. En effet, ils nécessitent des connaissances bio-informatiques importantes et la configuration des ordinateurs disponibles au CHRU n'offre pas la possibilité de les utiliser pleinement. Nous avons donc utilisé le site <https://usegalaxy.eu> qui est une plateforme bio-informatique en ligne qui propose des programmes qui permettent d'effectuer les différentes étapes bio-informatiques nécessaires à la bonne analyse du séquençage. Seule l'étape d'alignement des séquences obtenues sur un génome de référence était nécessaire à la visualisation des séquences. La littérature explique que plusieurs programmes d'alignement sont exploitables pour les données de séquençage long-fragment (ref). Nous avons donc utilisé le programme Minimap2, disponible sur la plateforme Usegalaxy, pour aligner les séquences.

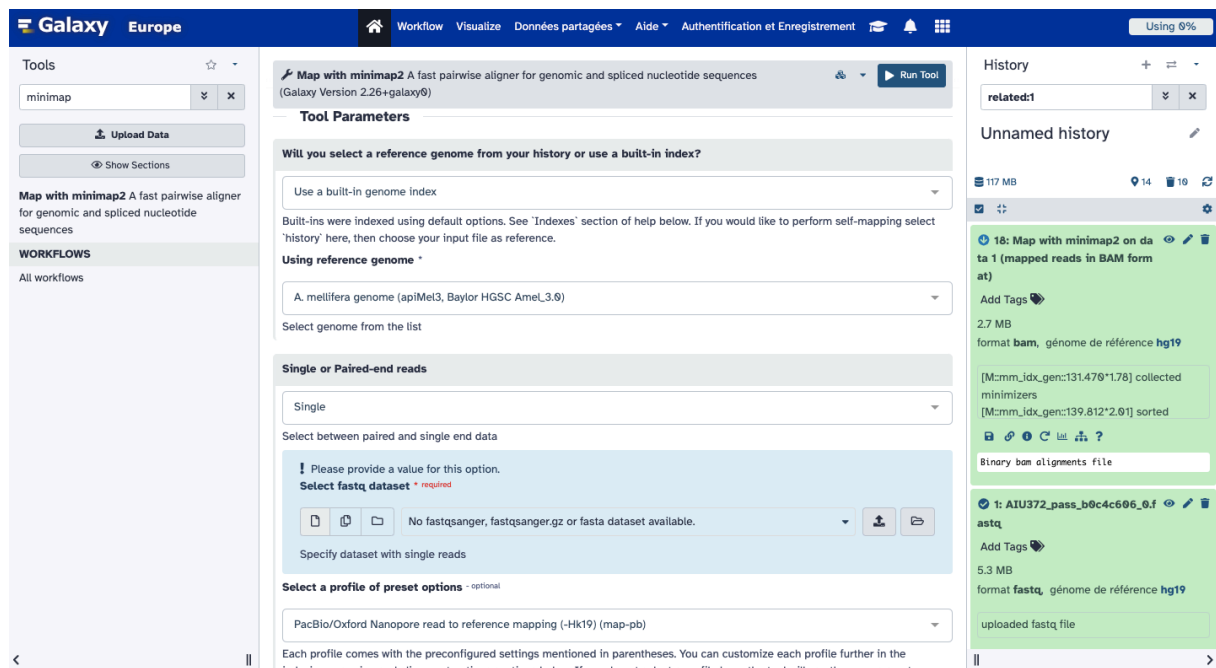


Figure 14 : Vue de la plateforme bio-informatique UseGalaxy

La capacité de séquençage du MINion est fonction de deux paramètres que nous pouvons modifier : le temps de séquençage et le nombre de pores disponibles sur la flow-cell. Pour mettre en place le protocole, nous avons donc opéré à plusieurs changements afin de faire varier ces paramètres entre chaque essai. Avec des flow-cell dites « Flongle » (flow-cell de recherche), le nombre de pores est limité (entre 50 et 100 pores), et varie d'une flow-cell « Flongle » à l'autre. Le premier essai a permis de découvrir les différentes manipulations techniques et la technologie de séquençage. Pour le deuxième essai, nous avons augmenté le temps de séquençage pour augmenter le nombre de « reads » séquencés, ce qui n'a pas été fructueux. Lors du 3ème essai, nous avons décidé de retourner à un séquençage moins long et de se concentrer sur les manipulations. Malgré un séquençage moins long, la quantité d'ADN séquencé a été similaire à l'essai précédent. Le 4ème essai a montré une augmentation de la quantité d'ADN séquencé, possiblement due à une maîtrise technique meilleure. Nous avons essayé, pour le 5ème essai, d'utiliser une autre flow-cell qui avait déjà été utilisée par le service de biochimie dans le cadre de la surveillance épidémiologique COVID. Cette flow-cell était la FLO-MIN106D. Elle possède initialement 800 pores mais après plusieurs utilisations et plusieurs lavages, nous avons pu l'utiliser avec 123 pores restants. Cette augmentation du nombre de pores aurait pu également permettre une augmentation de la quantité de séquençage. Au fur et à mesure des essais et des changements techniques nous avons réussi à doubler le nombre de « reads » séquencés par rapport au premier essai. Les

deux graphiques suivants montrent l'évolution du nombre de « reads » séquencés au cours des différents essais en fonction du temps de séquençage et du nombre initial de pores présents sur la flow-cell.

Figure 15 : Tableau montrant les différents paramètres utilisés pour les différents essais de séquençage

	Temps de séquençage en heures	Nombre initial de pores présents sur la flow-cell	Nombre de reads séquencés
Run 1	6	57	411
Run 2	19	82	527
Run 3	5	68	573
Run 4	6	62	886
Run 5	6	123	856

Figure 16 : Nombre de « reads » séquencés en fonction du nombre initial de pores viables dans la flow-cell

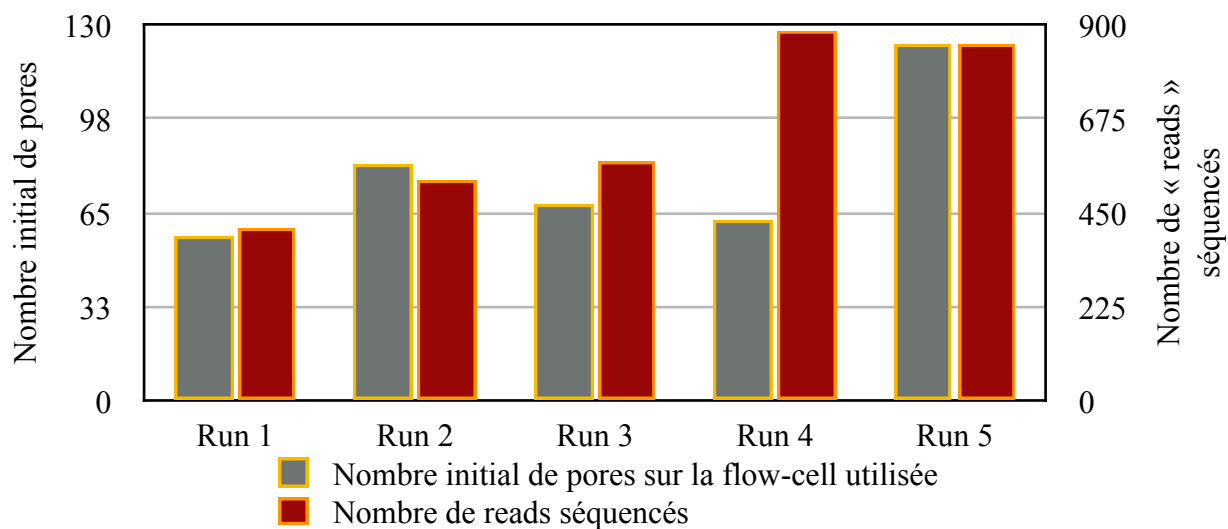
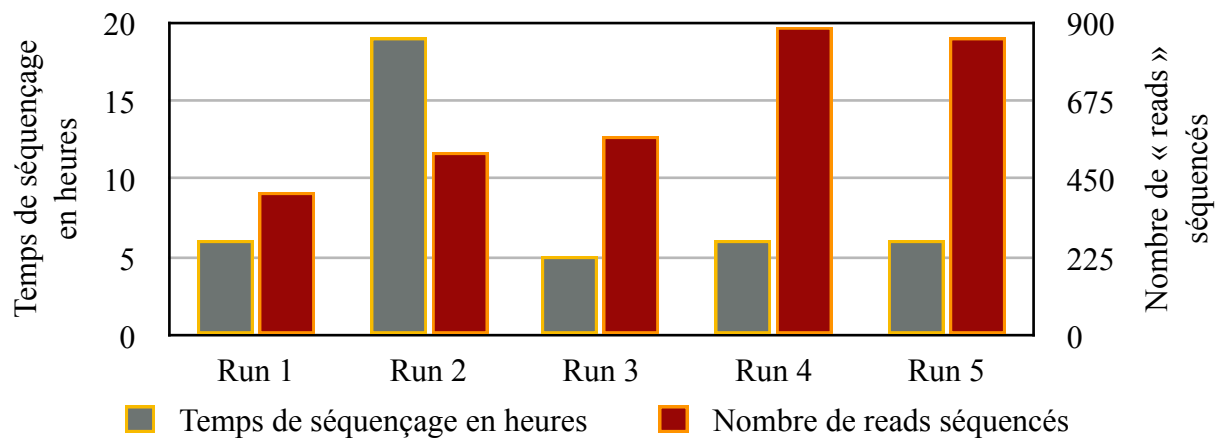


Figure 17 : Nombre de « reads » séquencés en fonction du temps de séquençage



Après plusieurs essais où différents changements opératoires ont été effectués, nous n'avons pas réussi à obtenir le séquençage des différentes régions génomiques sélectionnées.

En effet, les différents alignements des données de séquençage n'ont pas montré de données au niveau des différentes régions ciblées. D'autres fragments d'ADN, dans d'autres régions génomiques, ont par ailleurs pu être séquencés. Pour visualiser les séquences ADN obtenues, nous avons utilisé le logiciel Integrated Genome Browser.

Figure 18 : Exemple de données d'alignement sur le chromosome 19 :

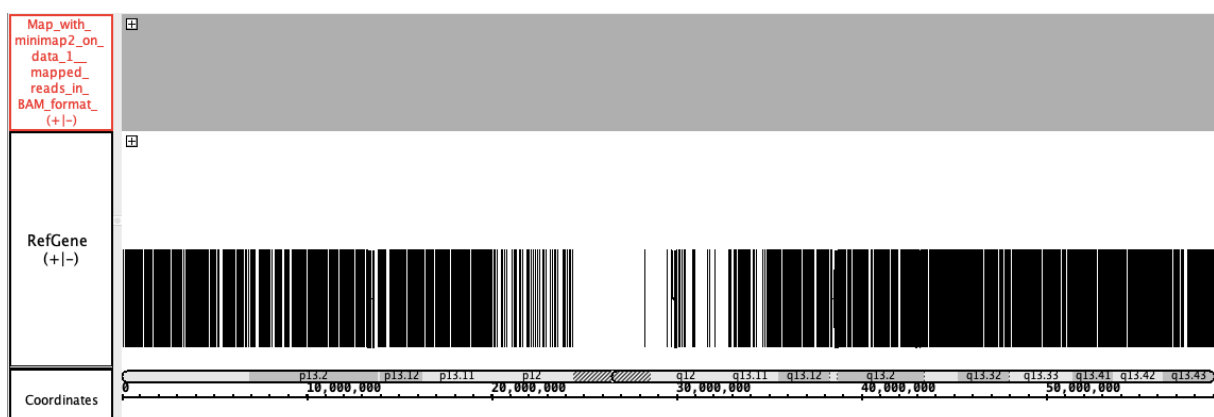
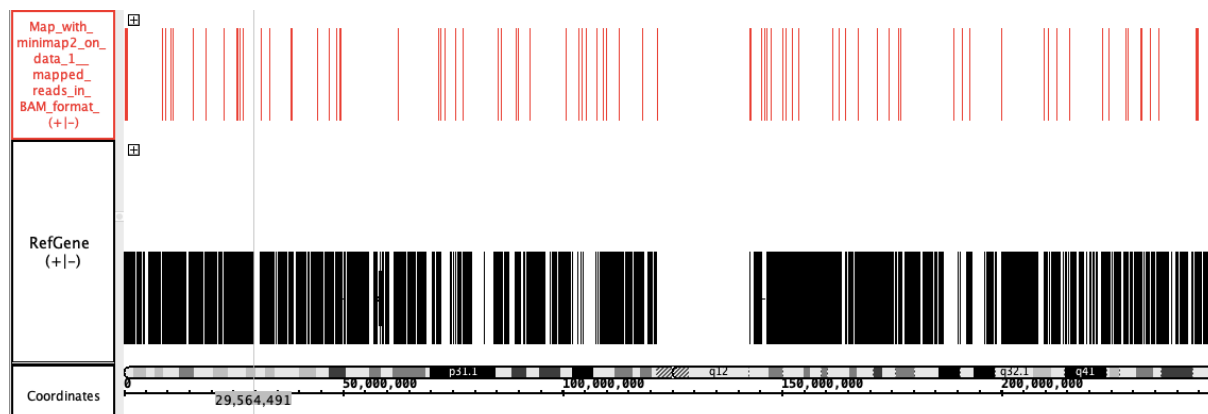


Figure 19 : Exemple de données d'alignement sur le chromosome 1 :



Données de qualité du séquençage :

Malgré des résultats primaires infructueux, nous avons quand même réussi à obtenir des données importantes à présenter.

On voit d'abord un résumé général des données de séquençage, avec des données brutes et des données après l'étape de basecalling (étape qui correspond à l'attribution des bases nucléotidiques aux changements de courant ionique résultant du passage de ces nucléotides). On peut donc voir que ce séquençage a duré 5 heures. 2,49 mégabases ont été séquencées. 573 lectures de fragments d'ADN ont été obtenues.

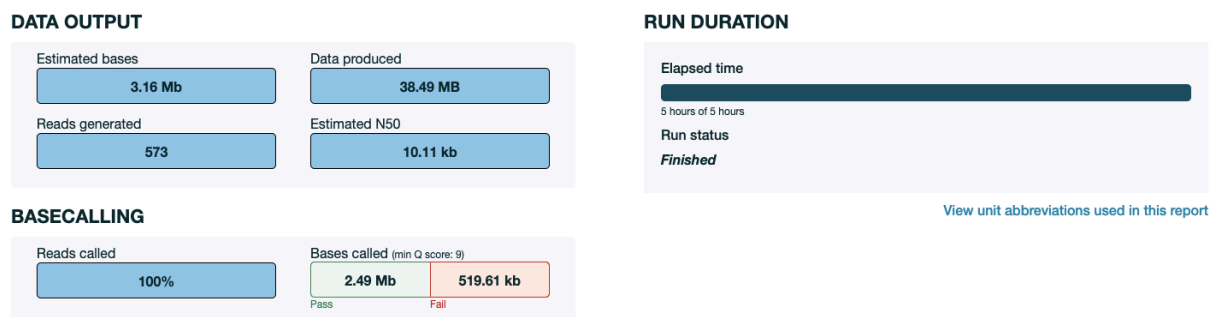


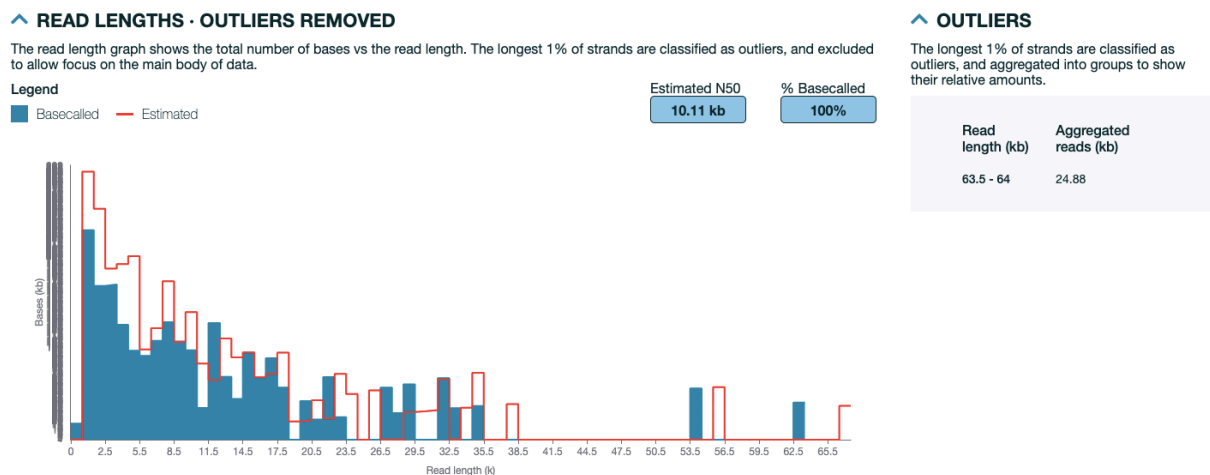
Figure 20 : Exemple de données générales d'un séquençage

Nous pouvons également observer des données plus spécifiques au séquençage en lui-même :

1 - Longueur des « reads » séquencés :

On peut voir sur le graphique ci-dessous la longueur des « reads » séquencés. Ils peuvent faire plus de plusieurs dizaines de kilobases. La médiane de la longueur des « reads » séquencés est de 10,11kb. Le plus long « read » séquencé mesure 62,5 kilobases.

Figure 21



2 - Quantité de bases et de reads et séquencés au cours du temps de séquençage :

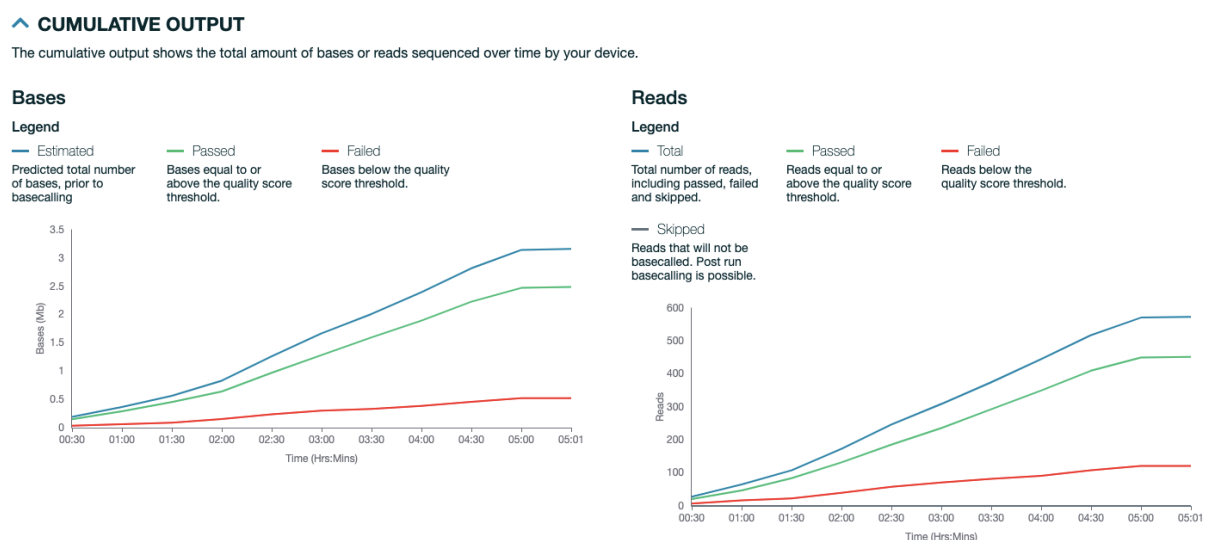


Figure 22 : Ces graphiques nous montrent l'évolution, pendant le temps de séquençage, de la quantité de bases et de « reads » séquencés.

3 - Score de qualité

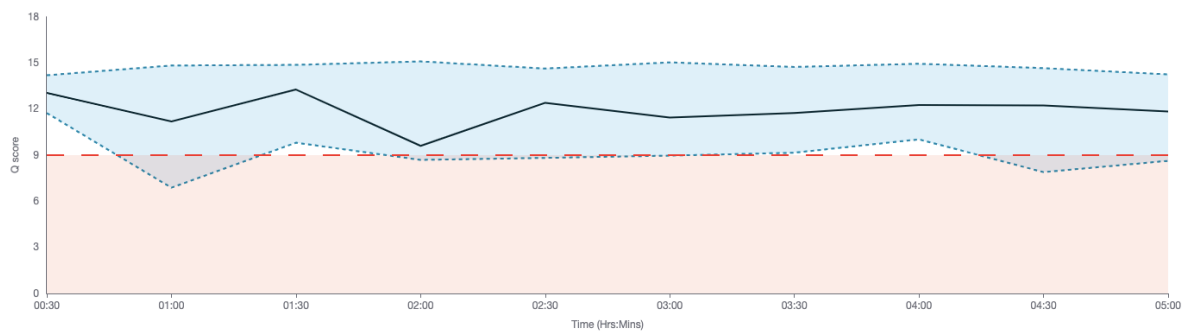
Figure 23 : Ce graphique indique que la majeure partie des bases séquencées avait un score de qualité supérieur au score minimum pour passer l'étape de « basecalling ».

^ QUALITY SCORE

The quality score is calculated as basecalling is performed on your device. Reads that fall below the minimum value of 9 will be classified as failed reads. You can alter the accepted minimum quality score in MinKNOW.

Legend

- Mode
The most frequent quality score of reads in the run.
- Spread
The spread of quality scores, found by calculating full width half maximum.
- Min. quality score
Minimum quality score to be accepted as a passed read.



4 - Evolution de l'activité des pores de séquençage au cours du temps

Figure 24 : On voit sur ce graphique l'évolution de l'activité des différents nanopores au cours du séquençage. On observe une diminution progressive de leur nombre au cours du temps. Le dernier histogramme correspond à l'arrêt du séquençage.

^ PORE ACTIVITY

The Pore activity graph shows the performance of your sample as it is being sequenced during a run.

Legend

- Sequencing
Pore currently sequencing
- Adapter
Pore currently sequencing adapter
- Pore available
Pore available for sequencing
- Unavailable
Pore unavailable for sequencing
- Active feedback
Channel ejecting analyte
- No pore
No pore detected in channel
- Out of range-high
Current is positive but unavailable for sequencing
- Out of range-low
Current is negative but unavailable for sequencing
- Multiple
Multiple pores detected. Unavailable for sequencing.
- Saturated
The channel has switched off as current levels exceed hardware limitations
- Zero
Pore currently unavailable for sequencing
- Channel disabled
Channel is disabled and awaiting another pore scan
- Unclassified
Pore status unknown

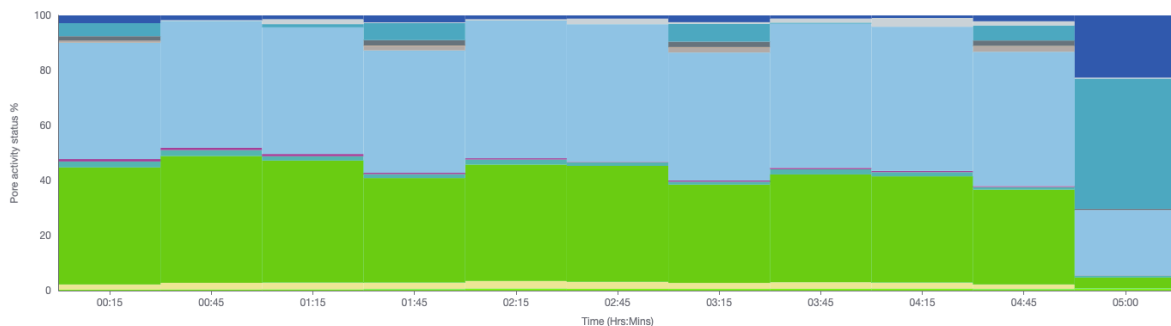


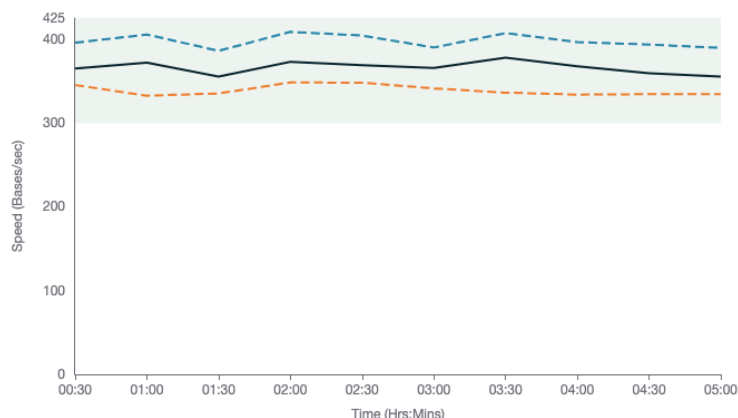
Figure 25 : Un graphe sur l'évolution de la vitesse de séquençage au cours du temps est également disponible. On observe une vitesse de séquençage constante pendant tout le séquençage. La vitesse médiane se situe aux alentours de 350 bases séquencées par seconde.

^ TRANSLOCATION SPEED

The translocation speed is the rate at which DNA/RNA travels through pores as it is being sequenced.

Legend

— Median - - - 75% quartile - - - 25% quartile □ Accepted range



Les résultats montrent un séquençage non spécifique de la région ciblée, c'est-à-dire qu'aucun fragment d'ADN correspondant aux gènes ciblés n'a pu être aligné.

De l'ADN était présent dans l'échantillon et a été séquencé. Le séquenceur n'a pas montré de dysfonctionnement. Les données de qualité sont en accord avec les recommandations du fabricant.

DISCUSSION :

Le séquençage long-fragment et le système CRISPR/Cas9 sont des technologies récentes et nous sommes à leurs prémices concernant leur utilisation. La mise au point d'un tel protocole est une tâche ardue, d'autant plus que les protocoles mis à disposition par les fournisseurs et la littérature peuvent ne pas être en accord. Pour autant la mise en place d'une telle technique de séquençage est motivée par l'accès au séquenceur au CHRU de Tours et le besoin d'améliorer le diagnostic des maladies à expansion de triplets car actuellement les techniques actuelles sont soit fastidieuses soit insuffisantes et demandent de multiplier les examens pour décrire moléculairement la pathologie.

Nous avons donc testé la faisabilité du séquençage long-fragment associé à la technologie CRISPR/Cas9 dans le diagnostic des maladies à expansion de triplets.

Les résultats des différentes manipulations ont montré la présence de séquences d'ADN, cependant ce n'était pas les séquences attendues. Ces résultats sont plutôt cohérents avec la littérature. En effet comparativement à la littérature, trop peu d'ADN a été séquencé. Plusieurs références montrent un nombre total de séquences obtenues bien plus important que nos expériences et l'ADN séquencé ne contient pas uniquement les séquences d'ADN ciblées mais surtout beaucoup d'ADN dit « off-target » (16). Il se peut donc que nous n'ayons pas assez séquencé d'ADN pour obtenir des résultats positifs. Mais compte tenu des données de qualité des différents séquençages, il semble que le séquençage fonctionne mais que la technique d'enrichissement (le Crispr/Cas9) ne soit pas efficace.

Figure 26 : tableau montrant le nombre total de lectures alignées, les lectures sur cible, le pourcentage sur cible et la couverture médiane dans chacune des dix régions enrichies.

10 sites (12-24kb)				
	GM12878			Breast Tumor
	one cut each side	multiple cuts (2 or 3) each side		
	minION	minION	flongle	minION
aligned reads	75822	235328	37715	611992
on-target reads	1370	10854	790	2702
on-target percent	1.81%	4.61%	2.09%	0.44%
median coverage	80	680	34	93

Gilpatrick T et al., Targeted nanopore sequencing with Cas9-guided adapter ligation. Nat Biotechnol. 2020 Apr; 38(4):433-438. doi:10.1038/s41587-020-0407-5.

L'utilisation des ARN-guide CRISPR a été effectuée selon les recommandations du fournisseur, c'est-à-dire que nous avons utilisé un couple d'ARN-guide, sélectionné via le site <https://chopchop.cbu.uib.no/> (voir Annexe), pour un même locus. Nous avons choisis ceux qui nous semblaient les plus pertinents c'est-à-dire ceux qui remplissaient des critères de qualité et de spécificité suffisants et qui permettaient de couper en amont et en aval des gènes à explorer. Dans la littérature, on retrouve l'utilisation de ce même kit de façon différente. Des auteurs suggèrent l'utilisation de deux couples d'ARN guide pour un même locus, afin d'augmenter la spécificité de l'enrichissement (16, 17). De plus, la sélection des ARN, via la site internet chopchop est une étape particulièrement compliquée. En effet l'interface est assez peu intuitive et le choix en ARN-guide est limité aux exigences de la technologie CRISPR/Cas9.

Le protocole mis en place pour enrichir nos échantillons est un protocole (Annexe) mis à disposition par Oxford Nanopore. Il contient beaucoup d'étapes. Chaque étape est minutieuse et les quantités de réactifs utilisés sont très faibles. Ceci implique certainement un taux d'erreur assez important lors des manipulations. Néanmoins des résultats de séquençage ont été obtenus et seulement des fragments d'ADN avec adaptateur peuvent être séquencés. Ceci oriente également vers l'hypothèse que l'utilisation de la technologie CRISPR/Cas9 est difficile.

La mise en place d'un pipeline bio-informatique a été une tâche compliquée. Les logiciels mis à disposition par Nanopore n'offrent pas une expérience utilisateur très simple et le matériel informatique à disposition n'est pas suffisant pour supporter ces programmes. Le pipeline a été mis en place à façon via la plateforme en ligne Galaxy. Aucun mode d'emploi n'était disponible pour la mise en place de ce mode opératoire via cette plateforme. Nous avons quand même aligné les séquences obtenues via le programme minimap2 disponible sur Galaxy qui permet d'aligner des « reads » de séquençage long-fragment (18). Mais classiquement, un pipeline bio-informatique ne contient pas seulement une étape d'alignement mais aussi d'autres programmes qui permettent de trier et classer les reads obtenus par le séquençage. Ces programmes n'étaient pas disponibles au moment des manipulations. Des reads qu'il n'était pas possible d'aligner sur le génome de référence, faisaient également partie des données obtenues par le séquençage. Il est donc possible que des reads avec les allèles contenant les expansions de triplet recherchés n'aient pas été

alignés. Pour autant, cette hypothèse semble peu crédible car nous aurions également retrouvé des séquences avec l'allèle non muté.

Afin d'améliorer ce protocole que nous voulons mettre en place, nous pourrions suggérer plusieurs solutions.

Une solution serait d'ajouter une étape préliminaire de prédigestion enzymatique afin d'utiliser la technologie CRISPR/Cas9 sur de l'ADN déjà enrichi. Cela augmenterait théoriquement la spécificité des ARN-guide, en diminuant la possibilité de off-target.

On pourrait également ajouter un autre couple d'ARN-guide mais nous avons vu dans la littérature que cette méthode nécessitait un temps de séquençage important et que le taux de off-target était encore très important.

Récemment une équipe de chercheurs a mis en place un protocole de diagnostic des maladies à expansion de triplets en utilisant la fonction « Read Until » de certains séquenceurs nanopore. La fonction « Read Until » est une fonction *in silico* permettant de changer la polarité des nanopores en fonction de la molécule d'ADN qui passe en temps réel via un fichier BED qui contient les régions génomiques ciblées (19). C'est donc une technique *in silico* qui permet grâce au séquençage en temps réel permis par Oxford Nanopore de sélectionner les brins d'ADN que nous voulons cibler. Cette technique semble prometteuse mais un séquenceur avec cette fonction intégrée ou bien un GPU (Graphics processing unit, processeur de calculs informatiques) plus puissant disponible est nécessaire. De plus, pour utiliser cette méthode, le fichier bed doit contenir 3 Mb de régions génomiques au minimum, c'est à dire qu'elle n'est pas adaptée pour un diagnostic ciblé mais plutôt pour un panel de gènes.

Un autre défaut du protocole que nous voulions mettre en place est qu'il s'agit d'une stratégie solo. Il n'est pas possible actuellement de multiplexer les échantillons en utilisant cette méthode. La possibilité de multiplexer les échantillons est une force du génome court-fragment et récemment un nouveau module bio-informatique de recherche d'expansion nucléotidique a été mis en place. Ce module permet de détecter si les plages d'expansion de nucléotides sont de taille normale. Ainsi il permet d'établir un diagnostic d'élimination des

pathologies à expansion de triplets. Ceci est également une piste quant à une refonte des stratégies de diagnostic de ces maladies (20).

En outre, la Fédération Hospitalo-Universitaire GenOMedS a lancé un projet utilisant le séquençage génomique T2T de PacBio Sciences pour les génomes à lecture courte négative, ce qui pourrait également ouvrir de nouvelles voies d'exploration.

Pour conclure la convergence des technologies, telle que l'association du séquençage long-fragment avec le mécanisme CRISPR/Cas9, représente probablement l'avenir du diagnostic génétique, notamment pour les maladies à expansion de triplets du fait des différentes possibilités techniques théoriquement envisageables citées. Si notre exploration initiale n'a pas abouti aux résultats escomptés, elle met en évidence les zones de défi tout en ouvrant la voie à de futures améliorations et à des explorations plus approfondies. Le paysage technologique évolue rapidement, et l'adaptabilité est essentielle. L'importance de ce travail ne réside pas seulement dans les résultats obtenus, mais aussi dans la richesse des enseignements tirés. En avançant, il est impératif de rester à la pointe des innovations technologiques et de tirer parti des expériences et des découvertes pour affiner et perfectionner les approches diagnostiques.

CONCLUSION :

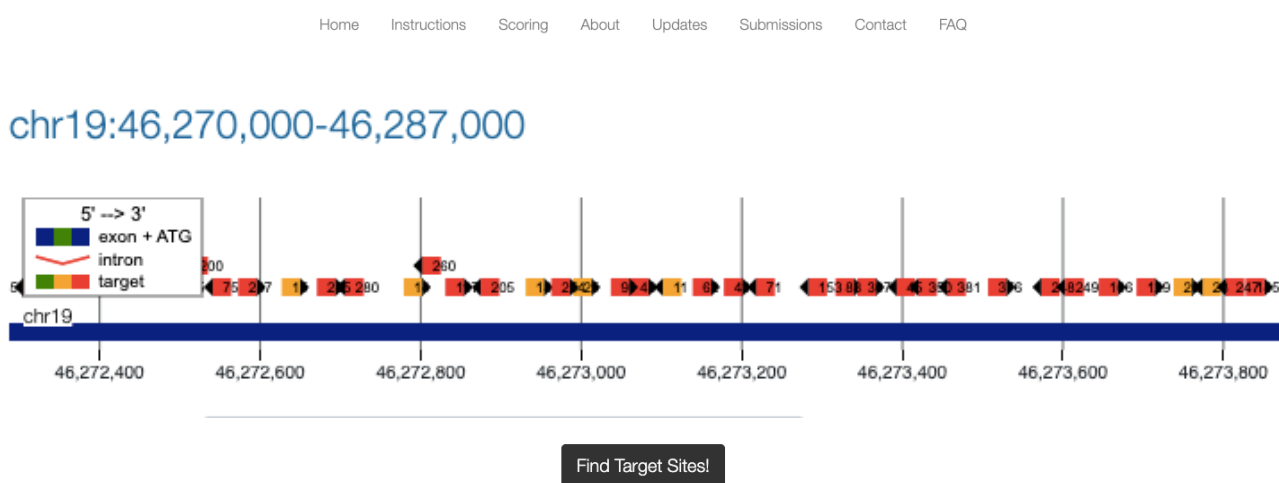
Malgré des expériences peu fructueuses, le séquençage long-fragment et la technologie CRISPR/Cas9 semblent être des technologies d'avenir concernant le diagnostic des maladies à expansion de triplets et d'autres maladies génétiques.

Les maladies à expansion de triplets sont très importantes dans les maladies humaines, en particulier dans le domaine de la neuro-génétique. Les méthodes établies pour diagnostiquer ces troubles sont lourdes et prennent du temps. Cependant, le séquençage à lecture longue offre peut-être la possibilité de transformer le diagnostic des maladies à expansion de triplets, en permettant un génotypage plus rapide et plus précis. Cela permettrait de mieux comprendre la physiopathologie moléculaire de ces maladies. Des recherches supplémentaires sont nécessaires pour surmonter les obstacles techniques et exploiter pleinement le potentiel du séquençage long-fragment.

ANNEXES :

Mode d'emploi Chopchop :

1. Accéder à <https://chopchop.cbu.uib.no>.
2. Saisir le gène ou les coordonnées génomiques qui nous intéressent dans le champ « target ».
3. Choisir le génome qui nous intéresse dans le champ « In ».
4. Choisir « CRISPR/Cas9 » dans le champ « Utilisation ».
5. Choisir « nanopore enrichment » dans le champ « For ».
6. Appuyez sur « Find targets sites ».
7. Dans la vue du locus génomique des résultats, identifier un ARNg qui correspond à notre objectif.



8. Vue tableau. Le classement par défaut est basé sur le score. Le score est défini principalement par le nombre de hors-cibles, l'efficacité étant la deuxième caractéristique la plus importante.

Rank	Target sequence	Genomic location	Strand	GC content (%)	Self-complementarity	MM0	MM1	MM2	MM3	Efficiency
1	TAGAAGCGCGCCATCTCGGCCGG	chr19:45778550	+	65	0	0	0	0	0	50.64
2	CTTGCCATAGGTCTCCGCGTGG	chr19:45777682	+	65	0	0	0	0	1	67.86
3	CGAAGTCTGTTGTCCGTGCGG	chr19:45771775	+	55	0	0	0	0	1	66.85
4	CATGGAGTATTACGTGGCGGGG	chr19:45778611	-	55	0	0	0	0	1	53.98
5	CAGTCAACTACGCGAGGAGAGG	chr19:45771646	-	60	1	0	0	0	1	53.38
6	CCGACTCGCTGACAGGCTACAGG	chr19:45769846	-	65	3	0	0	0	1	48.70
7	ACCTTCGAATCCGGTGTAAAGG	chr19:45777393	+	45	0	0	0	0	1	41.48
8	TGGGCGAGCGCCCTTCTACGCGG	chr19:45777708	-	65	0	0	0	0	2	69.14
9	GCAGCTGCGTGATCCACGCCGG	chr19:45779293	+	70	0	0	0	0	2	59.13
10	TAGGGTTCAGGAGCGCGGCGG	chr19:45770482	+	70	0	0	0	0	2	56.20
11	CCGGCCAACTACCGCAGTCTGG	chr19:45770518	-	70	1	0	0	0	2	48.94
12	CCTTTTATTCGCGAGGCTCGGGG	chr19:45769734	+	55	0	0	0	0	2	41.87
13	CATCGTGGTGAGGCTTAAGGAGG	chr19:45779840	-	55	2	0	0	0	3	62.08
14	GTCACAGTGCATCCAAACGTGG	chr19:45769805	+	50	0	0	0	0	4	64.19
15	GCGGTCCAGCAGGATGTTGTCGG	chr19:45778187	+	60	0	0	0	0	4	52.00
16	GACGTGAGTGGCAAACTGCAGG	chr19:45770137	-	60	0	0	0	0	4	50.77
17	CGGCCGTGCTCACCTTGTAGTGG	chr19:45777654	+	65	0	0	0	0	4	49.71

9. S'assurer que le contenu du GC correspond à nos besoins et que l'auto-complémentarité n'est pas trop élevée.

BIBLIOGRAPHIE :

1. Depienne C, Mandel JL. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *The American Journal of Human Genetics*. 2021 May;108(5):764–85.
2. Gymrek M, Willems T, Guilmatre A, Zeng H, Markus B, Georgiev S, Daly MJ, Price AL, Pritchard JK, Sharp AJ, Erlich Y. Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat Genet*. 2016 Jan;48(1):22-9. doi: 10.1038/ng.3461. Epub 2015 Dec 7. PMID: 26642241; PMCID: PMC4909355.
3. Sulovari A, Li R, Audano PA, Porubsky D, Vollger MR, Logsdon GA; Human Genome Structural Variation Consortium; Warren WC, Pollen AA, Chaisson MJP, Eichler EE. Human-specific tandem repeat expansion and differential gene expression during primate evolution. *Proc Natl Acad Sci U S A*. 2019 Nov 12;116(46):23243-23253. doi: 10.1073/pnas.1912175116. Epub 2019 Oct 28. PMID: 31659027; PMCID: PMC6859368.
4. Quilez J, Guilmatre A, Garg P, Highnam G, Gymrek M, Erlich Y, Joshi RS, Mittelman D, Sharp AJ. Polymorphic tandem repeats within gene promoters act as modifiers of gene expression and DNA methylation in humans. *Nucleic Acids Res*. 2016 May 5;44(8):3750-62. doi: 10.1093/nar/gkw219. Epub 2016 Apr 7. PMID: 27060133; PMCID: PMC4857002.
5. Paulson H. Repeat expansion diseases. *Handb Clin Neurol*. 2018;147:105-123. doi: 10.1016/B978-0-444-63233-3.00009-9. PMID: 29325606; PMCID: PMC6485936.
6. Malik I, Kelley CP, Wang ET, Todd PK. Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat Rev Mol Cell Biol*. 2021 Sep;22(9):589-607. doi: 10.1038/s41580-021-00382-6. Epub 2021 Jun 17. Erratum in: *Nat Rev Mol Cell Biol*. 2021 Jul 6; PMID: 34140671; PMCID: PMC9612635.
7. Hunter JE, Berry-Kravis E, Hipp H, et al. FMR1 Disorders. 1998 Jun 16 [Updated 2019 Nov 21]. In: Adam MP, Mirzaa GM, Pagon RA, et al., editors. *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2023. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK1384/>
8. Stone WL, Basit H, Shah M, et al. Fragile X Syndrome. [Updated 2023 Feb 5]. In: *StatPearls* [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK459243/>
9. Bidichandani SI, Delatycki MB. Friedreich Ataxia. 1998 Dec 18 [Updated 2017 Jun 1]. In: Adam MP, Mirzaa GM, Pagon RA, et al., editors. *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2023. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK1281/>
10. Williams CT, De Jesus O. Friedreich Ataxia. [Updated 2023 Jun 27]. In: *StatPearls* [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK563199/>

11. Bird TD. Myotonic Dystrophy Type 1. 1999 Sep 17 [Updated 2021 Mar 25]. In: Adam MP, Mirzaa GM, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2023. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK1165/>
12. Kaplan JC, Delpech M. Biologie moléculaire et médecine. 3^e édition. Flammarion Médecine-Sciences; 2007.
13. Seitz H. La technologie CRISPR/Cas9. revue-cmh [Internet]. 2018 Sep 30 [cited 2023 Oct 14];(15). Available from: <https://revues-msh.uca.fr/revue-cmh/index.php?id=711>
14. Fabien Montel Séquençage de l'ADN par nanopores - Résultats et perspectives; Med Sci (Paris), 34 2 (2018) 161-165; Published online: 2018-02-16; DOI: <https://doi.org/10.1051/medsci/20183402014>
15. Oxford Nanopore Technologies, Advancing human genomics with nanopore sequencing, White paper, <https://nanoporetech.com/sites/default/files/s3/white-papers/human-genetics-research-white-paper.pdf>
16. Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, et al. Targeted nanopore sequencing with Cas9-guided adapter ligation. Nat Biotechnol. 2020 Apr;38(4):433–8.
17. Giesselmann P, Brändl B, Raimondeau E, Bowen R, Rohrandt C, Tandon R, et al. Analysis of short tandem repeat expansions and their methylation state with nanopore sequencing. Nat Biotechnol. 2019 Dec;37(12):1478–81.
18. Becht C, Schmidt J, Blessing F, Wenzel F. Comparative analysis of alignment tools for application on Nanopore sequencing data. Current Directions in Biomedical Engineering. 2021 Oct 1;7(2):831–4.
19. Stevanovski I, Chintalaphani SR, Gamaarachchi H, Ferguson JM, Pineda SS, Scriba CK, Tchan M, Fung V, Ng K, Cortese A, Houlden H, Dobson-Stone C, Fitzpatrick L, Halliday G, Ravenscroft G, Davis MR, Laing NG, Fellner A, Kennerson M, Kumar KR, Deveson IW. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. Sci Adv. 2022 Mar 4;8(9):eabm5386. doi: 10.1126/sciadv.abm5386. Epub 2022 Mar 4. PMID: 35245110; PMCID: PMC8896783.
20. Dolzhenko E, Deshpande V, Schlesinger F, Krusche P, Petrovski R, Chen S, et al. ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. Birol I, editor. Bioinformatics. 2019 Nov 1;35(22):4754–6.

Vu, le Directeur de Thèse

**Vu, le Doyen
De la Faculté de Médecine de Tours
Tours, le**

Civit Antoine

64 pages – 26 figures

Résumé :

Introduction. Les maladies génétiques à expansion nucléotidique sont un groupe de maladies génétiques causées par l'expansion anormale d'une séquence de nucléotides dans un gène. La séquence est ainsi anormalement longue ayant pour conséquences diverses pathologies. Le diagnostic et la compréhension moléculaire de ces maladies se sont avérés difficiles en raison des limites des techniques conventionnelles de séquençage. Les avancées technologiques continuent de fournir de nouvelles méthodes et outils pour le diagnostic de ces maladies complexes. Ainsi le séquençage long-fragment semble être une technologie d'avenir dans le diagnostic et la caractérisation des maladies de nucléotides.

Matériels et méthodes. Cette thèse explore l'utilisation du séquençage long-fragment « Nanopore » associé à la technique CRISPR/Cas9 comme méthode innovante pour étudier les maladies à expansion de triplets et plus particulièrement le syndrome de l'X fragile, la maladie de Steinert et l'ataxie de Friedreich. Nous avons donc évalué la faisabilité du séquençage long-fragment « Nanopore » associé à la technique CRISPR/Cas9 pour détecter et quantifier les expansions de triplets dans ces trois maladies.

Résultats. Nos résultats montrent une technologie de séquençage long-fragment intéressante. Cependant la technologie CRISPR/Cas9 est difficile à utiliser. D'autres méthodes sont à expérimenter.

Conclusion. D'autres propositions méthodologiques associées au séquençage long-fragment paraissent plus performantes. Ainsi il semblerait que l'utilisation de ce type de séquençage ouvre la voie à un diagnostic plus précis et à une meilleure compréhension des mécanismes sous-jacents des maladies à expansion nucléotidique.

Mots clés : Maladies à expansion de triplets ; Microsatellites ; Courtes répétitions en tandem ; Séquençage long-fragment ; CRISPR/CAS9

Jury :

Président du Jury : Professeur Annick TOUTAIN

Directeur de thèse : Docteur Paul GUEGUEN

Membres du Jury : Docteur Anne TALLET
Madame Dr Marie-Laure WINTER VUILLAUME

Date de soutenance : 26 octobre 2023