



UNIVERSITÉ  
CAEN  
NORMANDIE



**DOSSIER UNIQUE DE CANDIDATURE AUX ALLOCATIONS  
DOCTORALES D'ETABLISSEMENT ET ALLOCATIONS  
DOCTORALES 100% REGION NORMANDIE**

A compter de la campagne 2024, les universités et écoles d'ingénieurs Normandes mettent en place un processus **unique** de soumission et d'évaluation pour les allocations d'établissement et les allocations 100% région, accompagné d'un calendrier unique de soumission et d'évaluation.

**Merci de cocher impérativement au moins une case (possibilité de cocher les deux cases en cas de double soumission)**

**La demande concerne :**

- Une allocation d'établissement  
 Une allocation régionale 100 %

Merci de remplir pour toutes les demandes la partie commune (pages 2-3).  
Dans le cas d'une demande d'allocation régionale (ou de double demande), remplir également la section correspondante (pages 4-6).

**Date limite de dépôt en ligne sur la plateforme de Normandie Université : 29 Janvier 2024**

**PARTIE COMMUNE : à remplir dans tous les cas (soumission à l'établissement ou à la région)**

**INFORMATIONS SUR LA THESE**

---

Acronyme de la thèse	Intitulé	Établissement employeur	Laboratoire d'accueil
LREAD-PhD	Long Read Early-onset Alzheimer Disease-PhD : exploitation de données de génomes long-read à haute fidélité pour explorer le déterminisme génétique de la maladie d'Alzheimer jeune	Université de Rouen Normandie	Inserm UMR1245 – Cancer and Brain Genomics (CBG) – équipe 3 "Genomics for Brain Disorders" (G Nicolas)

5 mots-clés associés au projet : Alzheimer, genomics, long-read, bioinformatics, missing heritability

**Informations sur le laboratoire d'accueil :**

Nom du Laboratoire d'accueil (nom, acronyme, numéro UR/UMR) : Cancer and Brain Genomics, CMG, Inserm UMR1245 (G Nicolas)

---

Équipe de recherche (si existante) : équipe 3 – Genomics for Brain Disorders (G Nicolas)

Si laboratoire sur plusieurs sites, précisez le lieu : Ville / Campus / Établissement hébergeur

Université de Rouen Normandie – UFR Santé

Adresse : Batiment Recherche UFR Santé

N° - Libellé de la voie : 22, boulevard Gambetta

Code postal : 76000

Commune : Rouen

**École doctorale de rattachement du directeur de thèse :**

**Calendrier :**

Période d'exécution du projet de thèse : **EDNBISE**

Du 02/11/2024 au 01/11/2027, soit 36 mois de projet.

---

### **Résumé vulgarisé et diffusable du projet en Français et en Anglais (4000 caractères maximum chaque) :**

*Description à destination d'un public non averti/novice sur le sujet. Ce paragraphe pourra être utilisé dans des documents de communication en cas de subventionnement du projet.*

#### Version Française

Environ 5% des malades atteints de la maladie d'Alzheimer présentent un début précoce, marqué par des premiers symptômes apparaissant avant l'âge de 65 ans. Parmi ces patients jeunes, seulement une minorité (environ 13%) présentent une forme purement héréditaire causée par un changement unique dans l'ADN, qui est alors nécessaire et suffisant pour développer la maladie. Pour tous les autres malades, atteints de formes précoces mais aussi de formes plus tardives, les facteurs génétiques jouent un rôle très important mais beaucoup plus complexe. Des différences génétiques individuelles multiples, certaines avec un effet fort, d'autres avec un effet plus modeste, combinent leur effet à celui de facteurs individuels non génétiques. Le domaine de la génétique a fait des progrès considérables ces 10 dernières années grâce à l'avènement de nouvelles technologies, permettant d'identifier des facteurs de risque génétiques avec un large spectre d'effets. Malgré ces progrès, un grand nombre de facteurs génétiques reste à identifier. Nous avons par exemple montré que 32% des patients avec une maladie d'Alzheimer jeune ne portent aucun des facteurs de risque génétiques modérés à fort actuellement connus. Ces malades restent donc sans explication sur la cause de leur maladie, même partiellement. Nous faisons l'hypothèse que certains facteurs génétiques jouant un rôle fort sur le développement de la maladie sont passés inaperçus jusque-là du fait de limites dans les technologies de séquençage utilisées. Nous souhaitons appliquer une toute nouvelle technologie permettant de lire de longues molécules d'ADN avec une qualité de séquençage inégalée (PacBio HiFi), permettant de voir quasiment tout type de variations génétiques, même parmi les plus complexes, inexplorées jusqu'à ce jour. Nous appliquerons cette nouvelle technologie à 230 patients avec une maladie d'Alzheimer jeune et sans facteur génétique au moins modéré connu parmi la série de patients du Centre National de Référence Malades Alzheimer Jeunes à Rouen, et à 170 témoins sélectionnés parmi la cohorte Insight de l'hôpital de la Pitié Salpêtrière à Paris. Cette étude, pionnière en France et en Europe, aura le potentiel de mettre au jour de nouveaux mécanismes génétiques impliqués dans la maladie d'Alzheimer, et ainsi mieux la comprendre et pouvoir permettre à l'avenir une prévention ciblée de cette maladie neurodégénérative. Le/la doctorant(e) sera en charge de développements méthodologiques bioinformatiques et biostatistiques et d'analyser ces données, permettant d'identifier de nouveaux déterminants génétiques forts de la MA et de les répliquer dans des séries indépendantes.

#### Version Anglaise

Around 5% of Alzheimer's patients have an early onset, with symptoms appearing before the age of 65 years of age. Of these younger patients, only a minority (around 13%) have a purely hereditary form of the disease caused by a single change in DNA, which is then necessary and sufficient to develop the disease. For all other patients, with both early and later forms, genetic factors play a very important but far more complex role. Multiple individual genetic differences, some with a strong effect, others with a more modest one, combine their effect with that of individual non-genetic factors. The field of genetics has made considerable progress over the last 10 years, thanks to the advent of new technologies, making it possible to identify genetic risk factors with a broad spectrum of effects. Despite this progress, a large number of genetic factors remain to be identified.

We have shown, for example, that 32% of patients with early-onset Alzheimer's disease do not carry any of the currently known moderate to strong genetic risk factors. These patients therefore remain without any explanation as to the cause of their disease, even partially.

We hypothesize that certain genetic factors playing a strong role on the development of the disease have so far gone unnoticed due to limitations in sequencing technologies. We will apply a completely new technology of long DNA molecules sequencing with high fidelity quality (PacBio HiFi), making it possible to see virtually any type of genetic variation, even the most complex, previously unexplored. We will be applying this new technology to 230 patients with early-onset Alzheimer's disease and no known risk factor among the moderate to strong ones, from the patient series of the National Reference Center for Young Alzheimer Patients (Rouen), and to 170 cognitively normal controls selected from the Insight cohort at the Pitié Salpêtrière Hospital in Paris as having normal amyloid PET imaging.

This study, a pioneer in France France and Europe, has the potential to uncover new genetic mechanisms involved in Alzheimer's disease, thereby improving our understanding and, in future, enable targeted prevention of this neurodegenerative disease. The PhD student will be responsible for developing bioinformatics and biostatistics pipelines, analyze the data, identify new strong genetic determinants of AD and replicate them in independent series.

**Contexte et Objectifs (2 pages maximum) :** précisez s'il s'agit d'une recherche incrémentale (poursuite, renforcement d'un thème existant) ou innovante (nouveau sujet), positionnement local-national-international, bibliographie

Les facteurs génétiques jouent un rôle majeur dans l'étiologie de la maladie d'Alzheimer (MA)<sup>13</sup>. **La MA jeune (MAJ, premiers symptômes <65 ans)** est considérée comme étant encore plus génétiquement déterminée. Cependant, **<15 % des patients atteints de MAJ présentent un variant pathogène dans l'un des gènes autosomiques dominants** connus de la MA, *APP*, *PSEN1*, ou *PSEN2*<sup>16,20</sup>, expliquant à eux seuls la survenue de la maladie. La plupart des patients avec MAJ non monogénique et ceux avec MA tardive, la maladie est causée par l'interaction de multiples facteurs génétiques et non génétiques, les facteurs génétiques jouant un rôle majeur. **Les facteurs de risque génétiques** peuvent être classés en fonction de leur fréquence et de leur effet sur le risque de MA. Le principal facteur de risque génétique de la MA est **l'allèle APOE4, à la fois fréquent** (~10-20 % dans diverses populations) et associé à un **effet modéré à élevé**, avec des odds ratios (OR) de 3 à 4 et jusqu'à 14 chez les porteurs hétérozygotes et homozygotes, respectivement<sup>14</sup>. **D'autres variants fréquents** (fréquence allélique >1 %), identifiés par **les études d'association pangénomique (GWAS) sur puces d'ADN**, sont associés à des **OR modestes** (<1,3)<sup>2</sup>. Bien que leur effet puisse être évalué à l'aide de scores de risque polygénique (PRS) combinant les faibles effets de multiples variants fréquents en un seul score, de telles approches manquent une grande partie de l'information génétique individuelle. Plus récemment, l'avènement du **séquençage parallèle massif ou du séquençage à lecture courte (SR, dit de nouvelle (NGS) ou deuxième génération)**, a permis l'analyse de presque toutes les régions codantes du génome (séquençage de **l'exome, ES**, ~1,5 % du génome). **Les variants codants rares de 5 gènes** sont désormais confirmés comme des facteurs de risque de MA avec des **OR allant de 1,5 à ~30**<sup>15</sup>. Notre équipe a joué un rôle pionnier dans ce domaine en (i) identifiant le rôle des variants rares dans *SORL1*<sup>25,21</sup>, (ii) contribuant à décrypter le rôle des variants *ABCA7* et *TREM2*<sup>3,17</sup> et, plus récemment, (iii) co-dirigeant un consortium européen (ADES, Alzheimer Disease Europe Sequencing) qui a identifié deux nouveaux gènes, *ATP8B4* et *ABCA1*<sup>15</sup>.

**Le séquençage du génome entier (WGS)** permet d'étudier les régions non explorées et certains types de variations génétiques peu étudiés. Les variants non codants dans les régions non traduites des gènes, les introns ou les régions régulatrices peuvent affecter l'expression ou l'épissage des gènes. Hormis les polymorphismes fréquents issus des GWAS, à faible effet, **les variants non codants rares ont été peu étudiés dans la MA**. Notre équipe et d'autres avons montré que ces différentes catégories de variants peuvent avoir un impact très fort, allant jusqu'à une hérédité monogénique, dans diverses maladies (par exemple<sup>8,9,5</sup>. **Dans la MA, les variants structurels (SV), les répétitions d'ADN et les insertions d'éléments transposables n'ont pas été largement étudiés**, représentant ainsi une source importante de nouvelles découvertes. Les SV affectent la structure de l'ADN génomique et peuvent être équilibrés ou non (variants du nombre de copies, CNVs, suppressions et duplications). Très peu d'études ont été menées dans la MA, au-delà des CNVs monogéniques rares (duplications d'*APP* et délétions partielles de *PSEN1*, identifiées dans notre groupe<sup>18,26</sup>). Nous analysons actuellement les CNVs à partir des données d'ES dans le cadre du consortium ADES. Bien que le WGS basé sur la technologie de lecture courte (**SR-WGS**, utilisant la technologie Illumina comme pour l'ES, séquençage par paire de fragments d'ADN de ~200-400 pb en lectures généralement de 150 pb) représente une stratégie puissante pour évaluer les variants non codants et les SV, une telle technologie (i) passe à côté de certaines régions génomiques (y compris des régions contenant des gènes)<sup>22</sup> et (ii) est associée à un taux élevé de faux positifs parmi différents outils de SV, tout en en manquant également beaucoup (par exemple, 47 % des délétions et près de 78 % des insertions manquées par Illumina SR-WGS<sup>6,19</sup>). De plus, le SR-WGS n'est pas bien conçu pour évaluer les répétitions en tandem (VNTR, comprenant

des répétitions en tandem courtes (STR, microsatellites, répétitions de 1 à 9 pb) et des minisatellites, répétitions de 10 à 99 pb ; les satellites représentant des répétitions >100 pb). Malgré cela, certains outils (par exemple, Expansion Hunter) sont capables de génotyper certaines répétitions, avec une précision limitée par la longueur de la répétition et une fiabilité moindre à l'échelle du génome entier. Les répétitions d'ADN représentent une source de polymorphisme et peuvent provoquer des maladies neurodégénératives (maladie de Huntington ou démence frontotemporale / sclérose latérale amyotrophique liées aux expansions introniques hexanucléotidiques *C9ORF72* ou identification plus récente des causes de l'ataxie telles que *FGF14*<sup>23,24</sup>, *RFC1*<sup>7</sup> ou d'autres maladies neurologiques (par exemple, *NOTCH2NLC*)<sup>10</sup>. La plupart des découvertes de répétitions en tandem utilisant le SR-WGS ont été rendues possibles grâce à une focalisation sur un locus donné à la suite d'une analyse de liaison dans des familles avec une hérédité monogénique suspectée. Dans la MA, une telle stratégie ne peut pas être appliquée directement dans les familles en raison du risque de phénocopies et parce que nous n'attendons pas une pénétrance complète dans les familles ayant eu un séquençage négatif pour *APP*, *PSEN1* et *PSEN2*. Cependant, la puissance des GWAS peut être utilisée car de telles études peuvent pointer vers des *loci* d'intérêt. Par exemple, grâce à un séquençage ciblé utilisant de nouvelles techniques de séquençage décrites ci-dessous, un groupe belge a découvert qu'une augmentation de la longueur de répétition d'un VNTR à côté d'une région d'épissage dans *ABCA7* était associée à un défaut d'épissage et augmentait significativement le risque de MA chez certains porteurs de l'haplotype à risque trouvé dans les GWAS, avec un OR beaucoup plus élevé de 4,5<sup>11</sup>. Cela illustre clairement le potentiel d'une analyse à l'échelle du génome avec cette technologie de séquençage.

**Ici, nous proposons une analyse à l'échelle du génome de tout type de variation génétique (variants codants et non codants, variants structuraux, répétitions) dans une approche de séquençage unique utilisant le séquençage de 3<sup>ème</sup> génération du génome.** Ce séquençage, dit à **lecture longue (LR)** repose sur de longs fragments d'ADN. Deux technologies sont désormais disponibles, par Oxford Nanopore (ONT) et **PacBio**. ONT, avec de très longs fragments d'ADN (N50 de 10 à 200 kb, signifiant que 50 % des bases séquencées sont dans des lectures de cette taille ou plus), montre de très bonnes performances pour détecter tout SV, mais une qualité plus faible pour les indels au sein de régions répétées. D'autre part, le séquençage PacBio permet l'analyse de lectures longues (N50 de 10 à 20 kb) et est également associé à une qualité d'appel extrêmement élevée pour les SNV/indels lors de **l'utilisation du mode Haute Fidélité (HiFi)** par séquençage de lectures de consensus circulaires en molécules uniques (CCS)<sup>28</sup>. Cette technologie combine les avantages des longues molécules avec un taux d'erreur très faible par base (>99 % d'exactitude des SNV et indels, supérieur à SR-WGS par Illumina<sup>19</sup>). Elle peut ainsi détecter **virtuellement tous les types de variants dans une seule expérience**, aboutissant à un **"génome presque parfait"**. De plus, la nouvelle machine **PacBio Revio** permet maintenant de réduire drastiquement les coûts, avec un coût d'environ ~1130€ pour un WGS de 30x en réactifs (avec une cellule SMRT Revio par échantillon) utilisant la technique HiFi. Cette technologie sera bientôt disponible à l'université de Rouen grâce au projet CPER-FEDER « GEMM » (Génétique des maladies et des microbes), faisant de notre centre l'un des 4 centres français équipé, le premier dans un contexte de recherche médicale.

## Projet détaillé (détails scientifiques sur le sujet, questions ou verrous abordés, programme et déroulement prévu) et perspectives pour l'équipe/le laboratoire (3 pages maximum) :

Malgré la réduction des coûts, les séquençages SR-WGS et LR-WGS restent plus coûteux que les puces ADN et les ES, limitant la taille des études cas-témoins réalisables. Pour résoudre ce problème, nous allons utiliser une **stratégie d'échantillonnage par phénotypes extrêmes**, qui augmente la puissance statistique lorsque la taille de l'échantillon est limitée<sup>1</sup>. En adoptant cette stratégie, nous avons pu identifier une association significative de variants rares codants de *SORL1*, l'un des principaux gènes dans la MA non monogénique, en utilisant un échantillon de taille limitée d'environ ~205 patients MAJ familiaux et ~500 témoins<sup>21</sup>. Cette approche nous a permis de mener une étude nationale à une époque où les ES étaient bien plus coûteux, et d'être les premiers à identifier ce résultat maintenant largement répliqué au niveau international. Nous proposons **d'appliquer cette stratégie au séquençage de 3<sup>ème</sup> génération**.

**Notre objectif est d'identifier de nouveaux déterminants génétiques de la MAJ** (et potentiellement de toutes les formes de MA, comme l'ont montré les résultats antérieurs par la suite généralisés à la MA) en utilisant un phénotypage extrême. Nous prévoyons de créer l'un des **premiers jeux de données HiFi LR-WGS en séquençant le génome de 230 patients MAJ hautement sélectionnés et de 170 témoins** appariés selon le sexe, les origines géographiques et le statut APOE.

**Sélection des patients MAJ.** Le centre de référence national maladies Alzheimer Jeunes (CNRMAJ) - Inserm U1245 (CHU de Rouen et Université de Rouen Normandie) est le seul **centre national de diagnostic et recherche en génétique de la MAJ** depuis plus de deux décennies. Au cours des dix dernières années, la plupart des patients MAJ ont eu un ES, soit rétrospectivement, soit prospectivement depuis 2017. Des échantillons de sang sont reçus dans notre centre à partir des consultations mémoire de tous les CH et CHU français, avec les informations cliniques, paracliniques et les arbres généalogiques, qui sont examinés par les neurologues du CNRMAJ. Cela nous a permis d'analyser des **données d'ES provenant de >2 200 patients MAJ non apparentés**. Parmi eux, 68% portent au moins un facteur de risque génétique connu de modéré à élevé ( $1,5 < OR < 30$ )<sup>20</sup>. Nous sélectionnerons les 230 patients MAJ les plus jeunes avec un diagnostic validé par biomarqueurs du LCR, aucun variant pathogène connu dans les gènes de démence monogénique, aucun facteur de risque génétique parmi les allèles *APOE4* et des variants rares prédits comme délétères et tronquants dans *SORL1*, *TREM2*, *ABCA7*, *ABCA1*, *ATP8B4* selon les catégories préalablement établies<sup>15</sup>. Après contrôle de qualité de l'ADN, nous sélectionnerons les patients en fonction des antécédents familiaux pour enrichir en déterminants génétiques forts et les appairerons avec les témoins déjà sélectionnés.

**Sélection des témoins.** Les témoins ont été sélectionnés au sein de la **cohorte INSIGHT** (âge > 70 ans), parmi ceux avec tests cognitifs normaux à l'inclusion et après 5 ans, ainsi qu'une TEP amyloïde négative<sup>12</sup>, soit 183 individus, dont 170 non-porteurs *APOE4*.

**Extraction et préparation des ADN.** Pour tous les patients, nous disposons de sang stocké au CRB du CHU de Rouen. L'ADN a été isolé à partir d'un tube pour des études précédentes, tandis qu'un deuxième tube ou le reste a été aliquoté et stocké directement à -80°C sans cycle de décongélation/recongélation. Les échantillons de sang des témoins ont été prélevés à l'hôpital de la Pitié-Salpêtrière, aliquotés et stockés directement à -80°C dans l'heure suivant le prélèvement, sans cycle de décongélation/recongélation. Une nouvelle extraction d'ADN sera réalisée à partir des échantillons stockés à l'aide de kits spécifiques pour obtenir de l'ADN de haut poids moléculaire, et un contrôle qualité sera effectué (Femtopulse). Les résultats obtenus à l'aide de 3 kits différents confirment que nous obtenons ainsi de l'ADN de haute qualité et de haut poids moléculaire.

**Séquençage.** Le séquençage sera réalisé en utilisant le mode CCS HiFi sur un séquenceur **PacBio REVIO**, qui sera installé sur la plateforme de génomique de l'Université de Rouen

Normandie (SCG, Service commun de Génomique) début 2024 (Inserm U1245), suite au financement par la région Normandie, l'état (CPER, contrat plan état région) et les fonds européens de développement régional FEDER. Étant donné l'expertise du SCG dans la mise en œuvre de nouvelles technologies de séquençage, cette technique devrait être opérationnelle dès début 2024. Lors du démarrage du projet du doctorant, nous estimons que les 100 premiers génomes auront été séquencés par les équipes du CNRMAJ/InsermU1245-SCG.

**Bioinformatique et statistiques.** Des pipelines bioinformatiques spécifiques aux données HiFi (SMRT analysis software) ainsi que des outils open source seront testés et comparés, notamment pour l'alignement sur GRCh38 (ex. minimap-2) et la détection des variants de type SNV/indels, SV équilibrés ou non, insertion d'éléments mobiles et détection de STR (ex. : deepvariant, HiCanu (duplications segmentaires, satellites), pbsv, sniffles2 (SV), etc.) ainsi que pour la visualisation des résultats. L'annotation des données sera effectuée à l'aide de scripts maison basés sur SNPEff/SNPSift pour les petits variants et AnnotSV pour les SV, qui comprend une liste non exhaustive d'annotations telles que les informations de la population gnomAD, les scores REVEL et CADD, un pipeline d'annotation non codant évaluant l'épissage (SpliceAI, SPIP), UTRannotator pour les 5'UTR, des outils spécifiques pour les 3'UTR, GeneHancer pour les régions régulatrices et les domaines topologiquement associés (TAD). Ce projet permettra des développements bioinformatiques afin de mettre en place les pipelines requis pour ce type d'analyse, en suivant l'état de l'art et les avancées du groupe international HiFi Solves.

**Les 80 loci de MA connus feront l'objet de la première analyse.** Tout comme les stratégies ciblées dans les familles avec informations de liaison, et comme précédemment réalisé au locus *ABCA7* dans la MA<sup>11</sup>, nous extrairons les variants de tous types qui affectent des loci GWAS<sup>2</sup> ou les gènes de formes monogéniques, chez des individus avec ou sans l'haplotype à risque. En effet, comme le montrent les variants rares ciblant des gènes également trouvés indépendamment comme des hits GWAS, et comme évalué dans notre article de consortium récent<sup>15</sup>, ces régions sont une source potentielle de nouvelles découvertes importantes. Ceci sera suivi d'une **analyse exploratoire à l'échelle du génome** basée sur (i) une analyse des variants codants et non codants triés en fonction d'effets prédits sur l'épissage ou l'expression génique, (ii) nous nous appuyerons sur le développement des quatre dernières années (travail de thèse d'O. Quenez) pour réaliser des analyses basées sur les transcrits pour les SV, ainsi que des analyses cas-témoins spécifiques aux SV, afin de ne pas passer à côté d'événements moins rares qui n'auraient pas été capturés préalablement pour des raisons techniques. Pour les répétitions d'ADN, nous nous concentrerons sur les longueurs de répétitions présentes chez moins de 1% des témoins et évaluerons des enrichissements spécifiques chez les cas.

**Réplication.** Le séquençage de longues lectures est essentiel pour découvrir des SV et des répétitions d'ADN avec une excellente précision ; cependant, la plupart des SV peuvent être génotypés par la technique SR-WGS avec des analyses bioinformatiques ciblées. Cette dernière stratégie a une précision beaucoup plus faible pour les grandes expansions et les SV, mais elle permet un génotypage ciblé et une confirmation reste possible par analyse moléculaire ciblée. A partir des régions identifiées après analyse de génomes HiFi, et **nous tenterons de répliquer le signal dans les données des consortiums ADES-ADSP, comprenant >36 000 SR-WGS** (<https://dss.niagads.org/datasets/ng00067/>). Nous sommes également informés d'un autre projet basé sur la même technologie HiFi par un groupe collaborateur proche (H. Holstege) aux Pays-Bas, et nous collaborerons pour répliquer les découvertes respectives. Dans certaines familles, nous évaluerons la **ségrégation** en génotypant les variants identifiés par technique ciblée basée sur PCR et séquençage et/ou le séquençage PacBio ou ONT adaptatif (Minlon, SCG), en fonction des types de variants chez les membres atteints et non atteints, ceci sera rendu possible grâce à notre réseau national

permettant l'accès aux ADN d'apparentés<sup>27</sup>. De plus, étant donné que certains variants non codants, SV, ou répétitions peuvent affecter l'expression génique ou l'épissage, nous effectuerons des **analyses ciblées d'ARNm**, soit en utilisant le séquençage PacBio d'ADNc, le séquençage ARN ONT ciblé, ou des techniques basées sur PCR. L'ARNm isolé du sang des patients pourra être disponible grâce à notre réseau national pour les patients et des données de RNAseq sont disponibles pour des témoins. A noter que les variants des gènes connus comme associés à travers les variants rares codants seront traités séparément, étant donné la sélection des cas contre les variants codants de ces gènes.

**Bénéfices attendus.** Nous explorerons des **régions et types de variants presque inexplorés** pour éclairer la composante génétique inexpliquée de la MAJ. Notre étude présente des limites de puissance; cependant, l'échantillonnage par phénotypes extrêmes permettra de prioriser des variants avec impact fort. La réplication sera recherchée dans des séries plus larges séquencées avec des techniques classiques, ainsi que dans des approches parallèles qui émergeront dans les années à venir (ex., collaboration H. Holstege). L'approche à l'échelle du génome pourrait révéler des **découvertes entièrement nouvelles**, comme ce fut le cas pour *C9ORF72* dans la DFT-SLA et d'autres maladies neurologiques<sup>10</sup>. Le décryptage des déterminants génétiques de la MA contribuera non seulement à expliquer l'étiologie de la MAJ, et servira également le **conseil génétique et la prévention** de la MA à l'avenir. En effet, les mécanismes de la MA commencent es décennies avant les premiers symptômes, la prévention de la MA par anti-amyloïdes (seuls ou en combinaison) pourrait bien être efficace à l'avenir. Le potentiel de thérapies modifiant la maladie est particulièrement prometteur pour les porteurs présymptomatiques de variants autosomiques dominants de la MA en tant que stratégie préventive, comme illustré par le premier essai clinique combiné anti-amyloïde et anti-tau en cours (DIAN-TU NexGen (E2814)). Au-delà des familles rares avec MA autosomique dominante, la **prédiction génétique** du risque de MAJ pourrait offrir une **prévention personnalisée** à un plus grand nombre d'individus. Cela nécessite d'abord l'identification des déterminants génétiques de la MAJ permis par ce projet.

**Rôle du doctorant et financement du projet.** Le(la) doctorant(e) sera en charge des **développements bioinformatiques, de l'analyse et de l'interprétation des résultats**. Il/elle devra être titulaire d'un diplôme de Master en Bioinformatique ou équivalent. Il/elle sera sous la direction du Pr Gaël Nicolas et sous la supervision du Dr Camille Charbonnier (MCU-PH en Biostatistiques, PhD) et du Dr Olivier Quenez (Ingénieur Bioinformaticien, PhD). Il/elle travaillera en interaction avec le Pr David Wallon et le Dr Aline Zarea, neurologues (analyses de ségrégation, prélèvements complémentaires), le Dr François Lecoquierre (analyse des données génomiques complexes) et Stéphane Rousseau (technicien CNRMAJ/InsermU1245). Ce projet bénéficie d'un financement de la Fondation Recherche Alzheimer permettant le séquençage des 300 premiers génomes et le reste sera financé par le CNRMAJ (CHU de Rouen, budget disponible). Le financement des analyses ciblées d'ADN (ségrégation, confirmations) et d'ARNm sera possible grâce à nos dotations récurrentes de l'Inserm et de l'Université de Rouen Normandie.

#### Références (\*issues de l'équipe)

1. Barnett et al., 2013 PMID: 23184518
- 2.\* Bellenguez et al., 2022 PMID: 35379992
- 3.\* Bellenguez et al., 2017 PMID: 28789839
- 4.\* Campion et al., 2019 PMID: 30911827
- 5.\* Cassinari et al., 2020 PMID: 32506582
6. Chaisson et al., 2019 PMID: 30992455
7. Cortese et al., 2020 PMID: 32040566
- 8.\* Coursimault et al., 2022a PMID: 35842780
- 9.\* Coursimault et al., 2022b PMID: 35446447
10. Depienne & Mandel, 2021 PMID: 33811808
11. De Rooeck et al., 2018 PMID: 29589097
- 12.\* Dubois et al., 2018 PMID 29500152
13. Gatz M, et al., 2006 PMID: 16461860
- 14.\* Génin et al., 2011 PMID: 21556001
- 15.\* Holstege et al., 2022 PMID: 36411364
16. Jia L, et al. 2020. PMID: 31914229
- 17.\* Le Guennec et al., 2017a PMID: 27037229\*Le 18. Guennec et al., 2017b PMID: 28461250
19. Logsdon et al., 2020 PMID: 32504078
- 20.\* Nicolas et al., in press (Genetics in Medicine)
- 21.\* Nicolas et al., 2016 PMID: 26303663
22. Nurk et al., 2020 PMID: 32801147
23. Rafehi et al., 2022 PMID: 36493768
24. Pellerin et al., 2022 PMID: 36516086
- 25.\* Pottier et al., 2012 PMID: 22472873
- 26.\* Rovelet-Lecrux et al., 2006 PMID: 16369530
- 27.\* Schramm et al., 2022 PMID: 35761418
28. Wengler et al., 2019 PMID: 31406327

**Actions prévues dans le cadre de la diffusion de la Culture Scientifique, Technique et Industrielle (CSTI)** en supplément des actions Fêtes de la Science (1 page maximum) :

**Diffusion scientifique** : Les résultats et développements de ce projet seront présentés dans le cadre de groupes de travail nationaux sur cette thématique émergente, y compris dans le cadre du plan France Médecine Génomique 2025, ainsi que dans des congrès scientifiques nationaux et internationaux, par le/la doctorant(e), et bien sûr à travers des publications scientifiques. Nos travaux sont également régulièrement repris dans la presse ou dans des articles généralistes à destination de médecins généralistes ou de neurologues non spécialistes du domaine.

**Diffusion grand public** : en plus de la fête de la science, le/la doctorant(e) participera activement à l'organisation des visites de laboratoire organisées régulièrement pour des associations de malades (maladies rares, Alzheimer, Cancers) ainsi que les journées du patrimoine (grand public), et pourra également participer aux actions de sensibilisation aux questions de génétique dans le cadre du Téléthon. Il/elle participera à la semaine du Cerveau organisée par Science Action à travers une vulgarisation auprès du grand public intéressé par les maladies du cerveau.

Il/elle participera aux réunions avec les associations de malades avec maladies neurologiques rares qui se tient tous les deux ans (prochaine édition en 2025) à Rouen : échanges avec les associations et vulgarisation de résultats.

Il/elle participera aux manifestations organisées par l'EDnBISE (ou le collège des ED), et de l'Université de Rouen Normandie, en particulier à travers des échanges et présentations de travaux auprès des étudiants plus jeunes.

## ENVIRONNEMENT DU PROJET

---

### Directeur de thèse

Nom et Prénom : NICOLAS Gaël

Établissement / Organisme de rattachement : URN

**Avis motivé du directeur de thèse** sur la recherche envisagée en ce qui concerne les objectifs scientifiques et technologiques poursuivis et leur place dans les orientations de recherche du laboratoire d'accueil :

Il s'agit d'un projet phare de l'équipe qui jouera un rôle pionnier et novateur à travers de multiples aspects : générer et analyser des génomes longues molécules haute-fidélité pour la première fois dans le laboratoire et en Normandie en utilisant les nouvelles technologies de séquençage, ces technologies permettront d'explorer tout une composante du génome jusque-là très largement inexploree, en faisant ainsi une source potentielle de grandes avancées scientifiques dans le domaine de la génétique de la maladie d'Alzheimer, domaine d'excellence locale. Notre projet se base sur l'une des cohortes mondiales les plus riches grâce à notre recrutement national et notre proximité soin-recherche, organisation permettant un recrutement à la fois large et de très grande qualité. Nos précédentes études

ont toujours montré que les données du CNRMAJ/InsermU1245 sont parmi les plus significatives comparé aux autres jeux de données, du fait de ce recrutement. Ceci place ce projet comme extrêmement prometteur et arrivant à point nommé, l'arrivée du nouveau séquenceur et du financement pour ce projet débutant dans une fenêtre de temps parfaite pour nous permettre des d'acquérir des résultats particulièrement originaux.

**Date : 24/01/2024**

**Nom, Prénom : NICOLAS, Gaël**

**Signature**



**Le directeur du laboratoire indiquera l'adéquation de ce projet** avec la stratégie du laboratoire d'accueil

Le directeur du laboratoire devra fournir (de façon séparée du dossier) **un interclassement (avec justification) des demandes du laboratoire** (incluant les allocations établissement et 100% région) *validé par son conseil d'Unité. Merci de ne pas indiquer d'interclassement dans ce document de soumission*

Ce projet sera un catalyseur pour l'ensemble de l'unité, le service commun de génomique, et plus largement le développement des compétences et connaissances techniques, bioinformatiques, mais aussi conceptuelles en ce qui concerne les variations du génome. En effet, malgré des avancées majeures dans les connaissances de régions codantes et des variations « simples » du génome depuis plus de 10 ans, les connaissances sur les variations et régions complexes du génome restent encore à développer, comme en témoigne la complétion du séquençage du génome humain réellement en entier seulement en 2022, grâce à cette technologie « HiFi » de Pacbio, justement. Le développement de ces compétences et connaissances permettra des avancées scientifiques propres à ce projet, lui-même d'envergure internationale, et catalysera d'autres projets basés sur cette même technologie.

Le/la doctorant(e) évoluera dans un domaine multidisciplinaire avec composante médicale, génétique moléculaire, bioinformatique et biostatistique, en développant des compétences dans le domaine de la bioinformatique et des biostatistiques plus précisément, compétences requises pour le développement de la médecine de précision à l'avenir.

**Date : 24/01/2024**

**Nom, Prénom : NICOLAS, Gaël**

**Signature**



**PARTIE SPECIFIQUE ALLOCATIONS REGIONALES (ne pas remplir si la demande concerne uniquement une allocation établissement)**

 <p>RÉGION NORMANDIE</p>	<p><b>Normandie Recherche</b> Dossier de candidature « Soutien Doctorants 100% »</p>	
	<p><b>Thème : Recherche et innovation</b></p>	
	<p><b>Objectif Stratégique</b></p>	<p>Pour une économie normande dynamique, attractive et innovante</p>
	<p><b>Mission</b></p>	<p>Accompagner la recherche et l'innovation, levier de développement économique</p>
	<p><b>Territoire</b></p>	<p>Normandie</p>
	<p><b>Type d'aide</b></p>	<p>Subvention</p>

Le dispositif Normandie Recherche - Soutien Doctorants 100%, doit répondre à l'ambition 2 du SRESRI en contribuant à soutenir les ambitions des étudiants, enseignants-chercheurs, chercheurs et institutions ESRI normands pour favoriser leur réussite.

Avant toute soumission, veuillez consulter le règlement du dispositif concerné disponible au lien suivant : [Normandie Recherche | Région Normandie](#)

**Actions éligibles :**

Le soutien de la Région porte sur la mise en place et la réalisation de projets de recherche. L'aide régionale versée permet l'accueil de doctorants dans les laboratoires académiques pendant 36 mois maximum.

La Région soutient de jeunes chercheurs préparant une thèse et désirant obtenir un doctorat d'Université. Il s'agit d'un financement de thèse à 100%.

Les doctorants financés par la Région devront être employés par une université normande et participer à des actions de CSTI.

**Modalités d'instruction et d'attribution :**

- 1 – La Région définit le nombre d'allocations allouées à chacune des trois Universités normandes,
- 2 – Dans le cadre de leur Accord de Partenariat Stratégique, la Région Normandie et chaque Université bénéficiaire définissent les thématiques et les modalités d'expertises respectant les critères d'éligibilité définis ci-dessous,
- 3 – Les Universités déposent leurs demandes sur la plateforme régionale ainsi que les deux expertises, avant la date communiquée annuellement par la Région
- 4 – La Région assurera l'instruction et la proposition en Commission permanente.

### **Critères d'éligibilité :**

Le projet de thèse devra répondre aux critères suivants :

- qualité scientifique du projet, laquelle sera vérifiée par la transmission de deux expertises externes à la structure (hors Normandie),
- retombées et perspectives attendues pour le territoire en termes de développement territorial, de valorisation du projet, d'attractivité, de rayonnement des équipes normandes,
- développement ou renforcement de collaborations régionales, interrégionales et internationales,
- inscription dans un des domaines de la S3 ou dans les thématiques spécifiées dans l'accord de partenariat Région-Université.

Le candidat ou la candidate devra :

- être accueilli dans un laboratoire situé en Normandie,
- être inscrit dans une École Doctorale normande,
- porter un projet de thèse en adéquation avec les axes de recherche du laboratoire d'accueil et la stratégie de l'établissement,
- s'engager chaque année dans des actions de culture scientifique, technique et industrielle, coordonnées par la Région Normandie ou l'un de ses partenaires.

**Ce dossier de candidature constituera un support dans le cadre d'une instruction régionale.**

Le projet s'inscrit dans la stratégie de spécialisation intelligente (S3) du Programme Opérationnel régional FEDER-FSE/IEJ Normandie 2021-2027, et en particulier dans le ou les domaines de spécialisation suivants :

#### **Domaine S3 principal :**

- Préserver et transformer durablement les ressources agricoles, marines, sylvicoles et les systèmes de production*
- Développer un mix énergétique vers zéro émission carbone*
- Transformer les process pour une industrie performante, durable et digitale*
- Développer de nouvelles solutions de mobilités bas-carbone efficaces et sécurisées*
- Accélérer les synergies et l'innovation au service d'une médecine 5P humaine et animale*
- Faire de la Normandie un territoire résilient par la maîtrise des risques*

#### **Domaine S3 secondaire (le cas échéant) :**

- Préserver et transformer durablement les ressources agricoles, marines, sylvicoles et les systèmes de production*
- Développer un mix énergétique vers zéro émission carbone*
- Transformer les process pour une industrie performante, durable et digitale*
- Développer de nouvelles solutions de mobilités bas-carbone efficaces et sécurisées*
- Accélérer les synergies et l'innovation au service d'une médecine 5P humaine et animale*
- Faire de la Normandie un territoire résilient par la maîtrise des risques*

Vous pouvez consulter la Stratégie de Spécialisation Intelligente (S3) de la Région Normandie [en cliquant ici](#).

**Dans le cas où le sujet de thèse ne correspondrait à aucun domaine de la S3, veuillez faire référence aux éléments de l'Accord de Partenariat Stratégique signé entre la Région et l'établissement porteur, justifiant la présente demande.**

*(médecine 5P)*

**Pour tout projet, vous préciserez son impact sur le territoire normand :**

Notre équipe a :

- un positionnement national à travers son rôle de centre de référence unique pour la génétique de la maladie d'Alzheimer Jeune en France, avec un impact dans le cadre de l'offre de soin et la recherche translationnelle, nos travaux de recherche étant traduits en offre de soin dans un délai généralement rapide après publication et validation, au bénéfice des patients et pour une meilleure prise en charge par leurs médecins,
- un rayonnement international à travers nos résultats scientifiques comme le montre par exemple notre rôle moteur dans le consortium européen ADES (co-investigateur principal).

Ce projet a un potentiel de découvertes totalement nouvelles s'appuyant à la fois sur la force de notre recrutement national dans un modèle unique, et sur une nouvelle technologie qui a une qualité inégalée démontrée mais actuellement encore peu disponible, du fait de limitations de débits. Le nouveau séquenceur, en cours d'acquisition sur fonds CPER/FEDER, permet d'augmenter considérablement le débit, et son acquisition va placer notre centre et ainsi la Normandie à la pointe de la recherche en génomique. Cette première application large sur la maladie d'Alzheimer aura potentiellement des résultats à haut impact.

Ce projet sera donc le premier projet de grande envergure utilisant la technologie HiFi de PacBio en Normandie. Ceci nous permettra d'acquérir une expérience significative dans ce domaine très novateur, puisque seulement 4 centres sont/seront équipés d'un tel séquenceur de type Revio en France à notre connaissance en 2024, alors que la technologie traditionnelle Illumina est implémentée dans chaque CHU/Université. Cette technologie est une avancée considérable et ce projet permettra de maintenir la région Normandie à la pointe de l'expertise en génomique au niveau national. Nous espérons que ce projet nous permettra d'intégrer un groupe de travail international sur l'analyse des génomes long read HiFi (HiFi Solves), regroupant les centres experts mondiaux, permettant d'établir des guidelines en vue de l'implémentation plus large à l'avenir. Ceci nous permettra d'être nous-mêmes une ressource pour d'autres projets au niveau national au minimum.