



**HAL**  
open science

# Variants rares et analyse d'exomes : application à la maladie d'Alzheimer du sujet jeune

Kilan Le Guennec

► **To cite this version:**

Kilan Le Guennec. Variants rares et analyse d'exomes : application à la maladie d'Alzheimer du sujet jeune. Psychiatrie et santé mentale. Normandie Université, 2017. Français. NNT : 2017NORMR148 . tel-01892656

**HAL Id: tel-01892656**

**<https://theses.hal.science/tel-01892656>**

Submitted on 10 Oct 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Normandie Université

# THÈSE

**Pour obtenir le diplôme de doctorat**

**Spécialité Aspects moléculaires et cellulaires de la biologie**

**Préparée au sein de l'Université de Rouen Normandie**

**Variants rares et analyse d'exomes :  
Application à la maladie d'Alzheimer du sujet jeune**

**Présentée et soutenue par  
Kilan LE GUENNEC**

**Thèse soutenue publiquement le 7 Juin 2017  
devant le jury composé de**

Mme Elisabeth TOURNIER-LASSERVE	Professeur des Universités – Praticien Hospitalier, Inserm U1161 Université Paris Diderot	Rapporteur
Mr Jean-Charles LAMBERT	Directeur de Recherche, Inserm U1167 Université de Lille	Rapporteur
Mme Isabelle LE BER	Praticien Hospitalier (HDR), Inserm U1127 Université Pierre et Marie Curie Paris 06	Examineur
Mr Thierry FREBOURG	Professeur des Universités – Praticien Hospitalier, Inserm U1245 Université de Rouen Normandie	Examineur
Mr Dominique CAMPION	Praticien Hospitalier (HDR), Inserm U1245 Université de Rouen Normandie	Directeur de thèse

**Thèse dirigée par M. le Dr Dominique Campion,  
Inserm U1245 – Genomic and Personalized Medicine in Cancer and Neurological Disorders**



## Remerciements

*Je tiens tout d'abord à remercier sincèrement Madame le Professeur Elisabeth Tournier-Lasserre, Madame le Docteur Isabelle Le Ber et Monsieur le Docteur Jean-Charles Lambert d'avoir accepté de juger mon travail de thèse.*

*Monsieur le Docteur Dominique Campion, je tenais à vous remercier pour tout ce que vous m'avez apporté durant ces trois années de thèse. Nos échanges et votre vision de la génétique humaine ont constamment enrichi mes connaissances scientifiques et m'ont permis d'aborder les différentes problématiques avec passion. Merci également pour le temps que vous m'avez consacré dans l'aide à l'écriture de cette thèse.*

*Je tiens également à exprimer toute ma gratitude à Monsieur le Professeur Thierry Frébourg. J'ai conscience de la chance que vous m'avez donnée en me permettant d'intégrer ce laboratoire pluridisciplinaire. Je tenais également à vous remercier de tous les conseils que vous avez su me prodiguer et qui ont permis d'approfondir mes travaux de recherche.*

*Merci également à Monsieur le Professeur Didier Hannequin et Monsieur le Docteur David Wallon de m'avoir fait découvrir les aspects cliniques des maladies neurodégénératives. Les différentes discussions durant les réunions d'équipe ont beaucoup contribué à renforcer mes connaissances.*

*Anne, je voulais te remercier pour tout le temps que tu m'as consacré depuis ma première année de Master. Tu as su m'amener jusqu'en thèse pour me permettre de participer à des travaux passionnants. Je te remercie également d'avoir fait preuve de patience dans l'apprentissage des différentes techniques de laboratoire et d'avoir pris du temps pour relire ma thèse.*

*Gaël, je te remercie pour tout le soutien et l'implication que tu m'as apporté au cours ma thèse. Tu as su m'aiguiller et me rediriger pour me permettre de réaliser avec efficacité ces travaux de thèse. Nos discussions m'auront constamment apporté un regard nouveau sur la génétique. J'espère sincèrement que nous serons dans le futur amené à travailler sur des projets communs.*

*Je tiens également à remercier le Professeur Bart De Strooper, le Docteur Lucía Chávez-Gutiérrez et Sarah Veugelen. Vous avez su m'accueillir et me former dans les meilleures conditions au cours de ces deux mois passés dans votre équipe à Leuven. Merci pour votre implication.*

*Je tiens à remercier sincèrement le Professeur Patrice Lerouge qui m'a permis de faire du monitorat durant ma dernière année de thèse. J'ai ainsi pu confirmer ma passion pour l'enseignement.*

*Je tiens à remercier toute l'équipe neuro de l'unité U1245 : Stéphane, merci pour ta gentillesse et ton aide lorsque je débutais dans les manip surtout quand je n'avais pas confiance en moi ! Anne-Claire, on aura passé la dernière année dans le même bureau ! Olivier et Camille, tout ce travail serait impossible sans vous, on vous doit beaucoup ! Cyril, merci de m'avoir formé durant ma première année de Master. Aux drosophilistes Laetitia, Sébastien et Magalie merci d'avoir répondu à mes questions et merci pour tous vos précieux conseils. Marine tu es la prochaine du groupe neuro, que la force soit avec toi !*

*Merci à tous les doctorants de l'équipe. Hafid, ton coaching, ta patience et ta gentillesse auront été importants durant cette thèse. Omar, tu es la gentillesse à toi tout seul, merci pour ta bonne humeur tout au long de cette thèse. Estelle, on se côtoie depuis la licence, je te souhaite une super fin de thèse ! Pierre, merci pour tous ces bons moments partagés and may the force be with you ! Sabine, tu auras su me faire rire parfois sans le vouloir ! Bon courage pour ta thèse ! Alexandre, merci pour toutes ces discussions immunologiques, footballistiques... et personnelles, promis la prochaine fois je ne tombe pas dans les pommes...*

*Merci également à tous les membres de l'Unité Inserm U1245 : Ludivine, Raphaël, Sophie, Chirine, Jean-Baptiste, Pascaline, Xana, Françoise, Myriam, Aurélie, Gaëlle et Isabelle pour toute votre aide durant cette thèse. Nathalie, merci pour ton aide quotidienne et ton efficacité.*

*Merci à ma famille : Maman, Papa, Nana, Pedro et Tom d'avoir fait semblant d'écouter mes discours passionnés sur la génétique^^ et de m'avoir toujours soutenu. Merci également à Laulau, Rémi, Mamie Poule, Papy, Claire, Christophe, Sylvie, Franck, Tata Do et Gégé. Je sais que je pourrais toujours compter sur vous.*

*Merci à Toi, l'être qui depuis plus de 3 ans maintenant partage mes jours et mes nuits et me fait vibrer quotidiennement. J'espère qu'on tremblera ensemble pendant encore de nombreuses années et que même dans nos vieux jours, après m'avoir supporté toute cette vie, tu m'aimeras toujours autant. Notre avenir à deux se construit chaque jour et ne fait que renforcer mon amour... Merci pour tout ce que tu m'apportes jour après jour ton soutien, ton amour, ta tendresse. Je t'aime...*

## Résumé

L'avènement du séquençage haut débit permet actuellement d'étudier et d'analyser la part de la composante génétique des maladies complexes médiée par les variants rares. Cependant, leur interprétation représente un défi majeur. En effet, le séquençage de milliers d'exomes et de génomes a révélé la complexité du polymorphisme humain et notamment la surreprésentation de variants rares. Et malgré le développement de logiciels d'analyse ainsi que de différentes bases de données, la priorisation des variants rares reste difficile.

Dans le cadre de cette thèse, nous avons focalisé nos analyses sur les variations génétiques rares impliquées dans la maladie d'Alzheimer (MA). D'un point de vue génétique, la MA répond dans une majorité des cas à un déterminisme multifactoriel mais une minorité des cas sont des formes précoces à transmission autosomique dominante. La caractérisation des gènes *PSEN1*, *PSEN2* et *APP* responsables des formes mendéliennes de MA a permis de formuler l'hypothèse de la cascade amyloïde en plaçant le peptide amyloïde (A $\beta$ ) au centre du processus physiopathologique.

Afin de détecter de nouveaux facteurs de risque génétique dans la survenue de la MA, nous avons réalisé une étude d'association à partir de données de séquençage d'exomes de 522 cas atteints de formes précoces de MA et 584 contrôles. Les premières analyses ont porté sur les variants mononucléotidiques ainsi que les courtes insertions/délétions et ont permis de mettre en évidence un enrichissement en variants rares prédits délétères dans le gène *ABCA7* chez les individus malades. Notre attention s'est ensuite portée sur les variations du nombre de copies (CNVs). L'absence de récurrence à l'échelle d'un gène nous a amené à travailler sur une liste de gènes. En nous focalisant sur l'hypothèse amyloïdergique, nous avons construit une liste de 342 gènes impliqués dans le métabolisme et la toxicité du peptide A $\beta$ . Grâce à cette stratégie, nous avons ainsi réussi à mettre en évidence un enrichissement de CNVs rares intersectant ce réseau centré sur le peptide A $\beta$ .

Le résultat majeur de cette étude de CNVs a été la mise en évidence d'une duplication du locus 17q21.31 chez 5 patients atteints d'une maladie neurodégénérative similaire à une maladie d'Alzheimer. Les patients porteurs présentent un diagnostic clinique de MA, des biomarqueurs et une imagerie métabolique en faveur d'une neurodégénérescence de type Alzheimer. En revanche, l'imagerie amyloïde et l'analyse neuropathologique n'ont pas révélé de pathologie amyloïde et sont donc en faveur d'une tauopathie pure.

L'étude des CNVs a également révélé une délétion partielle du gène *PSEN1*, emportant les exons 9 et 10, pour laquelle nous avons pu réaliser des études fonctionnelles. Nous avons ainsi pu déterminer que la protéine mutante favorisait la production de peptides amyloïdes plus longs, ces derniers étant des médiateurs majeurs de la neurotoxicité d'A $\beta$ .

Mots clé : séquençage haut débit, exome, études d'association, maladie d'Alzheimer du sujet jeune, *ABCA7*, CNV, *MAPT*, 17q21.31, *PSEN1*



## Abstract

Next-generation sequencing allows studying and analyzing the genetic component part of complex diseases mediated by rare variants. However, their interpretation represents a major challenge. Indeed, the sequencing of thousands of exomes and genomes revealed the human polymorphism complexity and in particular the overrepresentation of rare variants. Despite the development of softwares and variant databases, the prioritization of rare variants remains arduous.

My thesis subject was focused on the involvement of rare variants in Alzheimer's disease (AD). From a genetic point of view, AD is caused, in most cases, by a multifactorial determinism, but a minority of cases are autosomal dominant early-onset forms (ADEOAD). The characterization of mutations in the *PSEN1*, *PSEN2* and *APP* genes as a cause of these Mendelian forms of AD led to the formulation of the amyloid cascade hypothesis, stating that the amyloid- $\beta$  peptide ( $A\beta$ ) is triggering the pathophysiological process.

In order to detect new genetic risk factors involved in AD, we performed an association study using exome sequencing data from 522 cases with early-onset Alzheimer Disease and 584 controls. The first analyzes focused on single nucleotide variants and short insertions / deletions, and revealed an enrichment in cases of variants that are predicted to be deleterious in the *ABCA7* genes. We then then focused on copy number variations (CNVs). The lack of recurrence at the gene-level incited us to work on a gene list. By focusing on the amyloidogenic hypothesis, we built a list of 342 genes involved in the metabolism and toxicity of the  $A\beta$  peptide. Thanks to this strategy, we found an enrichment of rare CNVs intersecting this  $A\beta$  network in cases.

The main result of this CNV study was the identification of a duplication of the 17q21.31 locus in 5 patients with a neurodegenerative disease similar to Alzheimer's disease. These patients have a clinical diagnosis of AD, as well as biomarkers and metabolic imaging consistent with an AD neurodegeneration. However, amyloid imaging and neuropathological analysis did not reveal any amyloid pathology, and were therefore pointing to a pure tauopathy.

This CNV study also revealed a partial deletion of the *PSEN1* gene, overlapping exons 9 and 10, for which we performed functional studies. We demonstrated that the mutant protein enhanced the production of longer amyloid peptides, the latter being major mediators of  $A\beta$  neurotoxicity.

Keywords : next-generation sequencing, exome, association study, early-onset Alzheimer Disease, *ABCA7*, CNV, *MAPT*, 17q21.31, *PSEN1*



## Plan

I Introduction.....	6
A. La révolution du séquençage haut débit.....	9
1. Une révolution technologique.....	9
a. Le séquençage d'exome .....	9
b. Capture de l'exome .....	10
c. Principe de la technologie Illumina.....	10
2. Détection et annotation des variants.....	12
a. Alignement et détection des variants .....	12
b. Annotation des variants .....	13
3. Exome et variants rares.....	13
a. Les bases de données de variants et indels.....	13
b. La tolérance aux variants perte de fonction .....	14
c. Que comporte un exome ? .....	15
4. Le défi de l'interprétation des variants .....	16
a. Qualité .....	16
b. Fréquence.....	16
c. Classification des variants.....	19
d. Bases de données de variants pathogènes .....	22
e. Expressions géniques et données fonctionnelles.....	23
f. Sélection finale des variants.....	23
5. Détection de CNVs à partir de données de séquençage d'exomes .....	24
a. Les mécanismes de production des CNVs .....	24
b. Détection de CNVs.....	27
c. Principales limites des techniques de détection de CNVs à partir de données NGS.....	30
d. Les bases de données de CNVs .....	33
6. Le défi de l'interprétation des CNVs.....	35
a. Qualité .....	35
b. Fréquence.....	36
c. Classification des CNVs .....	37
d. Sélection finale des CNVs .....	38
B. Les différentes stratégies d'analyse d'exomes dans les maladies humaines .....	40
1. Stratégie intrafamiliale .....	41
2. Les analyses de trio .....	41
3. Etude de récurrence.....	43
4. Les études d'association cas-contrôles .....	43
a. Puissance contre phénotypes extrêmes.....	44

b. Agrégation des variants à l'échelle du gène.....	46
c. De l'échelle du gène à l'échelle d'une liste de gènes .....	46
d. Quel seuil de significativité pour les CNVs ?.....	48
C. Maladie d'Alzheimer .....	49
1. Epidémiologie et aspects cliniques .....	49
2. Lésions anatomopathologiques .....	51
a. Lésions macroscopiques .....	51
b. Lésions microscopiques.....	51
3. Génétique de la maladie d'Alzheimer .....	65
a. Formes mendéliennes de la MA .....	65
b. Facteurs de risque fréquents.....	68
c. Facteurs de risque rares .....	70
d. CNVs et MA.....	75
e. Cascade amyloïde .....	76
f. La part de la composante génétique manquante .....	77
II Résultats .....	79
1. Etude d'association à l'échelle du gène : Enrichissement de variants rares perte de fonction et faux-sens prédits délétères du gène <i>ABCA7</i> dans les formes précoces de maladie d'Alzheimer.....	79
2. Elaboration d'une liste de gènes centrée sur le peptide amyloïde considérant son rôle majeur dans la physiopathologie.....	81
3. Recherche de CNVs rare détectés à partir de données de séquençage d'exomes.....	82
Un enrichissement à l'échelle d'une liste de gènes .....	82
La duplication du locus 17q21.31 dans la maladie d'Alzheimer .....	83
4. Caractérisation fonctionnelle d'une nouvelle délétion des exons 9 et 10 de <i>PSEN1</i> .....	85
III. Discussion .....	87
1. Etude d'association à l'échelle d'une liste de gènes, d'un réseau, et comment prioriser certains gènes ?.....	87
2. La duplication du gène <i>MAPT</i> est-elle responsable d'une maladie d'Alzheimer ou d'une tauopathie pure ?.....	93
Annexes .....	96
Références.....	114

## Liste des figures et tableau

Figure 1 : Répartition des variants génétiques impliqués dans les maladies humaines en fonction de leur fréquence et de leur effet .....	8
Figure 2 : Le séquençage nouvelle génération par la technologie Illumina .....	11
Figure 3 : <i>Pipeline</i> bioinformatique d'analyse de données de NGS .....	13
Figure 4 : Nombre moyen de variations génétiques par individu .....	15
Figure 5 : Répartition des individus par population et comparaison des individus d'ExAC, d'ESP et 1000 génomes .....	18
Figure 6 : Représentation des réarrangements génomiques récurrents et non-récurrents .....	26
Figure 7 : Principe de la puce à SNP d'Illumina .....	28
Figure 8 : Les différentes méthodes de détection de CNVs à partir de données de NGS.....	29
Figure 9 : Principales limites des techniques de détection de CNVs à partir de données d'exomes....	31
Figure 10 : Comparaison de quatre algorithmes basés sur la méthode <i>read-depth</i> .....	32
Figure 11 : Taille moyenne des CNVs inclus dans la DGV.....	33
Figure 12 : Les différentes stratégies d'analyse d'exomes.....	40
Figure 13 : Hétérogénéité génétique de la maladie d'Alzheimer .....	50
Figure 14 : Plaques amyloïdes (à gauche) et dégénérescences neurofibrillaires (à droite) .....	51
Figure 15 : Les six isoformes de Tau .....	52
Figure 16: Progression des DNFs dans les cerveaux de malades Alzheimer .....	55
Figure 17 : Métabolisme d'APP .....	56
Figure 18 : Structure tridimensionnelle du complexe $\gamma$ -sécrétase .....	58
Figure 19 : Clivage endopeptidase et carboxypeptidase d'APP par le complexe $\gamma$ -sécrétase .....	58
Figure 20 : Trafic cellulaire d'APP.....	60
Figure 21: Mécanismes neurotoxiques du peptide amyloïde.....	64
Figure 22 : Les variations d' <i>APP</i> , <i>PSEN1</i> et <i>PSEN2</i> .....	67
Figure 23 : Répartition de l'allèle $\epsilon 4$ de l'APOE .....	68
Figure 24 : Cascade amyloïde .....	76
Figure 25 : Répartition des variants génétiques impliqués dans la maladie d'Alzheimer en fonction de leur fréquence et de leur effet .....	78
Figure 26 : Exemple de la détection d'une délétion grâce à CANOES.....	83
Figure 27 : <i>Human disease Network</i> .....	89
Figure 28 : Réseaux A $\beta$ .....	90
Tableau 1 : Gènes identifiés par les études de GWAS.....	70

## Abréviations

A2M : *α-2-macroglobulin*

ABCA7 : *ATP-Binding Cassette transporter A7*

ADN : *Acide DésoxyRiboNucléiques*

ADEOAD : *Autosomal Dominant Early-Onset Alzheimer Disease*

AICD : *APP IntraCellular Domain*

AMPA : *α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid*

APH-1 : *Anterior Pharynx defective 1*

APOE : *Apolipoprotein E*

APP : *Amyloid Precursor Protein*

APP-CTF $\alpha$  : *APP $\alpha$  C-terminal fragment*

APP-CTF $\beta$  : *APP $\beta$  C-terminal fragment*

ARN : *Acide RiboNucléiques*

BAC : *Bacterial Artificial Chromosome*

BWA : *Burrow-Wheeler Alignment*

CGH : *Comparative Genomic Hybridization*

CHRNA7 : *Cholinergic Receptor Nicotinic Alpha 7 Subunit*

CLU : *Clusterin*

CNV : *Copy Number Variation*

CTF : *C-Terminal Fragment*

dNTPs : *déoxyNucléotides Tri Phosphate*

ExAC : *Exome Aggregation Consortium*

FTDP-17 : *Frontotemporal Dementia and parkinsonism linked to chromosome 17*

GATK : *Genome Analysis Tool Kit*

gnomAD : *Genome Aggregation Database*

GWAS : *Genome Wide Association Study*

Indel : *Insertion/délétion*

LCR : *liquide céphalo-rachidien*

LTD : *Long Term Depression*

LTP : *Long Term Potentiation*

MA : *Maladie d'Alzheimer*

MAPT : *Microtubule-Associated Protein Tau*

MCI : *Mild Cognitive Impairment*

MNV : *Multiple Nucleotide Variant*

MT : *Microtubules*

NAHR : *Non-Allelic Homozygous Recombination* ; *Recombinaison homologue non-allélique*

NCT : *Nicastrin*

NGS : *Next Generation Sequencing*

NINCDS-ADRDA : *National Institute of Neurological and Communicative Disorders and Strokes – Alzheimer's Disease and Related Disorders Association*

NMDA : *N-methyl\_D-aspartate*

NTF : *Amino-Terminal Fragment*

OR : *Odds Ratio*

Pen-2 : *PSEN enhancer 2*

PPI : *Protéine-Protéine Interaction*

PrPC : *cellular prion protein*

PSEN : *Presinilin*

QMPSF : *Quantitative Multiplex of Short Fluorescent fragment*

RH : *Recombinaison Homologue*

SIFT : *Sorting Intolerant From Tolerant*

SNP : *Single nucleotide polymorphism*

SNV : *Single Nucleotide Variant*

SORL1 : *Sortilin Related Receptor 1*

TREM2 : *Triggering Receptor Expressed on Myeloid cells 2*

UTR : *Untranslated Transcribed Region*

VPS : *Vacuolar Protein Sorting*

WT : *Wild-Type*

## I Introduction

Les maladies humaines sont le résultat d'une combinaison de facteurs environnementaux et génétiques. En fonction de la maladie étudiée, la part attribuable à la composante génétique ou environnementale varie. Les maladies mendéliennes sont entièrement expliquées par la composante génétique tandis que les maladies complexes sont sous l'influence d'une part variable de facteurs génétiques et environnementaux. Le graphique adapté de Manolio et collaborateurs permet de représenter le rôle des différentes classes de variants dans la composante génétique des maladies humaines (Manolio et al., 2009) (Figure 1 page 8). Les variants sont répartis selon leur fréquence dans la population générale en fonction de leur effet sur le risque de développer la maladie. Les variants responsables de maladies mendéliennes sont extrêmement rares dans la population générale et confèrent un risque maximal de développer la maladie. Dans le cas des maladies complexes, l'effet des facteurs génétiques est multiple et deux hypothèses co-existent : variants communs – maladie commune (*common variant-common disease*) et variants rares – maladie commune (*rare variant-common disease*). Le paradigme actuel considère que les maladies complexes s'expliquent par une combinaison de variants fréquents et rares (fréquence allélique inférieure à 1%) conférant un risque faible ou élevé de développer la maladie. Les études d'association pangénomique (*Genome Wide Association Study* : GWAS) ont permis d'élucider une part de la composante génétique liée aux variants fréquents tandis que l'avènement récent des techniques de séquençage haut débit rend possible la détection et l'analyse des variants rares.

### *Variants fréquents*

L'hypothèse « *Common disease, common variant* » a émergé durant les années 2000 et considère qu'une part importante de la composante génétique des maladies complexes est médiée par des facteurs de risque génétique fréquents (Gibson, 2012). En fait cette hypothèse reflétait les connaissances de l'époque sur le polymorphisme humain, alors sous-estimé. Cette hypothèse a pu être testée grâce à des analyses pangénomiques dites de GWAS qui reposent sur des études d'association cas-contrôles. Chaque individu est génotypé sur un grand nombre de variations nucléotidiques fréquentes suffisamment espacées sur le génome pour permettre de couvrir la totalité des blocs haplotypiques. Pour chacune de ces variations, la fréquence allélique est comparée entre les cas et les contrôles pour déterminer les allèles fréquents à risque ou protecteurs. En dehors de quelques exceptions, la plupart des signaux détectés sont associés à un risque relatif faible (approché par l'odds ratio (OR)), Les variants étudiés sont souvent intergéniques et distants de plusieurs gènes. L'hypothèse est que ces variants peuvent toucher des régions régulatrices d'expression ou être en déséquilibre de liaison avec un variant génique qui altère la fonction d'un

gène. Les GWAS mettent donc en évidence des *loci*, mais il est souvent difficile de déterminer quel est le polymorphisme ou même le gène précis qui est responsable du signal.

### *Variants rares*

L'hypothèse *common disease, rare variants* propose d'expliquer la part de la composante génétique des maladies complexes à partir de variants géniques rares qui conféreraient un risque important de développer la maladie. En effet, la théorie de l'évolution suppose que des variants délétères touchant les régions codantes sont soumis à la pression de sélection et donc maintenus à basse fréquence dans la population. Cependant, certains variants rares peuvent également avoir un effet faible dans la survenue de la maladie et sont plus difficiles à étudier. Depuis le début des années 2010, le développement important des études de séquençage nouvelle génération (*Next Generation Sequencing* : NGS) a rendu possible l'exploration de cette hypothèse avec notamment le séquençage d'exome qui permet d'accéder à l'ensemble des variations codantes d'un individu.

La problématique de cette thèse a été d'identifier, à partir de données de séquençage d'exomes, de nouveaux facteurs génétiques associés à la maladie d'Alzheimer du sujet jeune. Dans l'introduction, j'aborderai dans un premier temps le défi de l'interprétation des variations génétiques détectées à partir de données de séquençage d'exomes puis les stratégies d'études génétiques en focalisant mon propos sur les études d'association. Enfin, je reviendrai sur les aspects génétiques et physiopathologiques de la maladie d'Alzheimer.

Les résultats s'articuleront autour de trois articles résultant de l'analyse d'association cas-contrôles menée au laboratoire. Cette étude d'association nous a permis de mettre en évidence (i) un enrichissement de variants rares dans le gène *ABCA7* et (ii) un excès de remaniements génomiques intersectant une liste de gènes centrée sur le peptide amyloïde chez les cas, résultat duquel ont découlé deux études. La première étude a révélé l'implication de la duplication du locus 17q21.31 chez des patients atteints de démences précoces. La seconde étude a permis d'évaluer les conséquences fonctionnelles d'une nouvelle délétion génomique de *PSEN1*.

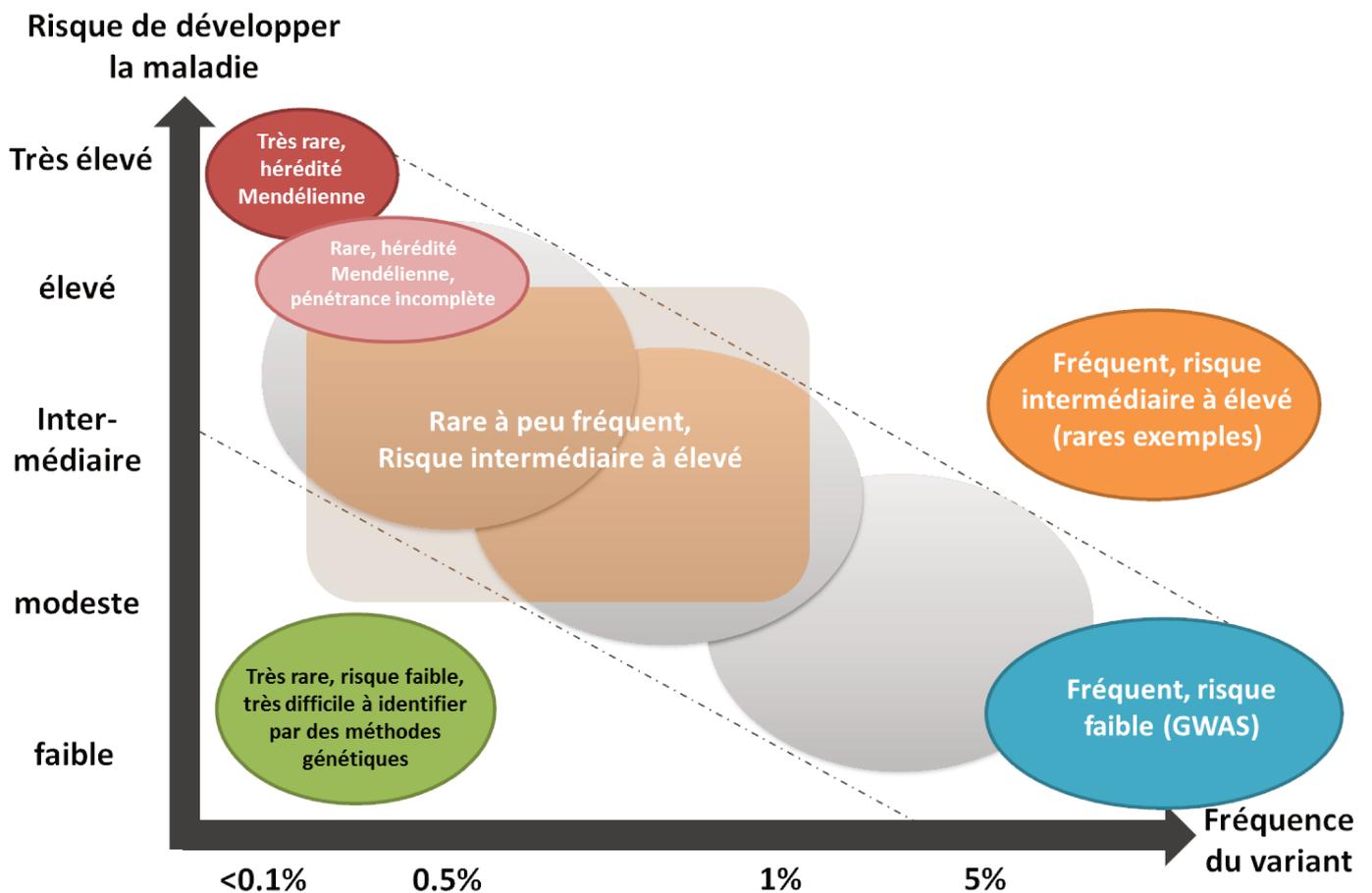


Figure 1 : Répartition des variants génétiques impliqués dans les maladies humaines en fonction de leur fréquence et de leur effet

Figure adapté de Manolio et al., 2009

Le graphe représente la répartition des variants en fonction de leur fréquence dans la population générale (en abscisse) et de leur risque sur la survenue de la maladie (en ordonnée). Les variants responsables de formes mendéliennes sont extrêmement rares et confèrent un risque maximal de développer la maladie (en haut à gauche sur le graphique). À l'autre extrémité on considère les variants fréquents qui confèrent un risque faible de développer la maladie. Exceptionnellement, certains variants fréquents peuvent également conférer un risque important de développer la maladie (exemple de l'allèle  $\epsilon 4$  de l'APOE (*Apolipoprotein E*) dans la maladie d'Alzheimer). Enfin, les variants rares confèrent soit un risque intermédiaire à élevé de développer la maladie soit un risque faible mais dans ce dernier cas ils sont plus difficiles à étudier.

## A. La révolution du séquençage haut débit

Initié en 1990 et achevé en 2004, le séquençage complet du génome humain a profondément changé notre vision de la génétique (International Human Genome Sequencing Consortium, 2004; Lander et al., 2001; Venter et al., 2001). Cette avancée majeure a servi de socle pour le développement d'analyses pangénomiques dont le séquençage à haut débit de l'exome (~1,6% du génome). S'il a fallu presque 15 ans pour obtenir le premier génome complet grâce au *shotgun sequencing* (séquençage de courts fragments subclonés), aujourd'hui quelques semaines suffisent pour accéder à l'ensemble des variations génétiques codantes d'un individu.

Cette révolution technologique a été nécessairement accompagnée d'un essor des outils bioinformatiques pour permettre l'analyse des données générées. La détection et l'annotation des variants nucléotidiques (*Single Nucleotide Variants* : SNV), des insertions/délétions et, plus récemment, des variations du nombre de copies (*Copy Number Variation* : CNV) à partir de ces données sont maintenant réalisées grâce à des logiciels bioinformatiques complexes qui nécessitent des compétences très spécifiques.

Cette avancée technologique a permis de franchir un cap en termes de connaissances des bases moléculaires de nombreuses maladies mendéliennes et complexes, mais également d'un point de vue conceptuel *via* le séquençage accru de nombreux génomes et exomes humains.

### 1. Une révolution technologique

#### a. Le séquençage d'exome

Si le séquençage Sanger suivi d'une électrophorèse capillaire est la technique de référence lorsqu'il s'agit de séquencer des régions cibles d'ADN de la taille d'un exon ou d'un gène, cette technique est néanmoins inadaptée au séquençage d'un exome (Sanger and Coulson, 1975; Sanger et al., 1977). En effet, l'exome contient 160 000 exons correspondant à environ 50 millions de bases soit 1,6% du génome humain. Le développement du séquençage à haut débit a permis de réduire considérablement le temps de séquençage d'un exome tout en présentant également un avantage économique par rapport au séquençage Sanger. Si l'on estime le prix auquel reviendrait le séquençage d'exome par la technique de Sanger, on avoisine les 5 à 30 millions d'euros pour un individu alors qu'aujourd'hui, il ne coûte que 500 à 1000 euros grâce aux technologies de séquençage haut débit (Bennett et al., 2005; Hert et al., 2008).

Le séquençage d'exome peut être réalisé grâce à différentes technologies (pour revue (Hert et al., 2008)). Dans le cadre de cette thèse, l'ensemble des exomes a été séquençé grâce au kit de capture *SureSelect* d'Agilent et à la technologie Illumina.

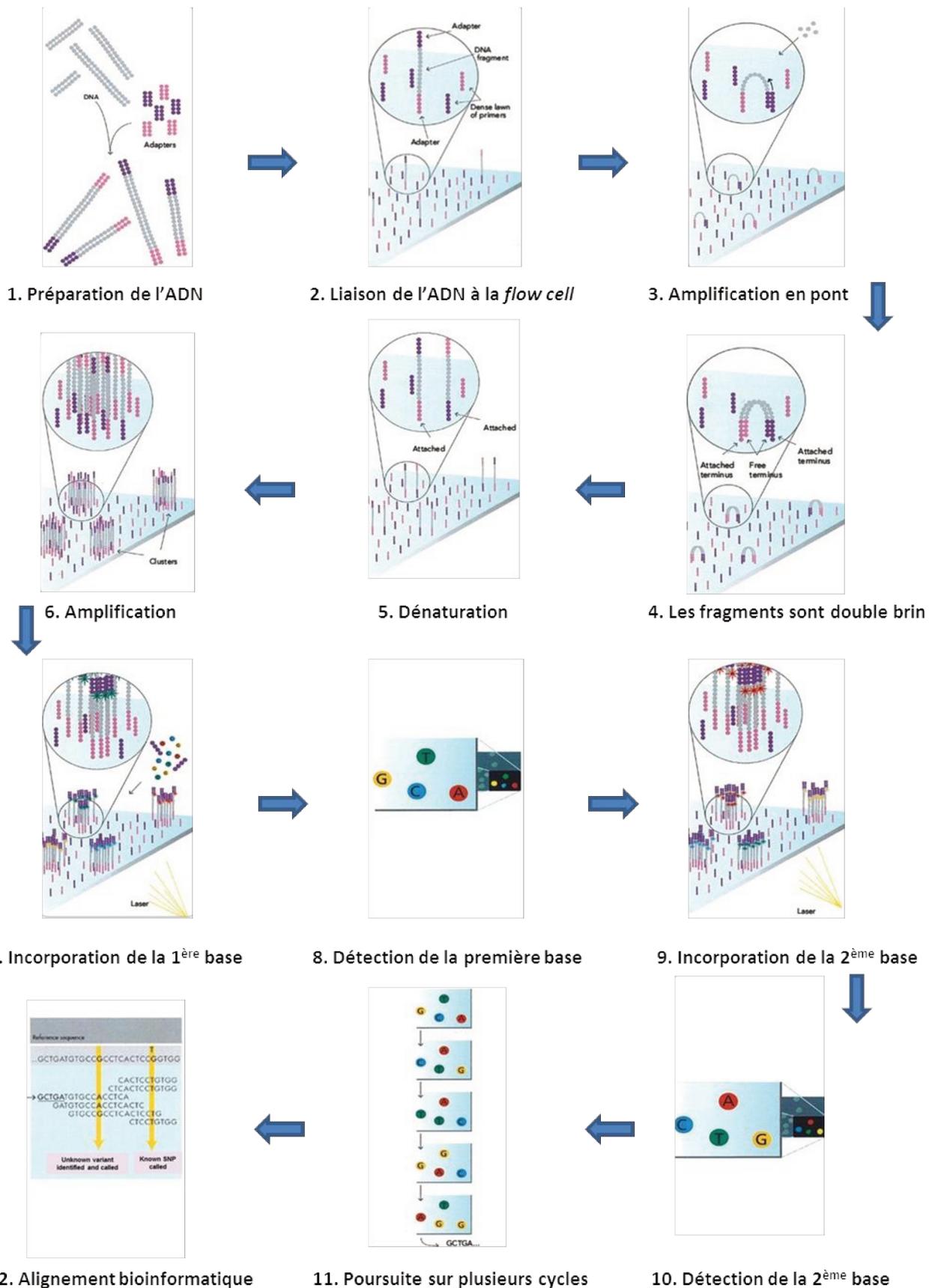
### **b. Capture de l'exome**

Dans le cas d'un séquençage ciblé, il faut pouvoir sélectionner les régions d'intérêt. En l'occurrence, dans le cas de l'exome, les régions codantes du génome. Dans notre étude, le séquençage d'exome a été réalisé à partir du kit de capture *SureSelect* d'Agilent. Cette capture est basée sur l'hybridation de molécules d'ARNs complémentaires biotinylées récupérées grâce à des billes magnétiques recouvertes de streptavidine. Les brins d'ADN sont ensuite libérés de leurs ARNs biotinylés et des billes magnétiques grâce à un champ magnétique et un lavage. Les ADN collectés servent de matrice au séquençage haut débit.

### **c. Principe de la technologie Illumina**

Cette technologie repose sur la chimie des terminateurs réversibles. Après fragmentation de l'ADN en fragments de 200 paires de bases (pbs) en moyenne par sonication, des adaptateurs sont ajoutés aux extrémités de ces fragments double brin pour permettre leur attachement sur une surface plane, appelée cellule d'écoulement (*flow cell*), sur laquelle sont liés de manière covalente des adaptateurs complémentaires. Après une amplification en groupe (*cluster*), les brins ainsi générés servent de matrice au séquençage effectué grâce à des déoxynucléotides tri phosphate (dNTPs) fluorescents réversibles. A chaque ajout de dNTPs fluorescents réversibles, une image est acquise afin de révéler quel dNTP a été incorporé. Le terminateur des dNTPs fluorescents est ensuite clivé supprimant son caractère bloquant et fluorescent pour permettre l'incorporation de la base suivante. Enfin, l'analyse bioinformatique permet de déterminer l'enchaînement des bases incorporées et par conséquent la séquence de chaque fragment d'ADN (Hert et al., 2008) (Figure 2 page 11).

Le séquençage Illumina repose sur le séquençage en extrémités appariées (*paired-end*). Dans un premier temps, chaque fragment d'ADN est séquençé aléatoirement à partir de l'extrémité sens ou antisens puis à partir de l'autre extrémité antisens ou sens respectivement. Ce séquençage *paired-end* génère donc des séquençages par paires d'extrémités d'un même fragment d'ADN.



**Figure 2 : Le séquençage nouvelle génération par la technologie Illumina**

Figure adaptée de [www.Illumina.com](http://www.Illumina.com).

La préparation de l'ADN (1) est effectuée par fragmentation de l'ADN par sonication ou digestion enzymatique puis l'ajout d'adaptateurs qui permettront leur attachement sur la *flow cell* (2). L'ADN est ensuite amplifié en pont grâce à des dNTPs non fluorescents (3). Les fragments double-brin (4) sont ensuite dénaturés (5). L'amplification continue en *cluster* (6). La première base fluorescente (dNTP fluorescent terminateur réversible) est incorporée (7) puis détectée (8). Le terminateur est ensuite clivé puis la deuxième base est incorporée et détectée (10,11) et ainsi de suite sur de multiples cycles. Les séquences générées sont ensuite alignées sur le génome de référence (12).

## 2. Détection et annotation des variants

Le séquençage d'exome ne se limite pas à la seule technologie chimique. Ce séquençage haut débit génère des données qu'il est impossible d'analyser sans une expertise bioinformatique préalable.

Après la détection des bases effectuée par le séquenceur lui-même, les séquences lues, appelées *reads*, sont alignées sur un génome de référence. Le nombre de *reads* obtenu correspond à la profondeur avec laquelle chaque base est lue et est couramment exprimée en « x ». Cette première étape d'alignement est suivie d'une détection des variants par comparaison avec ce génome de référence. La dernière étape consiste en l'annotation de ces variations permettant ainsi la priorisation ou la filtration des variants d'intérêt. L'ensemble de ces étapes correspond à un *pipeline* bioinformatique qui au cours de son processus produit une quantité importante de données (Figure 3 page 13). Ces données représentent une contrainte majeure en termes de stockage. A titre d'exemple, un exome avec une profondeur moyenne de 50x ou 100x génère respectivement 2,5 ou 5 Gigabases de données.

### a. Alignement et détection des variants

Le *pipeline* développé par les bioinformaticiens au sein du laboratoire est présenté dans la Figure 3. De manière simplifiée : après la récupération des fichiers FastQ fournis par le centre de séquençage, la première étape d'alignement a été réalisée au sein du laboratoire grâce au logiciel BWA (*Burrows-wheeler alignment tool*) (Li and Durbin, 2009) qui produit alors des fichiers BAM (*Binary Alignment Map*). La détection des variants est ensuite réalisée grâce au logiciel GATK (*Genome Analysis ToolKit*) (Van der Auwera et al., 2013). Ces deux logiciels sont aujourd'hui largement utilisés pour les analyses d'exome. Les fichiers VCF (*Variant Call Format*) ainsi produits servent de base au logiciel d'annotation des variants.

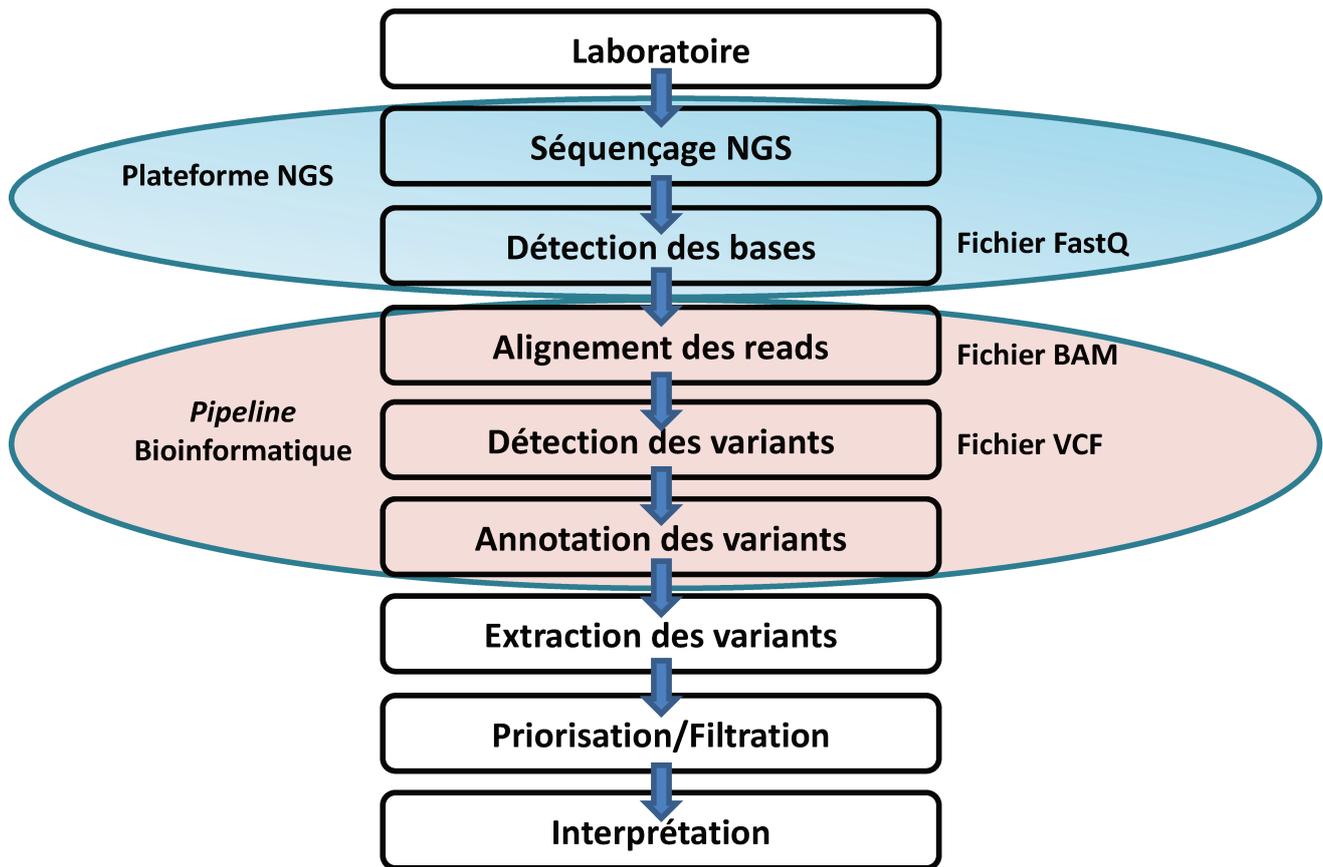


Figure 3 : Pipeline bioinformatique d'analyse de données de NGS

### b. Annotation des variants

La dernière étape bioinformatique est l'annotation des variants, elle peut être réalisée grâce à des logiciels tels que *ANNOVAR*, *ALAMUT batch* ou encore *SNPeff*. Cette étape permet de discriminer les variants en fonction de leur localisation (exoniques, introniques, dans le 5'UTR (*Untranslated Transcribed Region*) ou dans le 3'UTR) et de leur éventuelle conséquence sur la protéine (variation synonyme, faux sens, non-sens, insertion/délétion (indel) décalant le cadre de lecture (*frameshift*) ou non). L'annotation peut également permettre d'attribuer à chaque variant d'autres séries de caractéristiques : la fréquence de l'allèle mineur (*Minor Allelic Frequency* : MAF) provenant de différentes bases de données, les prédictions de pathogénicité des variants faux-sens, le score de qualité du variant défini par un score probabiliste calculé par GATK (voir paragraphe A.4.a page 16).

## 3. Exome et variants rares

### a. Les bases de données de variants et indels

L'essor et la facilité d'accès aux technologies de séquençage haut débit a permis le séquençage de centaines de milliers d'exomes (1000 Genomes Project Consortium et al., 2012; Dewey et al., 2016; Lek et al., 2016; Tennessen et al., 2012; UK10K Consortium et al., 2015). Ces données représentent

une source majeure d'information sur les variations du génome et de l'exome humain. Dans ce but, 60706 exomes ont été analysés à partir du même *pipeline* bioinformatique par l'*Exome Aggregation Consortium* (ExAC : <http://exac.broadinstitute.org/>) (Lek et al., 2016) devenant ainsi la plus large base de données d'exomes jamais recensée. Plus de 60 000 exomes et 15 000 génomes ont été récemment ajoutés à cette base de données qu'ils ont ainsi nommé gnomAD (*Genome Aggregation Database*) (<http://gnomad.broadinstitute.org/>).

A partir d'ExAC, les résultats obtenus sont édifiants. L'ensemble de ces exomes a permis d'identifier ~7,4 millions de variations distinctes, ce qui correspond à une variation toutes les 8 pbs (Lek et al., 2016). 99 % de ces variants ont une fréquence inférieure à 1 %, 54 % sont des singletons (observés chez seulement un individu) confirmant les résultats des précédentes études sur d'autres populations portant sur un nombre plus restreint d'individus telles que 1000G (1000 Genomes Project Consortium et al., 2015), UK10K (UK10K Consortium et al., 2015), NHLBI GO ESP (Tennessen et al., 2012) ou encore l'étude DiscovEHR portant sur 50 726 exomes (Dewey et al., 2016) pour ce qui est de la surreprésentation de variants rares. Il faut noter que les populations NHLBI GO ESP et 1000G sont inclus dans la population d'ExAC.

Cette proportion de variants rares a fortement changé notre regard sur le polymorphisme humain. Il était assez courant de considérer qu'une variation rare était un critère majeur dans l'interprétation et l'implication de cette variation dans une pathologie donnée. Cette surreprésentation de variants rares remet fortement en cause ce principe.

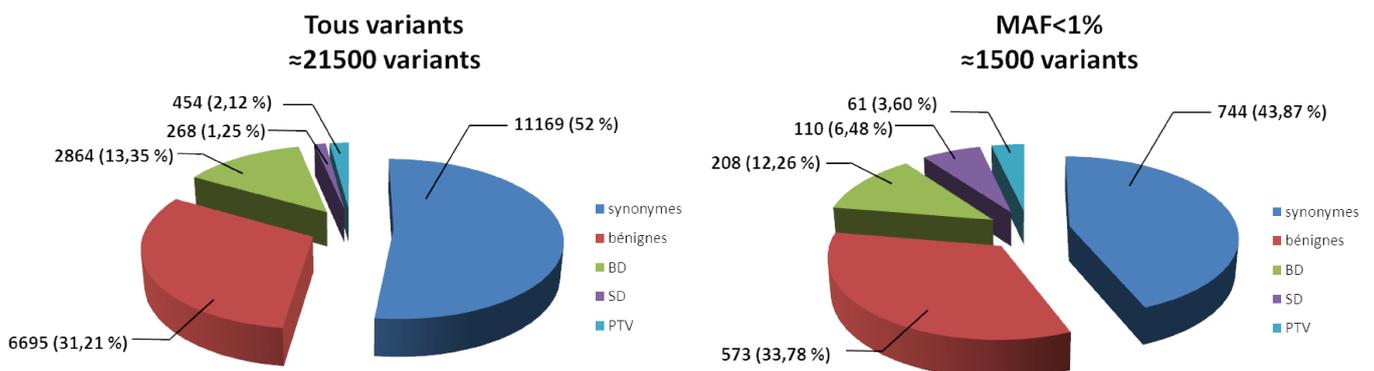
## **b. La tolérance aux variants perte de fonction**

L'interprétation des variants rares est principalement focalisée sur les catégories ou classes de variants dont l'effet sur la fonction de la protéine est plus facilement prévisible. Par conséquent, les études d'exomes ont principalement axé leurs recherches sur les variants communément appelés perte de fonction qui sont en fait des variations tronquant les protéines (*Protein Truncating Variant* : PTV) avec un effet prédit de perte de fonction de la protéine. Les PTVs regroupent les variations non-sens, les variations détruisant les sites canoniques d'épissage et les insertions/délétions (indels) décalant le cadre de lecture (*frameshift*). Une des leçons importantes à retirer de l'analyse de ces études d'exomes est la relative tolérance aux variants dits perte de fonction (Lek et al., 2016; Tennessen et al., 2012). Dans les données d'ExAC, ces variants PTVs sont les plus rares et représentent 2,4 % de l'ensemble des variants détectés. En moyenne, chaque individu présente dans la séquence codante de son génome 85 PTV à l'état hétérozygote et 35 à l'état homozygote (Lek et al., 2016). Cependant, la grande majorité des PTVs présents chez un individu sont communs (fréquence supérieure à 1 %) réduisant ainsi le nombre de PTVs rares (MAF inférieure à 1 %) à 18 par individu dont 2 variants singletons en moyenne (Lek et al., 2016).

D'après une étude réalisée par MacArthur et al. à partir d'individus sains, les gènes touchés par ces variations PTV sont relativement moins conservés au cours de l'évolution, ont plus de paralogues, impliquant donc une redondance fonctionnelle, et présentent une connectivité moins importante dans les interactions protéine-protéine ou gène-gène (MacArthur et al., 2012). De manière intéressante, ces mêmes gènes apparemment tolérants à la perte de fonction sont également plus souvent recensés comme non-haploinsuffisants (Huang et al., 2010) et n'apparaissent pas dans les listes de gènes définis comme essentiels à la survie (Hart et al., 2015). Dans le cadre d'études cas-témoins, la majorité des variants PTVs détectés est donc retrouvée indifféremment chez les individus sains ou malades. Par conséquent, le filtre sur la fréquence ainsi que sur les gènes définis comme essentiels ou haploinsuffisants peut permettre de limiter l'analyse à des variants rares à effet fort sur le risque de survenue de la maladie.

### c. Que comporte un exome ?

Selon ExAC, un individu porte en moyenne 21 451 variations exoniques (ce résultat peut varier en fonction du kit de capture utilisé) dont 1 500 avec une fréquence allélique inférieure à 1 %, environ la moitié étant des variations synonymes et l'autre moitié des variations non synonymes regroupant les PTVs et les variants faux-sens (Lek et al., 2016). Lorsque l'on recherche les bases moléculaires d'une maladie mendélienne ou complexe il faut pouvoir sélectionner et prioriser le ou les variants d'intérêt. La difficulté réside donc dans la filtration puis l'interprétation de l'ensemble de ces variations (Figure 4 page 15).



**Figure 4 : Nombre moyen de variations génétiques par individu**

Nombre de variants synonymes, faux-sens bénins, *Broadly Damaging* (BD), *Strictly Damaging* (SD) et *protein truncating variants* (PTV) retrouvées en moyenne par individu. Les variants SD sont des variants prédits pathogènes par les trois logiciels de prédiction de pathogénicité SIFT, Polyphen-2 et Mutation Taster. Les variants BD sont prédits pathogènes par au moins un des trois logiciels et les variants bénins sont prédits non-pathogènes par les trois logiciels. Les calculs ont été effectués en considérant tous les variants ou en filtrant les variants avec une fréquence allélique (MAF : *Minor Allelic Frequency*) supérieur à 1% dans une population de 1 971 individus séquencés avec les kits SureSelect V5 ou V5UTR d'Agilent. Les individus sont issus de Rouen, du French Exome project et de Lille (Equipe du Docteur Jean-Charles Lambert).

## 4. Le défi de l'interprétation des variants

Parmi les ~21 000 variations présentes, laquelle ou lesquelles sont impliquées dans le déterminisme génétique de la maladie chez le patient ? Différents filtres peuvent être appliqués pour essayer de prioriser les variants d'intérêts : en fonction de la qualité du variant détecté, de la nature de la variation (synonymes, non-sens, indels décalant le cadre de lecture ou non, faux-sens, touchant un site canonique d'épissage), de la fréquence allélique, de la conservation ainsi que des prédictions de pathogénicité *in silico* pour les variations faux-sens.

### a. Qualité

La première étape consiste à déterminer la qualité avec laquelle chaque variant est détecté. Les différents logiciels utilisés dans les *pipelines* bioinformatiques peuvent rendre compte de cette qualité grâce à des scores de qualité de génotype et de profondeur. Le logiciel GATK, utilisé au laboratoire, calcule des scores de qualité qui permettent d'estimer la probabilité qu'un variant soit vrai ou faux (Van der Auwera et al., 2013). Ce logiciel prend en compte les principaux biais rencontrés dans les analyses de données de NGS (Van der Auwera et al., 2013):

- un variant détecté dans un seul sens de lecture (sens ou antisens) est très souvent un faux-positif
- un variant uniquement présent en fin d'un *read* est également souvent un faux-positif
- la prise en compte de la qualité de l'alignement du *read* et la balance allélique par variant (basée sur la profondeur et le nombre de fois où le variant alternatif est lu par rapport à la référence)

Si les seuils définis sont trop lâches, on considérera un grand nombre de variants faux positifs, à l'inverse, des seuils trop stricts élimineront des variants réels de l'analyse. Les différents paramètres de qualité, génotype et profondeur sont généralement déterminés grâce à la confirmation par une technique indépendante (ex : le séquençage Sanger).

### b. Fréquence

Le graphique de Manolio et collaborateurs (Figure 1 page 8) permet de se rendre compte de l'implication potentielle de chaque variation en fonction de sa fréquence et de son effet, à savoir le risque de développer la maladie (Manolio et al., 2009).

La fréquence allélique est un filtre important dans l'analyse des 21 000 variations portées par chaque individu. Cette filtration peut être effectuée à partir des différentes bases de données citées précédemment (ExAC, NHLBI GO ESP ou encore 1000G) (1000 Genomes Project Consortium et al.,

2015; Lek et al., 2016; Tennessen et al., 2012). Cependant, leur utilisation peut être limitée par trois facteurs : la couverture, la qualité de chaque variant ainsi que l'origine ethnique des populations étudiées. En effet, certaines variations considérées comme rares sont simplement le reflet d'une mauvaise profondeur de séquençage et peuvent donc introduire un biais dans l'analyse. De plus, certaines variations génétiques peuvent être rares dans certaines populations quand elles sont fréquentes dans d'autres (Lek et al., 2016; Tennessen et al., 2012). Ainsi, des biais de stratification de population peuvent altérer l'interprétation des variants lorsqu'ils sont sélectionnés sur leur fréquence. Un moyen de contrer ce biais est de disposer d'une base de données de témoins strictement appariés aux patients étudiés pour l'origine ethnique.

Dans la base de données ExAC, différents groupes ethniques sont individualisés (Figure 5 page 18). La sous-population ExAC NFE inclut les individus Européens non-Finlandais et peut être potentiellement utilisée pour filtrer les variations fréquentes lorsque les études sont menées sur des échantillons européens.

### a. Répartition individus d'ExAC par population

Population	Hommes	Femmes	Total
Africain/ Afro-Américain	1 888	3 315	5 203
Latino-américain	2 254	3 535	5 789
Est Asiatique	2 016	2 311	4 327
Finlandais	2 084	1 223	3 307
Européen non Finlandais	18 740	14 630	33 370
Sud Asiatique	6 387	1 869	8 256
Autre	275	179	454
<b>Total</b>	<b>33 644</b>	<b>27 062</b>	<b>60 706</b>

### b. Comparaison des bases de données de séquençage haut débit

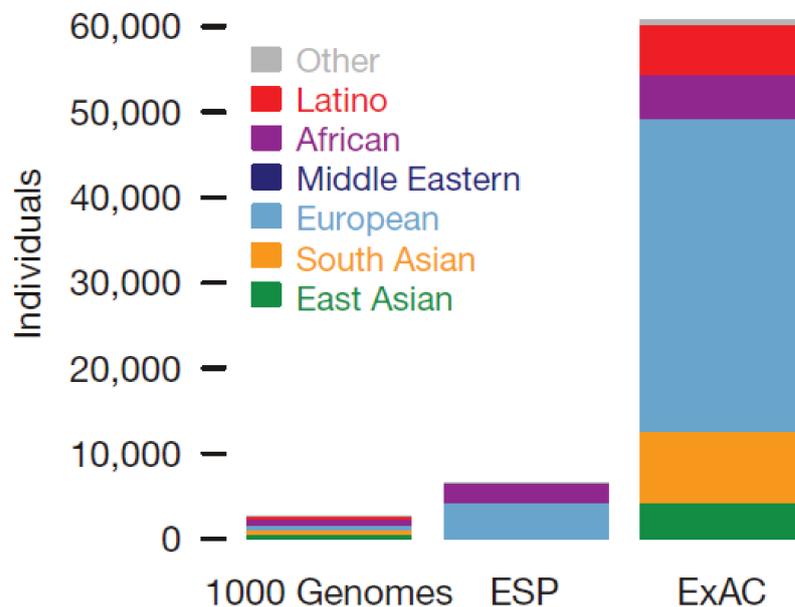


Figure 5 : Répartition des individus par population et comparaison des individus d'ExAC, d'ESP et 1000 génomes

Figure issue de Lek *et al.*, 2016.

Les ~60 000 individus ayant fait l'objet d'un séquençage d'exome dans l'étude d'ExAC ont été distingués en fonction de leur origine ethnique. On distingue 6 populations différentes : Les Africains et Afro-Américains, les Latino-Américains, les Asiatiques de l'Est, les Finlandais, les Européens non-Finlandais et les Asiatiques du Sud.

### c. Classification des variants

Outre la filtration des variants à partir de la fréquence, les variations nucléotidiques sont également classifiées selon l'effet prédit sur la protéine associée au gène. Dans les études d'exomes, on considère 5 classes de variants : les variations synonymes, les variations dites perte de fonction (non-sens, les indels décalant le cadre de lecture, variants situés sur les sites canoniques d'épissage), les variants non-synonymes (faux-sens, indels ne décalant pas le cadre de lecture), les variations introniques en dehors des sites canoniques d'épissage et les variations situées dans les régions non-traduites (3'UTR et 5'UTR).

#### *Les variations entraînant une perte de fonction de la protéine : non-sens, indels, affectant les sites canoniques d'épissage*

De par leur effet tronquant sur la protéine, ces variations sont considérées en priorité dans les études d'exomes. L'annotation simple rend compte du type de la variation, à savoir s'il s'agit d'une variation non-sens, indels *frameshift* ou d'une variation d'épissage. Cependant, les études récentes tentent de distinguer ces variants en excluant les variants PTV selon différents critères (Lek et al., 2016; MacArthur et al., 2012):

- lorsque l'allèle alternatif est en réalité l'allèle ancestral
- lorsque le variant est situé au début de la séquence codante : on suppose que des codons initiateurs alternatifs permettront de sauvegarder la fonction du gène
- lorsque le variant est situé à la fin de la séquence codante : on suppose que le gène sera transcrit et traduit pour produire une protéine fonctionnelle raccourcie
- lorsque le variant est situé en +2 ou -2 d'un exon mais l'exon concerné n'a pas de sites canoniques d'épissage

#### *Les variations faux-sens et leur prédiction de pathogénicité*

Les variations faux-sens sont plus difficiles d'interprétation car il est complexe de prédire leur effet sur la protéine. Il est donc important de pouvoir distinguer les variations faux-sens qui ont un effet fonctionnel sur la protéine de celles qui n'en ont pas. Pour cela, différents logiciels de prédiction de pathogénicité des variations faux-sens existent et leurs algorithmes sont basés sur plusieurs critères dont les principaux sont la conservation au cours de l'évolution de la base nucléotidique ainsi que de l'acide aminé (AA) étudié mais également la distance physico-chimique entre l'AA original et son remplaçant (Adzhubei et al., 2013; Ng and Henikoff, 2003; Schwarz et al., 2010). Ainsi, chaque logiciel définit une prédiction de pathogénicité pour chaque variant et le classifie dans différentes catégories prédites comme plus ou moins délétères pour la protéine.

Au sein du laboratoire, nous avons basé nos études sur les trois logiciels les plus utilisés (Castellana and Mazza, 2013; Richards et al., 2015) : Mutation Taster (Schwarz et al., 2014), SIFT (*Sorting Intolerant From Tolerant*) (Ng and Henikoff, 2003) et Polyphen-2 (Adzhubei et al., 2013).

### Mutation Taster

Accessible en ligne (<http://www.mutationtaster.org/>), cet algorithme se veut le plus complet car il définit un score pour tous les types de variants, synonymes ou non-synonymes. L'algorithme repose sur la conservation nucléotidique et protéique, les altérations des sites d'épissage, l'altération de certaines fonctions protéiques et les changements pouvant influencer la quantité d'ARNm produit. De plus, les altérations génétiques retrouvées plus de 4 fois dans les 1000G (1000 Genomes Project Consortium et al., 2012) ou dans HapMap (International HapMap 3 Consortium et al., 2010) sont automatiquement considérées comme neutres (*neutral*) et les variants rapportés dans la base de données ClinVar (recense les variations en relation avec les pathologies humaines) sont automatiquement prédits comme causant des maladies (*disease causing*). Mutation Taster a été entraîné sur les variants considérés comme polymorphiques à partir du projet 1000 génomes (1000 Genomes Project Consortium et al., 2012) et l'ensemble des variations considérées comme causales dans la base de données *HGMD Professionnal*.

### Polyphen-2

Polyphen-2 (<http://genetics.bwh.harvard.edu/pph2/>) définit 3 classes de variations faux-sens : bénignes (*benign*), possiblement délétères (*possibly damaging*) et probablement délétères (*probably damaging*) (Adzhubei et al., 2013). Son score est principalement défini grâce à différentes fonctions (11 en tout) basées sur la comparaison avec des homologues de séquences nucléotidiques et protéiques et sur des données de propriétés structurales des protéines. Le logiciel a été entraîné sur deux paires de jeux de données différents : HumDiv et HumVar. HumDiv est constituée d'allèles délétères, annotés dans la base de données UniProtKB comme causant des maladies mendéliennes, et d'allèles non-délétères, basés sur les différences observées entre les protéines humaines et leurs homologues mammifères proches (conservation faible). HumVar est constituée de l'ensemble des mutations causant des maladies humaines annotées par la base de données UniProtKB et des variations non-synonymes n'ayant jamais été impliquées dans les maladies humaines. Dans l'étude des maladies complexes, il est recommandé d'utiliser Polyphen-2 entraîné par HumDiv car des allèles mêmes faiblement délétères doivent être considérés comme délétères. Polyphen-2 entraîné par HumVar est plus adapté aux maladies mendéliennes qui requièrent une distinction forte entre les allèles délétères et les allèles faiblement ou modérément délétères.

## SIFT

Ce logiciel de prédiction est également disponible en ligne (<http://sift.bii.a-star.edu.sg/>). Cet algorithme repose uniquement sur la conservation protéique calculée à partir des homologies de séquences protéiques et sur la distance physico-chimique entre les deux acides aminés, sauvage et mutant. Il distingue ainsi deux types de variants : délétères (*Deleterious*) ou tolérés (*Tolerated*). SIFT a été entraîné à partir de bases de données de variants : SwissProt (variants causant des maladies mendéliennes) d'une part, et dbSNP (Ng and Henikoff, 2002) ainsi que des variants non-synonymes détectés chez des contrôles (Cargill et al., 1999) d'autre part.

## Les autres outils de prédiction de pathogénicité des variants faux-sens

De nombreux autres logiciels de prédiction de pathogénicité des variants sont régulièrement développés. Les derniers en date génèrent des scores dits « intégrés » en combinant différents logiciels de prédiction tels que Condel (González-Pérez and López-Bigas, 2011), CADD (Kircher et al., 2014) ou encore KGGSeq (Li et al., 2012b). Ces logiciels varient évidemment en fonction des combinaisons de logiciels utilisés. De plus, les développeurs de ces algorithmes testent leur logiciel à partir de jeux de données distinguant les variants dits « vrais variants pathogènes » des « vrais variants neutres ». Les jeux de données utilisés ne sont pas les mêmes entre les différentes études entraînant *in fine* une grande variabilité des prédictions de pathogénicité entre les différents logiciels. Différentes études ont tenté de comparer certains logiciels (Castellana and Mazza, 2013; Dong et al., 2015; Genovese et al., 2016; Liu et al., 2013; Thusberg et al., 2011) : les principales discordances proviennent des jeux de données « vrais variants pathogènes, vrais variants neutres » sur lesquels ont été entraînés les logiciels. En plus d'avoir utilisé des jeux de données différents, les bases de données sont polluées par des variants annotés à tort comme délétères entraînant nécessairement un biais lors des phases d'entraînement des différents algorithmes.

## Utilisation des logiciels de prédiction de pathogénicité

Grâce à notre expérience locale et de manière logique avec les comparaisons interlogicielles, nous avons décidé d'utiliser Mutation Taster, SIFT et Polyphen-2 qui sont complémentaires et les plus adaptés pour nos études. Nous avons défini plusieurs classes de variants faux-sens en fonction des prédictions *in silico* des différents logiciels :

- variations faux-sens strictement pathogènes : prédites pathogènes par les trois logiciels
- variations faux-sens probablement pathogènes : prédites pathogènes par au moins un des trois logiciels
- variations faux-sens bénignes : prédites non-pathogènes par les trois logiciels

### **Analyse des variants synonymes, introniques et faux-sens bénins**

Les variants synonymes, introniques ne touchant pas les sites canoniques d'épissage ou encore les variants faux-sens prédits bénins n'ont *a priori* pas d'effet sur la structure et la fonction de la protéine. Ces différents variants peuvent malgré tout avoir un effet sur l'épissage en altérant à distance la fonction des sites canoniques d'épissage ou des éléments régulateurs (région exonique/intronique promotrice ou inhibitrice d'épissage, *Exonic or Intronic Splicing Enhancer or Silencer*). Ces altérations peuvent avoir différentes conséquences entraînant le saut d'un ou plusieurs exons, la rétention d'une partie d'un intron ou encore l'utilisation de site donneur et accepteur d'épissage différent etc. Toutes ces variations peuvent notamment conduire à la formation d'une protéine tronquée. Ces variations considérées initialement comme bénignes peuvent donc être à l'origine de modifications importantes dans le rôle physiologique de la protéine. Différents logiciels de prédiction d'épissage existent (huit sont comparés dans la référence (Jian et al., 2014)). Cependant ces logiciels restent encore peu utilisés dans les analyses d'exomes. Leur développement est actuellement en cours et certaines recommandations existent mais ne s'appliquent dans la plupart des cas qu'à des gènes spécifiques (par exemple pour les gènes *BRCA1* et *BRCA2* (Houdayer et al., 2012)).

### **d. Bases de données de variants pathogènes**

Différentes bases de données recensent les variants impliqués dans les maladies mendéliennes. Deux types de bases de données de variants pathogènes existent : les bases de données spécifiques de gènes et celles non spécifiques.

Les bases de données spécifiques de gènes sont souvent mieux annotées que les bases de données non spécifiques car elles ont été établies par des spécialistes du ou des gènes d'intérêts. Dans le cas de la maladie d'Alzheimer à transmission autosomique dominante, les bases de données de mutations AD&FTD (<http://www.molgen.ua.ac.be/admutations/>) et d'Alzforum (<http://www.alzforum.org/mutations>) sont considérées comme étant fiables.

Les bases de données non-spécifiques telles que *HGMD professional*, *ClinVar*, *OMIM* et *SwissProt* sont souvent polluées de variants pour lesquels l'annotation n'est pas fiable.

Les auteurs du consortium ExAC ont tenté d'évaluer la pathogénicité des variants rapportés dans les bases de données *Clinvar* ou *HGMD* grâce aux ~60 000 exomes d'individus indemnes de maladies pédiatriques sévères. De manière intéressante, 22 765 variants annotés comme pathogènes dans ces bases de données ont également été détectés dans les exomes d'ExAC. Parmi ces 22 765 variants, 844 avaient une fréquence allélique supérieure à 1 % dans au moins une des populations incluses dans ExAC (Figure 5 page 18). En moyenne, les individus d'ExAC portent ~54 variants précédemment rapportés comme causaux de maladies mendéliennes mais ~47 d'entre eux ont une fréquence

allélique supérieure à 1 % dans au moins une des populations d'ExAC. Cette haute fréquence allélique est difficilement compatible avec une maladie mendélienne même dans un mode de transmission récessif. Lorsque finalement on filtre les variants rapportés comme pathogènes sur leur fréquence, il reste environ 6 variations rares (< 1%) annotées comme pathogènes par *Clinvar* ou *HGMD* correspondant à 0,89 variants pathogènes impliquant des gènes responsables de maladies à transmission autosomique dominante (Lek et al., 2016). Cela suggère finalement que la quasi-totalité des individus d'ExAC ont une maladie mendélienne à transmission dominante, ce qui est fortement improbable et soulève le problème de la pollution de ces bases par des variants non-causaux.

### e. Expressions géniques et données fonctionnelles

Différentes bases de données permettent de déterminer dans quels tissus le gène d'intérêt est exprimé. Le *Human Protein Atlas* (<http://www.proteinatlas.org/>) est très complet et les données sont basées sur l'expression d'ARNm et de protéines de différents tissus. De manière plus spécifique, *Allen Brain Atlas* (<http://brain-map.org/>) fournit des données d'expression d'ARN sur différents cerveaux humains en distinguant les différentes régions du cerveau. Ces bases de données reposent principalement sur des données de *microarrays*, d'hybridation *in situ* et sur l'hybridation d'anticorps spécifiques sur coupes histologiques.

### f. Sélection finale des variants

Lors de nos analyses d'exomes des patients Alzheimer, les différents variants de bonne qualité sont sélectionnés à partir de différents filtres :

- Une MAF inférieure à 1% dans notre échantillon de ~1000 patients et contrôles ou dans la base de données ExAC NFE qui considère uniquement la sous-catégorie des Européens non Finlandais pour éviter d'éventuel biais de stratification
- Un taux de données manquantes sur la base nucléotidique analysée inférieur à 5 % déterminé à partir de la fonction de *variant calling multiple* de GATK. Cette fonction permet de déterminer si la base nucléotidique d'intérêt a été correctement séquencée chez l'ensemble des individus
- Exclusion des variants dans les UTRs et des variants introniques en dehors des sites canoniques d'épissage
- Interrogation des bases de données non spécifiques *HGMD*, *ClinVar* et *Swissprot* dans le cas de diagnostic différentiel mais également spécifiques, *Alzforum* et *AD&FTD*
- Dans le cas spécifique de la maladie d'Alzheimer (MA), les variants non référencés d'*APP*, *PSEN1* et *PSEN2* sont interprétés grâce aux recommandations de l'algorithme de Guerreiro (Guerreiro et al., 2010) par les neurologues

Les patients porteurs de variants classifiés comme responsables de maladie d'Alzheimer ou d'autres maladies neurodégénérative font l'objet d'un compte rendu diagnostique et sont retirés de l'analyse finale. Au sein du laboratoire, les individus atteints de formes précoces de MA (<65 ans) et qui ont un apparenté d'une autre génération atteint de la MA jeune ainsi que les cas sporadiques avec un âge de début très précoce ( $\leq 50$  ans) font l'objet préalable d'un séquençage Sanger d'*APP*, *PSEN1* et *PSEN2*. Ainsi, 77% des cas familiaux précoces et 17,9% des cas sporadiques très jeunes étaient porteurs d'un variant causal (Nicolas et al., 2016a; Wallon et al., 2012). Les individus non porteurs de mutations causales et les cas ne répondant pas aux critères précédents font l'objet d'un séquençage d'exome. De façon notable, parmi les 264 individus qui n'ont pas fait l'objet d'un séquençage Sanger d'*APP*, *PSEN1* et *PSEN2*, uniquement 4 individus étaient porteurs d'un variant soit causal, soit possiblement ou probablement causal (1,5 %) (Nicolas et al., 2016a).

## 5. Détection de CNVs à partir de données de séquençage d'exomes

La variabilité du génome humain ne se limite pas aux SNV et aux insertions/délétions (indels) de petites tailles, les variations du nombre de copies (CNVs) représentent également une source importante de polymorphismes (Alkan et al., 2011). De manière arbitraire, les CNVs ont été définis comme des variations structurales dont la taille est supérieure à 50 pbs (MacDonald et al., 2014; Zarrei et al., 2015). Idéalement, la caractérisation complète d'un CNV repose sur trois principaux critères : le nombre de copies, la région touchée par le réarrangement et enfin les informations de positions (Hastings et al., 2009). Les CNVs varient en fonction de leur taille et proviennent de différents types de variations structurales : les délétions, les translocations, les inversions, les éléments mobiles, les duplications et les insertions nouvelles. Leur détection au niveau pangénomique a longtemps reposé sur les techniques de *microarrays* dont la résolution dépend de la densité de sondes. Ces dernières années, l'amélioration de la qualité du séquençage haut débit et le développement d'algorithmes spécifiques ont permis de détecter des CNVs de très petites tailles à partir de données de séquençage d'exomes ou de génomes (Hehir-Kwa et al., 2015).

### a. Les mécanismes de production des CNVs

Les variations structurales appartiennent à deux catégories différentes : les variations récurrentes et non récurrentes (Carvalho and Lupski, 2016). Les réarrangements récurrents sont caractérisés par le fait qu'ils ont la même taille et affectent la même région génomique chez deux individus non apparentés. Les réarrangements non-récurrents quant à eux ont une taille différente et un contenu différent à un même locus chez deux individus non-apparentés (Figure 6 page 26) (Carvalho and Lupski, 2016).

La caractérisation des points de cassures des variations du nombre de copies permet d'appréhender les mécanismes de formation des variations structurales.

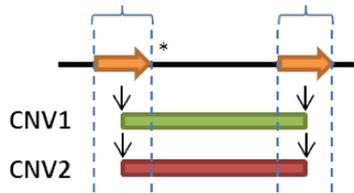
La plupart des points de cassures des CNVs récurrents sont retrouvées dans des zones d'au minimum 1kb présentant une forte homologie de séquence (plus de 95%) appelées les *Low Copy repeats* (LCR) ou duplications segmentaires. Les duplications segmentaires sont à l'origine du principal mécanisme de recombinaison entraînant la formation de CNVs, à savoir la recombinaison homologue non-allélique ou NAHR (*non-allelic homologous recombination*) (Hastings et al., 2009).

La recombinaison homologue (RH) survient principalement au cours de la méiose et assure trois principales fonctions : la réparation des cassures de l'ADN, la ségrégation correcte des chromosomes et la diversité de l'espèce par échange d'ADN. La RH s'appuie principalement sur les homologies de séquences et de préférence sur celles existant entre les deux chromatides sœurs. Contrairement à la recombinaison homologue qui a lieu entre deux chromatides soeurs, la NAHR utilise comme homologues de séquence les zones de duplications segmentaires qui peuvent donc être situées sur des chromosomes différents (Carvalho and Lupski, 2016). Par conséquent, si la RH permet de réaliser des transferts équilibrés il n'en est pas de même pour la NAHR qui a la possibilité de changer la structure des chromosomes et donc d'entraîner la formation de CNVs. A titre d'exemple, les haplotypes H1 et H2 du locus 17q21.31 sont bordés de zones de duplication segmentaire entraînant une sensibilité de cette région aux réarrangements récurrents médiés par la NAHR. La délétion de ce locus est notamment responsable du syndrome de Koolen–de Vries (Koolen et al., 2006).

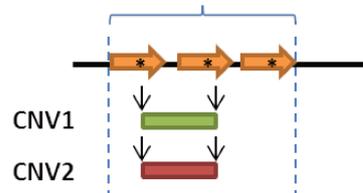
Les CNVs non-récurrents présentent des points de cassure à la fois sur des *Low Copy Repeat* mais également sur des zones courtes (de 2 à 15 pbs) et de faible homologie, voire même sans aucune homologie. Le mécanisme n'est donc pas toujours dépendant d'une homologie de séquence (Carvalho and Lupski, 2016; Hastings et al., 2009). La génération de ces CNVs dépend de plusieurs mécanismes dont la ligature non-homologue des extrémités, la ligature par microhomologie des extrémités, le cycle cassure-fusion-pont, le glissement de réplication, le décrochage de la fourche de réplication ou encore la réparation des cassures induites par la réplication grâce à des microhomologies (pour revue (Hastings et al., 2009)).

## A. Réarrangements récurrents

### Aa. Unités répétées intercalées

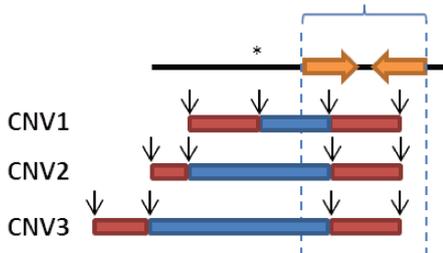


### Ab. Unités répétées en tandem

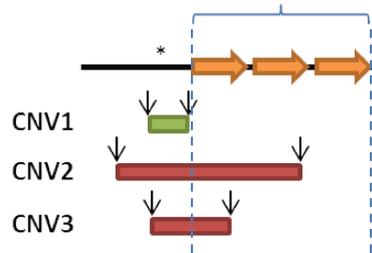


## B. Réarrangements non-récurrents

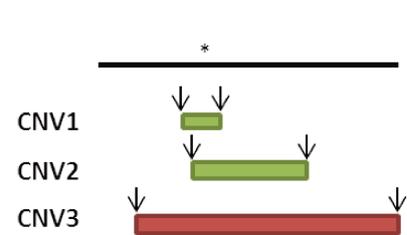
### Ba. DUP-TRP/INV-DUP



### Bb. DUP et DEL



### Bc. DUP et DEL sans implication des LCR



**Figure 6 : Représentation des réarrangements génomiques récurrents et non-récurrents**

Figure adaptée de Carvalho *and* Lupski, 2016.

Les lignes noires représentent les segments génomiques d'un locus donné. Les gènes sont représentés par des astérisques. Les LCR sont représentés par des flèches orange. Les CNVs sont représentés en rouge pour les duplications, en bleu pour les triplications et en vert pour les délétions. Les points de cassures sont indiqués par des flèches verticales. Les crochets déterminent les régions comportant les groupes de points de cassure. A. Les remaniements récurrents ont la même taille et la même composition génomique chez des individus non-apparentés. Aa. Les points de cassures des duplications et délétions se situent dans les régions hautement homologues (duplication segmentaire, *Low copy repeat* (LCR)) et sont médiés par la NAHR affectant les gènes situés entre ces deux zones. Ab. Les remaniements récurrents médiés par des zones répétées en tandem affectant les gènes situés sur ces régions. B. Les remaniements non-récurrents ont une taille et une composition génomique unique entre deux individus non-apparentés. Ba. Les Duplication-Triplication inversé-Duplication (DUP-TRP/INV-DUP) sont des évènements non-récurrents. Bb. Les points de cassures sont localisés en de multiples positions dans des régions répétées. Bc. Certains locus sont soumis à des réarrangements différents sans zones communes de points de cassure.

## b. Détection de CNVs

### *Les techniques de puces CGH (Comparative Genomic Hybridization) et puces à SNP*

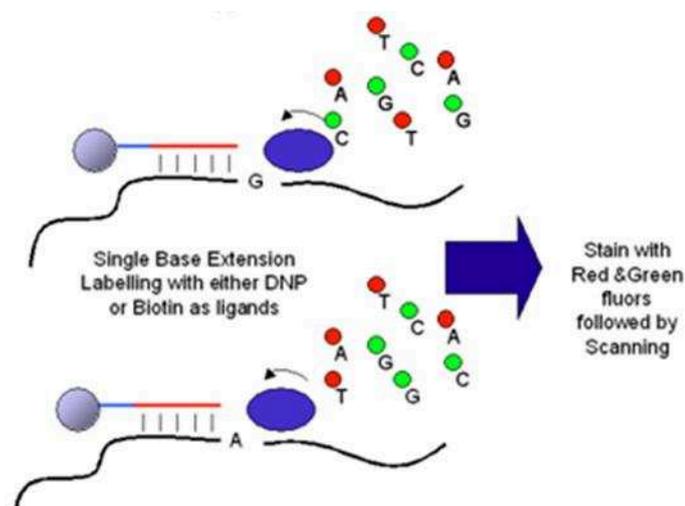
#### Les puces à CGH

Après la FISH (*fluorescence in situ hybridization*), les premières analyses des variations de structure génomique ont été réalisées grâce aux techniques de puces à ADN (Iafrate et al., 2004; Redon et al., 2006; Sebat et al., 2004). Ces puces à ADN sont des petits supports solides sur lesquels sont fixés des séquences d'ADN, appelées sondes, spécifiques d'une séquence du génome humain. Le principe de la puce CGH consiste en l'hybridation compétitive de l'ADN d'un individu à tester et d'un ADN témoin ou d'un pool d'ADN témoin tous les deux marqués par des fluorophores différents (Pinkel and Albertson, 2005). La lecture de la fluorescence par un scanner permet ensuite de déterminer les régions génétiques déséquilibrées chez l'individu testé. Les puces étaient initialement composées de sondes de 40 à 200 kbs, les BACs (*Bacterial Artificial Chromosome*), permettant de couvrir l'intégralité du génome humain mais restreignant fortement la précision avec laquelle les points de cassures étaient détectés. La technique de CGH a constamment évolué en développant des sondes de plus petites tailles (oligomères de 60 nucléotides pour la puce Agilent) afin d'augmenter la résolution mais également en augmentant le nombre de sondes sur la lame (1 million pour la puce Agilent) (Pinkel and Albertson, 2005). La meilleure résolution obtenue à ce jour est la détection de CNVs d'un minimum de 450 pb grâce à des puces à CGH contenant 2,1 millions sondes oligonucléotidiques espacées en moyenne de 56 pbs (Conrad et al., 2010). Cependant, les sondes oligonucléotidiques ne couvrent pas les régions répétées du génome humain car leur alignement sur le génome de référence serait impossible.

#### Les puces à SNP (*Single nucleotide polymorphism*)

Ces puces à ADN ont été initialement conçues pour étudier les SNPs. Cette technique permet de discriminer des variations nucléotidiques en hybridant des fragments d'ADN d'un même individu sur des sondes spécifiques d'un polymorphisme (pour revue LaFramboise, 2009). Par conséquent, la comparaison ne se fait pas entre deux individus comme c'est le cas pour la CGH, mais entre les deux allèles d'un même individu. A titre d'exemple, la technologie Illumina repose sur l'hybridation des fragments d'ADN sur des billes portant les sondes d'un locus spécifique. Ces sondes sont composées de la séquence nucléotidique précédant le polymorphisme d'intérêt. Le fragment d'ADN hybridé sur la sonde sert de matrice à la polymérase qui ajoute un dNTP fluorescent à l'extrémité de la sonde donc au niveau du polymorphisme étudié. La détection de fluorescence permet d'établir si l'individu est homozygote ou hétérozygote pour ce locus (Figure 7 page 28). Bien qu'ayant été initialement

développées pour l'analyse des SNPs, ces puces peuvent également permettre de détecter des CNVs (LaFramboise, 2009). Pour cela, il suffit de comparer les intensités de fluorescences entre deux individus afin d'évaluer le nombre de copies associé au SNP étudié. De la même manière que la technique de CGH, cette technique est limitée par le nombre de sondes utilisées mais permet de couvrir les régions répétées du génome. En pratique, les puces généralement utilisées pour ces études comprennent jusqu'à 1 million de sondes.



**Figure 7 : Principe de la puce à SNP d'Illumina**

Figure issue de <http://www.ipc.nxgenomics.org/newsletter/no8.htm>

Après fragmentation, l'ADN s'hybride sur les sondes fixées à des billes. Une ADN polymérase permet l'extension de ces sondes en prenant comme brin matrice les fragments hybridés. La détection de fluorescence de la base incorporée permettra de définir le statut homozygote ou hétérozygote.

### CNVs détectés à partir de données de séquençage haut débit

Les algorithmes développés pour la détection de variations structurales à partir de données de NGS ne sont pas tous basés sur les mêmes méthodes. On distingue quatre types d'approche : *read-pair* discordant, *read-depth*, *split-read* et l'assemblage *de novo* (Figure 8 page 29) (pour revue (Hehir-Kwa et al., 2015)).

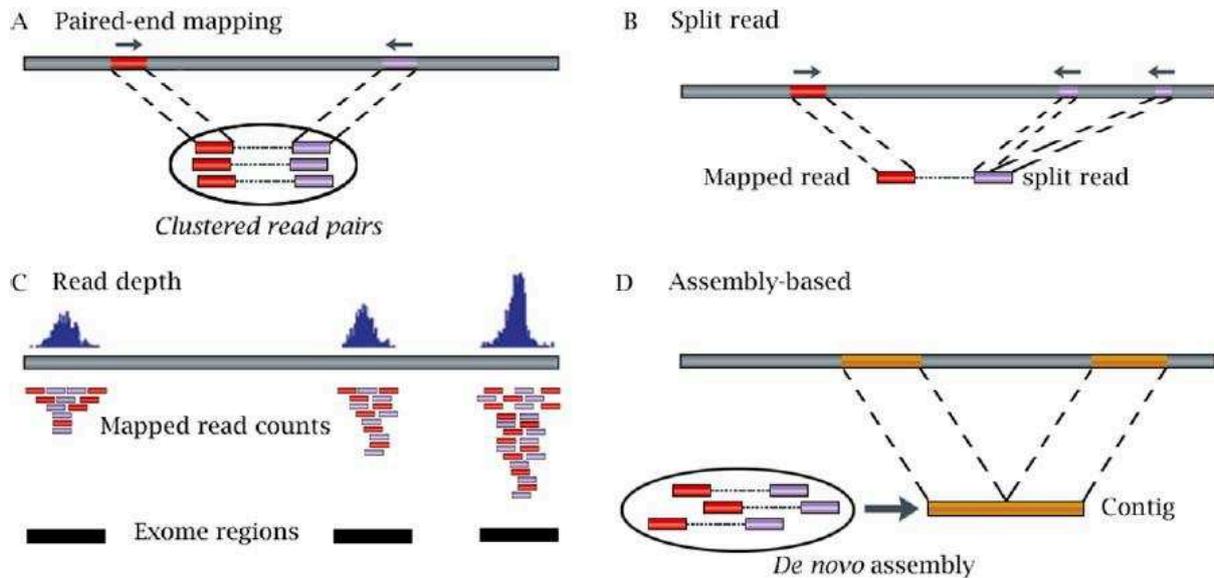


Figure 8 : Les différentes méthodes de détection de CNVs à partir de données de NGS  
Figure issue de la revue de Zhao *et al.*, 2013.

### Read-pair discordant

Cette méthode, également appelée *Paired-end mapping*, repose sur l'alignement des séquences évoquées précédemment (voir paragraphe A.1.c) et notamment sur l'alignement discordant de ces séquences par rapport au génome de référence. Cette technique permet de détecter toutes les variations de structures possibles. En effet, si les *reads* obtenus à partir du même fragment d'ADN en sens et en antisens s'alignent sur le génome de référence de manière trop éloignés, ceci indique la présence d'une délétion chez l'individu séquençé (Zhao et al., 2013). *A contrario*, les *paired-end* qui se retrouvent rapprochés lors de leur alignement signent la présence d'une insertion. De la même manière, des incohérences dans l'orientation des *reads* peuvent permettre de délimiter les inversions ainsi que certaines duplications en tandem (Alkan et al., 2011). De plus, les paires de *reads* pour lesquelles seul l'un des *reads* a été aligné sur le génome de référence servent à détecter de nouvelles insertions.

### **Split-read**

Comme son nom l'indique, cette approche utilise les *reads* dont l'alignement sur le génome de référence peut se faire sur deux régions physiquement éloignées (Figure 8 page 29). Ces *reads* sont généralement retirés de l'étape d'alignement sur le génome de référence car il s'aligne sur deux régions différentes. La technique *split-read* consiste à récupérer ces *reads* car on suppose que la discordance provient d'un réarrangement génomique. En somme, le *read* contiendrait les points de cassures de la variation structurale. En théorie, cette méthode permet également comme la technique *paired-end* de détecter l'ensemble des variations structurales possibles.

### **Assemblage de novo**

Cette approche consiste en l'assemblage *de novo* sans alignement sur le génome de référence de l'ensemble des *reads* séquencés (Figure 8 page 29). Ce nouveau génome ainsi créé sera ensuite comparé au génome de référence afin d'en détecter les discordances.

### **Read-depth**

Cette dernière approche est basée sur l'idée simple que les régions dupliquées auront un nombre de *reads* plus important et les régions délétées en auront moins (Figure 8 page 29). Cette méthode compare les divergences de distribution des profondeurs de séquençage chez un individu par rapport à un modèle aléatoire défini par le logiciel utilisé.

## **c. Principales limites des techniques de détection de CNVs à partir de données NGS**

La caractérisation d'un remaniement repose sur trois critères : le nombre de copie, la région touchée et les informations de position. Ces différentes stratégies ont chacune leurs avantages et leurs inconvénients mais aucune n'est parfaite (Figure 9 page 31) (Hehir-Kwa et al., 2015). Si la méthode *read-pairs* permet de détecter tous les types de variations structurales (inversions et translocations comprises), elle reste limitée par la longueur des *reads* et est donc restreinte à la détection de CNVs de petites tailles. La technique *split-read* est très précise concernant la localisation des points de cassures mais est également limitée, pour la même raison que la technique *read-pair*, aux petits CNVs. En théorie, l'assemblage *de novo* est la technique la plus résolutive pour détecter les variations structurales. Cependant, sa mise en application n'est pas évidente notamment à cause des longues séquences répétées présentes dans le génome pour lesquelles l'assemblage est plus difficile surtout dans le cas où les *reads* sont courts (Chaisson et al., 2015). L'assemblage *de novo* serait plus efficace si les *reads* obtenus étaient suffisamment longs pour permettre de séquencer les régions adjacentes à ces régions répétées. Cependant, ces techniques de séquençage de longs *read* sont encore peu développées dans les laboratoires à cause de leur coût et de leur complexité technologique. Enfin,

l'approche *read-depth* permet de détecter le nombre absolu de copies du CNV avec le plus de précision mais ne permet généralement pas de définir les points de cassure et se limite seulement aux duplications et délétions.

De nombreux algorithmes ont également été conçus en combinant ces différentes méthodes afin de bénéficier de l'ensemble des avantages de chaque méthode comme MGP-HMM ou LUMPY à titre d'exemple (Layer et al., 2014; Malekpour et al., 2016). L'algorithme LUMPY combine les techniques de *split-read*, *paired-end* et *read-depth* tandis que MGP-HMM intègre uniquement les techniques *read-depth* et *paired end*.

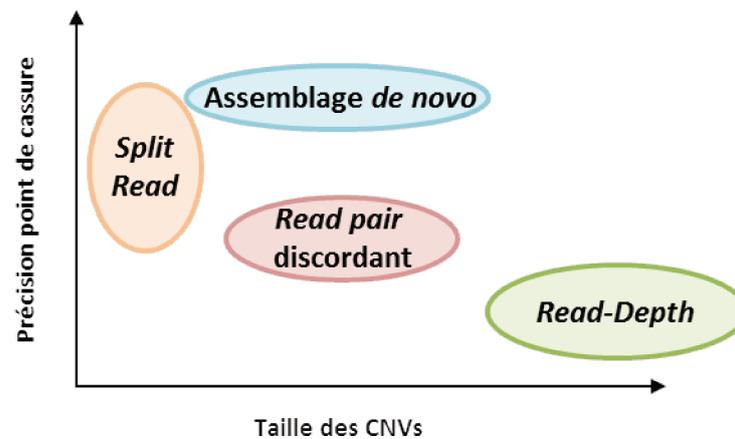
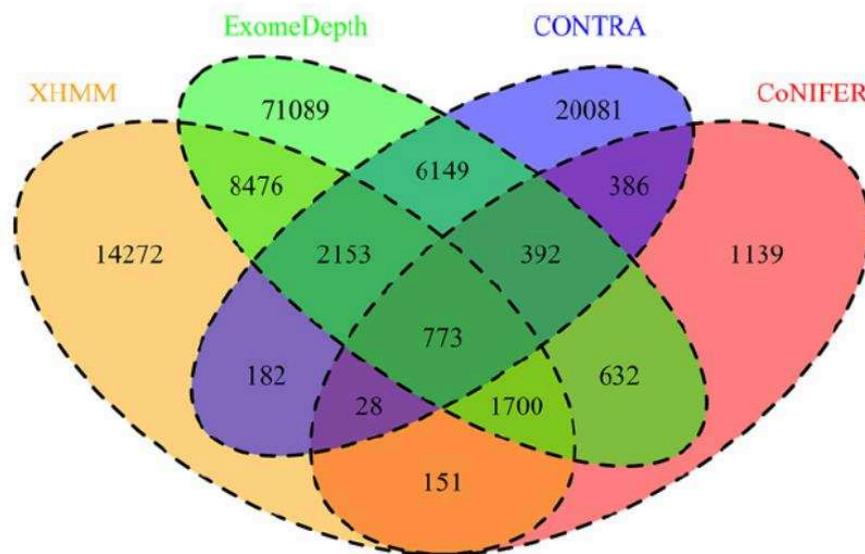


Figure 9 : Principales limites des techniques de détection de CNVs à partir de données d'exomes  
Figure adaptée de Hehir-Kwa et al., 2015

Ces quatre techniques sont adaptées au séquençage de génome. Cependant, la méthode *read-depth* est la seule utilisable lors de l'analyse de données d'exomes. En effet, cette méthode ne requiert pas des données continues, contrairement à l'assemblage *de novo*, et ne repose pas sur la détection des points de cassure des CNVs, contrairement à l'approche *split-read*. Lors du séquençage d'exome, les données obtenues sont fragmentées ne ciblant que les différents exons. De plus les points de cassures sont très généralement retrouvés dans les séquences intergéniques ou introniques, séquences qui ne sont donc pas capturées lors d'un séquençage d'exome. Par conséquent, la méthode *read-depth* ne permet de détecter que les duplications ou délétions. Il est impossible de détecter des variations structurales plus complexes telles que les inversions.

Au cours des dernières années, de nombreux algorithmes basés sur la méthode de *read-depth* ont été développés (Backenroth et al., 2014; Fromer et al., 2012; Krumm et al., 2012; Plagnol et al., 2012; Samarakoon et al., 2016). De nombreuses études ont également tenté d'évaluer quel logiciel obtenait la meilleur sensibilité et spécificité (Guo et al., 2013; Kadalayil et al., 2015; Nam et al., 2016; Tan et al., 2014). Dans l'étude de Tan et collaborateurs, les auteurs ont comparé 4 logiciels différents basés sur la méthode *read-depth*. Les quatre logiciels présentent des discordances entre eux comme

le prouve le diagramme de Venn (Figure 10 page 32). Les auteurs ont ainsi pu démontrer que les logiciels XHMM (Fromer et al., 2012) et CoNIFER (Krumm et al., 2012) présentait une meilleure spécificité mais une moins bonne sensibilité que ExomeDepth (Plagnol et al., 2012) et CONTRA (Li et al., 2012a). Toutes les études de comparaison de CNVs s'accordent à dire que les algorithmes développés présentent tous leurs avantages et leurs inconvénients en termes de spécificité et de sensibilité et qu'il est nécessaire de développer des algorithmes plus robustes (Guo et al., 2013; Kadalayil et al., 2015; Nam et al., 2016; Tan et al., 2014).



**Figure 10 : Comparaison de quatre algorithmes basés sur la méthode *read-depth***

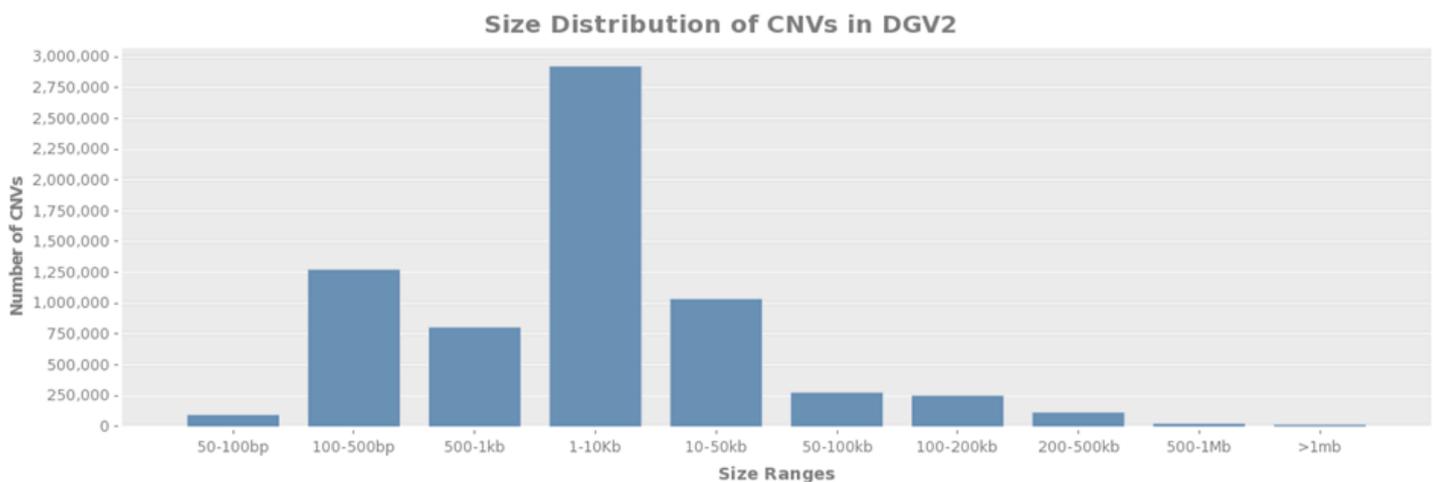
Figure issue Tan et al., 2014

Diagramme de Venn décrivant le nombre de CNV à l'échelle de l'exon détecté en commun par 4 logiciels suivant : ExomeDepth, CONTRA, CoNIFER et XHMM. Les données ont été obtenues à partir de 32 individus

#### d. Les bases de données de CNVs

##### *Database of Genomic Variants (DGV)*

Une étude récente a tenté d'établir une carte des CNVs du génome humain à partir des données de la DGV (*Database of Genomic Variants*) qui regroupent 55 études différentes (MacDonald et al., 2014) (<http://dgv.tcag.ca/dgv/app/home>). Parmi ces 55 études, les auteurs ont conservé celles qu'ils considéraient comme étant les plus résolutes. Les études de FISH, de puces à ADN de basse résolution (puces de moins de 1 million de sondes) ont donc été écartées de l'analyse. Au final, 26 études de haute résolution ont été sélectionnées dont des études de CGH-array, SNP-array des études à partir de données de NGS (Zarrei et al., 2015). Cette carte des CNVs du génome humain a permis de recenser à partir de 2 647 individus plus de 2 millions de CNVs dont plus de 90% sont des délétions. La majorité des CNVs détectés ont une taille située entre 1 et 10kb ce qui correspond aux limites de détection des technologies utilisées (Figure 11 page 33).



**Figure 11 : Taille moyenne des CNVs inclus dans la DGV**

Figure issue de la base de données DGV (<http://dgv.tcag.ca/dgv/app/home>)

##### *Les CNVs détectés à partir de la population d'ExAC*

Les techniques de NGS permettent de détecter des remaniements de plus petite taille, bien inférieure aux plus résolutes des études de *microarrays* (Hehir-Kwa et al., 2015; Zhao et al., 2013). La plus grande étude réalisée à partir de données d'exomes est celle de la base de données ExAC regroupant près de 60 000 individus (Ruderfer et al., 2016). Les CNVs ont été détectés à partir du logiciel XHMM (Fromer et al., 2012). Cet algorithme élimine systématiquement les effets des individus présentant un nombre de CNVs aberrants dû notamment à un mauvais séquençage d'exome, les effets de batch causés par les biais de technique de séquençage haut débit, ou les effets de cibles, à savoir lorsque celles-ci sont riches en GC car ces régions sont amplifiées de manière variable. Dans l'étude d'ExAC, ce système d'élimination est réalisé à partir des 219 347 cibles de

l'exome et des 60 642 individus. Un modèle statistique dit de Markov caché permet de détecter les CNVs à partir du modèle créé sur l'ensemble de ces individus. Les données finales recensent 126 771 CNVs exoniques touchant 19 430 gènes autosomiques retrouvés chez 59 898 individus correspondant en moyenne à 2,1 CNVs rares (inférieur à 1 %) par individus incluant 0,82 délétion et 1,29 duplication. Dans les données d'ExAC, la définition d'un CNV rare est assez imprécise car la fréquence de 0,5 % utilisée comme filtre prend en compte plusieurs paramètres (Ruderfer et al., 2016). Deux CNVs se chevauchant d'au moins 50 % sont considérés comme identiques et les gènes contenant au moins un CNV fréquent sont retirés de l'analyse. Après filtration, 15 734 gènes ont été analysés. Au final le nombre de CNVs exoniques détectés par individu est biaisé. Il faut également noter que ces études ne permettent pas de détecter les variations structurales plus complexes telles que des inversions et ne permettent pas non plus de définir de manière précise les positions exactes des CNVs détectés.

Ces techniques de détection à partir de données de NGS sont encore en développement et la spécificité et sensibilité de ces techniques restent encore imparfaites. Afin d'évaluer la qualité de détection du logiciel XHMM dans l'étude d'ExAC, 241 trios de patients schizophrènes (Fromer et al., 2014; Kirov et al., 2012) ont été analysés pour confirmer que le taux de CNV *de novo* et la transmission parent-enfant était conforme aux règles de ségrégation mendélienne. Brièvement, une mauvaise sensibilité résulterait d'une baisse de la transmission parent-enfant attendue et une baisse de la spécificité serait déterminée par une augmentation des variations survenues *de novo*.

Pour donner plus de poids à la qualité des CNVs détectés, un ensemble de 10 091 individus inclus dans cette étude d'ExAC avait aussi fait l'objet d'une analyse grâce à des puces à SNP de haute résolution (700 000-1 000 000 de sondes) (Szatkiewicz et al., 2014). La comparaison entre ces deux techniques à partir du même panel d'individus a permis de mettre en évidence les avantages et les inconvénients de ces deux méthodes. Les principaux résultats ont permis de déterminer que la détection de CNVs à partir d'exomes permettait d'identifier 2,2 fois plus de CNVs que les techniques de *microarrays* (certainement dû aux limites de résolution en termes de taille de ces techniques) mais que les CNVs détectés à partir de peu de cibles par *microarrays* étaient plus difficilement détectables à partir des données d'exomes (Ruderfer et al., 2016).

Ces deux techniques, *microarrays* et exomes, sont imparfaites mais semblent complémentaires. Les techniques de puces permettent de couvrir les régions géniques et intergéniques contrairement à la détection à partir de données d'exomes. Elles permettent de détecter à la fois des CNVs rares et fréquents. Cependant, les puces sont très souvent limitées à la détection de CNVs de taille supérieure à 5-10 kb contrairement aux technologies de NGS. A l'inverse, le séquençage d'exome permet de

mettre en évidence des remaniements de très petite taille. En revanche, les logiciels d'analyse tels que XHMM sont basés sur la comparaison des profondeurs de lecture entre un individu donné et un panel de référence comprenant 30 à 50 sujets. Ceci limite donc la sensibilité à la des remaniements fréquents. C'est pour cette raison que les données d'ExAC sont restreintes aux CNVs rares (< 1%). Enfin, les données d'exomes et de puces ne permettent de détecter que des duplications ou délétions contrairement aux données de génomes. Le principal avantage de la détection des CNVs à partir de données de séquençage NGS est qu'à partir des mêmes données la détection à la fois des variants et des CNVs est réalisable réduisant ainsi le coût et le temps de l'analyse. Les études de CNVs identifiés à partir de puces ou de séquençage de génomes ont permis d'établir une carte des CNVs polymorphiques du génome humain (1000 Genomes Project Consortium et al., 2015; Zarrei et al., 2015).

Finalement, les CNVs exoniques rares sont moins nombreux par individu par rapport aux variants et indels rares (2,1 vs 1 500) mais affectent un plus grand nombre de bases nucléotidiques. Cette rareté rend les CNVs particulièrement intéressants lors de leur découverte chez des individus malades (Conrad et al., 2010; Redon et al., 2006; Zarrei et al., 2015). Cependant, la majorité de ces CNVs touchant l'exome sont relativement grands, 73 kb en moyenne d'après ExAC (Ruderfer et al., 2016) et n'emportent pas qu'un seul gène. Dans ce contexte, déterminer la pathogénicité d'un remaniement est difficile.

## 6. Le défi de l'interprétation des CNVs

Afin de déterminer l'implication d'un CNV dans une maladie spécifique, plusieurs critères sont à considérer : la qualité avec lequel le CNV est détecté, sa fréquence en population générale et les prédictions au niveau protéique de ces variations structurales

### a. Qualité

Les différents logiciels de détection de CNV sont encore en développement et leur spécificité et sensibilité sont encore relativement imparfaites complexifiant la détection exhaustive de l'ensemble des remaniements par ces techniques (Hehir-Kwa et al., 2015). De plus les principales études pour déterminer la spécificité et la sensibilité des différents logiciels (Guo et al., 2013; Kadalayil et al., 2015; Nam et al., 2016; Tan et al., 2014), ont été réalisées en comparaison avec des techniques de *microarrays*, qui nous le savons permettent de détecter des variations structurales de tailles supérieures à 5-10kb, ou avec des logiciels de détection adaptés au séquençage de génomes qui sont également en cours de développement. Leur vraie sensibilité et spécificité reste donc encore à déterminer. Par conséquent, l'ensemble des variations structurales détectées doit être dans un

second temps validé par une technique indépendante. Ceci permet de déterminer la valeur prédictive positive d'un logiciel mais ne permet pas d'évaluer le taux de faux-négatifs.

## **b. Fréquence**

La DGV rapporte la fréquence des CNVs détectés. En effet, elle recense de nombreuses études permettant d'identifier les remaniements polymorphiques (MacDonald et al., 2014). Cependant, la détection d'un CNV peut être imprécise car les points de cassures de ces remaniements sont souvent mal localisés à cause des limites de résolution des technologies utilisées. La DGV peut ainsi considérer que deux CNVs sont différents alors qu'ils sont identiques, simplement à cause d'un positionnement imprécis des points de cassure. De plus, de nombreux CNVs sont apparus grâce aux mécanismes de recombinaison homologue (paragraphe A.5.a) au niveau de zones de duplications segmentaires. Ces séquences hautement homologues ne sont pas classiquement capturées par les techniques de *microarrays* car il est très difficile, de par leur forte homologie, de distinguer une copie de l'autre et de les positionner correctement sur le génome. Leur implication majeure dans le mécanisme de formation des CNVs compromet l'identification correcte de nombreux CNVs identiques.

Pour essayer de contourner ce problème, des régions appelées CNV *regions* (CNVR) ont été délimitées. Ce sont des *loci* dans lesquels de nombreux CNVs sont retrouvés (Zarrei et al., 2015). Bien que ces régions soient fortement touchées, tous les CNVs inclus ne sont pas obligatoirement identiques et ne concernent pas nécessairement les mêmes gènes. L'effort de sélection des CNVs identifiés à partir de technologies hautement résolutive (Zarrei et al., 2015) a permis de limiter ce phénomène et de cartographier au mieux les CNVs fréquents du génome humain.

Il est donc difficile de comparer nos données de CNVs détectés à partir de données de séquençage d'exomes avec des données de *microarrays* en termes de fréquence. Cependant, les analyses sont possibles lorsque l'on compare les CNVs de taille supérieure à environ 5-10 kb (limite de résolution des puces) et lorsque ceux-ci impliquent des portions suffisamment importantes de régions codantes (limite de l'exome).

Au sein du laboratoire, la stratégie a été de définir la fréquence d'une variation structurale à partir de 1106 individus participant à l'étude d'association cas-contrôles et en les analysant avec le même logiciel bioinformatique considérant ainsi que le mauvais positionnement des points de cassure est identique sur l'ensemble des CNVs détectés. Les bornes des différents CNVs étaient finalement définies par les exons les plus proches des bornes de cassure. À partir du moment où les mêmes régions codantes étaient emportées et dans le même sens (duplication ou délétion), nous considérons ces événements comme identiques pour nous permettre d'établir une fréquence des CNVs détectés au sein de notre échantillon. Nous avons également éliminé de l'analyse les gènes pour lesquels la fréquence des CNVs était supérieure à 1 %. Les gènes situés sur des zones de

duplication segmentaire ont également été retirés de l'analyse car les logiciels de détection rencontrent des difficultés pour distinguer ces CNVs fréquents entraînant un nombre de faux-positifs important.

### c. Classification des CNVs

#### *CNVs et prédictions*

La prédiction au niveau fonctionnel de l'effet d'un CNV n'est pas simple et l'interprétation des délétions est souvent plus évidente que celle des duplications. La délétion complète d'un gène seul amène à penser que le mécanisme physiopathologique en cause est associé à une perte de fonction. L'implication de variants PTV de ce gène dans la maladie étudiée ou lorsque les données fonctionnelles convergent vers une hypothèse de perte de fonction sont autant d'arguments supplémentaires. Cependant, cette interprétation devient plus difficile si la zone délétée emporte plusieurs gènes. A titre d'exemple, la délétion du locus 17q21.31, initialement détecté en 2006, emporte 4 gènes : *MAPT*, *CRHR1*, *STH* et *KANSL1* et est responsable du syndrome de Koolen-de Vries caractérisé par un retard du développement, une hypotonie du jeune enfant, une dysmorphie faciale et un comportement amical caractéristique (Shaw-Smith et al., 2006). Ce n'est qu'en 2012 que deux études ont rapporté des mutations tronquantes du gène *KANSL1* chez d'autres patients présentant un phénotype similaire. Cette perte de fonction du gène *KANSL1* permet ainsi d'expliquer la majeure partie du phénotype associé à ce syndrome (Koolen et al., 2012; Zollino et al., 2012).

De la même manière, une délétion emportant un à quelques exons d'un gène n'entraînant pas un décalage du cadre de lecture n'est pas nécessairement associée à une perte de fonction d'un gène comme le démontre la délétion de l'exon 9 de *PSEN1* dans la maladie d'Alzheimer qui aboutit à une dysfonction du complexe  $\gamma$ -sécrétase (Chávez-Gutiérrez et al., 2012).

Les duplications sont encore plus difficiles d'interprétation. En effet, si la duplication concerne le gène en entier, on peut supposer qu'elle entraînera une augmentation d'expression de ce gène. C'est le cas de la duplication d'*APP* dans la maladie d'Alzheimer associée à une angiopathie amyloïde qui entraîne effectivement une augmentation d'expression de l'ARNm d'*APP* (Pottier et al., 2012a; Rovelet-Lecrux et al., 2006). Cependant, si la duplication ne concerne pas la région promotrice elle ne constituera pas un gain de fonction. Enfin, les duplications partielles de gènes sont particulièrement compliquées à analyser car à partir des données d'exomes il est impossible de définir le sens de la duplication (tandem, inversion, ailleurs sur le génome) et donc de prédire les conséquences précises sur la séquence codante du gène.

Des travaux récents réalisés par Chiang et collaborateurs ont apporté un éclairage en ce qui concerne l'effet des CNVs sur l'expression génique en combinant le séquençage de génome et l'analyse de

séquençage d'ARN dans 13 tissus différents chez 147 individus (Chiang et al., 2017). Environ 500 remaniements fréquents touchant les régions codantes ont été analysées. Ces remaniements semblent entraîner, de manière cohérente, une augmentation ou une diminution d'expression des gènes selon le type de variation, duplication ou délétion respectivement. Les études des CNVs rares ont également permis d'établir que les duplications entraînaient une augmentation d'expression et que les délétions entraînaient une diminution d'expression. Cependant, certains remaniements ne semblent pas respecter cette règle (Chiang et al., 2017; Schlattl et al., 2011). Concernant les variations intergéniques, l'effet observé est plus aléatoire car on suppose que les zones touchées comportent des éléments régulateurs qui peuvent être activateurs ou inhibiteurs de l'expression des gènes.

### ***Bases de données de CNVs pathogènes***

De même que pour les variants nucléotidiques, il existe des bases de données de CNVs non-spécifiques et spécifiques de pathologie. Les bases de données spécifiques concernant la maladie d'Alzheimer sont les mêmes que précédemment (AD&FTD et Alzforum). Concernant les bases de données non spécifiques de CNVs pathogènes, la base de données DECIPHER (*Database of genomic variation and Phenotype in Humans using Ensembl Resources* ; <https://decipher.sanger.ac.uk/>)(Firth et al., 2009) est la plus complète et attribue une pathogénicité pour chaque CNV collecté à partir de 246 projets différents regroupant ainsi près de 40 000 patients différents. Pour chaque CNV détecté, la base de données évalue le niveau de certitude avec lequel elle considère la variation comme pathogène et renvoie à la littérature sur laquelle elle s'est appuyée. Ces bases de données peuvent être une première approche pour évaluer le rôle d'un CNV dans une pathologie.

En conclusion, les CNVs sont moins nombreux que les variants et indels, représentent une portion plus importante du génome humain mais sont d'interprétation difficile. Dans les études futures, les techniques de RNAseq seront nécessaires pour élucider l'effet de ces remaniements sur les aspects qualitatifs et quantitatifs de la transcription des gènes.

### **d. Sélection finale des CNVs**

Dans notre étude, nous avons utilisé le logiciel CANOES dont l'algorithme s'appuie sur la méthode *read-depth* (Backenroth et al., 2014). CANOES utilise un modèle statistique de Markov caché. Le logiciel crée un modèle différent pour chaque individu à partir d'un groupe d'individus séquencés. La comparaison des profils de profondeur et de couverture par rapport à ce modèle permet de détecter les CNVs rares. En effet, les CNVs fréquents seront majoritairement éliminés grâce au modèle créé. Comparé à XHMM ou CoNIFER, CANOES a créé son modèle de référence pour permettre une

meilleure détection des remaniements génomiques de petites tailles situés sur les zones de faibles profondeurs de séquençage. Les CNVs ont été sélectionnés de la manière suivante :

- Filtre qualité sur les individus porteurs d'un nombre trop important de CNVs reflétant une mauvaise qualité de séquençage d'exome
- Filtration des CNVs lorsque la fréquence était supérieure à 1 % à l'échelle d'un remaniement ou d'un gène chez les 1106 cas et individus contrôles séquencés dans cette étude
- Filtration des gènes situés sur les zones de duplications segmentaires
- Validation par une technique de PCR semi-quantitative des CNVs détectés

Les patients porteurs d'une duplication d'*APP* ou d'une délétion de l'exon 9 de *PSEN1*, responsables de formes mendéliennes de maladie d'Alzheimer, sont rendus au titre du diagnostic. L'ensemble des variations structurales restantes ont fait l'objet de nos recherches.

## B. Les différentes stratégies d'analyse d'exomes dans les maladies humaines

Dans notre étude, la sélection des variations nucléotidiques, indels et variations structurales a permis de restreindre l'analyse à des variations rares à effet potentiellement fort dans le déterminisme génétique de la maladie. Chaque individu portait en moyenne 61 variations PTVs, 991 faux-sens et 6,4 CNVs rares. En fonction de la maladie étudiée, différentes stratégies permettent de prioriser les variants d'intérêt afin de restreindre le nombre de variants à étudier.

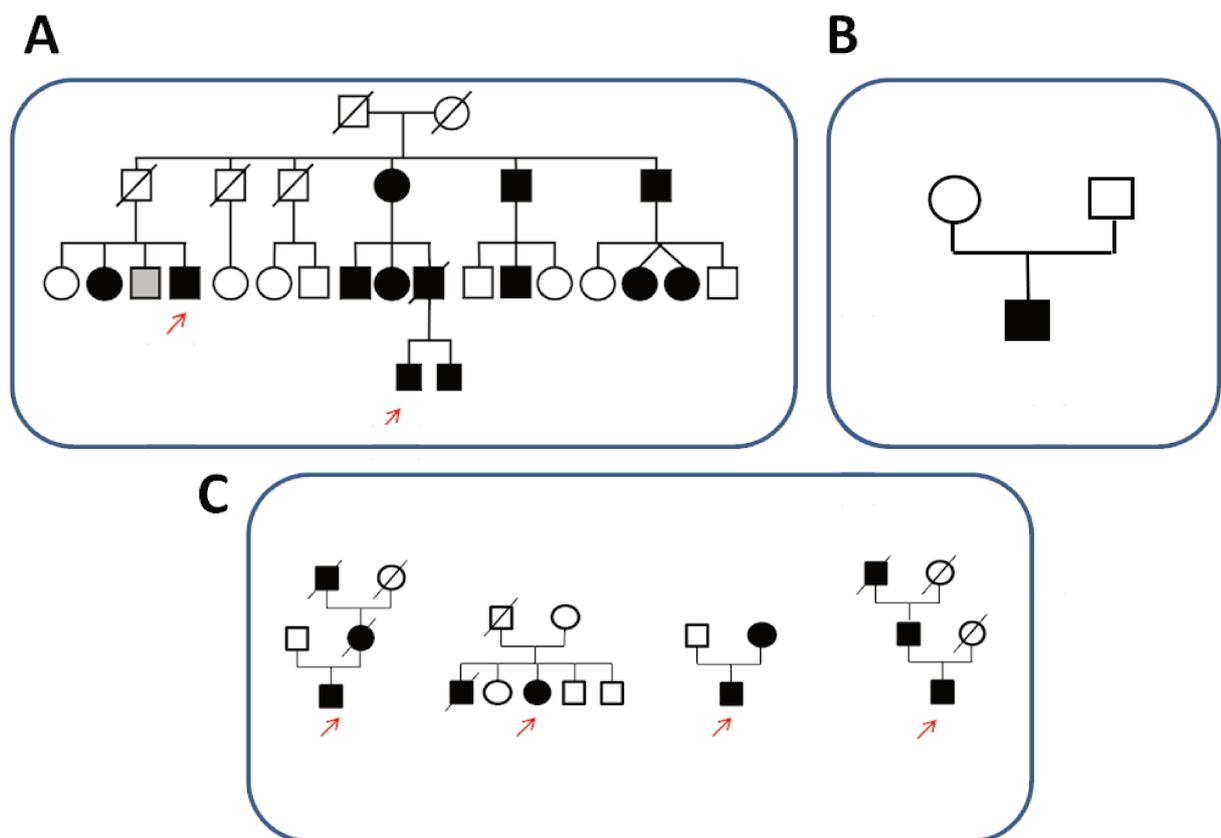


Figure 12 : Les différentes stratégies d'analyse d'exomes

A. Stratégie intrafamiliale à transmission autosomique dominante ; B. Etude de récurrence; C. Analyse de trios

## 1. Stratégie intrafamiliale

Selon le mode de transmission étudiée (dominant, récessif, liée à l’X), il est possible d’utiliser la stratégie intrafamiliale qui consiste à séquencer plusieurs individus apparentés malades et sains dans le but d’étudier la ségrégation d’une variation génétique (Figure 12 page 40). Dans le cadre de maladies récessives, on recherchera un variant à l’état homozygote surtout si une consanguinité est observée, ou des variants à l’état d’hétérozygotes composites. Les études de liaisons peuvent permettre de limiter l’analyse des variants à une région d’intérêt plutôt que de considérer l’intégralité du génome ou de l’exome. Dans le cadre d’une transmission autosomique dominante, il est possible de séquencer deux apparentés éloignés malades. Par exemple, deux cousins partagent au maximum 1/8 de leurs variations génétiques. Dans ce cas, cette stratégie peut permettre de diviser le nombre de variants rares à analyser par 8 soit  $1\,500/8$  soit  $\sim 187$  dont  $61/8$  soit  $\sim 8$  pour les PTVs. Plus les individus apparentés sont génétiquement éloignés, moins ils partagent de variants en communs. Le nombre de variants à analyser peut donc considérablement diminuer. La ségrégation des variants identifiés peut ensuite s’étudier par séquençage Sanger sans nécessité de réaliser d’autres séquençages d’exomes.

## 2. Les analyses de trio

Cette analyse est adaptée uniquement aux cas sporadiques. Le génome d’un individu provient pour moitié du génome de son père et pour l’autre de sa mère. En plus de ce patrimoine génétique transmis, chacun porte dans son génome des variations nouvelles, les variations *de novo*, survenues durant la formation des gamètes ou très précocement après la fécondation. En moyenne, un individu porte 44 à 82 variations *de novo* dans son génome avec seulement une à deux variations intersectant les régions codantes (Acuna-Hidalgo et al., 2016). 80 % des mutations *de novo* proviendraient de l’allèle paternel en raison du grand nombre de divisions cellulaires au cours de la vie des gamètes males par rapport aux oocytes. De plus, un âge avancé du père accroît le risque de survenue de mutations *de novo* (Acuna-Hidalgo et al., 2016; Kong et al., 2012). Lors de la conception de l’enfant, chaque année supplémentaire du père conduit à l’apparition de 2 variations *de novo* supplémentaires en moyenne chez l’enfant.

En ce qui concerne les CNVs, les estimations sont moins claires. D’après Veltman et al. (Veltman and Brunner, 2012), il y aurait entre 0,01 à 0,02 CNV *de novo* par génération pour les événements structuraux supérieurs à 100 kb, les estimations pour les remaniements plus petits n’ont pu être effectuées de par les limitations technologiques. De manière intéressante, les CNVs *de novo* non-récurrents seraient principalement hérités du père alors que les CNVs récurrents principalement

médiés par les systèmes de recombinaisons homologues seraient majoritairement hérités de la mère (pour revue Acuna-Hidalgo et al., 2016).

La stratégie des trios repose sur cette rareté des variations *de novo* survenues dans l'exome d'un individu. Elle consiste à séquencer l'exome des parents non atteints et du cas index. On effectue ensuite une étape de soustraction des variations présentes chez les parents par rapport à celles présentes chez le cas index. Cette stratégie conduit donc à restreindre l'analyse en moyenne à 1 variation exonique *de novo*. Au sein du laboratoire, nous avons pu constater que le nombre de variations exoniques *de novo* peut être compris entre 0 et 4 (Rovelet-Lecrux et al., 2015) (article en annexe).

Pour établir l'implication d'un gène touché par des variations *de novo* dans le déterminisme d'une maladie, il est nécessaire de trouver une récurrence à l'échelle de la variation, du gène ou encore à l'échelle d'un réseau ou d'une liste de gènes. Afin d'évaluer la significativité d'une telle récurrence, la mutabilité du gène ou des gènes d'une liste doit pouvoir être évaluée. En effet, il faut pouvoir vérifier que la récurrence observée n'est pas dépendante de différentes caractéristiques du ou des gènes impliqués. Plusieurs paramètres permettent de déterminer de manière théorique la mutabilité d'un gène (Campbell and Eichler, 2013):

- la taille : les gènes plus longs ont une probabilité plus importante de porter des variations *de novo*
- les caractéristiques intrinsèques du gène : leur séquence nucléotidique, taux de GC, marqueurs épigénétiques

Le taux de mutation *de novo* attendues par gène peut être estimé à partir :

- des observations de mutations *de novo* dans les grandes séries de quartets de génomes (Michaelson et al., 2012)
- des différences homme-chimpanzé en comparant les variations fixées chez l'homme ou chez le chimpanzé par rapport à celle d'un ancêtre commun

Cette stratégie des trios a notamment été utilisée à partir de grands échantillons de patients autistes en recrutant plus de 2500 trios (Iossifov et al., 2014) et a permis de mettre en évidence un enrichissement de variants *de novo* perte de fonction chez ces patients.

De plus, les étapes de sélection des variants évoquées dans les parties précédentes peuvent apporter un niveau de preuve supplémentaire. Dans certains cas, générer des données fonctionnelles peut s'avérer absolument nécessaire notamment lorsque le gène a été peu étudié.

Au sein du laboratoire, nous avons étudié la récurrence de variations nucléotidiques et de variants structuraux *de novo* à l'échelle d'une liste de gènes chez des cas atteints de formes précoces de maladie d'Alzheimer (Rovelet-Lecrux et al., 2015). Exceptés la variation *de novo* c.869-2A>G de

*PSEN1* et la duplication *de novo* d'*APP* pour lesquelles la causalité est clairement établie, une variation de *VPS35* et une variation de *MARK4* avaient fait l'objet d'analyses fonctionnelles permettant de statuer sur leur rôle causal dans la physiopathologie de la maladie (Rovelet-Lecrux et al., 2015) (article en annexe).

### 3. Etude de récurrence

Cette stratégie consiste à détecter une récurrence à l'échelle d'un variant ou d'un gène chez différents cas index présentant la même pathologie. Cette stratégie avait été utilisée pour la première fois par Ng et collaborateurs et avait permis de mettre en évidence une récurrence de mutations du gène *MYH3* chez des patients atteints du syndrome de Freeman-Sheldon (Ng et al., 2009). Cette stratégie avait été utilisée au laboratoire pour l'analyse de 14 formes familiales de MA à début précoce, compatibles avec une transmission autosomique dominante. Il avait alors été mis en évidence une récurrence modérée (5/14) de variants rares de *SORL1* (Pottier et al., 2012b).

### 4. Les études d'association cas-contrôles

Outre les stratégies familiales ou l'analyse de trios, les stratégies d'analyses de variants rares reposent essentiellement sur les études d'association. Elles consistent à comparer la fréquence des variants détectés dans une population de malades avec celle présente dans une population de témoins. Ces études sont en fait des études de récurrence avec une puissance statistique plus importante.

Les premières études d'association ont été réalisées grâce aux GWAS. Les études de GWAS, basées sur des variations fréquentes, requièrent souvent pour être statistiquement significatives des échantillons très importants (excepté les cas de variants fréquents à effet fort). Ces études supposent que le génotypage d'environ 1 million de variations fréquentes correctement sélectionnées permet de couvrir l'ensemble des blocs haplotypiques du génome. Cependant, tester un million de variations revient à réaliser un million de tests et par conséquent à ajuster le seuil de significativité à  $0,05/10^6=5 \times 10^{-8}$ . Par conséquent, les études de GWAS nécessitent souvent des échantillons de cas et de contrôles très importants : 74 046 individus par exemple pour la méta-analyse du *consortium* Alzheimer caractérisant 20 *loci* (Lambert et al., 2013a). Les variants fréquents à effet fort, tels que l'allèle  $\epsilon 4$  d'*APOE* dans la MA (Genin et al., 2011), sont les seuls à pouvoir être associés avec des échantillons plus petits.

Quand on analyse de la même manière des variants rares, la taille de l'échantillon doit être nécessairement encore plus grande. Cette principale limite de puissance a fortement restreint les

analyses d'association de variants rares et seules quelques études ont été capables d'établir de telles associations (quelques exemples seront présentés dans la suite du manuscrit).

Différentes approches ont été suggérées afin d'augmenter la puissance des études d'association avec les variants rares.

### **a. Puissance contre phénotypes extrêmes**

La sélection rigoureuse des cas et des contrôles sur des critères diagnostics stricts peut permettre d'augmenter la puissance statistique de l'étude.

#### *Patients phénotypes extrêmes*

Les études de GWAS ont réussi à détecter des signaux en utilisant des effectifs importants parfois au détriment de l'inclusion des patients qui répondaient à des critères diagnostics plus lâches. A titre d'exemple, dans la MA les critères de diagnostics de 1984 définis par McKhann suffisaient (McKhann et al., 1984). Or, les critères d'inclusion ont fortement évolué et une mise à jour a été publiée en 2011 (McKhann et al., 2011). Ainsi ces études ont parfois altéré leur puissance en incluant des patients pour lequel le diagnostic n'est pas aujourd'hui considéré comme suffisamment fondé. La sélection rigoureuse des cas basée sur des critères diagnostiques stricts est une première façon d'augmenter la puissance statistique en réduisant l'hétérogénéité génétique induite par l'implication de patients possiblement atteints d'une autre pathologie. Une possibilité additionnelle est de restreindre les analyses d'association à des phénotypes extrêmes (Barnett et al., 2013; Lee et al., 2014). L'hypothèse est que les individus présentant un phénotype extrême concentre des variants rares à fort effet (Barnett et al., 2013; Lee et al., 2014). Dans le cas de la MA, la sélection des individus sur leur âge de début, avant 65 ans pour les formes précoces, permet de s'affranchir d'une part de la composante environnementale. En effet, le principal facteur de risque dans la MA est l'âge avec l'accumulation de multiples atteintes d'origine environnementale (Hirtz et al., 2007). On estime ainsi que les facteurs stochastiques impliqués dans la MA représentent une part plus limitée dans les formes précoces de MA. Par conséquent, la part attribuable aux facteurs génétiques est nécessairement plus grande. De plus, il est estimé que l'hétérogénéité génétique des maladies complexes serait moins importante chez les patients présentant un phénotype extrême commun (Barnett et al., 2013). En se focalisant sur ces phénotypes extrêmes de MA, cette stratégie nous a permis de détecter une association de variants rares de *SORL1* dans des formes précoces familiales de la maladie d'Alzheimer (Nicolas et al., 2016b).

### *Sélection des contrôles*

La sélection des contrôles est également très importante dans les études d'association génétique. Premièrement, faut-il utiliser des contrôles de la population générale (appariés pour l'âge) ou des super-contrôles, c'est à dire des sujets ayant traversé la période à risque pour la maladie en étant indemne ? Dans le cadre de la MA, il a ainsi été proposé d'utiliser des centenaires comme contrôles (Garagnani et al., 2013). En théorie, cette stratégie doit augmenter la puissance mais cette méthode peut s'avérer dangereuse. En effet, des individus très âgés présentent une résistance à plusieurs pathologies et donc une association mise en évidence à partir de ces sujets peut traduire la présence de facteurs protecteurs n'ayant aucun rapport avec la MA.

Deuxièmement, il est très important de considérer l'origine ethnique de nos cas et de nos contrôles. Si ceux-ci ne sont pas appariés, on introduit un biais dit de stratification de telle sorte que la variation de fréquence constatée entre des cas et des témoins n'est pas due à la présence de la maladie mais résulte d'une structure génétique différente de chaque population (Lee et al., 2014; Wang et al., 2015a).

Troisièmement, les groupes de témoins pourraient être constitués à partir des individus présents dans les différentes bases de données citées précédemment. Ces bases de données peuvent représenter un intérêt majeur notamment en termes de coûts évitant de séquencer des individus contrôles en parallèle des patients. Cependant, des problèmes supplémentaires apparaissent. Tout d'abord, les informations obtenues à partir de ces bases de données sont des données de fréquence, il n'est donc pas toujours possible de déterminer quels sont les individus porteurs de variations. Cela empêche alors les analyses à l'échelle du gène qui considèrent le nombre de porteurs de variants et non le nombre de variants absolu. Par ailleurs, il n'est pas possible de définir le taux de données manquantes pour chaque position à partir de ces bases de données bien que les profondeurs moyenne soient indiquées. Par conséquent, le résultat peut être biaisé par une absence de variants due à des problèmes de qualité de séquençage (kit de capture différent, profondeur de couverture etc.).

Pour éviter ces biais, il est plus approprié de séquencer des individus contrôles appariés pour l'âge et l'origine ethnique en parallèle des individus malades (Lee et al., 2014). De cette façon, la sélection des témoins est faite à partir des prérequis spécifiques nécessaires aux différentes analyses. Dans le cas de nos études, nous avons pu bénéficier du *French Exome Project* (FREX, France génomique, E Génin, JF Deleuze, R Redon, JC Lambert, D Champion) qui regroupe aujourd'hui environ 600 contrôles nés en France de parents nés en France.

## **b. Agrégation des variants à l'échelle du gène**

Devant l'extrême rareté de chaque variant, il est nécessaire d'agréger ces variants à l'échelle du gène afin d'augmenter notre puissance statistique en cumulant les MAFs de chaque variant. En effet, cela peut permettre de cumuler l'effet de plusieurs variants sur un même gène. Ces études d'association à l'échelle du gène requièrent cependant de considérer différents paramètres. En effet, lorsque l'on réalise des tests à l'échelle du gène il est nécessaire de corriger le seuil de significativité par le nombre de gènes étudiés soit  $0,05/20\ 000=2,5\times 10^{-6}$  (Lee et al., 2014; Wang et al., 2015a). De plus, en fonction du type de variations étudié le test utilisé sera différent. Dans le cas de nos études, nous avons principalement axé nos recherches sur les variants PTVs et les variants faux sens prédits pathogènes par les trois logiciels de prédiction de pathogénicité (Mutation Taster, SIFT et Polyphen-2) (paragraphe A.4.c) dans une hypothèse de perte de fonction. Par conséquent, en comparant les cas et les témoins nous cherchions à détecter des variants rares ayant une MAF cumulée plus forte chez les cas que chez les contrôles. Les tests statistiques les plus appropriés dans ce type d'analyse sont les tests de *Burden* ou *Collapsing test* qui cumulent ces MAFs sans distinction entre les variants qu'ils considèrent tous comme ayant le même effet sur la protéine (Lee et al., 2014). Cependant, certaines variations d'un gène peuvent être protectrices quand d'autres peuvent être à risque et par conséquent les effets au niveau des MAFs cumulées chez les cas et chez les contrôles s'annuleront, reflétant ainsi la surreprésentation de variants rares protecteurs chez les contrôles et de variants à risque chez les cas. Certains tests reposant sur les variances permettent de prendre en compte la présence de variants protecteurs ou à risque tel que le test SKAT (Lee et al., 2014).

Cependant, la stratégie d'association à l'échelle du gène, restreinte à des phénotypes extrêmes et des individus contrôles correctement sélectionnés, analysée à l'aide tests statistiques appropriés n'est pas toujours suffisante et une part de la composante génétique de nombreuses maladies complexes reste encore incomplète.

## **c. De l'échelle du gène à l'échelle d'une liste de gènes**

Afin de contourner le problème de puissance statistique insuffisante, l'association à l'échelle d'une liste de gènes peut permettre là encore d'agréger l'effet d'un plus grand nombre de variants. Cependant, cette analyse ajoute un niveau de complexité car elle demande de prendre en considération l'effet de chaque gène, contenu dans cette liste, dans la maladie.

### ***Association à l'échelle d'une liste de gènes***

Cette analyse à l'échelle d'une liste de gènes se révèle nécessaire quand les analyses à l'échelle du gène ne permettent pas d'obtenir des résultats significatifs avec des variants très rares. Cette stratégie permet d'augmenter ainsi la puissance statistique en agrégeant un nombre plus important

de variants. Ces analyses à l'échelle d'une liste de gènes ont notamment été utilisées dans la schizophrénie grâce à une étude d'association cas-contrôles (Genovese et al., 2016) mais également à partir de l'analyse de trios (Fromer et al., 2014).

De nombreuses bases de données permettent de définir des listes de gènes tels que KEGG (<http://www.genome.jp/kegg/pathway.html>), Reactome (<http://www.reactome.org/>) ou encore PantherDB (<http://www.pantherdb.org/pathway/>). Cependant, ces bases de données ne sont pas exhaustives et ne sont pas régulièrement mises à jour. C'est également le cas pour les termes de *Gene Ontology* (GO) qui distinguent par exemple les protéines grâce à leur localisation cellulaire ou par leur fonction catalytique.

Dans le cadre de cette thèse, une liste de gènes a été construite à partir d'une bibliographie extensive centrée sur le peptide amyloïde et excluant les gènes dont la caractérisation fonctionnelle serait trop imprécise. A partir de cette liste de gènes, nous avons notamment réussi à mettre en évidence un enrichissement de variants *de novo* et un excès de CNVs chez des patients atteints de formes précoces de MA (Le Guennec et al., 2016; Rovelet-Lecrux et al., 2015). Cependant, cette liste de gènes, régulièrement mis à jour, n'est pas forcément exhaustive et restreint nos analyses aux voies biologiques connues. Il n'est donc pas possible à partir de cette liste de détecter de nouvelles voies biologiques potentiellement impliquées dans la MA.

### *Quels tests statistiques ?*

Dans le cadre des études d'association à l'échelle d'une liste de gènes, il est nécessaire de déterminer le test statistique à utiliser ainsi que le seuil de significativité. En effet, la significativité de l'association à l'échelle du gène est corrigée sur le nombre de gènes présents dans le génome humain soit 20 000. Lorsqu'on étudie une liste de gènes, les principales études corrigent le seuil de significativité par rapport au nombre de listes de gènes qu'ils testent. A titre d'exemple, une étude réalisée sur plus de 11 000 patients schizophrènes comparés à plus de 16 000 contrôles a testé 134 listes de gènes différentes et donc corrigée leur significativité en fonction de ces 134 tests (Pocklington et al., 2015). La détermination du seuil de significativité à partir d'une liste de gènes n'est, à ce jour, pas clairement définie dans la communauté scientifique (pour revue voir (Mooney and Wilmot, 2015)). Les associations à l'échelle d'une liste de gènes soulèvent également un problème majeur car elles peuvent potentiellement diluer l'effet de variants portés par quelques gènes qui se retrouvent confondus avec d'autres gènes ne portant pas de variations à risque ou protectrices.

#### d. Quel seuil de significativité pour les CNVs ?

Si les différents seuils de significativité ont pu être établis en ce qui concerne les études de GWAS ( $5 \times 10^{-8}$ ) ainsi que pour les associations à l'échelle du gène ( $2,5 \times 10^{-6}$ ) pour les variants et indels, le seuil de significativité à l'échelle des CNVs rares est difficile à déterminer. La principale limite est, de nouveau, de déterminer le nombre de tests effectués et donc la correction statistique à appliquer. En effet, à cause des limites technologiques qui ne définissent pas correctement les différents points de cassure des CNVs, le nombre de CNVs différents analysés dans une étude est toujours imprécis. Afin de traiter cette question, les études d'association tentent de grouper les CNVs grâce à deux stratégies. La première consiste à regrouper les CNVs lorsqu'ils partagent une même partie de séquence génomique. La seconde stratégie regroupe les CNVs lorsqu'ils touchent le même gène. Malgré de nombreuses approximations, Levinson *et al.* parviennent à la conclusion que cela revient à réaliser 2000 tests différents (Levinson et al., 2011). Par conséquent, le principal seuil de significativité retenu pour associer un CNV à une maladie dans une étude cas-contrôles a été corrigé à  $2,5 \times 10^{-5}$  puis arrondi à  $10^{-5}$ .

## C. Maladie d'Alzheimer

La maladie d'Alzheimer est un exemple de pathologie au déterminisme génétique hétérogène. La composante génétique de cette maladie est expliquée par des variants nucléotidiques, des indels et des variations structurales, rares ou fréquents (Cacace et al., 2016; Cuyvers and Sleegers, 2016). A travers cette partie, j'aborderai les aspects épidémiologiques et cliniques ainsi que les principaux mécanismes physiopathologiques de la maladie d'Alzheimer et décrirai ensuite les différentes causes et facteurs de risque génétiques de cette maladie.

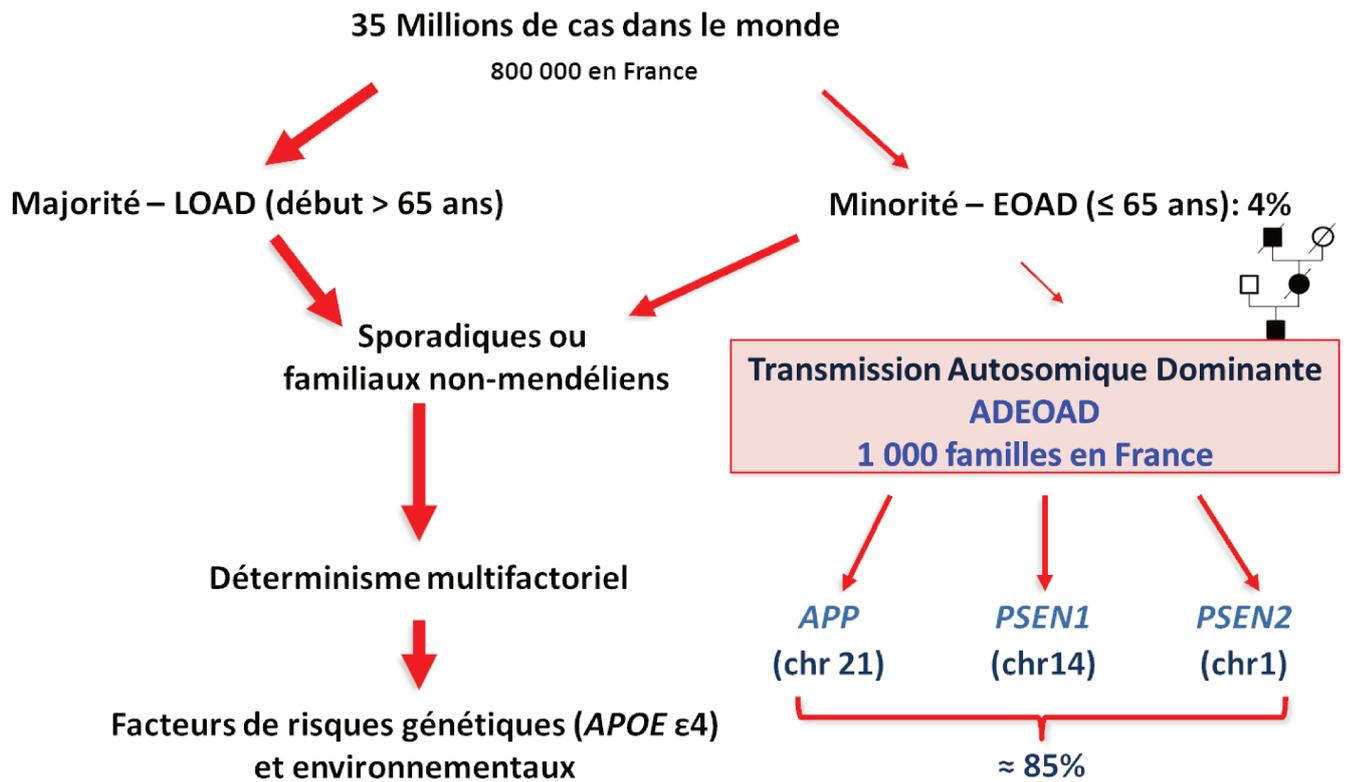
### 1. Epidémiologie et aspects cliniques

Pour la première fois décrite par Aloïs Alzheimer en 1907, la maladie d'Alzheimer est la maladie neurodégénérative la plus fréquente dans le monde touchant aujourd'hui plus de 30 millions de personnes. La MA se caractérise initialement par des troubles de la mémoire associés à un fort impact sur l'autonomie. Elle conduit généralement au décès entre 3 et 15 ans après le diagnostic. Le principal facteur de risque est l'âge avec une incidence qui double tous les 5 ans après 65 ans (Hirtz et al., 2007).

Au sein de la maladie d'Alzheimer, on distingue les formes précoces avec un âge de début avant 65 ans et les formes tardives (Figure 13 page 50). Parmi les formes précoces, une minorité des cas répond à un déterminisme mendélien de transmission autosomique dominante (*Autosomal Dominant Early-Onset Alzheimer Disease* : ADEOAD). La vaste majorité des cas sont donc atteints de MA sporadiques ou avec antécédents familiaux mais sans transmission mendélienne. La découverte des gènes responsables des formes mendéliennes de MA a permis de formuler des hypothèses physiopathologiques.

D'un point de vue clinique, le diagnostic de la maladie d'Alzheimer est établi à partir de différents critères établis par le NINCDS-ADRDA (*National Institute of Neurological and Communicative Disorders and Strokes – Alzheimer's Disease and Related Disorders Association*) (McKhann et al., 2011). Ces critères reposent sur des examens neuropsychologiques standardisés, des examens d'imagerie cérébrale, des analyses de biomarqueurs du liquide céphalo-rachidien (LCR) et sur des analyses anatomopathologiques *post-mortem* permettant ainsi de différencier une MA d'Alzheimer possible, d'une MA probable ou certaine (par ordre de certitude).

L'analyse anatomopathologique *post-mortem* du cerveau des patients reste le seul examen qui permet avec certitude d'établir le diagnostic.



**Figure 13 : Hétérogénéité génétique de la maladie d'Alzheimer**

La majorité des patients atteints de MA, qu'il s'agisse de formes précoces (EOAD) ou tardives (LOAD), répondent à un déterminisme multifactoriel dont le principal facteur de risque génétique est l'allèle ε4 de l'APOE. Une minorité des cas sont des formes précoces à transmission autosomique dominante (AEOAD, *Autosomal Dominant Early-Onset Alzheimer Disease*). Les mutations d'*APP*, *PSEN1* et *PSEN2* sont responsables d'environ 85 % de ces cas (Campion et al., 1999 ; Wallon et al., 2012).

## 2. Lésions anatomopathologiques

### a. Lésions macroscopiques

Au niveau macroscopique, les principales lésions dans la MA consistent en une atrophie globale du cerveau accompagnée d'une dilatation des ventricules cérébraux, d'un élargissement des sillons corticaux ainsi que d'une perte neuronale diffuse (Duyckaerts et al., 1985; Whitehouse et al., 1981). Cependant, ces caractéristiques macroscopiques ne sont pas suffisamment discriminantes pour caractériser une MA. L'examen microscopique est donc indispensable et met en évidence les lésions caractéristiques de la maladie décrites par Alois Alzheimer au début du siècle dernier (Alzheimer et al., 1995 (version en anglais)).

### b. Lésions microscopiques

Au niveau microscopique, les dégénérescences neurofibrillaires et les plaques séniles sont les deux types de lésions définissant une MA.

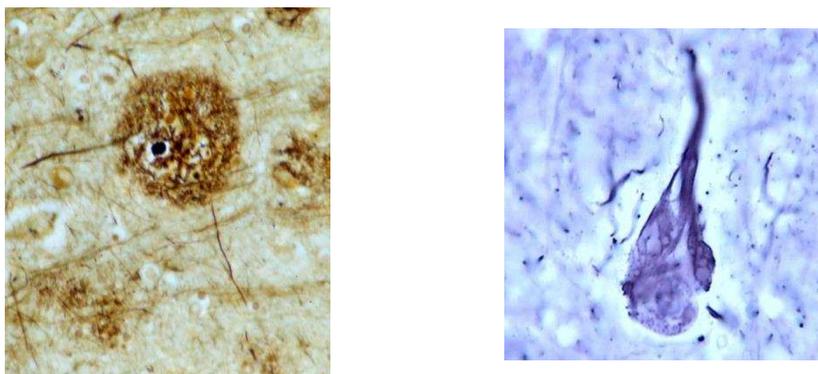
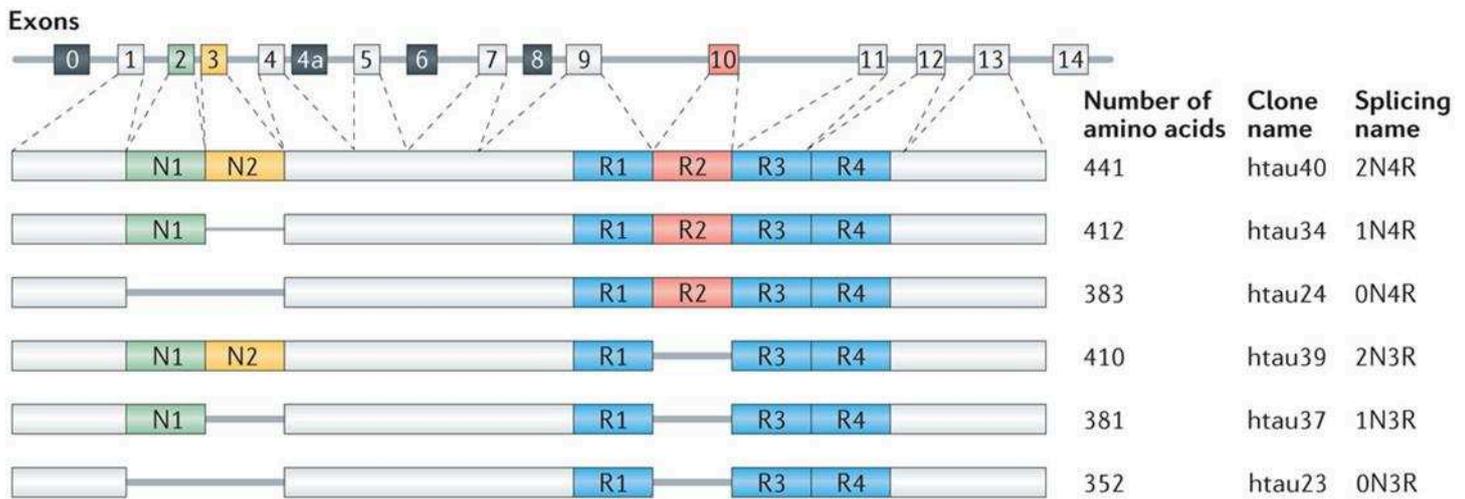


Figure 14 : Plaques amyloïdes (à gauche) et dégénérescences neurofibrillaires (à droite)

### c. Les dégénérescences neurofibrillaires

Les dégénérescences neurofibrillaires (DNF) sont des lésions intraneuronales constituées de protéines Tau hyper et anormalement phosphorylées qui s'associent en paire de filaments hélicoïdaux (Figure 14 page 51). Ces lésions sont caractéristiques de la MA bien qu'elles ne soient pas spécifiques de cette dernière. En effet, de telles lésions ont été également retrouvées dans de nombreuses maladies sans pathologie amyloïde regroupées sous le terme de tauopathies : les dégénérescences lobaires fronto-temporales, les paralysies supranucléaires progressives, la dégénérescence cortico-basale ou encore la maladie de Pick (Kovacs, 2015). Contrairement à ces tauopathies, les DNFs retrouvées dans les cerveaux de malades Alzheimer sont constituées de deux isoformes de Tau nommées 3R et 4R (Espinoza et al., 2008). Le gène *MAPT* (*Microtubule-associated protein tau*) code pour 6 isoformes de la protéine Tau et se différencie par l'épissage alternatif des

exons 2, 3 et 10 (NM\_005910) qui contiennent les séquences des domaines de liaisons aux microtubules (MT) (Figure 15 page 52). Ces deux isoformes de Tau diffèrent par l'inclusion (4R, quatre domaines de liaisons aux MT) ou non (3R, trois domaines de liaisons aux MT) de l'exon 10 (Wang and Mandelkow, 2016). Les autres tauopathies citées présentent des inclusions de Tau 3R ou Tau 4R (Kovacs, 2015).



**Figure 15 : Les six isoformes de Tau**

Figure issue de la revue de Wang et al., 2016.

Les 6 isoformes de Tau dépendent de l'épissage alternatif des exons 2, 3 et 10. L'inclusion des exons 2 et 3 définit les formes 0N, 1N ou 2N tandis que l'exon 10 définit le nombre de domaines de liaisons aux MT 3R ou 4R.

### Primary Age-Related Tauopathy

De façon notable, ces DNFs sont retrouvées de façon quasi constante au cours du vieillissement au niveau du cortex entorhinal et de l'hippocampe. Le terme de *Primary Age-Related Tauopathie* (PART) a été proposé par Cray et collaborateurs pour décrire ce type de lésions chez des individus âgés, généralement asymptomatiques (Crary et al., 2014). Les DNFs sont similaires à celles retrouvées dans la maladie d'Alzheimer, composées d'une accumulation de formes 3R et 4R de Tau. De plus, les individus ne portent pas ou peu de plaques amyloïdes et ne peuvent donc pas être classés comme malades Alzheimer. Le terme PART définit en réalité un continuum qui inclut des individus âgés présentant des DNFs mais cognitivement sains jusqu'à des patients atteints de démences couramment nommées « *tangle only dementia* », « *tangle-predominant senile dementia* » « *preferential development of NFT without senile plaques* » ou encore « *senile dementia of the neurofibrillary tangle type* ».

### Progression des DNF

Chez la majorité des sujets, les lésions restent confinées dans les zones entorhinales et hippocampiques. Toutefois, lorsque des dépôts de peptide A $\beta$  sont présents dans les régions

corticales, les DNFs envahissent progressivement le neocortex selon une progression stéréotypée entre des neurones connectés synaptiquement, on est alors dans le cadre d'une MA (Braak and Braak, 1991; Thal et al., 2002). Les différents stades de la progression des DNFs sont décrits par les stades de Braak (Braak and Braak, 1991; Delacourte et al., 1999) (Figure 16 page 55). La première structure touchée par les DNFs est le cortex trans-entorhinal (stade I et II de Braak). La deuxième structure affectée est l'hippocampe (stade III et IV de Braak). Ces lésions sont observées indistinctement chez des individus sains et malades. Enfin, les deux derniers stades V et VI de Braak ne sont jamais observés chez des individus cognitivement sains et décrivent l'atteinte par les DNFs du cortex. Ces différents stades corrélerent avec le développement de la maladie : une absence de symptômes (I, II) puis des troubles cognitifs modérés (MCI : *Mild Cognitive Impairment*) (III, IV) jusqu'à la maladie lorsque les zones corticales sont touchées (V et VI).

### Mécanismes neurotoxiques de Tau

Tau est une protéine soluble associée aux MT (MAP : *Microtubules-Associated Protein*). Il a été initialement prouvé qu'elle intervenait dans la stabilisation du réseau de MT (Cleveland et al., 1977; Weingarten et al., 1975). Toutefois, Tau est majoritairement localisée au niveau axonal mais est également retrouvée au niveau dendritique (Wang and Mandelkow, 2016). Cette localisation axonale et dendritique a permis d'impliquer Tau dans les mécanismes de maturation neuronale et de plasticité synaptique. Dans un mécanisme pathologique, Tau est hyper et/ou anormalement phosphorylée. Cette conformation anormale de Tau causée par des mutations est suffisante pour déclencher la pathologie chez les patients atteints de FTDP-17 (*Frontotemporal Dementia and parkinsonism linked to chromosome 17*) (Vogelsberg-Ragaglia et al., 2000).

Il a d'abord été suggéré que la neurotoxicité des formes hyperphosphorylées de Tau serait due à une perte de fonction de la protéine. La protéine Tau hyper et anormalement phosphorylée se détacherait des MT et entraînerait une déstabilisation du réseau de MT (revue (Wang and Mandelkow, 2016)). Cette perte de fonction aurait pour principale conséquence d'altérer la structure des neurones et leur activité synaptique. Cependant, Tau n'est pas la seule protéine à se lier aux MT et plusieurs études ont démontré que cette perte de fonction de Tau pouvait être compensée par d'autres MAPs (Dawson et al., 2001; Fujio et al., 2007; Harada et al., 1994; Tucker et al., 2001).

Il a ensuite été suggéré que Tau hyperphosphorylée agirait par un gain de fonction toxique. En effet, Tau hyper et anormalement phosphorylée est décrochée des microtubules et posséderait alors des propriétés biophysiques favorisant son agrégation. Cette agrégation initierait la formation des DNFs, lésions qui entraîneraient la mort des neurones. Cependant, les neurones atteints de DNFs ne meurent pas tous et certains resteraient fonctionnels (de Calignon et al., 2009; Morsch et al., 1999; Spires-Jones et al., 2008).

Ces dernières années plusieurs points ont été établis, renouvelant la compréhension de la pathologie Tau. Premièrement, Tau dendritique joue un rôle important dans les mécanismes de plasticité synaptique (Frandemiche et al., 2014). Deuxièmement, Tau hyper et anormalement phosphorylée se comporte comme une protéine prion capable de (i) changer par contact la conformation d'une protéine native pour favoriser son agrégation (*seeding*) et (ii) d'être transférée de cellules à cellules (*spreading*) (Holmes and Diamond, 2014).

#### **d. Les plaques séniles**

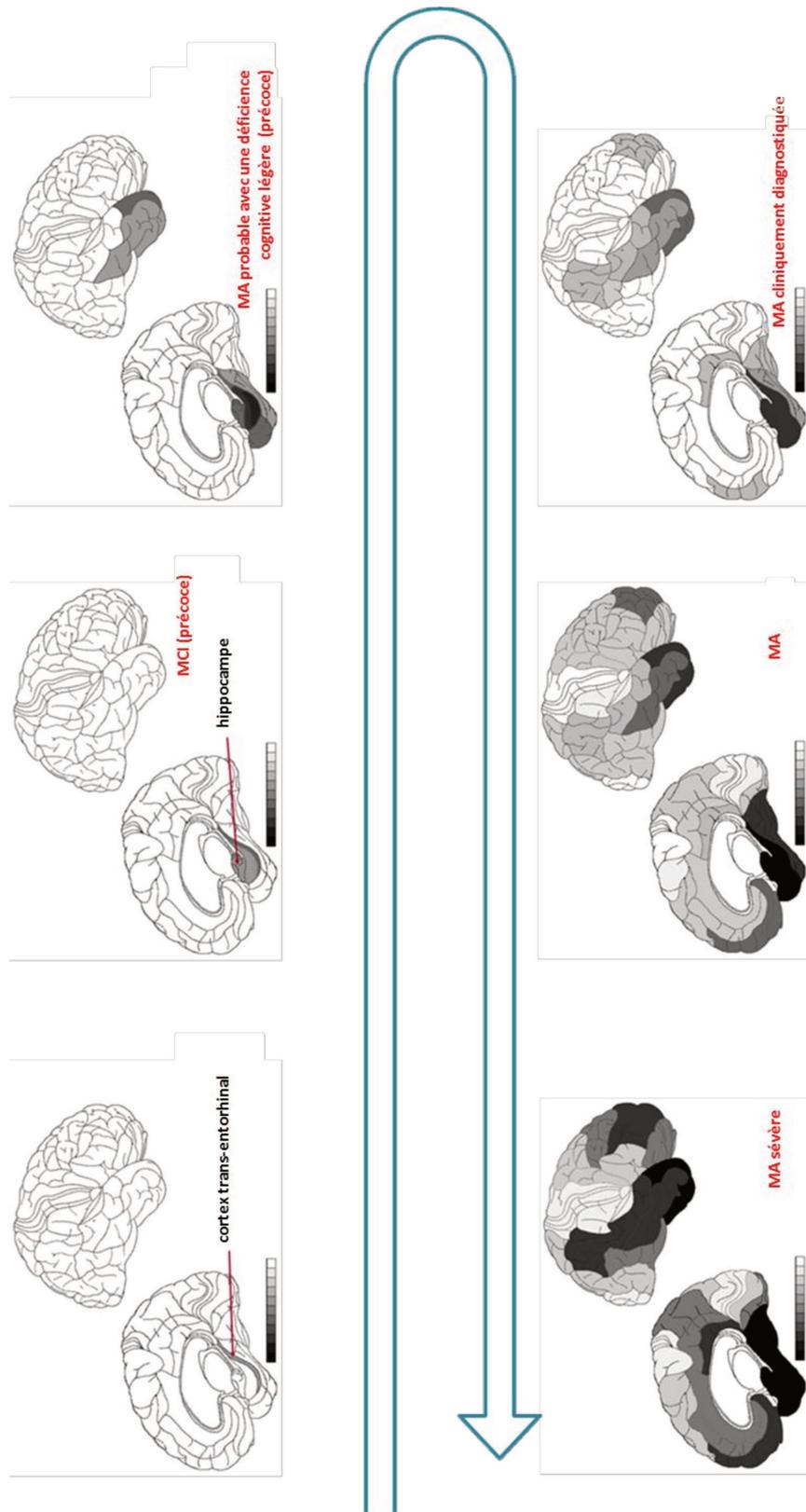
Décrites par Alois Alzheimer (Alzheimer et al., 1995 (version en anglais)), les plaques séniles sont des lésions extracellulaires constituées majoritairement de peptides amyloïdes ou A $\beta$  agrégés en feuillet  $\beta$ -plissé entouré de neurites dystrophiques. Ce peptide fibrillogène, A $\beta$ , purifié en 1984 (Glennner and Wong, 1984) est produit majoritairement sous deux formes A $\beta$ 40 et A $\beta$ 42 se différenciant par la présence de 2AA supplémentaires pour A $\beta$ 42. Les plaques séniles sont composées d'un cœur hydrophobe de peptide A $\beta$ 42 entouré d'une couronne de peptide A $\beta$ 40 (Figure 14 à gauche page 51). Ces lésions extracellulaires sont spécifiques de la MA.

Si les plaques séniles sont les plus fréquentes, d'autres types de plaques amyloïdes ont également été documentés et notamment les *cotton wool plaques*. Ces lésions particulières, associées à certaines mutations de PSEN1, consistent en de larges dépôts sans cœur amyloïde (Crook et al., 1998).

L'angiopathie amyloïde est une lésion observée chez certains patients Alzheimer, notamment ceux porteurs de duplication du gène *APP* (*Amyloid Precursor Protein*) (Rovelet-Lecrux et al., 2006), et correspond à l'agrégation du peptide A $\beta$  dans la paroi des vaisseaux cérébraux.

Par ailleurs, le développement d'anticorps dirigés contre A $\beta$  a permis de mettre en évidence des dépôts diffus de peptides amyloïdes présents indistinctement chez des patients et des individus cognitivement sains (Duyckaerts et al., 1999).

La progression de ces dépôts diffus au sein du tissu cérébral a été décrite par Thal et collaborateurs (Thal et al., 2002). Contrairement à la pathologie Tau, les dépôts diffus sont localisés initialement dans les zones corticales pour ensuite atteindre les structures subcorticales comme le cortex entorhinal et l'hippocampe.



**Figure 16: Progression des DNFs dans les cerveaux de malades Alzheimer**  
 La progression des DNFs suit un parcours stéréotypé depuis le cortex entorhinal jusqu'aux zones corticales. Les premières régions touchées sont le cortex entorhinal et l'hippocampe. Dans la majorité des cas, ces lésions restent confinées à ces zones. Dans le cadre de la MA, la progression continue pour atteindre les régions corticales. MCI : *Mild Cognitive Impairment*

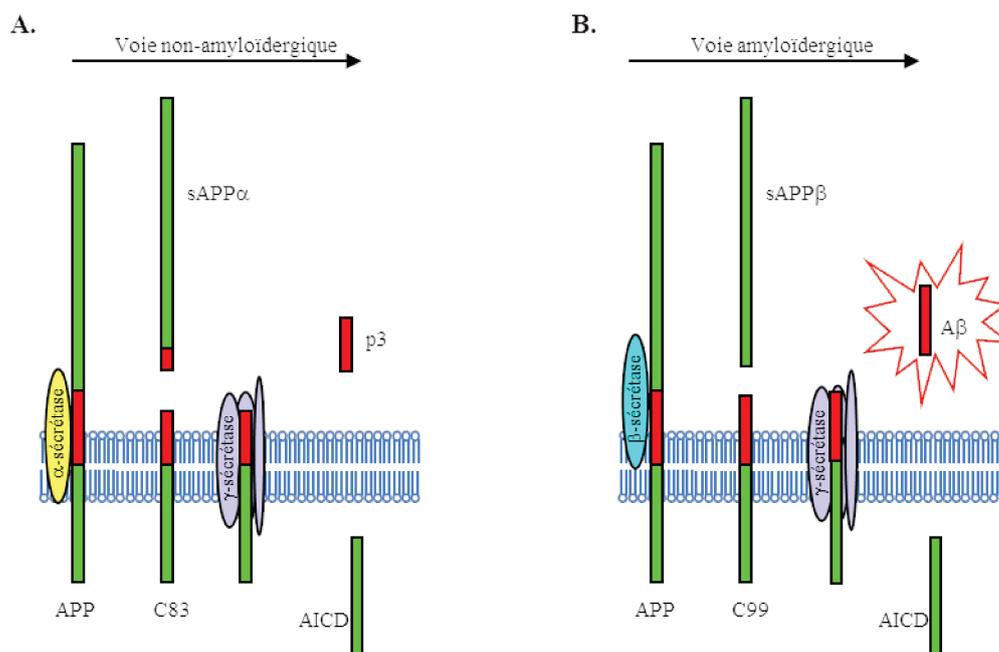
### e. Amyloïdogénèse

Le peptide A $\beta$ , composant majoritaire des plaques séniles et des dépôts amyloïdes est issu de son précurseur l'*Amyloid Precursor Protein* (St George-Hyslop et al., 1987).

Au cours de sa maturation, cette protéine est soumise à différentes coupures enzymatiques conduisant à deux voies distinctes : la voie amyloïdérique et la voie non amyloïdérique (Vetrivel and Thinakaran, 2006).

#### Voie amyloïdérique

Durant la voie amyloïdérique, APP est successivement clivée par la  $\beta$ -sécrétase, qui correspond à l'enzyme BACE1 ( *$\beta$ -site APP Cleaving Enzyme 1*), puis par le complexe  $\gamma$ -sécrétase composé de 4 sous-unités distinctes : *Presinilin* (PSEN), *Nicastrin* (NCT), *Anterior Pharynx defective 1* (APH-1) et *PSEN enhancer 2* (Pen-2). Ces clivages successifs aboutissent à la formation de trois fragments que sont le fragment APP soluble  $\beta$ , l'AICD (*APP IntraCellular Domain*) et le peptide A $\beta$  (Figure 17 page 56).



**Figure 17 : Métabolisme d'APP**

Figure adaptée de Vetrivel and Thinakaran, 2006.

APP suit deux voies métaboliques distinctes : la voie non-amyloïdérique (A) et la voie amyloïdérique (B). La voie non-amyloïdérique permet le clivage d'APP par l' $\alpha$ -sécrétase et le complexe  $\gamma$ -sécrétase entraînant la production du fragment sAPP $\alpha$ , le peptide p3 et l'AICD. Au cours de la voie amyloïdérique, APP est successivement clivée par la  $\beta$ -sécrétase (BACE1) et le complexe  $\gamma$ -sécrétase produisant le fragment sAPP $\beta$ , le peptide A $\beta$  et l'AICD.

Bien qu'ubiquitaire, l'enzyme BACE1 est majoritairement exprimée dans le cerveau et le pancréas (Vassar, 2004). Cette enzyme peut générer différents fragments car elle peut cliver APP en deux positions distinctes : au niveau de l'acide aspartique 1 (Asp-1), le premier acide aminé du peptide amyloïde ou au niveau de l'acide glutamique 11 (Glu-11) (Vassar et al., 1999). La coupure au niveau de l'Asp-1 est majoritaire et génère le fragment C99 aussi appelé APP-CTF $\beta$  (*APP $\beta$  C-terminal fragment*) et le fragment APP soluble  $\beta$ . La coupure au niveau de Glu-11 est plus minoritaire et entraîne la formation d'un fragment APP-CTF $\beta$  plus court et par conséquent d'un peptide amyloïde plus court (Vassar, 2004).

### *Complexe $\gamma$ -sécrétase*

Le fragment C99, majoritairement produit, est ensuite pris en charge par le complexe  $\gamma$ -sécrétase qui aboutit à la formation de l'AICD et du peptide A $\beta$ .

Il existe deux homologues de la préséniline, PSEN1 et PSEN2. Durant leur maturation, ces deux protéines d'environ 50 kDa sont clivées en deux fragments d'environ 30 et 20 kDa générant respectivement le fragment NTF (*Amino-terminal fragment*) et CTF (*Carboxy-terminal fragment*) (Thinakaran et al., 1996) tous deux impliqués dans la formation du complexe. Plusieurs études ont réussi à démontrer que la préséniline était la sous-unité responsable de l'activité catalytique du complexe  $\gamma$ -sécrétase (De Strooper et al., 1998; Wolfe et al., 1999). Cependant, cette enzyme ne peut agir seule et nécessite le recrutement des autres sous-unités que sont NCT, Pen-2 et APH-1 (Figure 18 page 58) (Bai et al., 2015).

Le clivage du fragment C99 par le complexe  $\gamma$ -sécrétase s'effectue en deux temps. Le premier clivage, également appelé clivage endopeptidase ( $\epsilon$ -clivage) libère l'AICD ainsi que le peptide A $\beta$  de 48 (A $\beta$ 48) ou 49 (A $\beta$ 49) AA selon la position exacte de la coupure (Takami et al., 2009). Après ce premier clivage, l'activité du complexe  $\gamma$ -sécrétase prend en charge le long peptide amyloïde et raccourcit consécutivement celui-ci selon deux voies distinctes : A $\beta$ 49-> A $\beta$ 46-> A $\beta$ 43-> A $\beta$ 40 ou A $\beta$ 48-> A $\beta$ 45-> A $\beta$ 42-> A $\beta$ 38 (Takami et al., 2009). Suite à ce clivage *carboxypeptidase-like*, les deux formes majoritairement produites sont les peptides A $\beta$ 40 et A $\beta$ 42 (Figure 19 page 58).

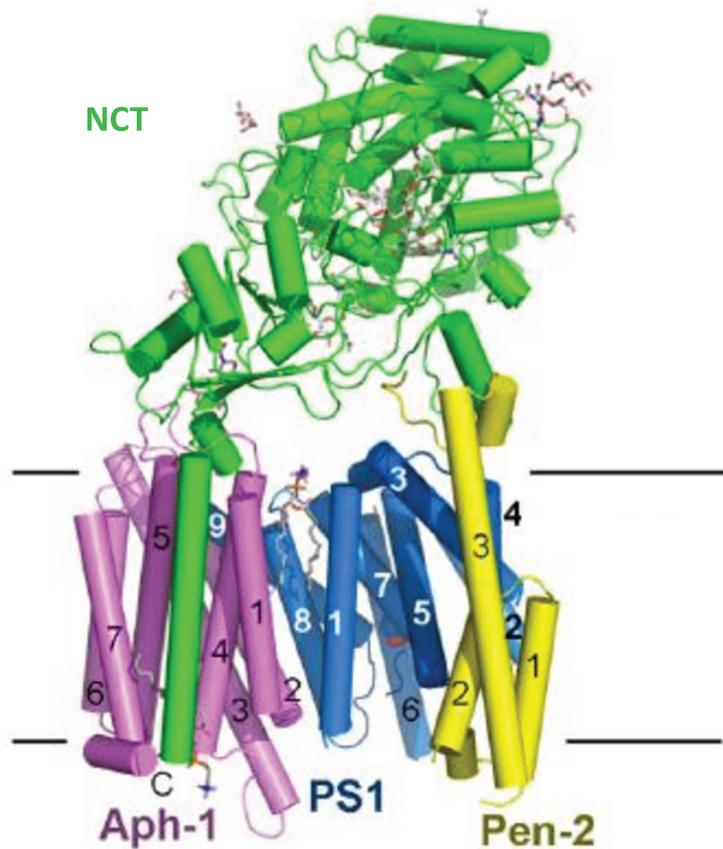


Figure 18 : Structure tridimensionnelle du complexe  $\gamma$ -sécrétase

Figure issue de Bai et al., 2016.

PS1 : Presenilin 1 ; NCT : Nicastrin ; Aph-1 : Anterior Pharynx-defective ; Presenilin enhancer 2 (Pen-2)

### Clivage endopeptidase



### Clivage carboxypeptidase

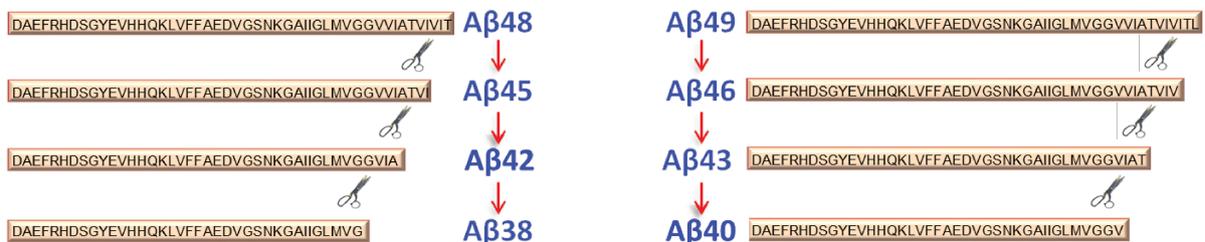


Figure 19 : Clivage endopeptidase et carboxypeptidase d'APP par le complexe  $\gamma$ -sécrétase

Le clivage endopeptidase permet de libérer un peptide Aβ de 48 ou 49 acides aminés ainsi que l'AICD (APP Intracellular Domain). Le peptide Amyloïde est ensuite successivement raccourci grâce au clivage carboxypeptidase. Les deux formes majoritairement produites sont les peptides Aβ40 et Aβ42.

## Voie non-amyloïdergique

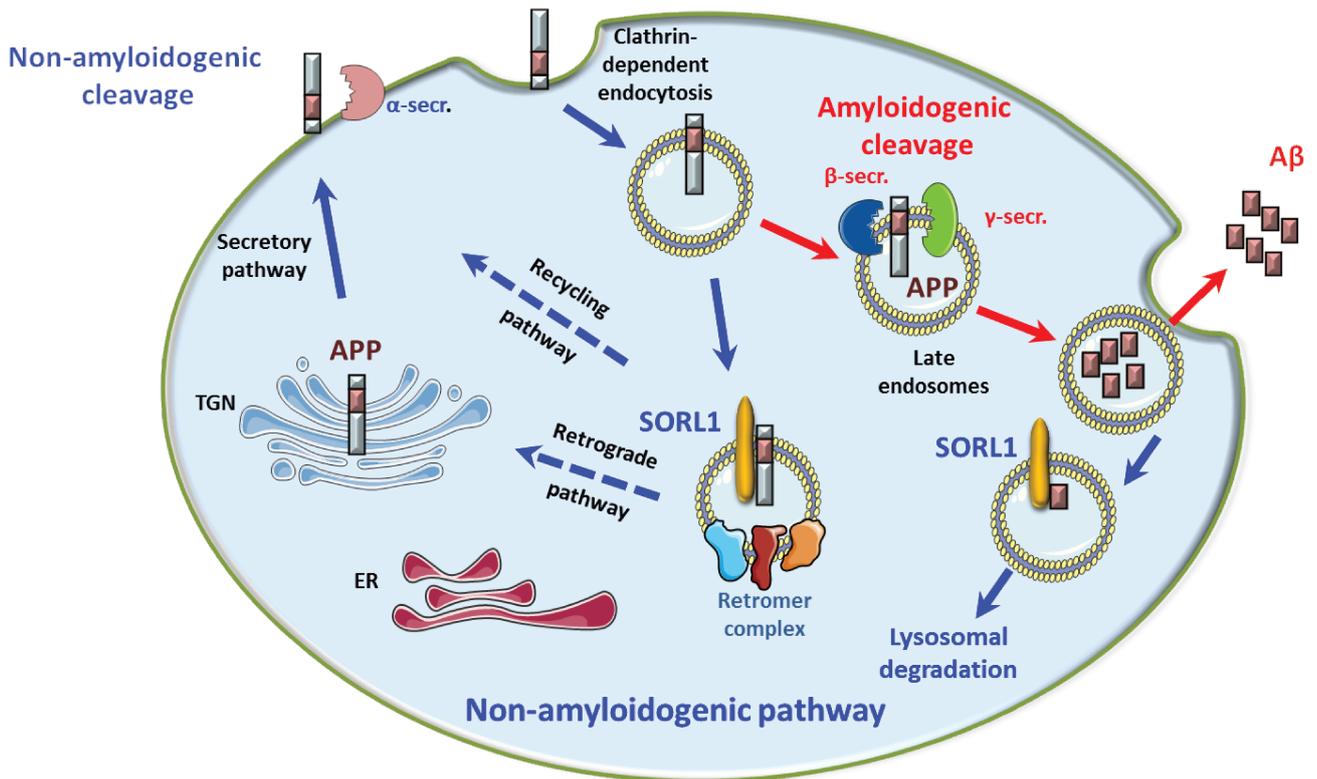
La voie non-amyloïdergique est la voie majoritaire de maturation d'APP et aboutit à la formation de trois produits distincts : le fragment APP soluble  $\alpha$ , l'AICD et le peptide p3. Ces produits sont formés grâce à l'activité successive de deux enzymes : l' $\alpha$ -sécrétase et le complexe  $\gamma$ -sécrétase (Figure 17 page 56).

L'activité  $\alpha$ -sécrétase a été attribuée à différentes protéines mais il semblerait que l'acteur majoritaire soit la protéine ADAM10 (Kuhn et al., 2010). Ce clivage s'effectue entre les résidus Lys-16 et Leu-17 du domaine A $\beta$  et produit deux fragments : C83 ou APP-CTF $\alpha$  (*APP $\alpha$  carboxy-terminal fragment*) et le fragment APP soluble  $\alpha$  (Sisodia et al., 1990). Le fragment C83 est alors clivé par le complexe  $\gamma$ -sécrétase libérant ainsi le peptide p3 et l'AICD.

## Trafic cellulaire d'APP

Les différentes étapes conduisant à la formation du peptide amyloïde à partir de son précurseur protéique sont localisées dans des régions subcellulaires particulières (pour revue (Haass et al., 2012)). La production d'A $\beta$  a essentiellement lieu au niveau du système endosomal bien que certaines études à la fin des années 90 aient documenté une présence d'A $\beta$  au niveau du reticulum endoplasmique et du réseau trans-golgien (Greenfield et al., 1999; Hartmann et al., 1997).

Après maturation d'APP au sein du reticulum endoplasmique et de l'appareil de golgi, APP suit la voie de sécrétion et est adressée à la membrane plasmique. À ce niveau, APP suit majoritairement la voie non-amyloïdergique et est clivée successivement par l' $\alpha$ -sécrétase et le complexe  $\gamma$ -sécrétase. Cependant, la présence de la séquence consensus YENPTY dans le domaine intracellulaire d'APP permet également son endocytose dans les endosomes précoces. À ce niveau, APP peut être redirigée vers la voie non-amyloïdergique en empruntant la voie de recyclage vers la membrane plasmique ou le réseau trans-golgien via le complexe du rétromère (Seaman, 2012). Le rétromère est constitué d'une combinaison de cinq protéines différentes dont un trimère de protéine appelé VPS (*Vacuolar Protein Sorting* VPS35, VPS26 et VPS29) et de deux sous-unités SNX (*Sorting Nexin* SNX1, SNX2, SNX5 ou SNX6) (Seaman, 2012). SORL1 est un acteur majeur de cette voie de recyclage. En effet, il permet de faire le lien entre APP et le rétromère (pour revue (Reitz, 2015)). Ainsi, APP est recyclée, retourne à la membrane plasmique ou au niveau du réseau trans-golgien pour suivre la voie non-amyloïdergique. Dans le cas contraire, APP est délivrée aux endosomes précoces, organelles majeurs, grâce à leur pH acide, pour l'activité de la  $\beta$ -sécrétase qui clive alors APP et libère le C99. Le fragment C99 est ensuite dirigé depuis les endosomes précoces vers les endosomes tardifs puis les lysosomes. Le lieu précis d'activité du complexe  $\gamma$ -sécrétase n'est pas clairement défini mais interviendrait principalement dans les endosomes tardifs et/ou lysosomes. Le peptide amyloïde est ensuite sécrété au niveau extracellulaire (Figure 20 page 60) (pour revue (Haass et al., 2012)).



**Figure 20 : Trafic cellulaire d'APP**

Figure issue de Campion *et al.*, 2016.

APP suit un parcours précis au sein de la cellule. Après maturation dans le reticulum endoplasmique et l'appareil de Golgi, la protéine APP est adressée à la membrane plasmique par la voie de sécrétion. Au niveau de la membrane plasmique, APP peut suivre la voie non-amyloïdergique et être clivée successivement par l' $\alpha$ -sécrétase puis la  $\gamma$ -sécrétase. Elle peut également être endocytée dans les endosomes précoces. Le complexe du rétromère, *via* SORL1, peut lier APP dans ces endosomes pour le recycler vers le réseau trans-golgien ou la membrane plasmique prévenant ainsi le clivage amyloïdergique. Dans le compartiment endosomal et lysosomal, APP est successivement clivée par la  $\beta$ -sécrétase puis la  $\gamma$ -sécrétase aboutissant à la production et sécrétion du peptide amyloïde dans le milieu extracellulaire.

## Clairance dégradation

Afin de prévenir son agrégation en plaques séniles, le peptide amyloïde peut être éliminé par différentes voies : par protéolyse extracellulaire grâce à différentes enzymes (MME, IDE, MMP2 et MMP9, CTSB, BACE2...), par transport à travers la barrière hémato-encéphalique (BHE), par endocytose puis dégradation dans le système endosomal-lysosomal ou encore par phagocytose grâce à la microglie. La clairance d'A $\beta$  à travers la BHE s'effectue grâce à plusieurs molécules. L'APOE, la *clusterin* (CLU) ou encore l' *$\alpha$ -2-macroglobulin* (A2M) peuvent se lier à APOE pour le délivrer à la BHE. Les complexes ainsi formés se lient aux récepteurs LRP1 (*low density lipoprotein receptor-related protein-1*) ou LRP2 permettant le passage du peptide amyloïde à travers la BHE. Ces voies seront décrites plus en détails dans les résultats partie 2.

## Agrégation

L'agrégation physiopathologique du peptide amyloïde peut résulter de plusieurs dysfonctions dans les différents mécanismes évoqués précédemment. Les plaques séniles ont été les premières lésions décrites et ne sont retrouvées que dans les cerveaux de patients Alzheimer. Cependant, le paradigme actuel considère que les oligomères d'A $\beta$  sont les responsables de la toxicité physiopathologique observée dans la MA et non les fibrilles retrouvées dans les plaques séniles (Benilova et al., 2012) Ces voies seront décrites plus en détails dans les résultats partie 2.

## **f. Mécanismes neurotoxiques du peptide amyloïde**

La toxicité d'A $\beta$  serait médiée par différentes voies affectant la plasticité synaptique (Laurén et al., 2009; Li et al., 2011; Shankar et al., 2008). La capacité des neurones à maintenir une homéostasie synaptique serait fortement corrélée aux fonctions de mémoires et d'apprentissage. Les deux principaux mécanismes de cette plasticité synaptique sont la LTP (*Long Term Potentiation*) et LTD (*Long -Term Depression*). Il est maintenant majoritairement admis que les oligomères d'A $\beta$  ont la capacité de bloquer la LTP (Li et al., 2011) et il semblerait qu'ils soient capables d'entraîner la LTD (Guntupalli et al., 2016; Shankar et al., 2008). Les oligomères d'A $\beta$  seraient capables de se lier à plusieurs récepteurs cellulaires tels que PrP<sup>C</sup> (*cellular prion protein*) (Laurén et al., 2009), ou CHRNA7 (*Cholinergic Receptor Nicotinic Alpha 7 Subunit*) (Wang et al., 2000) mais seraient également capable de s'agréger à la surface cellulaire au niveau des rafts lipidiques créant ainsi des pores membranaires (Kawarabayashi et al., 2004; Nisbet et al., 2015). L'interaction des oligomères avec ces récepteurs auraient plusieurs conséquences. Elles aboutiraient *in fine* à l'internalisation des récepteurs NMDA (*N-methyl\_D-aspartate*) et AMPA ( *$\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid*) entraînant un mécanisme de LTD et en conséquence un rétrécissement et une rétraction des épines dendritiques (voir revue Nisbet et al., 2015). De manière intéressante, le processus d'internalisation

des récepteurs AMPA serait médié par une activation des récepteurs NMDA induite par les oligomères d'A $\beta$  permettant une entrée massive de calcium qui entraînerait plusieurs cascades de régulation impliquant des phosphatases (Calcineurine, PP1 and STEP61) et des kinases (CDK5, PKC et GSK3 $\beta$ ) (Guntupalli et al., 2016). Les phosphatases calcineurine et PP1 entraîneraient l'internalisation des récepteurs AMPA en déphosphorylant la serine 845 de la sous-unité GluA1 de ce récepteur, (Beattie et al., 2000; Ehlers, 2000), site crucial dans le maintien de la stabilité de ces récepteurs au niveau de la synapse ainsi que la LTP (He et al., 2009; Oh et al., 2006). En résumé, les oligomères d'A $\beta$  entraîneraient de manière indirecte une activation des récepteurs NMDA entraînant l'internalisation des récepteurs AMPA *via* une entrée de calcium intracellulaire et une déphosphorylation de ces récepteurs entraînant un mécanisme de LTD. A $\beta$  aurait donc un rôle majeur dans le maintien de l'homéostasie synaptique en prévenant un excès d'activité synaptique qui serait néfaste pour la cellule. D'autre part, ces oligomères auraient également la capacité *via* leur liaison à CHRNA7 et PrP<sup>c</sup> d'augmenter l'internalisation des récepteurs NMDA (Snyder et al., 2005). Ce processus serait médié notamment par l'activation de la kinase Fyn qui augmente la phosphorylation de la sous unité NR2b des récepteurs NMDA entraînant de manière transitoire leur activation (revue (Guntupalli et al., 2016)). Cette activation des récepteurs NMDA serait rapidement suivie d'une augmentation des niveaux de la phosphatase de STEP61 qui déphosphoryle la sous-unité NR2B et entraîne l'internalisation de ces récepteurs. En résumé, les mécanismes de régulation synaptique induits par le peptide amyloïde semblent extrêmement complexes et impliquent plusieurs voies qui tentent d'établir une homéostasie synaptique (Figure 21 page 64).

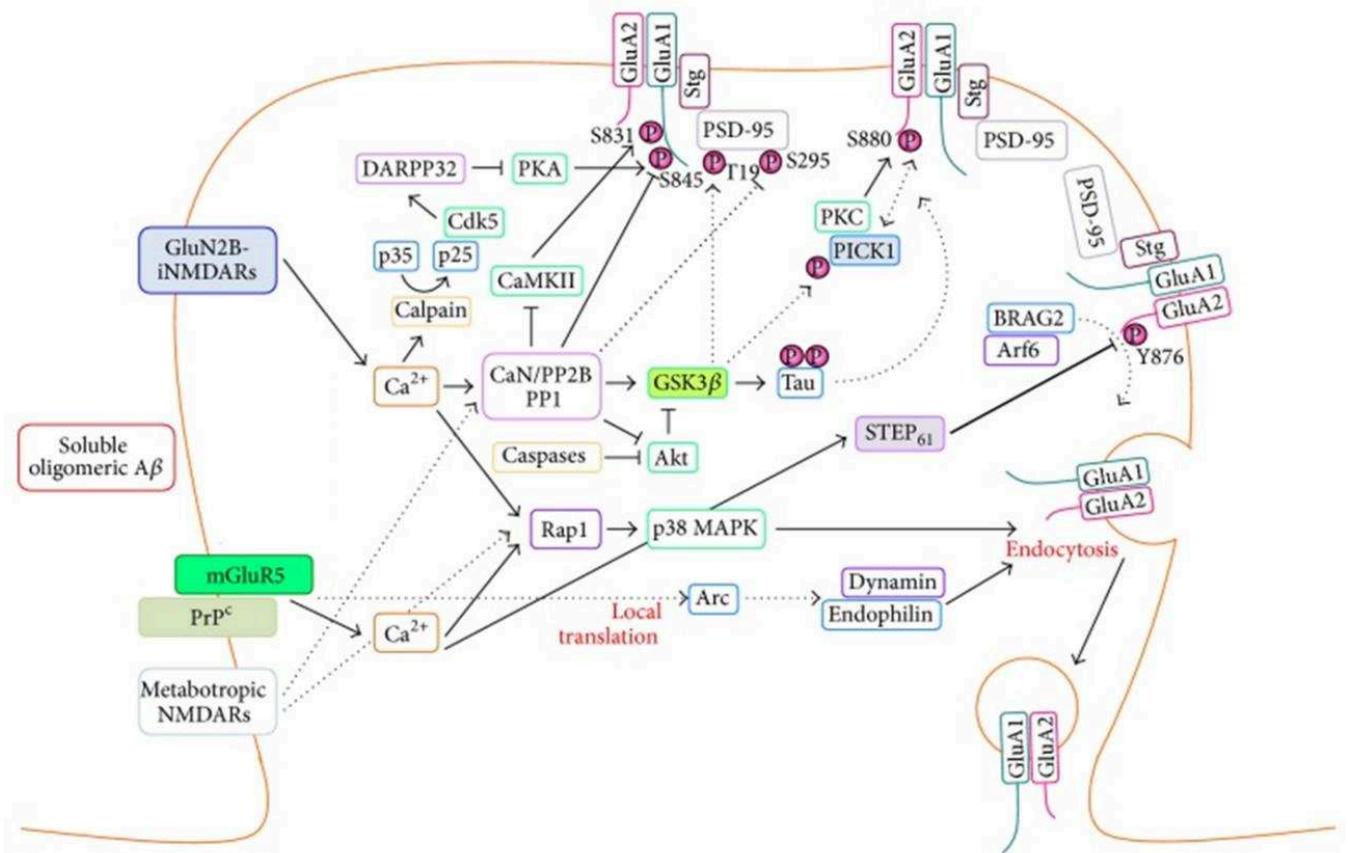
### ***g. Interaction Tau-A $\beta$***

La spécificité des plaques amyloïdes dans la maladie d'Alzheimer et la forte corrélation des DNFs avec les symptômes cliniques de la maladie ont longtemps alimenté la controverse pour définir qui était le déclencheur de la pathologie (Morris et al., 2014). Le paradigme majoritaire actuel consiste à définir, de façon imagée, A $\beta$  comme la gâchette et Tau la balle qui tue les neurones (Bloom, 2014). La toxicité entraînée par A $\beta$  aurait des conséquences post-traductionnelles sur Tau changeant ainsi sa conformation et son décrochage physiologique des MT et facilitant ainsi son agrégation sous formes de DNF pour aboutir à la mort neuronale (Nisbet et al., 2015). Les mécanismes intrinsèques reliant le peptide amyloïde à la protéine Tau ne sont pas encore totalement élucidés. De manière intéressante, le cortex trans-entorhinal est la première zone touchée par les DNFs et sont observées à la fois chez les individus cognitivement sains et atteints de MA (Braak and Braak, 1991). La progression de ces DNFs peut même atteindre l'hippocampe chez certains individus âgés ne présentant pas de pathologie amyloïde (PART) (Crary et al., 2014). D'autre part, les dépôts diffus de peptides amyloïdes sont observés dans le cerveau de personnes saines et sont majoritairement produits au niveau

cortical (Duyckaerts et al., 1999; Thal et al., 2002). Cette progression lente et limitée des DNFs au sein du cerveau de personnes saines s'accompagne en parallèle de la production de dépôts diffus dans des zones physiquement éloignées. Dans un processus physiopathologique, on imagine que ces deux phénomènes finissent par se rencontrer favorisant réciproquement leurs activités dégénératives.

Plusieurs études ont essayé de comprendre cette interaction pathogène entre Tau et A $\beta$ . Les premières études ont permis de démontrer que la toxicité d'A $\beta$  était dépendante de Tau. En effet, l'exposition de neurones hippocampiques de souris transgéniques à des fibrilles d'A $\beta$  aboutissait à une dégénérescence des neurones (Rapoport et al., 2002). Cependant, cette dégénérescence n'était pas observée lorsque les cellules n'exprimaient pas la protéine Tau (Rapoport et al., 2002). Les mécanismes intrinsèques de cette relation entre A $\beta$  et Tau ont fait l'objet de nombreuses études (pour revue Nisbet et al., 2015).

Dans la MA et d'autres tauopathies, l'hyperphosphorylation de Tau entrainerait son décrochage des MT, altérant sa localisation subcellulaire et facilitant son agrégation. Cette hyperphosphorylation de Tau dépend de nombreuses kinases telles que GSK3 $\beta$ , CDK5, AMPK, ou encore les MARKs (revue Guntupalli et al., 2016). Lors de la régulation synaptique, l'activation des récepteurs d'A $\beta$  induit plusieurs kinases impliquées dans l'hyperphosphorylation de Tau (Hartigan and Johnson, 1999; Lee et al., 2000; Mairet-Coello et al., 2013; Yu et al., 2012) participant finalement à son décrochage des MT et accompagnant ainsi la rétraction des épines dendritiques initiée par l'internalisation des récepteurs AMPA et la LTD. On peut alors émettre l'hypothèse que cette disparition progressive des dendrites en condition pathologique, induite par les oligomères d'A $\beta$ , entrainerait *in fine* une perte de la synapse (Shankar et al., 2007). Ce mécanisme physiologique permettant de réguler l'excitabilité d'une synapse serait déstabilisé en condition pathologique aboutissant à une perte massive des synapses accompagnée d'une hyperphosphorylation et d'une agrégation de la protéine Tau conduisant à terme à la mort des neurones.



**Figure 21: Mécanismes neurotoxiques du peptide amyloïde**

Figure adaptée de la revue Guntupalli et al., 2016

Les oligomères solubles du peptide amyloïde activent les récepteurs NMDA entraînant l'augmentation du calcium ( $\text{Ca}^{2+}$ ) intracellulaire. L'entrée du  $\text{Ca}^{2+}$  dans la cellule active de nombreuses cascades de transduction impliquant des kinases et des phosphatases qui modulent l'état de phosphorylation des récepteurs AMPA. L'activation de ces voies, en lien avec les mécanismes de LTD, entraînent l'internalisation des AMPARs et une dépression synaptique.

Les flèches en pointillé indiquent les voies de signalisation découlant des études sur la LTD mais qui ne sont pas directement activées par  $\text{A}\beta$ . Les flèches continues sont les voies biologiques courantes.

### 3. Génétique de la maladie d'Alzheimer

La maladie d'Alzheimer répond dans une majorité des cas à un déterminisme multifactoriel dont le principal facteur de risque est l'âge (Hirtz et al., 2007). Cependant, une forme mendélienne rare à transmission autosomique dominante de MA touche environ 1 000 personnes en France. La caractérisation des gènes responsables de cette forme de MA a permis de formuler l'hypothèse de la cascade amyloïde et a placé le peptide amyloïde au centre du processus physiopathologique.

#### a. Formes mendéliennes de la MA

##### *Mutations et duplication d'APP*

À ce jour, 27 variations pathogènes du gène *APP* sont recensées dans la base de données de mutations d'Alzforum (Figure 22 page 67). La plupart des mutations d'*APP* sont des variations faux-sens retrouvées dans les exons 16 et 17 qui codent pour le peptide amyloïde. Les principales conséquences de ces variations sont multiples. Elles peuvent entraîner une augmentation de production du peptide amyloïde en favorisant le clivage  $\beta$ -sécrétase (Cai et al., 1993; Citron et al., 1992). D'autres variations aboutissent à une augmentation des capacités d'agrégation du peptide amyloïde en altérant le clivage  $\gamma$ -sécrétase favorisant la production de peptides amyloïdes plus longs (De Jonghe et al., 2001). De la même manière, des mutations dans la séquence codante du peptide  $A\beta$  entraînent également des modifications des propriétés biophysiques du peptide favorisant son agrégation (Nilsberth et al., 2001).

Les duplications d'*APP* décrites pour la première fois en 2006 (Rovelet-Lecrux et al., 2006) entraînent une augmentation d'expression d'*APP* et *in fine* une surproduction du peptide amyloïde (Pottier et al., 2012a).

##### *Mutations et délétions de PSEN1*

Avec 221 variations pathogènes, *PSEN1* est le gène majoritairement impliqué dans les formes mendéliennes de MA (Figure 22 page 67). Il représente à lui seul 64 % des variations responsables de ces formes mendéliennes de MA dans notre échantillon comportant 170 familles ADEOAD (Lanoiselée et al., 2017). On retrouve principalement des variations faux-sens mais également des délétions de l'exon 9 résultant d'un réarrangement génomique et d'une mutation ponctuelle affectant l'épissage (Crook et al., 1998; Dumanchin et al., 2006; Perez-Tur et al., 1995; Rovelet-Lecrux et al., 2015; Sato et al., 1998; Smith et al., 2001). Les variations se répartissent sur l'ensemble de la séquence de *PSEN1* mais ont pourtant toutes la même conséquence fonctionnelle sur l'activité  $\gamma$ -sécrétase (Chávez-Gutiérrez et al., 2012; Szaruga et al., 2015; Veugelen et al., 2016). En effet, si seulement certaines affectent le clivage endopeptidase, il semble que l'ensemble des variations

pathogènes de *PSEN1* affectent le clivage carboxypeptidase (Figure 19 page 58) (Chávez-Gutiérrez et al., 2012; Szaruga et al., 2015; Veugelen et al., 2016). Cette dysfonction du complexe  $\gamma$ -sécrétase est donc caractérisée par une perte de fonction de l'activité carboxypeptidase entraînant une augmentation de la production de formes plus longues d'A $\beta$  tels que les peptides amyloïdes 42 ou 43. Concrètement, le raccourcissement successif du peptide amyloïde réalisé par le clivage carboxypeptidase s'arrête précocement produisant ainsi des formes plus longues de peptides amyloïdes (Chávez-Gutiérrez et al., 2012). Cette dysfonction peut être évaluée par l'analyse des ratios A $\beta$ 40/A $\beta$ 43 ou A $\beta$ 38/A $\beta$ 42. L'effet des variations pathogènes *PSEN1* sur le complexe  $\gamma$ -sécrétase caractérisé par une perte de fonction a fait l'objet de controverses au sein de la communauté scientifique spécialiste de cette enzyme (Chávez-Gutiérrez et al., 2012; Kelleher and Shen, 2010; Shen and Kelleher, 2007; Veugelen et al., 2016; Xia et al., 2016). En effet, des souris *knock-in* contenant des mutations de *PSEN1* responsables d'EOAD présentaient un phénotype similaire à celui de souris déficientes pour *PSEN1* (Xia et al., 2015). Le complexe  $\gamma$ -sécrétase clive d'autres protéines telles que le récepteur Notch (Logeat et al., 1998). Ce récepteur joue un rôle dans la fonction neuronale et le développement (Xia et al., 2015). Ces études suggèrent donc un mécanisme physiopathologique en faveur d'une perte de fonction globale de *PSEN1* et peut amener à considérer l'agrégation du peptide amyloïde comme un épiphénomène (Kelleher and Shen, 2010). Cependant, des variations perte de fonction telles que des mutations non-sens de *PSEN1* ne sont pas responsables de MA mais d'une autre pathologie dermatologique sans lien avec une maladie neurodégénérative nommée acné inversée (Wang et al., 2010). Par ailleurs, les mutations et duplications d'*APP* affectent directement le peptide amyloïde sans conséquence sur l'activité globale du complexe  $\gamma$ -sécrétase. L'hypothèse centrée sur l'accumulation du peptide amyloïde plutôt que sur une perte de fonction de *PSEN1* reste donc la plus probable.

### **Mutations de *PSEN2***

Les variations de *PSEN2* sont plus rares, on en recense 18 à l'heure actuelle (Figure 22 page 67) d'après la base de données Alzforum. De la même manière que pour *PSEN1* ces variations semblent entraîner une augmentation de production des formes plus longues établie par le ratio A $\beta$ 42/A $\beta$ 40 qui bien qu'incorrect d'un point de vue fonctionnel (les ratios A $\beta$ 40/A $\beta$ 43 ou A $\beta$ 38/A $\beta$ 42 sont plus appropriés) traduit une dysfonction du clivage carboxypeptidase du complexe  $\gamma$ -sécrétase (Citron et al., 1997; Oyama et al., 1998; Sawamura et al., 2000; Scheuner et al., 1996). Une étude récente a montré que *PSEN2* était localisé majoritairement au niveau du compartiments endosomal tardif et du compartiment lysosomal contrairement à *PSEN1* qui est plus largement distribuée dans la cellule (Sannerud et al., 2016). Cette localisation subcellulaire spécifique entrainerait la formation d'un pool

intracellulaire d'Aβ qui pourrait être un évènement précoce dans la cascade physiopathologique de la MA.

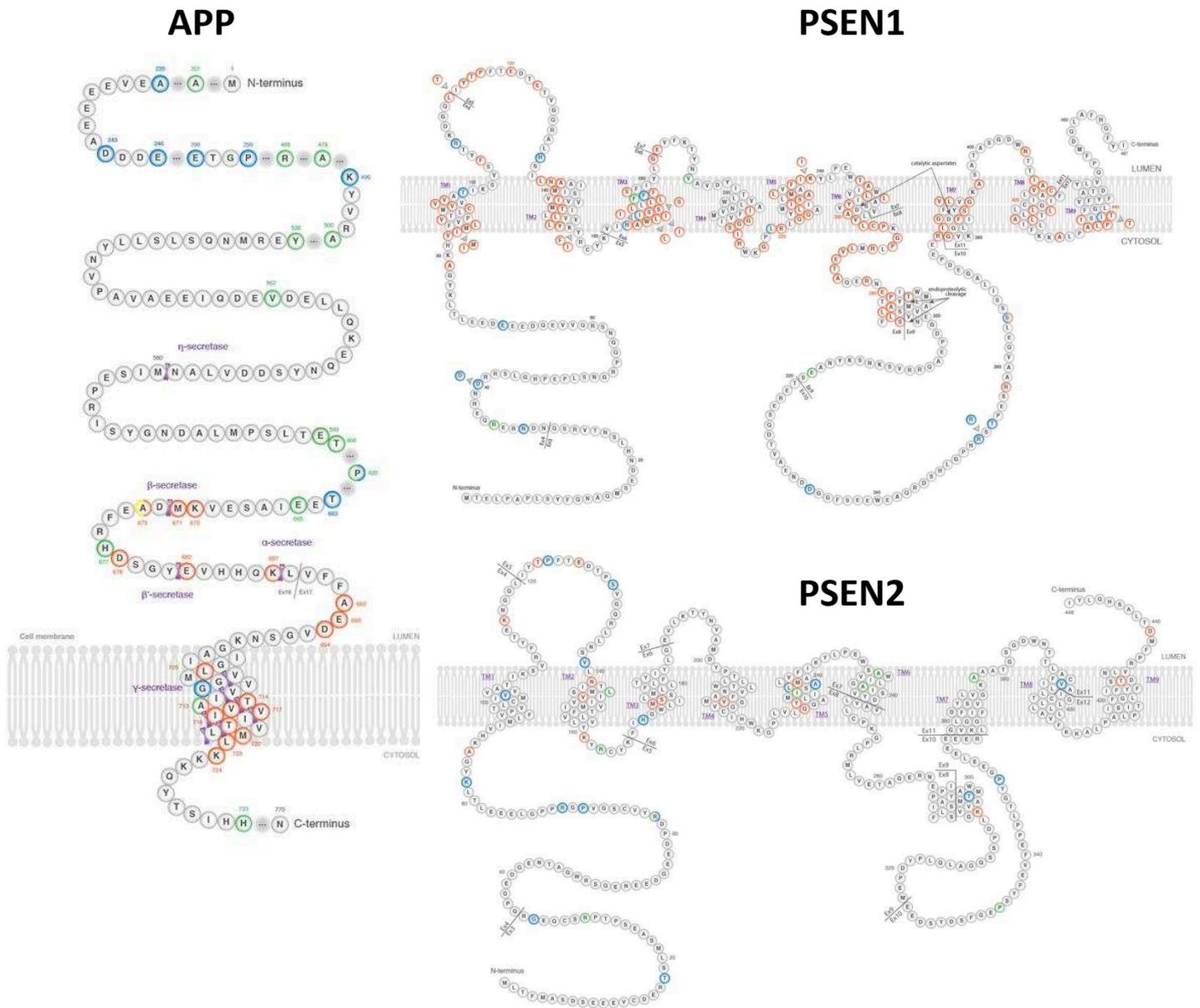


Figure 22 : Les variations d' APP, PSEN1 et PSEN2

Figure issue de la base de données Alzforum

Ensemble des variations pathogènes (en rouge), non pathogènes (en vert), protectrices (en jaune) et de signification inconnue (en bleu) dans APP, PSEN1 et PSEN2

## b. Facteurs de risque fréquents

Les mutations des gènes responsables d'EOAD ne permettent pas de caractériser l'ensemble de l'hétérogénéité génétique de la maladie d'Alzheimer. Une part de la composante génétique de la MA est attribuée à des facteurs de risque fréquents. Les variants fréquents confèrent principalement un risque faible de développer la maladie. De manière exceptionnelle, un facteur de risque fréquent, l'allèle  $\epsilon 4$  de l'APOE, est associé à un risque important de développer la MA.

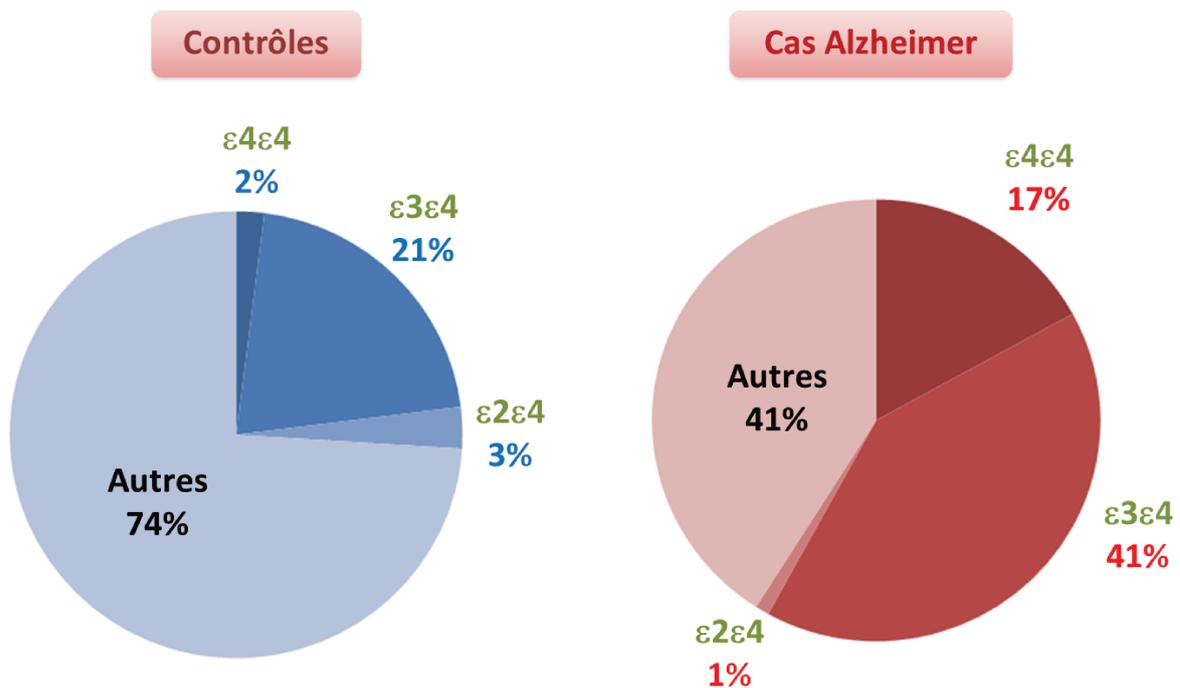


Figure 23 : Répartition de l'allèle  $\epsilon 4$  de l'APOE

La fréquence de l'allèle  $\epsilon 4$  de l'APOE a été calculée à partir d'un échantillon de 7 351 patients atteints de MA et d'un échantillon de 10 132 contrôles sains.(Gélin et al., 2011.)

### APOE

L'allèle  $\epsilon 4$  de l'apolipoprotéine E (APOE) est le facteur de risque majeur de la maladie d'Alzheimer. Initialement détecté en 1987 (Schellenberg et al., 1987), cet allèle est caractérisé par deux SNPs en déséquilibre de liaison sur APOE (NM\_000041)(rs429358 (c.388T>C) et rs7412 (c.526C>T)). Ces deux polymorphismes définissent ainsi 3 allèles différents :  $\epsilon 2$  (rs429358-T ; rs7412-T),  $\epsilon 3$  (rs429358-T ; rs7412-C) et  $\epsilon 4$  (rs429358-C ; rs7412-C). L'allèle  $\epsilon 4$  est relativement fréquent en population générale avec une MAF de 12,3 % (Lambert et al., 2013a) (Figure 23 page 68). Cet allèle est donc un exemple exceptionnel de variant fréquent conférant un risque élevé de développer la maladie. Le risque associé aux porteurs homozygotes  $\epsilon 4\epsilon 4$  et hétérozygotes  $\epsilon 3\epsilon 4$  par rapport aux porteurs homozygotes  $\epsilon 3\epsilon 3$  est respectivement de OR=14,4 et de OR=3,63 (Genin et al., 2011). De plus, si l'on ne considère pas le génotype APOE le risque de développer la maladie à 85 ans est de 11 % pour les hommes et de 14 % pour les femmes par contre il est de 23 % et 31 % pour les porteurs

hétérozygotes  $\epsilon 3\epsilon 4$  et de 51 % et 60 % pour les porteurs homozygotes (Genin et al., 2011). Il est intéressant de noter également que l'allèle  $\epsilon 2$  est protecteur atteignant un OR de 0,56. Les études pour relier le génotype d'APOE à la physiopathologie de la maladie d'Alzheimer sont extrêmement nombreuses. La caractérisation extensive de ce transporteur lipidique a permis de l'impliquer dans différentes voies métaboliques impliquant le peptide amyloïde. Il interviendrait ainsi principalement dans son agrégation, sa clairance (pour revue voir (Yu et al., 2014) et également dans la maturation d'APP et la sécrétion d'A $\beta$  (Huang et al., 2017). APOE jouerait également un rôle dans divers mécanismes indépendants du peptide amyloïde : la plasticité synaptique, la neuroinflammation, la neurotoxicité, le métabolisme des lipides, la dysfonction mitochondriale etc. (pour revue (Yu et al., 2014)). Le nombre de fonctions attribuées à APOE complexifie la définition de son rôle précis dans le mécanisme physiopathologique de la MA.

### **Etude de GWAS**

Les études de GWAS ont permis de mettre en évidence plus de 20 *loci* associés avec la MA (Harold et al., 2009; Hollingworth et al., 2011; Jun et al., 2016; Lambert et al., 2009, 2013b, 2013a, Naj et al., 2011, 2017; Seshadri et al., 2010; Tosto et al., 2015) (Tableau 1 page 70) atteignant une significativité pangénomique ( $5 \cdot 10^{-8}$ ). Cependant, les odds ratio compris entre 1,08 et 1,66 sont sans commune mesure avec celui de l'APOE (OR APOE  $\epsilon 3\epsilon 4$  =3,63 et OR  $\epsilon 4\epsilon 4$  =14,4). Ces facteurs de risque définis par les *loci* sont majoritairement des polymorphismes intergéniques ou introniques reliés avec le gène le plus proche. Des études fonctionnelles ont permis de documenter l'implication physiopathologique des gènes associés aux *loci* des études de GWAS (Chapuis et al., 2016; Delay et al., 2016; Dourlen et al., 2016; Sottejeau et al., 2015; Ubelmann et al., 2017; pour revue Van Cauwenberghe et al., 2016). Les gènes interagissant avec Tau, APP ou les deux à la fois sont indiqués dans le Tableau 1 (page 65). Bien que ces variants fréquents représentent individuellement un risque faible de développer la maladie, leur addition chez un même individu peut représenter un risque important notamment si les interactions génétiques de ces variants sont synergiques. De manière intéressante, certains gènes tels que *SORL1* et *ABCA7* associés à des facteurs de risque fréquents ont également été impliqués dans les études d'association de variants rares. Néanmoins, des facteurs de risque fréquents ne peuvent être en déséquilibre de liaison avec des variants rares. On estime donc que la part attribuée aux facteurs de risque fréquents sur ces gènes n'est pas la même que celle attribuée aux variants rares.

SNP	FAR	Gène associé	Localisation	OR	P valeur
rs6733839	0,41	<i>BIN1</i>	intergénique	1,22 [1,18-1,25]	6,9.10 <sup>-44</sup>
rs10792832	0,64	<i>PICALM</i>	intergénique	1,15 [1,12-1,18]	9,3.10 <sup>-26</sup>
rs9331896	0,62	<i>CLU</i>	intronique	1,16 [1,12-1,19]	2,8.10 <sup>-25</sup>
rs6656401	0,20	<i>CR1</i>	intronique	1,18 [1,14-1,22]	5,7.10 <sup>-24</sup>
rs983392	0,60	<i>MS4A6A</i>	intergénique	1,11 [1,08-1,15]	6,1.10 <sup>-16</sup>
rs4147929	0,19	<i>ABCA7</i>	intronique	1,15 [1,11-1,19]	1,1.10 <sup>-15</sup>
rs11218343	0,60	<i>SORL1</i>	intronique	1,30 [1,22- 1,39]	9,7.10 <sup>-15</sup>
rs28834970	0,37	<i>PTK2B</i>	intronique	1,10 [1,08-1,13]	7,4.10 <sup>-14</sup>
rs11771145	0,66	<i>EPHA1</i>	intergénique	1,11 [1,08-1,15]	1,1.10 <sup>-13</sup>
rs9271192	0,28	<i>HLA-DRB5</i>	intergénique	1,11 [1,08-1,15]	2,9.10 <sup>-12</sup>
rs10948363	0,27	<i>CD2AP</i>	intronique	1,10 [1,07-1,13]	5,2.10 <sup>-11</sup>
rs1476679	0,61	<i>ZCWPW1</i>	intronique	1,10 [1,07-1,13]	5,6.10 <sup>-10</sup>
rs75002042	0,92	<i>FBXL7</i>	intronique	1,66 [1,43-1,96]	4,7.10 <sup>-9</sup>
rs2718058	0,63	<i>NME8</i>	intergénique	1,08 [1,05-1,11]	4,8.10 <sup>-9</sup>
rs10498633	0,78	<i>SLC24A4/RIN3</i>	intronique	1,10 [1,07-1,13]	5,5.10 <sup>-9</sup>
rs2732703	0,87	<i>MAPT</i>	intergénique	1,37 [1,23-1,53]	5,8.10 <sup>-9</sup>
rs17125944	0,09	<i>FERMT2</i>	intronique	1,14 [1,09-1,19]	7,9.10 <sup>-9</sup>
rs10838725	0,31	<i>CELF1</i>	intronique	1,08 [1,05-1,11]	1,1.10 <sup>-8</sup>
rs7274581	0,92	<i>CASS4</i>	intronique	1,14 [1,09-1,19]	2,5.10 <sup>-8</sup>
rs35349669	0,49	<i>INPP5D</i>	intronique	1,08 [1,05-1,11]	3,2.10 <sup>-8</sup>
rs190982	0,59	<i>MEF2C</i>	intergénique	1,08 [1,05-1,11]	3,2.10 <sup>-8</sup>

**Tableau 1 : Gènes identifiés par les études de GWAS**

Ce tableau a été créé à partir de la base de données GWAS *catalog* (<https://www.ebi.ac.uk/gwas/>).

SNP : *Single nucleotide polymorphism* ; FAR : Fréquence de l'allèle à risque ; OR : odds ratio. Gènes en lien avec le peptide Aβ (en rouge), en lien avec Tau (orange) et lien avec Aβ et Tau (en bleu).

### c. Facteurs de risque rares

A côté de ces études de GWAS qui ont permis d'associer des variants fréquents avec la MA, le développement des techniques de séquençage haut débit a permis d'étudier l'implication de variants rares dans cette pathologie. Différentes études d'associations ont ainsi mis en lumière une partie de la composante génétique de la maladie d'Alzheimer médiée par des variants rares et qui confèrent un risque intermédiaire à élevé de développer la maladie. En fin d'année 2012, deux articles ont démontré l'association d'un variant rare à risque de *TREM2* (*Triggering Receptor Expressed on Myeloid cells 2*) (Guerreiro et al., 2013; Jonsson et al., 2013) qui faisait suite à la découverte d'un variant protecteur d'APP (Jonsson et al., 2012). Les études menées sur la population islandaise qui ont permis de détecter des associations significatives sur ces deux variants (Jonsson et al., 2012, 2013) ont également pu, grâce à une analyse singulière, impliquer des variants pertes de fonction cette fois-ci à l'échelle du gène codant la protéine ABCA7 (Steinberg et al., 2015). En parallèle, notre équipe était la première à identifier des variants rares de *SORL1* comme contribuant probablement au déterminisme de la MA du sujet jeune (Pottier et al., 2012b).

## TREM2

Deux études ont rapporté en même temps l'association d'un variant rare de *TREM2* p.R47H comme facteur de risque associé à la MA (Guerreiro et al., 2013; Jonsson et al., 2013). Bien qu'utilisant des techniques de NGS, les deux études ont appliqué des stratégies différentes. La première étude réalisée par l'équipe islandaise est la plus complexe. Dans un premier temps, ils ont détecté 191 777 variants non-synonymes, *frameshifts* ou affectant les sites canoniques d'épissage à partir de données de séquençage de génomes de 2 261 islandais contrôles. Dans un second temps, ils ont imputé ces 191 777 variants chez 3 550 patients atteints de MA et plus de 100 000 contrôles. Cette stratégie a permis de mettre en évidence deux variants associés avec la MA. Le premier est situé sur la séquence codante d'*APP* (A673T) et est protecteur atteignant un OR de 0,236 quand l'analyse est réalisée à partir de l'ensemble de la population contrôle et 0,189 quand elle est restreinte aux individus âgés de plus de 85 ans. Au niveau fonctionnel, ce variant affecte la production de peptide amyloïde en défavorisant le clivage  $\beta$ -sécrétase diminuant ainsi de près de 40 % la production de peptide amyloïde dans un modèle *in vitro* (Jonsson et al., 2012). Le second variant est une variation faux-sens du gène *TREM2* (p.R47H) qui confère un risque de développer la MA atteignant un OR de 2,92, lorsque l'on considère l'ensemble des individus contrôles, et de 4,66 quand l'analyse est restreinte aux individus contrôles de plus de 85 ans. L'étude de réplification dans une population européenne a abouti un OR similaire de 2,83. Cette association du variant rare de *TREM2* a été découverte au même moment par une équipe anglaise. Cette équipe, pilotée par John Hardy, a pu mettre en évidence à partir à la fois de données de génomes, d'exomes et de séquençage Sanger ciblé sur l'exon 2 de *TREM2* une association de ce variant rare en comparant environ 1 000 cas et 1 000 contrôles avec un OR de 4,51. Au sein du laboratoire, nous avons rapidement répliqué ce résultat chez 726 EOAD et 783 contrôles (OR=4,07) (Pottier et al., 2012b). D'autres études ont par la suite confirmé ce résultat (Rosenthal et al., 2015 à titre d'exemple) et certaines ont été capables de détecter une association à partir d'un autre variant de *TREM2*, p.R62H, ou encore en agrégeant les variants à l'échelle du gène (Jin et al., 2014). Cependant, ces deux dernières associations nécessitent une réplification dans des études cas-contrôles indépendantes. D'un point de vue fonctionnel, le mécanisme par lequel la protéine *TREM2* interviendrait dans la physiopathologie de la MA serait multiple. Cette protéine est un récepteur présent à la surface cellulaire des cellules microgliales et jouerait un rôle dans la dégradation d'A $\beta$  en entraînant la phagocytose du peptide amyloïde par ces mêmes cellules (Kleinberger et al., 2014). Le lien entre ce variant et la physiopathologie reposerait sur une dysfonction microgliale. Cependant, les résultats obtenus à partir de différents modèles souris de MA sont discordants. Certaines études n'observent aucun effet sur la charge en plaques amyloïdes chez des modèles souris de MA déficientes pour *TREM2* alors que d'autres études semblent indiquer le contraire (Jay et al., 2015; Ulrich et al., 2014; Wang et al., 2015b). Récemment,

une étude a mis en évidence une fonction de protection de la microglie qui formerait une barrière microgliale autour des plaques et dépôts amyloïdes favorisant leur compaction et prévenant ainsi leur effet neurotoxique (Condello et al., 2015). L'haploinsuffisance de *TREM2* et les analyses de cerveaux de patients malades porteurs de la variation R47H semble démontrer que cette barrière protectrice est altérée (Yuan et al., 2016) (Wang et al., 2016). *TREM2* est aussi produit sous une forme soluble qui favoriserait l'activation des cellules microgliales et l'inflammation, fonctions altérées lorsque la protéine est mutée (Zhong et al., 2017). Toutes les études s'accordent à dire que la variation p.R47H agirait comme une variation perte de fonction. Plus récemment, une méta-analyse réalisée par Alzforum a permis de mettre en évidence une élévation des niveaux de formes solubles de *TREM2* dans le liquide céphalo-rachidien de patients malades permettant éventuellement d'utiliser dans le futur ce biomarqueur pour établir le diagnostic de MA (Henjum et al., 2016; Heslegrave et al., 2016; Piccio et al., 2016; Suárez-Calvet et al., 2016).

### ***SORL1***

Grâce à une stratégie de récurrence, notre laboratoire a été le premier à identifier des variants rares de *SORL1* chez des patients atteints de formes familiales précoces de MA compatible avec un mode de transmission autosomique dominant (Pottier et al., 2012b). Récemment, nous avons réalisé une étude d'association à partir de données d'exomes obtenues chez 484 cas EOAD et 498 contrôles. L'analyse de ces données a permis de valider les résultats initiaux révélant un enrichissement de variants rares assimilés à des pertes de fonction dans le gène *SORL1* chez les cas, proche de la significativité nécessaire en exome (OR=5,03, p-valeur=7,49.10<sup>-5</sup>). L'analyse restreinte aux individus présentant une forme familiale d'EOAD avait permis de franchir ce seuil de significativité (OR=8,86, p-valeur=3,8.10<sup>-7</sup>) (Nicolas et al., 2016b). Une autre étude d'association réalisée chez des patients d'origine caribéenne-hispanique atteints de formes tardives de MA avait également détecté un signal similaire sur des variants de *SORL1* mais sans franchir le seuil de significativité de l'exome (p-valeur=9.10<sup>-4</sup>) (Vardarajan et al., 2015). Cependant, cette analyse d'association n'a pas été réalisée uniquement à partir de données de séquençage d'exomes. En effet, seuls les variants de *SORL1* détectés chez différents cas familiaux avaient été génotypés chez 498 contrôles non-apparentés. Cette analyse ne prend donc pas en compte les variants rares potentiellement portés par les individus contrôles qui pourraient entraîner un rééquilibrage de la charge. Très récemment, 3 cas d'une même famille, porteurs à l'état homozygote de l'allèle  $\epsilon 4$  d'*APOE*, portaient également un variant rare faux-sens de *SORL1* (Louwersheimer et al., 2017). La co-occurrence de l'allèle  $\epsilon 4$  d'*APOE* et du variant rare de *SORL1* influencerait l'âge de début chez ces patients. Cependant, cette hypothèse n'a été testée que dans une seule famille et nécessite d'être répliquée.

Au niveau fonctionnel, *SORL1* (*Sortilin Related Receptor 1*) code pour une protéine cargo impliquée dans le recyclage d'APP, via le complexe du rétromère, à la membrane plasmique ou au niveau du réseau trans-golgien favorisant ainsi la voie non-amyloïdérique (Figure 20 page 60) (pour revue (Andersen et al., 2016)). Des études menées sur des modèles de souris Alzheimer déficientes pour *SORL1* ont permis de démontrer que cette perte de fonction exacerbait la pathologie amyloïde (Andersen et al., 2005; Dodson et al., 2008). De façon plus récente, une équipe a impliqué *SORL1* dans l'adressage du peptide amyloïde au lysosome favorisant sa dégradation (Caglayan et al., 2014; Dumanis et al., 2015). Il existerait donc deux fonctions protectrices de *SORL1* impliquant la production du peptide amyloïde. Ces études fonctionnelles sont en accord avec les études cas-contrôles qui associent les variants PTVs et les variants faux-sens assimilés comme tels avec le risque de développer la MA (Nicolas et al., 2016b; Pottier et al., 2012a).

### **ABCA7**

Suite à l'étude des variants rares d'*APP* et *TREM2*, la méthodologie de l'équipe Islandaise, détection puis imputation, a permis d'associer des variants rares perte de fonction d'*ABCA7* avec la physiopathologie de la MA (Steinberg et al., 2015). Dans cette étude, aucun variant seul ne franchissait le seuil de significativité nécessaire. Pour contourner ce problème, la stratégie a consisté en l'agrégation des variants à l'échelle du gène en distinguant deux catégories : les variations rares PTV d'une part et les variations faux-sens et d'épissage en dehors des sites canoniques d'autre part. Cette méthodologie permettait de s'approcher du seuil de significativité. Concernant les variants perte de fonction, le risque estimé par l'OR était de 2,08 avec une significativité approchant celle pan-exomique ( $p\text{-valeur}=3,8.10^{-5}$ ). Le signal d'association obtenu grâce à l'analyse statistique de variance SKAT ( $p\text{-valeur}=2.10^{-4}$ ) était en réalité largement dû à un seul variant d'épissage, c.5570+5G>C ( $p\text{-valeur}=0,46$  sans ce variant). L'analyse de l'ARN des patients porteurs a permis de mettre en évidence que cette variation entraînait la rétention d'une partie de l'intron aboutissant à l'inclusion d'un codon stop dans la séquence et a pu donc être recatégorisée en variant perte de fonction. L'analyse ainsi réalisée à partir des variants perte de fonction et de ce variant d'épissage a permis aux auteurs d'obtenir une significativité pan-exomique (OR=2,12,  $p\text{-valeur}=2,2.10^{-13}$ ). L'association a été également répliquée en génotypant la population européenne précédemment utilisée dans les études sur *TREM2* aboutissant à un OR combiné de 2,03 ( $p\text{-valeur}=6,8.10^{-15}$ ). Cette très large population islandaise séquencée permet d'obtenir une très grande puissance statistique. Cependant, la méthodologie limite l'analyse à un nombre restreint de variants découverts dans leur échantillon d'origine composé d'individus sains. Le séquençage ciblé par techniques de NGS réalisé par une équipe belge a permis de mettre en évidence l'implication de l'ensemble des variants rares PTV d'*ABCA7* dans la MA (RR=4,03,  $p\text{-valeur}=2.10^{-4}$ ) (Cuyvers et al., 2015). Cette association n'est pas

significative d'un point de vue de l'exome. Au cours de ma thèse, nous avons alors tenté de réaliser une méta-analyse en combinant nos données avec celles de l'étude belge pour augmenter le nombre d'individus et la puissance statistique de notre analyse afin de détecter un signal significatif sur les variants rares perte de fonction et faux-sens prédits pathogènes (Résultats Partie 1).

*ABCA7* (*ATP-Binding Cassette transporter A7*) est une protéine exprimée dans le cerveau et majoritairement dans les cellules microgliales et les neurones (Kim et al., 2006). Cette protéine transmembranaire est impliquée dans le transport de différents substrats (principalement des lipides) à travers la membrane plasmique et la membrane des organelles (Vasiliou et al., 2009). D'un point de vue physiopathologique, la déplétion d'*ABCA7* dans des souris modèles de MA exacerbe la pathologie en doublant la quantité d'A $\beta$  insoluble et le nombre de plaques amyloïdes positives à la thioflavine-S (Kim et al., 2013). Cet effet protecteur d'*ABCA7* pourrait être expliqué par ses fonctions de phagocytose du peptide amyloïde (Fu et al., 2016), son implication dans le transport des lipides à l'instar de son homologue *ABCA1* (Koldamova et al., 2010) ou encore dans le *processing* d'APP (Satoh et al., 2015). Toutes ces études convergent vers l'idée qu'une perte de fonction d'*ABCA7* conduit à une augmentation de la charge en A $\beta$  et représente donc un risque dans un mécanisme physiopathologique de MA.

### ***PLD3***

L'implication de variants rares de *PLD3* dans la maladie d'Alzheimer est très discutable. L'étude initiale publiée dans *Nature* par Cruchaga et collaborateurs (Cruchaga et al., 2014) avait permis d'établir la ségrégation d'un variant rare de *PLD3* (p.V232M) dans 2 familles sur 14 atteintes de formes tardives de MA. L'étude d'association réalisée chez 4 998 cas et 6 356 contrôles suggérait que ce variant rare était à risque (OR=2,1, p-valeur=2,93.10<sup>-5</sup>). Le séquençage de la partie codante de *PLD3* chez 2 363 cas et 2 024 contrôles avait ainsi révélé un enrichissement de variants rares à l'échelle du gène chez les cas comparés aux contrôles passant le seuil de significativité de l'exome (OR=2,75, p-valeur=1,44.10<sup>-11</sup>). Cependant, la majorité des études qui ont tenté de répliquer cette association à l'échelle du variant ou du gène ont échoué (Heilmann et al., 2015; Hooli et al., 2015; Lambert et al., 2015; van der Lee et al., 2015). De plus, la méta-analyse focalisée sur le variant p.V232M réalisée par Hooli et collaborateurs incluant l'étude de Cruchaga et collaborateurs affaiblit les résultats de la première étude (OR=1,62, p-valeur=3,47.10<sup>-4</sup>). La méta-analyse réalisée sans les données de la première étude n'est pas significative (OR=1,29, IC[0,93-1,79]). Une autre étude plus récente a réussi à répliquer nominalement cette association à l'échelle du gène (Schulte et al., 2015) mais uniquement chez les LOAD tandis qu'une autre étude d'association chez des EOAD a de nouveau échoué (Cacace et al., 2015). Une méta-analyse plus récente regroupant l'ensemble de ces études, soit environ 40 000 individus, détecte un signal d'association à l'échelle du gène de 5,03.10<sup>-4</sup>

avec un OR de 1,53 (Zhang et al., 2016a). Cependant, cette méta-analyse inclut l'analyse d'association initiale réalisée par Cruchaga et collaborateurs (Cruchaga et al., 2014). Il aurait été intéressant de faire également cette méta-analyse sans les données de Cruchaga afin d'éliminer ce biais potentiel.

Des études fonctionnelles réalisées par Cruchaga et collaborateurs ont démontré que la surexpression ou l'inhibition de PLD3 dans un modèle cellulaire diminuait et augmentait respectivement les niveaux de sécrétion d'Aβ40 et Aβ42 (Cruchaga et al., 2014). Cependant, la variation p.V232M en elle-même n'avait pas fait l'objet d'analyses fonctionnelles. De plus, une étude récente a remis en cause cette corrélation entre le niveau de surexpression de PLD3 et la sécrétion d'Aβ (Fazzari et al., 2017). Les auteurs ont ainsi pu démontrer que l'effet observé était un artefact dû à la trop grande surexpression de PLD3. De plus, la déficience de PLD3 dans des modèles souris de MA n'affectent pas le *processing* d'APP et n'augmentent pas la charge en plaques amyloïdes (Fazzari et al., 2017). Cependant, la perte de fonction de PLD3 pourrait influencer la structure des lysosomes. Le mécanisme précis reliant PLD3 à la MA *via* les lysosomes reste encore à définir (Fazzari et al., 2017).

L'ensemble des études génétiques et fonctionnelles de PLD3 ne sont pas assez convaincantes pour permettre de statuer sur son rôle de facteur de risque dans la physiopathologie de la MA.

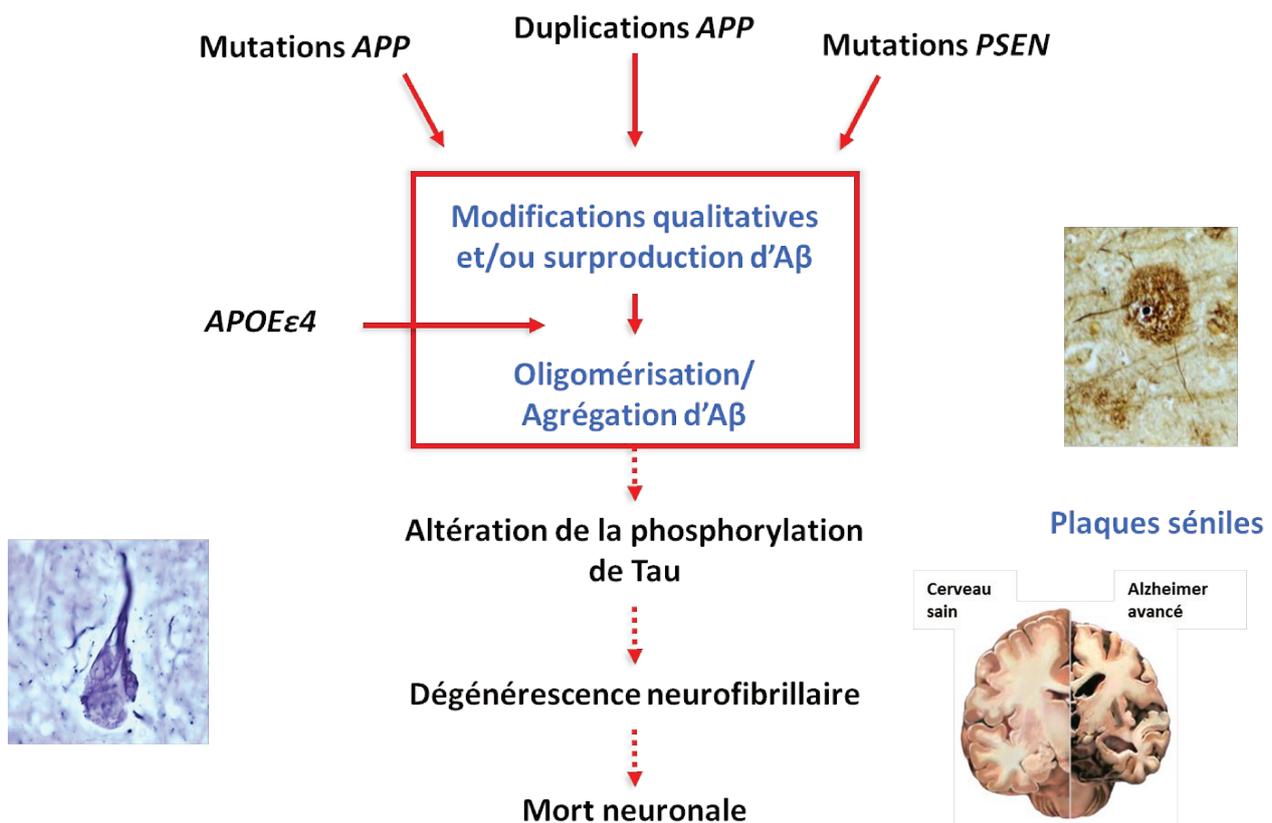
#### **d. CNVs et MA**

L'implication des CNVs dans la part de la composante génétique de la MA peut également s'intégrer dans le graphique adapté de Manolio et collaborateurs. Les duplications d'APP ainsi que les délétions de l'exon 9 de PSEN1 sont extrêmement rares et sont responsables de formes mendéliennes de MA. Des CNVs rares pour lesquels l'implication dans la maladie reste à déterminer ont été détectés chez des patients atteints de formes précoces de MA qu'il s'agisse de formes familiales ou sporadiques (Hooli et al., 2014; Rovelet-Lecrux et al., 2012). Ces CNVs rares, supposés à risque important, sont difficiles à interpréter de par l'absence de récurrence et de validation au niveau fonctionnel. Les études de puces à SNP ont également tenté de mettre en évidence des CNVs fréquents comme facteur de risque plus limité (Chapman et al., 2013; Guffanti et al., 2013; Li et al., 2012c; Swaminathan et al., 2012a, 2012b; Szigeti et al., 2013). Cependant, la majorité de ces études ont établi des associations avec des faibles valeurs de significativité et l'absence de réplique ne permet pas de clairement les impliquer comme facteurs de risque. L'étude réalisée initialement dans une population belge et répliquée dans une population française a permis de mettre en évidence que les individus porteurs de 3 copies de LCR dans le gène CR1, définissant l'isoforme CR1 S, ont un risque plus important de développer la maladie (OR=1,32, p-valeur=0,0025) (Brouwers et al., 2012). Il est important de noter que ce gène a également été associé dans la MA grâce à des études de GWAS

(Lambert et al., 2009; Zhu et al., 2015). L'implication de CR1 dans la physiopathologie de la MA n'est pas clairement établie mais des études fonctionnelles lui attribuent un rôle dans le système de clairance du peptide amyloïde *via* son activité dans le système immunitaire inné (Zhu et al., 2015).

### e. Cascade amyloïde

La caractérisation des gènes responsables de formes mendéliennes de MA a permis de formuler l'hypothèse de la cascade amyloïde plaçant le peptide amyloïde au centre du processus physiopathologique (Figure 24 page 76). Cette théorie décrite pour la première fois en 1992 (Hardy and Higgins, 1992) considère que l'élément déclencheur de la pathologie est l'accumulation du peptide A $\beta$  au sein des plaques séniles qui entraînerait une toxicité neuronale se traduisant par une phosphorylation anormale de la protéine Tau formant ainsi les DNFs qui seront responsables de la mort des neurones. Ce concept a été redéfini pour porter l'accent plus précisément sur les oligomères d'A $\beta$  plutôt que sur les plaques séniles leur attribuant un effet plus neurotoxiques (Walsh and Selkoe, 2007).



**Figure 24 : Cascade amyloïde**

La caractérisation des gènes responsables de formes mendéliennes de MA ont permis de formuler l'hypothèse de la cascade amyloïde. Cette théorie considère que l'élément déclencheur de la maladie est l'oligomérisation et/ou agrégation du peptide amyloïde en plaques séniles. Ces évènements auraient pour conséquence l'altération de la phosphorylation de Tau conduisant à terme aux dégénérescences neurofibrillaires et à la mort neuronale.

#### **f. La part de la composante génétique manquante**

La majorité des formes précoces de MA ne sont pas des formes mendéliennes. Chez ces patients, le paradigme actuel considère que la survenue de la maladie résulterait d'une combinaison de variants fréquents et rares conférant un risque faible ou élevé de développer la maladie (Figure 25 page 78). Les études de GWAS ont réussi à mettre en évidence le rôle de variants fréquents dans la composante génétique des variants fréquents dans le risque de développer la MA. Lorsque l'on s'intéresse aux variants rares, l'association est plus difficile à détecter de par leur faible fréquence dans la population générale amenant à considérer d'autres stratégies d'analyse. Dans le cadre de nos études, afin de contourner ce problème de puissance statistique nous avons décidé de réaliser une étude d'association en combinant plusieurs approches. Les cas sélectionnés présentaient un phénotype extrême (avec un âge de début avant 65 ans) et répondaient à des critères diagnostiques stricts. Les contrôles étaient de même origine géographique que les cas. Les variants ont ensuite été agrégés à l'échelle du gène et priorisés à partir de données de fréquence et de prédictions de pathogénicité.

Outre l'implication des variants rares dans la part de la composante génétique de la MA, nous nous sommes également intéressés à l'implication des CNVs rares dans cette composante grâce aux nouveaux outils de détection de CNVs à partir de données de séquençage d'exomes. Cependant, de par leur rareté même lors de l'agrégation des CNVs à l'échelle d'un gène, il nous a fallu considérer ces CNVs à l'échelle d'une liste de gènes centrée sur le peptide amyloïde.

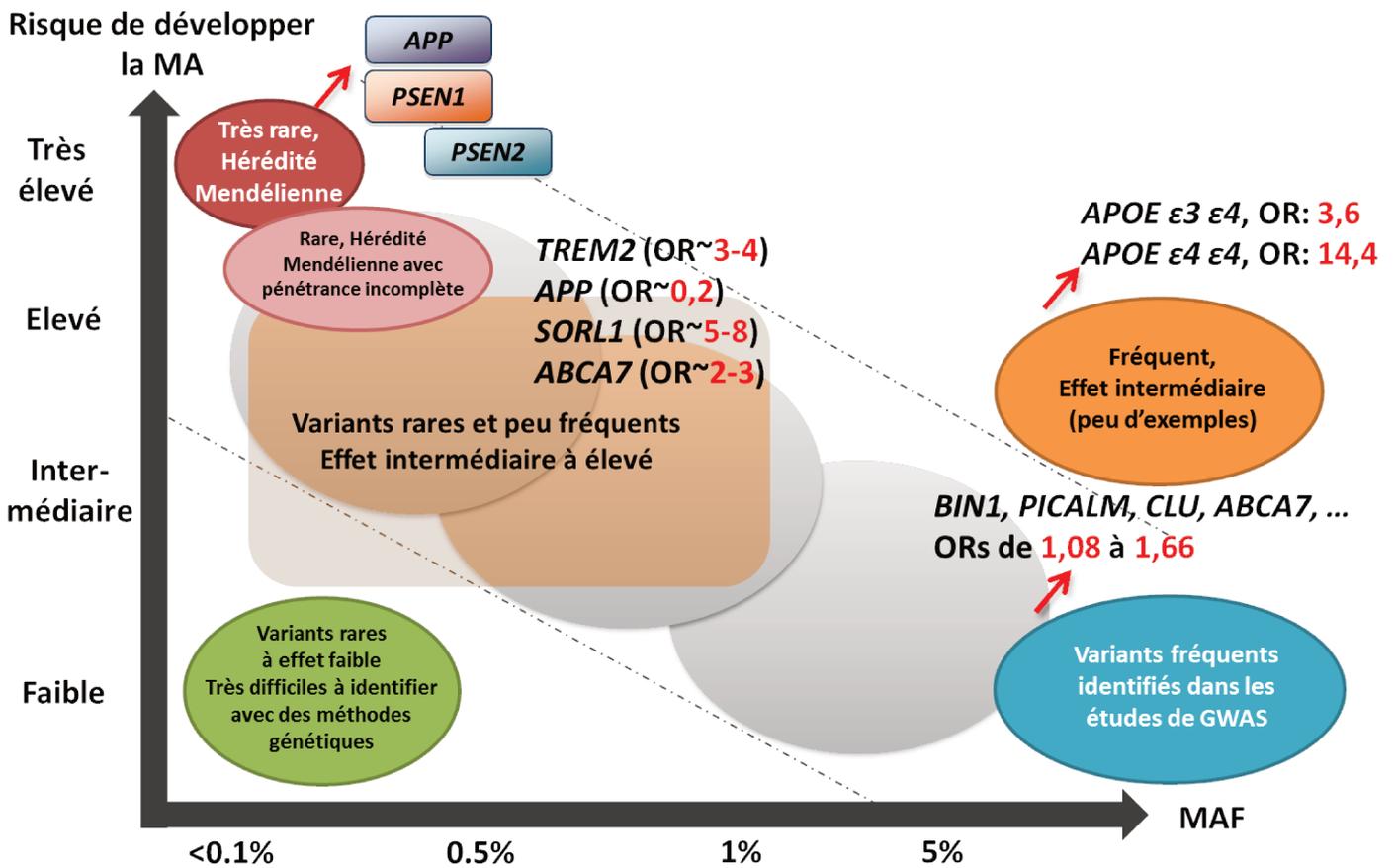


Figure 25 : Répartition des variants génétiques impliqués dans la maladie d'Alzheimer en fonction de leur fréquence et de leur effet

Figure adaptée de Manolio *et al.*, 2009.

Génin *et al.*, 2011 ; Lambert *et al.*, 2013 ; Jonsson *et al.*, 2012, 2013 ; Guerreiro *et al.*, 2012 ; Steinberg *et al.*, 2015 ; Cuyvers *et al.*, 2015 ; Nicolas *et al.*, 2015 ;

## II Résultats

### 1. Etude d'association à l'échelle du gène : Enrichissement de variants rares perte de fonction et faux-sens prédits délétères du gène *ABCA7* dans les formes précoces de maladie d'Alzheimer

Afin d'étudier la part de la composante génétique de la MA médiée par les variants rares, nous avons réalisé une étude d'association cas-contrôles à partir de données de séquençage d'exomes. Dans ce but, nous avons sélectionné 484 cas en nous focalisant sur des phénotypes extrêmes c'est-à-dire des patients atteints de MA avec un âge de début précoce (avant 65 ans) et incluant une majorité de patients très jeunes (52.9 % avant 55 ans). Nous avons également sélectionné de manière rigoureuse 590 contrôles qui proviennent du French Exome Project (E Génin, JF Deleuze, R Redon, L Letenneur, D. Campion, JC Lambert, France Génomique). Cette étude était composée d'individus sains provenant de 6 villes françaises. Par conséquent, l'ensemble des individus contrôles étaient de la même origine ethnique que les cas. Nous avons focalisé nos études sur les variants rares en les agrégeant à l'échelle du gène et en réalisant un test de charge (*Burden*). Le seuil de significativité nécessaire dans ce type d'association est de  $0,05/20\ 000=2,5 \times 10^{-6}$ . Nous avons restreint l'analyse aux variants rares (MAF < 1 %) perte de fonction (non-sens, indels décalant le cadre de lecture et site canonique d'épissage) et aux variants faux-sens classés « strictement délétères (SD) », c'est-à-dire prédits pathogènes par trois logiciels de prédiction de pathogénicité (SIFT, Polyphen-2 HumDiv et Mutation Taster). Nous n'avons pas inclus les variants bénins ou probablement pathogènes pour éviter une dilution potentielle de l'effet par des variants sans conséquences fonctionnelles sur la protéine. Cette stratégie avait permis de mettre en évidence un enrichissement de variants rares prédits comme perte de fonction dans le gène *SORL1* (OR=5,03,  $p=7,49 \cdot 10^{-5}$ ) (Nicolas et al., 2016b). La restriction aux cas familiaux permettait d'augmenter la significativité et de passer le seuil de significativité de  $2,5 \cdot 10^{-6}$  (OR=8,86,  $p=3,82 \cdot 10^{-7}$ ). Le deuxième signal d'association détecté suite à cette analyse était porté par le gène *ABCA7*. Nous avons ainsi identifié 20 patients porteurs de variants perte de fonction du gène *ABCA7* contre 8 contrôles porteurs. En ajoutant les variants faux-sens SD, 32 patients étaient finalement porteurs d'une variation perte de fonction ou faux-sens SD contre 12 contrôles porteurs, aboutissant à un OR de 3,40,  $p=2 \cdot 10^{-4}$ . Il existe donc une tendance à l'enrichissement de variants rares d'*ABCA7* chez les cas EOAD mais le seuil de significativité en exome n'est pas atteint.

Afin d'augmenter la puissance de notre étude, nous avons souhaité augmenter la taille de notre échantillon. Deux précédentes études avaient récemment décrit un enrichissement de variants rares perte de fonction dans le gène *ABCA7* (Cuyvers et al., 2015; Steinberg et al., 2015). La stratégie d'étude de l'équipe islandaise de Steinberg et collaborateurs ne permettait que d'étudier 8 variants

PTV qu'ils avaient pu associer à la MA. Cuyvers et collaborateurs avait réalisé le séquençage complet du gène *ABCA7* et avait alors pu observer un enrichissement de variants rares perte de fonction à l'échelle du gène entier (OR=4,03), résultat qui n'était pas significatif d'un point de vue d'une analyse d'association en exome ( $p=2.10^{-4}$ ). Nous avons alors réalisé une méta-analyse en combinant nos données à celles obtenues par l'équipe belge afin d'augmenter la puissance statistique. Grâce à cette analyse combinée de 1 256 patients et 1 347 contrôles, nous avons pu obtenir un résultat significatif,  $p=3,6.10^{-7}$ , avec un OR=2,81.

L'étude réalisée est la première à répliquer l'enrichissement de variants rares perte de fonction du gène *ABCA7* dans des formes précoces de MA et également la première à impliquer les variations faux-sens qui pourraient également entraîner une perte de fonction de la protéine.

# ABCA7 rare variants and Alzheimer disease risk

Kilan Le Guennec, MSc  
 Gaël Nicolas, MD, PhD  
 Olivier Quenez, MSc  
 Camille Charbonnier, PhD  
 David Wallon, MD, PhD  
 Céline Bellenguez, PhD  
 Benjamin Grenier-Boley, MSc  
 Stéphane Rousseau, MSc  
 Anne-Claire Richard, MSc  
 Anne Rovelet-Lecrux, PhD  
 Delphine Bacq, MSc  
 Jean-Guillaume Garnier, MSc  
 Robert Olaso, PhD  
 Anne Boland, PharmD, PhD  
 Vincent Meyer, PhD  
 Jean-François Deleuze, PhD  
 Philippe Amouyel, MD, PhD  
 Hans Markus Munter, PhD  
 Guillaume Bourque, MSc  
 Mark Lathrop, PhD  
 Thierry Frebourg, MD, PhD  
 Richard Redon, MD, PhD  
 Luc Letenneur, PhD  
 Jean-François Dartigues, MD, PhD  
 Florence Pasquier, MD, PhD  
 Adeline Rollin-Sillaire, MD  
 Emmanuelle Génin, PhD  
 Jean-Charles Lambert, PhD  
 Didier Hannequin, MD, PhD

## ABSTRACT

**Objective:** To study the association between ABCA7 rare coding variants and Alzheimer disease (AD) in a case-control setting.

**Methods:** We conducted a whole exome analysis among 484 French patients with early-onset AD and 590 ethnically matched controls.

**Results:** After collapsing rare variants (minor allele frequency  $\leq 1\%$ ), we detected an enrichment of ABCA7 loss of function (LOF) and predicted damaging missense variants in cases (odds ratio [OR] 3.40, 95% confidence interval [CI] 1.68–7.35,  $p = 0.0002$ ). Performing a meta-analysis with previously published data, we found that in a combined sample of 1,256 patients and 1,347 controls from France and Belgium, the OR was 2.81 (95% CI 1.89–4.20,  $p = 3.60 \times 10^{-7}$ ).

**Conclusions:** These results confirm that ABCA7 LOF variants are enriched in patients with AD and extend this finding to predicted damaging missense variants. *Neurology*® 2016;86:1–4

## GLOSSARY

**AD** = Alzheimer disease; **CI** = confidence interval; **EOAD** = early-onset Alzheimer disease; **LOF** = loss of function; **OR** = odds ratio; **MAF** = minor allele frequency; **WES** = whole-exome sequencing.

In Alzheimer disease (AD), shifting from a common disease–common variant hypothesis to a common disease–rare variant paradigm has allowed to identify rare variants in several genes, including *TREM2*, *APP*, *UNC5C*, and *SORL1*, that affect AD risk.<sup>1–7</sup> An association with *PLD3* rare variants was also suggested<sup>3</sup> but not replicated in subsequent studies.<sup>8–12</sup> Recently, the role of loss of function (LOF) variants in *ABCA7*, a gene encoding an ATP-binding cassette transporter involved in lipid transport, has been highlighted in 2 reports.<sup>13,14</sup> First, using imputation of the whole-genome sequences of  $\sim 2,600$  Icelanders into  $\sim 3,400$  patients with AD and  $\sim 151,000$  population controls, Steinberg et al.<sup>13</sup> reported an association between 8 rare LOF variants and AD risk. They replicated this finding by genotyping additional patients and controls in datasets from Europe and the United states (combined odds ratio [OR] 2.3,  $p = 6.8 \times 10^{-15}$ ). Second, using targeted resequencing of a panel of candidate genes, Cuyvers et al.<sup>14</sup> identified an increase of *ABCA7* LOF variants in 772 Flanders-Belgian patients compared to 757 geographically matched controls (relative risk 4.03,  $p = 0.0002$ ). Analyzing our whole-exome sequencing (WES) data in 484 unrelated French patients with early-onset AD (EOAD) and 590 ethnically matched controls, we found that besides *SORL1*,<sup>7</sup> *ABCA7* was among the top hits for enrichment in rare (minor allele frequency [MAF]  $< 1\%$ ) predicted damaging variants. We present the results of the *ABCA7* study.

From INSERM (K.L.G., G.N., O.Q., C.C., D.W., S.R., A.C.R., A.R.-L., T.F., D.H., D.C.), U1079, IRIB, University of Rouen, Normandy University; Normandy Centre for Genomic Medicine and Personalized Medicine (K.L.G., G.N., O.Q., C.C., D.W., S.R., A.-C.R., A.R.-L., T.F., D.H., D.C.), Rouen; Department of Genetics (G.N., T.F., D.H.), CNR-MAJ (G.N., O.Q., C.C., D.W., S.R., A.-C.R., F.P., A.R.-S., D.H., D.C.), and Department of Neurology (D.W., D.H.), Rouen University Hospital; INSERM (C.B., B.G.-B., P.A., J.-C.L.), U1167, Lille; Institut Pasteur de Lille (C.B., B.G.-B., P.A., J.-C.L.); Université Lille-Nord de France (C.B., B.G.-B., P.A., J.-C.L.); Centre National de Génotypage (D.B., J.-G.G., R.O., A.B., V.M., J.-F. Deleuze.), Institut de Génétique, CEA, Evry; Fondation Jean Dausset (J.-F. Deleuze.), Centre d'Etudes du Polymorphisme Humain, Paris, France; McGill University and Génome Québec Innovation Centre (H.M.M., G.B., M.L.), Montréal, Canada; INSERM (R.R.), UMR 1087, l'Institut du Thorax, CHU Nantes; CNRS (R.R.), UMR 6291, Université de Nantes; INSERM (L.L., J.-F. Dartigues.), U897, Bordeaux; University of Bordeaux (L.L., J.-F. Dartigues.); Department of Neurology (F.P., A.R.S.), Lille University Hospital; INSERM (E.G.), UMR1078, CHU Brest, Université Bretagne Occidentale, Brest; and Department of Research (D.C.), Rouvray Psychiatric Hospital, Sotteville-lès-Rouen, France.

Coinvestigators are listed on the *Neurology*® Web site at [Neurology.org](http://Neurology.org).

Go to [Neurology.org](http://Neurology.org) for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

Author list continued on next page

Dominique Campion,  
MD, PhD  
On behalf of the CNR-  
MAJ collaborators

Correspondence to  
Dr. Campion:  
dominique.campion@univ-rouen.fr

Supplemental data at  
Neurology.org

**METHODS Patients.** Patient ascertainment is described in detail in reference 7. Briefly, we included 484 unrelated patients with EOAD (age at onset  $\leq 65$  years) from French ancestry recruited by the French National CNR-MAJ consortium. All diagnoses fulfilled the National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer’s Disease and Related Disorders Association criteria.<sup>15</sup> The clinical examination included personal medical and family history assessment, neurologic examination, neuropsychological assessment, and neuroimaging. In addition, CSF biomarkers indicative of AD were available for 249 patients.<sup>16</sup> Patients with CSF biomarkers not consistent with an AD diagnostic ( $n = 14$ ) were excluded.

Patients with *PSEN1*, *PSEN2*, or *APP* mutation (detected either by Sanger sequencing or by WES), *APP* duplication (detected by quantitative multiplex PCR of short fluorescent fragments),<sup>17</sup> or *C9ORF72* expansion (detected following the protocol described by DeJesus-Hernandez et al.<sup>18</sup>) were excluded. The number of *APOE*  $\epsilon 4$ -positive patients was 256 (53%).

**Controls.** A total of 595 controls of French ancestry were recruited. A total of 396 had an age at inclusion of  $\geq 55$  years and had normal Mini-Mental State Examination score, according to age and education level. The remaining controls were healthy blood donors. The mean age at inclusion ( $\pm$ SD) was  $66.4 \pm 17.5$  years. The number of *APOE*  $\epsilon 4$ -positive controls was 138 (23%).

**Standard protocol approvals, registrations, and patient consents.** All patients and controls gave informed, written consent for genetic analyses. This study was approved by our ethics committee.

**WES.** As previously described,<sup>7</sup> exomes were captured using Agilent SureSelect Human All Exon kits (Agilent Technologies, Santa Clara, CA). Final libraries were sequenced on a HiSeq2000 or 2500 (Illumina, San Diego, CA) with paired ends, 76 or 100 bp reads. We processed all exome samples following GATK 3.3–0 Best Practices recommendations.<sup>19</sup> We used BWA 0.7.5a to map reads to the 1,000 Genomes GRCh37 build.<sup>20</sup> Duplicate reads were flagged by using Picard Tools 1.101 (<http://picard.sourceforge.net>). We applied GATK for short insertion and deletions (indels) realignment, base quality score recalibration, and single nucleotide variants and indels discovery using the Haplotype Caller across all samples simultaneously. The joint variant calling file was annotated with RefGene gene regions, variant effects, and functional effect prediction tools using Annovar.<sup>21</sup> We then extracted high-quality (defined as variants with a variant quality score VQSLOD above  $-2$ ) exonic and canonical splice site variants (located  $\pm 2$  bp around each coding exon) with a MAF of less than 1% in our whole dataset within each RefGene gene region.

We filtered genotypes according to the following criteria: genotype read depth had to be above 6 as well as genotype quality above 70. Nonsense, frameshift indels, and canonical splice site variants were classified as LOF. Missense variants were considered as strictly damaging if they were simultaneously predicted damaging by the following 3 software programs: Polyphen2 HumDiv, Mutation Taster, and SIFT.<sup>21</sup>

**Sanger sequencing.** *ABCA7* rare variants included in this study were confirmed by Sanger sequencing.

**Quality checks.** As previously described,<sup>7</sup> we used recommended quality control steps for genetic case-control association studies to analyze our data.<sup>22</sup> Most checks were carried out with PLINK 1.9 (<https://www.cog-genomics.org/plink2>). We processed all individuals in the sample through the following

steps: (1) verifying concordant sex information using Plink sex check,<sup>23</sup> (2) discarding contaminated samples identified as such by significantly high heterozygosity rates and freemix contamination scores provided by the VerifyBamID software,<sup>24</sup> and (3) discarding samples when Plink *pi\_hat* relatedness estimation exceeded 18.5%. Out of 485 initially included cases and 595 controls, 484 cases and 590 controls passed these quality checks. We did not detect any individual of divergent ancestry with Plink neighbor function. To confirm ethnic match between cases and controls, we used principal component analysis on common variants ( $MAF > 5\%$ ) after exclusion of long-range linkage disequilibrium regions and variant pruning on linkage disequilibrium ( $r^2 > 0.2$ ).

We filtered out variants that (1) were missing in more than 5% of individuals, (2) showed a significant deviation from Hardy-Weinberg equilibrium, or (3) presented significantly different missing call rates between cases and controls, as confirmed by Plink test missing at a threshold of  $1.10^{-6}$ . None of these filters discarded any of the *ABCA7* variants with a MAF below 1%.

Finally, the following 2 filters were applied locally to each RefGene coding region: (1) samples that missed information for more than 50% of variants in the given region were excluded from analysis and (2) variants that remained missing in more than 5% of remaining individuals were excluded from analysis. None of these filters affected the number of samples and variants under study within the *ABCA7* gene.

**Statistics.** We imported all rare ( $MAF < 1\%$  on the whole case-control sample) LOF and strictly damaging missense *ABCA7* into Variant Association Tools.<sup>25</sup> We computed unadjusted gene-level collapsing tests using the CFisher function grouped by RefGene name, which corresponds here to a Fisher exact test comparing the proportions of variant carriers between cases and controls. We computed corresponding OR and confidence intervals (CIs) with R statistical software (<http://www.R-project.org/>). Single analysis results were obtained via the Fisher test function from package stats, while fixed effect meta-analysis results were obtained via the metabin function from package meta. We applied the same quality checks and gene-level association tests to every RefGene coding region in order to extract all possible gene-level *p* values and compute the genomic inflation factor. Using the *estlamba* function of the R statistical package genABEL, we found a value of  $\lambda = 0.88$  for this genomic inflation factor, thereby excluding possible confounding from fine-scale population stratification. Finally, we checked that the observed association signal was independent from *APOE* status in a logistic regression model using the *glm* function from the R statistical software, adjusting for *APOE*  $\epsilon 4$ -positive status.

**RESULTS** We first analyzed rare LOF variants (nonsense, frameshift indels, and splice site) and identified in the entire dataset 12 variants (3 nonsense, 6 frameshift indels, and 3 splice site) present in 20 patients and 5 variants (3 frameshift and 2 splice site) present in 8 controls (OR 3.13 [1.30–8.30],  $p = 0.006$ ) (table e-1 on the *Neurology*<sup>®</sup> Web site at Neurology.org). Besides canonical splice site variants, we included in this count the c.5570+5G>C variant, which has previously been shown to disrupt the splicing of exon 41.<sup>13</sup> Of note, 4 other intronic mutations, located close to splice sites (c.2067+4A>T, c.2684+5G>A, c.3472+5G>C, c.5279+4A>G),

were exclusively found in patients. Although they are predicted in silico as probably affecting splicing, we conservatively decided not to include them in the count. Overall, 4 of the 14 LOF mutations found in the whole dataset were already reported in the Icelandic study,<sup>13</sup> 6 were found in the Flanders-Belgian cohort,<sup>14</sup> and 8 were novel (table e-1). Four variants were simultaneously present in the 3 datasets from Iceland, Belgium, and France.

We then focused on rare strictly damaging missense variants, namely variants with a MAF below 1% in our entire dataset, which were predicted as damaging by 3 software programs: Polyphen2 HumDiv, Mutation Taster, and SIFT. We identified 10 rare strictly damaging missense variants present in 13 cases and 5 in 5 controls, resulting in an OR of 3.23 (95% CI [1.07–11.63],  $p = 0.03$ ) (table e-1). In total, 32 patients (6.6%) vs 12 controls (2.0%) carried at least one LOF or strictly damaging variant. Missense mutations were located throughout the sequence of *ABCA7*. No particular exon or functional domain was hit. The cumulative effect of LOF and strictly damaging missense variants in patients and controls resulted in an OR of 3.40 (95% CI [1.68–7.35],  $p = 0.0002$ ). Moreover, after adjustment for *APOE4+* status, the effect of *ABCA7* variants remained largely significant (OR 3.1, 95% CI [1.57–6.48],  $p = 0.0016$ ). Therefore, the observed signal of association between EOAD and *ABCA7* variants can be considered independent from *APOE4* status. However, no interaction between these 2 risk factors could be detected ( $p = 0.666$ ). Of note, 4 patients also carried a *SORL1* risk allele (table e-1).

We next performed a meta-analysis including the *ABCA7* published data. Due to differences in the study design, it was not possible to recover the full load of *ABCA7* rare variants in patients and controls of the Icelandic sample. Therefore, this analysis was restricted to the Flanders-Belgian sample (table e-2). In the combined study group of 1,256 patients and 1,347 controls from France and Belgium, the OR resulting from a fixed-effect meta-analysis was 2.89 (95% CI [1.78–4.70],  $p = 1.83 \times 10^{-5}$ ) for LOF variants and 2.81 (95% CI [1.89–4.20],  $p = 3.60 \times 10^{-7}$ ) for combined LOF and strictly damaging variants.

Finally, after the report that a frequent variant (p.Glu188Gly) had a nominally significant increased frequency in controls in the Flanders-Belgian sample,<sup>14</sup> we checked the allelic frequency of this variant in our dataset. The frequency of this variant was 0.41 in patients compared to 0.43 in controls (NS).

**DISCUSSION** In this study, we explored the common disease–rare variants paradigm in EOAD. We performed WES with high depth of coverage (mean  $\sim 120\times$ ) on 484 well-characterized French patients

with EOAD and 590 ethnically matched controls. There was a clear excess of cases carrying at least one LOF or a predicted damaging missense variant in the *ABCA7* gene compared to controls (6.6% vs 2.0%). Of note, about half of the controls were still in the age range for developing EOAD, and although this condition is rare we cannot exclude that a small subset of them will develop it. Therefore the association could have been slightly underestimated. Performing a meta-analysis of these data with the previously published Belgian series allowed reaching an exome-wide significant association signal, strongly supporting the role of *ABCA7* in AD determinism. It had initially been suggested that *ABCA7* played a role in AD by favoring amyloid clearance.<sup>26</sup> Recently, it has been shown that *ABCA7* also interferes with APP processing and reduces A $\beta$  production, possibly by regulating endocytic pathways.<sup>27</sup> In each case, suppression of *ABCA7* is predicted to result in increased A $\beta$  load, which is consistent with the burden of LOF mutations found in patients with AD. Interestingly, considering the genetic data presented here, it is likely that at least a subset of rare missense strictly damaging *ABCA7* mutations also results in *ABCA7* LOF.

Of note, among the 32 patients with an *ABCA7* LOF or missense variant, 4 additionally carried a *SORL1* risk allele, pointing toward a possible oligogenic model of inheritance. However, considering the rarity of the observed variants, performing an accurate polygenic score analysis including *SORL1*, *ABCA7*, and *TREM2* rare variants together with *APOE* status is still premature in terms of statistical power. Finally, the *ABCA7* frequent variant (p.Glu188Gly) that had a nominally significant increased frequency in controls of the Flanders-Belgian sample was not found to be associated in the present sample, although a same trend toward an increased frequency in controls was noted. Our negative result might reflect a lack of power.

## AUTHOR CONTRIBUTIONS

Principal investigator and study supervision: Dominique Campion. Manuscript draft: Kilan Le Guennec, Gaël Nicolas, Camille Charbonnier, Dominique Campion. Acquisition, collection, and interpretation of data: Kilan Le Guennec, Gaël Nicolas, Olivier Quenez, Camille Charbonnier, David Wallon, Céline Bellenguez, Benjamin Grenier-Boley, Stéphane Rousseau, Anne-Claire Richard, Anne Rovelet-Lecrux, Delphine Bacq, Jean-Guillaume Garnier, Robert Olasso, Anne Boland, Vincent Meyer, Jean-François Deleuze, Philippe Amouyel, Hans Markus Munter, Guillaume Bourque, Mark Lathrop, Thierry Frebourg, Richard Redon, Luc Letenneur, Jean-François Dartigues, Florence Pasquier, Adeline Rollin-Sillaire, Emmanuelle Génin, Jean-Charles Lambert, Didier Hannequin, Dominique Campion, and the CNR-MAJ collaborators.

## STUDY FUNDING

This study was funded by grants from the Clinical Research Hospital Program from the French Ministry of Health (GMAJ, PHRC 2008/067), the CNR-MAJ, the JPND PERADES, and France Génomique.

This work was supported by Labex GENMED ANR-10-LABX-0013, the National Foundation for Alzheimer's Disease and Related Disorders, the Institut Pasteur de Lille, the Centre National de Génotypage, INSERM, FRC (Fondation pour la Recherche sur le Cerveau), Rotary, and LABEX (Laboratory of Excellence Program Investment for the Future) and DISTALZ (Development of Innovative Strategies for a Transdisciplinary Approach to Alzheimer's Disease) grants.

## DISCLOSURE

The authors report no disclosures relevant to the manuscript. Go to [Neurology.org](http://Neurology.org) for full disclosures.

Received October 26, 2015. Accepted in final form February 29, 2016.

## REFERENCES

1. Jonsson T, Stefansson H, Steinberg S, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* 2013;368:107–116.
2. Guerreiro R, Wojtas A, Bras J, et al. TREM2 variants in Alzheimer's disease. *N Engl J Med* 2013;368:117–127.
3. Cruchaga C, Karch CM, Jin SC, et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* 2014;505:550–554.
4. Pottier C, Hannequin D, Coutant S, et al. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. *Mol Psychiatry* 2012;17:875–879.
5. Jonsson T, Atwal JK, Steinberg S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 2012;488:96–99.
6. Wetzell-Smith MK, Hunkapiller J, Bhangale TR, et al. A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. *Nat Med* 2014;20:1452–1457.
7. Nicolas G, Charbonnier C, Wallon D, et al. SORL1 rare variants: a major risk factor for familial early-onset Alzheimer's disease. *Mol Psychiatry* Epub 2015 Aug 25.
8. Heilmann S, Driichel D, Clarimon J, et al. PLD3 in non-familial Alzheimer's disease. *Nature* 2015;520:E3–E5.
9. Hooli BV, Lill CM, Mullin K, et al. PLD3 gene variants and Alzheimer's disease. *Nature* 2015;520:E7–E8.
10. Lambert JC, Grenier-Boley B, Bellenguez C, et al. PLD3 and sporadic Alzheimer's disease risk. *Nature* 2015;520:E1.
11. van der Lee SJ, Holstege H, Wong TH, et al. PLD3 variants in population studies. *Nature* 2015;520:E2–E3.
12. Cacace R, Van den Bossche T, Engelborghs S, et al. Rare variants in PLD3 do not affect risk for early-onset Alzheimer disease in a European Consortium Cohort. *Hum Mutat* 2015;36:1226–1235.
13. Steinberg S, Stefansson H, Jonsson T, et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nat Genet* 2015;47:445–447.
14. Cuyvers E, De Roeck A, Van den Bossche T, et al. Mutations in ABCA7 in a Belgian cohort of Alzheimer's disease patients: a targeted resequencing study. *Lancet Neurol* 2015;14:814–822.
15. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011;7:263–269.
16. Nicolas G, Wallon D, Charbonnier C, et al. Screening of dementia genes by whole-exome sequencing in early-onset Alzheimer disease: input and lessons. *Eur J Hum Genet* Epub 2015 Aug 5.
17. Rovelet-Lecrux A, Hannequin D, Raux G, et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet* 2006;38:24–26.
18. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 2011;72:245–256.
19. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–1303.
20. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26:589–595.
21. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
22. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc* 2010;5:1564–1573.
23. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;4:7.
24. Jun G, Flickinger M, Hetrick KN, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet* 2012;91:839–848.
25. Wang GT, Peng B, Leal SM. Variant association tools for quality control and analysis of large-scale sequence and genotyping array data. *Am J Hum Genet* 2014;94:770–783.
26. Kim WS, Li H, Ruberu K, et al. Deletion of Abca7 increases cerebral amyloid-beta accumulation in the J20 mouse model of Alzheimer's disease. *J Neurosci* 2013;33:4387–4394.
27. Satoh K, Abe-Dohmae S, Yokoyama S, St George-Hyslop P, Fraser PE. ABCA7 loss of function Alters Alzheimer amyloid processing. *J Biol Chem* 2015;290:24152–24165.

## 2. Elaboration d'une liste de gènes centrée sur le peptide amyloïde considérant son rôle majeur dans la physiopathologie

Les études d'association à partir de variants rares ou fréquents sont très difficiles à réaliser et se retrouvent très souvent confrontées à des problèmes de puissance statistique. L'analyse à l'échelle du variant étant inapplicable, nous avons dans un premier temps analysé les variants rares prédits comme perte de fonctions en mesurant la charge à l'échelle du gène et avons pu détecter un enrichissement de ces variations dans les gènes *SORL1* et *ABCA7*. Cependant, nous avons également détecté de nombreux signaux d'association relativement faibles pour lesquelles nous ne pouvions conclure quant à leur implication dans le risque de développer la MA. De nombreuses études ont montré la pertinence de l'agrégation des variants rares à l'échelle d'un réseau de gènes (Fromer et al., 2014; Iossifov et al., 2014; Kirov et al., 2012). Dans ce but, nous avons réalisé une bibliographie extensive afin de recenser les gènes pour lesquels la protéine produite présente un lien avec le peptide amyloïde considérant son rôle majeur dans la physiopathologie. Les gènes ont été inclus si les preuves expérimentales reposaient sur des modèles cellulaires ou animaux (phénotype résultant de la surexpression ou de l'inactivation de ces gènes). La liste de gènes est mise à jour régulièrement et nous recensons à ce jour 342 gènes (335 dans l'article) subdivisés en 4 grandes catégories : le routage d'APP et sa maturation, la clairance/dégradation d'A $\beta$ , l'agrégation du peptide amyloïde et une dernière catégorie englobant les médiateurs de la toxicité du peptide A $\beta$ . Chaque grande catégorie contient plusieurs sous-catégories (Table en Annexe). Chaque catégorie et sous-catégorie de la liste de gènes a fait l'objet de discussions détaillées permettant de caractériser les différents mécanismes physiopathologiques de la MA. Nous avons également indiqué pour chaque gène s'il était capable d'interagir physiquement avec APP/A $\beta$ , PSEN1 ou BACE1 (colonne PPI), si des variants fréquents étaient associés par étude de GWAS et également si le gène avait déjà été impliqué dans une maladie mendélienne. De manière intéressante, parmi les 342 gènes, 86 ont déjà été impliqués dans des maladies mendéliennes d'après la base de données d'OMIM (*Online Mendelian Inheritance in Man*). Ce constat supporte l'hypothèse selon laquelle les maladies génétiques ont, en partie, une origine commune et que certaines voies biologiques sont essentielles. Dans un souci d'interprétation, nous avons essayé de définir pour chaque gène si une variation perte ou gain de fonction serait potentiellement à risque ou protectrice dans un mécanisme physiopathologique de MA. Cependant, cette interprétation n'est pas réalisable pour tous les gènes notamment car il existe des discordances dans les résultats obtenus à partir de différents modèles expérimentaux. Par conséquent, 144 gènes sur 342 se retrouvent sans annotation d'interprétation (Table en Annexe).

## EXPERT REVIEW

# Alzheimer disease: modeling an A $\beta$ -centered biological network

D Champion<sup>1,2,3,4</sup>, C Pottier<sup>1,4</sup>, G Nicolas<sup>1,2,4,5</sup>, K Le Guennec<sup>1,4</sup> and A Rovelet-Lecrux<sup>1,4</sup>

In genetically complex diseases, the search for missing heritability is focusing on rare variants with large effect. Thanks to next generation sequencing technologies, genome-wide characterization of these variants is now feasible in every individual. However, a lesson from current studies is that collapsing rare variants at the gene level is often insufficient to obtain a statistically significant signal in case–control studies, and that network-based analyses are an attractive complement to classical approaches. In Alzheimer disease (AD), according to the prevalent amyloid cascade hypothesis, the pathology is driven by the amyloid beta (A $\beta$ ) peptide. In past years, based on experimental studies, several hundreds of proteins have been shown to interfere with A $\beta$  production, clearance, aggregation or toxicity. Thanks to a manual curation of the literature, we identified 335 genes/proteins involved in this biological network and classified them according to their cellular function. The complete list of genes, or its subcomponents, will be of interest in ongoing AD genetic studies.

*Molecular Psychiatry* advance online publication, 29 March 2016; doi:10.1038/mp.2016.38

## INTRODUCTION

Alzheimer disease (AD) is defined by the presence of two types of abnormal protein deposits composed of: (i) hyperphosphorylated forms of the microtubule-associated protein tau constituting intracellular aggregates named neurofibrillary tangles (NFT) and (ii) the 39–42 amino acid amyloid beta (A $\beta$ ) peptide mainly present in extracellular senile plaques. The A $\beta$  peptide results from the sequential cleavage of a transmembrane protein, the amyloid precursor protein (encoded by *APP*) by  $\beta$ - and  $\gamma$ -secretases. Pathogenic mutations causing autosomal dominant early onset AD have been described within *APP* and the *PSEN1* and *PSEN2* genes, which encode presenilins, a major component of the  $\gamma$ -secretase complex. Tau is encoded by the *MAPT* gene. With very few exceptions,<sup>1,2</sup> *MAPT* mutations are never encountered in pathologically-confirmed AD but cause another tauopathy, namely frontotemporal dementia with Parkinsonism linked to chromosome 17 (FTDP-17), which is characterized by NFT without A $\beta$  pathology. Crossing mice that exhibit A $\beta$  plaques due to a double *APP* and *PSEN1* mutation, with mice that exhibit NFT due to overexpression of a mutant form of *MAPT*, has demonstrated enhanced tau pathology caused by A $\beta$  in the absence of any effect on A $\beta$  caused by mutant tau.<sup>3</sup> These experiments strongly support the amyloid cascade hypothesis, which states that in AD, A $\beta$  functions upstream of tau and is the primary driver of the entire AD pathology. Since its first formulation,<sup>4</sup> the amyloid cascade hypothesis has been somewhat refined. The emphasis has been put on soluble A $\beta$  oligomers rather than insoluble aggregates as the most neurotoxic species.<sup>5</sup> It has also been acknowledged that tau pathology is a key determinant of synaptic dysfunction and neuronal death. In normal aging, NFT are commonly observed in the entorhinal cortex without any

obvious clinical symptoms. They slowly spread along synaptically connected circuits to the hippocampal formation and then to neocortical regions according to Braak stages.<sup>6</sup> If they encounter A $\beta$  oligomers in these structures, the A $\beta$  peptide boosts tau pathology, the severity of dementia being finally correlated with the number of NFTs in the neocortex.<sup>7,8</sup> During this process, it is now believed that rather than a strictly hierarchical mode of interaction, there is an interplay between A $\beta$  and tau whereby they amplify each other's toxic effects.<sup>9,10</sup> In that sense, AD can be defined as a disease in which A $\beta$  triggers, facilitates or amplifies tau pathology in cortical structures,<sup>11,12</sup> although tau is probably the essential killer. Owing to its key role in AD pathogenesis, A $\beta$  has arguably become the most studied mammalian peptide in cell biology.

Contrasting with this massive effort that has allowed the identification of several hundreds of proteins interfering with A $\beta$  production, clearance, aggregation or toxicity, geneticists have only characterized a few genes involved in AD determinism. First, rare mutations involving *APP* or *PSENs*, which are directly linked to the amyloidogenic processing of *APP*, have been shown to cause autosomal dominant early onset AD. Second, a frequent DNA polymorphism, the  $\epsilon 4$  allele of the apolipoprotein E gene (*APOE*), has been associated with AD risk. At the age of 85, the lifetime risk associated with the  $\epsilon 4$  allele ranges from 23% for *APOE*  $\epsilon 3\epsilon 4$  male carriers to 30% for *APOE*  $\epsilon 3\epsilon 4$  female carriers, and from 51% for *APOE*  $\epsilon 4\epsilon 4$  male carriers to 60% for *APOE*  $\epsilon 4\epsilon 4$  female carriers.<sup>13</sup> The apolipoprotein-E is involved both in the clearance and the aggregation of the A $\beta$  peptide. Third, recent genome-wide association studies have identified 19 other loci harboring frequent variants, each with minimal impact on AD risk (odds ratios ranging from 0.77 to 1.22).<sup>14</sup> To pursue the search for missing heritability in AD, the focus has now been put on the

<sup>1</sup>Inserm U1079, Rouen University, Institute for Research and Innovation in Biomedicine, Normandie University, Rouen, France; <sup>2</sup>National Reference Center for Young Alzheimer Patients (CNR-MAJ), Rouen, France; <sup>3</sup>Department of Research, Rouvray Psychiatric Hospital, Sotteville-lès-Rouen, France; <sup>4</sup>Normandy Center for Genomic Medicine and Personalized Medicine, Rouen, France and <sup>5</sup>Department of Genetics, Rouen University Hospital, Rouen, France. Correspondence: Dr D Champion, Inserm U1079, Faculté de Médecine et de Pharmacie, Institute for Research and Innovation in Biomedicine, Normandie University, 22 boulevard Gambetta, Rouen 76183, France. E-mail: dominique.champion@univ-rouen.fr

Received 6 November 2015; revised 16 February 2016; accepted 18 February 2016

discovery of rare variants with large effect.<sup>15</sup> According to this common disease-rare variants paradigm, the genetic heterogeneity of the disease is high, and a large number of genes, each contributing to a small subset of cases must be involved. In addition, it can be hypothesized that genes exerting only a moderate effect on AD risk can operate interactively. Thanks to next-generation sequencing technologies, it is now possible to identify roughly 20–25 000 variants in the coding sequence of any individual, ~1000 of them being non-synonymous variants with a minor allele frequency below 1%. Owing to their rarity and the necessity to account for large-scale multiple statistical testing, it is extremely challenging to conduct case–control studies on an individual variant basis. In most cases, given achievable sample sizes, even collapsing rare variants at the gene level will be insufficient to reach genome-wide statistical significance. Therefore, in complement to gene-based analyses, it may be useful in some cases to jointly analyze a set of genes showing high *a priori* probability of being involved in the pathogenesis. Promising methods are currently developed, which tackle the crucial issue of controlling type 1 error for pathway-level analysis.<sup>16</sup> Of note, the study that first isolated *SORL1* in AD used a specific *a priori* hypothesis about retromer-related genes to perform an association analysis.<sup>17</sup> Performing rare variants studies at the level of a list of functionally relevant genes is an almost unavoidable approach when dealing with *de novo* variants or copy number variants due to the very low recurrence of these variants. It has recently proved to be fruitful in complex diseases. For example, when jointly analyzing a group of 842 genes previously defined as regulated by fragile X mental retardation (*FMR1*) protein in autism spectrum disorder lossifov *et al.*<sup>18</sup> found a significant excess in *de novo* disrupting mutations. Similarly, we have recently reported an enrichment in *de novo* mutations at the level of this A $\beta$  network in early onset AD patients.<sup>19</sup> Overall, there is a long history of dispute between defenders of ‘hypothesis free’ versus ‘hypothesis driven’ genetic studies but the two approaches should be viewed as complementary.

Bearing that in mind, we decided to define, as a resource for geneticists, a network of genes whose protein products are known to interact with A $\beta$  production, aggregation, clearance or synaptic toxicity. We undertook a systematic literature search by using PubMed ([www.ncbi.nlm.nih.gov/pubmed](http://www.ncbi.nlm.nih.gov/pubmed)) to collate all relevant information. The search terms included APP, Abeta and amyloid beta, combined with the different terms listed in Table 1. We considered all peer-reviewed, original articles and reviews, published in English language up to October 2015. We only selected genes for which experimental evidence based on cell or animal models had been published. Direct protein–protein interaction was not a prerequisite but was noted, when documented. The search yielded 335 hits (Supplementary Table). Note that several genes are involved in different subcomponents of the network. In the following sections we will discuss the role of the major contributors from each subclass in light of physiological data. Each gene is identified according to the usual name or symbol of its protein product retrieved from UniProtKB ([www.uniprot.org/uniprot/](http://www.uniprot.org/uniprot/)) and, in parentheses, to its HUGO Gene Nomenclature Committee (HGNC, <http://www.gene.names.org/>) symbol, unless the protein symbol and its associated gene symbol are the same. In each subheading, whenever possible, we preferentially referenced reviews rather than individual reports.

#### APP trafficking and processing

Knowledge of the cellular pathways in which APP traffics and is processed is central for the formulation of the amyloid cascade hypothesis, and has recently been the subject of several reviews.<sup>20–22</sup> Basically, APP can undergo either amyloidogenic or non-amyloidogenic processing, occurring in different cellular compartments. Figure 1 recapitulates the intracellular trafficking of APP.

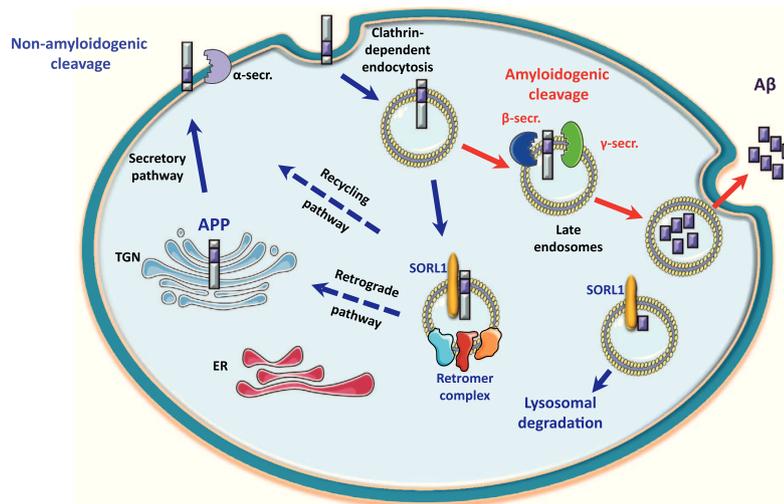
**Table 1.** Categories and subcategories of the A $\beta$  network.

Category	Subcategory
APP processing and trafficking	Endoplasmic reticulum ERAD/proteasome Early secretory pathway Axonal transport $\alpha$ -secretase cleavage Clathrin/caveolin-dependent endocytosis Internalization of APP Endosomes/retromer $\beta$ -secretase cleavage $\gamma$ -secretase cleavage Cholesterol metabolism Miscellaneous
Clearance and degradation of A $\beta$	Enzymatic degradation of A $\beta$ ERAD/proteasome Autophagy Blood–brain barrier Microglia
Aggregation of A $\beta$	Miscellaneous Chaperones Heparan sulfate proteoglycan
Mediator of A $\beta$ toxicity	Synaptic toxicity Tau toxicity Mitochondria Calcium homeostasis Miscellaneous

Abbreviations: APP: amyloid precursor protein; ERAD: endoplasmic reticulum-associated protein degradation.

*ER quality control.* Upon synthesis, APP translocates to the endoplasmic reticulum (ER) where it undergoes structural maturation. Although in physiological conditions the bulk of APP localizes to more distal compartments (that is, the Golgi apparatus and the trans-Golgi network (TGN)), ER chaperones, such as the 78 kDa glucose-regulated protein (*HSPA5*) or calnexin (*CANX*), which contribute to the protein quality control in the ER, may retain APP at this location.<sup>23</sup> This retention leads to an alteration of its processing and to a reduction of A $\beta$  production. Several molecular components of the ER-associated degradation system (ERAD) have also been shown to interfere with APP processing. The ERAD system recognizes misfolded proteins and retrotranslocates them across the ER membrane to the cytosol, where they become a substrate for the ubiquitin–proteasome degradation system. Therefore the ERAD system, by promoting the degradation of unfolded APP, regulates the availability of APP for A $\beta$  generation in distal compartments of the early secretory pathway. Two molecules acting through this mechanism are the Serine protease HTRA2 and the E3 ubiquitin–protein ligase synoviolin (*SYVN1*). HTRA2 shuttles between the outer ER membrane and the proteasome<sup>24</sup>, and has also been involved in Parkinson disease and essential tremor genetic determinism. Overexpression of synoviolin results in decreased A $\beta$  production, whereas reduced expression induces increased A $\beta$  generation.<sup>25</sup> The E3 ubiquitin–protein ligase CHIP (*STUB1*) is also involved in APP ubiquitination, mediating subsequent degradation by the proteasome.<sup>26</sup> The role of these proteins is particularly important in AD patients, because as discussed below, the depletion of ER calcium stores caused by the A $\beta$  peptide toxicity induces an accumulation of unfolded proteins, including APP, in the ER lumen.

*Traffic through the secretory pathway and axonal transport.* After initial maturation within the ER, APP is then transferred to post-ER compartments of the early secretory pathway for further processing. In the Golgi apparatus, APP encounters several



**Figure 1.** Schematic representation of amyloid precursor protein (APP) trafficking (adapted from<sup>20</sup>). APP is routed from the endoplasmic reticulum (ER) to the cell membrane through the early secretory pathway. At the cell surface, it can be processed by enzymes with  $\alpha$ -secretase activity (and then by the  $\gamma$ -secretase complex, not shown) leading to a non-amyloidogenic pathway. Alternatively, APP can be reinternalized in a clathrin-dependant manner and be directed to the endosomal compartments. In late endosomes, pH conditions and lipid composition of the membranes allow for  $\beta$ -secretase BACE1 and  $\gamma$ -secretase activity, leading to the production and secretion of the amyloid beta (A $\beta$ ) peptide. Before secretion, a subset of nascent A $\beta$  peptides can bind SORL1, which directs it to the lysosome, leading to its degradation. Before amyloidogenic processing in late endosomes, APP can also bind SORL1 and be redirected to non-amyloidogenic pathways through the retromer complex, back to the Golgi (retrograde pathway) or back to the membrane (recycling pathway).

interactors such as the transmembrane protein 59 (*TMEM59*), which modulates its *N*- and *O*-glycosylation and induces its retention in the Golgi apparatus, preventing A $\beta$  generation.<sup>27</sup> It has been suggested that two other sets of APP-interacting proteins, the scaffolding Homer protein homologs 2 and 3 (*HOMER2*, *HOMER3*),<sup>28</sup> and the reticulon-4 receptors (*RTN4R*) (also known as Nogo receptors), also affect cellular trafficking of APP and reduce the amount delivered to the cell surface.<sup>29</sup> Another Golgi-localized molecule strongly interfering with APP trafficking is stathmin-2 (*STMN2*), which binds to the KFFEQ sorting motif of APP, allowing its exit from the Golgi and subsequent transportation to the cell surface.<sup>30</sup> Stathmin-2 promotes non-amyloidogenic processing of APP.

Only a small portion of nascent APP is transported to the plasma membrane, whereas the bulk of APP molecules remains in the TGN. In neurons, APP traffics down the axons and dendrites via the anterograde transport machinery to reach synapses, which are a key site for A $\beta$  production and release. The anterograde transport of APP-containing vesicles along the microtubules is kinesin-1-dependent. Interfering with the kinesin light chain 1 (*KLC1*) gene expression modifies A $\beta$  production.<sup>31</sup> Moreover, reducing calsynenin-1 (*CLSTN1*)<sup>32</sup> and huntingtin-associated protein 1 (*HAP1*),<sup>33</sup> two ligands of *KLC1*, increases A $\beta$  production.

**Alpha-secretase cleavage.** At the cell membrane, APP is cleaved within the A $\beta$  domain between residues Lys16 and Leu17 as part of the non-amyloidogenic pathway.<sup>34</sup> This cleavage is performed by a set of metalloproteases termed  $\alpha$ -secretases, the principal member of them being ADAM10.<sup>35</sup> ADAM10 is localized in the Golgi as a pro-enzyme and then matured in the secretory pathway. Recently, two rare missense variants in the prodomain of ADAM10 that attenuate enzyme activity have been described in AD patients.<sup>36</sup> ADAM10 is controlled at the transcriptional level by the NAD-dependent protein deacetylase sirtuin-1 (*SIRT1*).<sup>37</sup> Its trafficking and maturation are regulated by proteins, such

as tetraspanin-12 (*TSPAN12*),<sup>38</sup> nardilysin (*NRD1*)<sup>39</sup> or synapse-associated protein 97 (*DLG1*),<sup>40</sup> whereas its activity is negatively regulated by the metalloproteinase inhibitors 1 and 3 (*TIMP1*, *TIMP3*).<sup>41</sup> The signaling pathways involving different cell membrane receptors and protein kinases may increase  $\alpha$ -secretase cleavage through incompletely characterized mechanisms, which implicate complex and very indirect mechanisms. Another metalloproteinase, ADAM17, also known as TACE, is responsible in a large part for the regulated  $\alpha$ -secretase cleavage.<sup>42</sup>

**Beta-secretase cleavage.** The amyloidogenic pathway, which is the prominent pathway of APP metabolism in neurons,<sup>20</sup> requires the sequential cleavage of APP by the membrane-bound enzyme beta-secretase 1 (*BACE1*) and the  $\gamma$ -secretase enzymatic complex. The aspartyl protease BACE1 is synthesized as a proenzyme, which is matured by furin-like mediated cleavage during its trafficking to the plasma membrane through the secretory pathway.<sup>43</sup> BACE1 then cycles between the cell membrane and endosomes in which it encounters an acidic pH optimum for its functioning. So far, ~25 proteins have been identified that specifically interfere with  $\beta$ -secretase cleavage of APP and A $\beta$  production. Although for most of them the mechanism by which they produce their effect remains uncharacterized, some are suspected to block BACE1-APP enzyme-substrate interaction, such as reticulons 1-4 (*RTN1-4*),<sup>44</sup> spondin-1 (*SPON1*)<sup>45</sup> or BRI2 (*ITM2B*) and BRI3 (*ITM2C*).<sup>46,47</sup> Other proteins regulate BACE1 trafficking within the secretory pathway. A subset, such as the Ras-related proteins Rab-11A and B,<sup>48</sup> the PH and SEC7 domain-containing protein 1 (*PSD*),<sup>49</sup> the ADP-ribosylation factor 6 (*ARF6*),<sup>49</sup> the EH domain-containing proteins (*EHD1* and 3),<sup>50</sup> and the ADP-ribosylation factor-binding protein GGA1,<sup>51</sup> plays a role in BACE1 endosomal sorting. Others, for example seladin-1 (*DHCR24*)<sup>52</sup> and GGA3,<sup>53</sup> target it for lysosomal degradation. The deubiquitinating enzyme ubiquitin carboxyl-terminal hydrolase isozyme L1 (*UCHL1*)<sup>54</sup> accelerates BACE1 degradation. Finally, the

copper chaperone for superoxide dismutase (CCS) binds to the intracellular domain of BACE1 and may deliver copper to it.<sup>55</sup> The role of copper in BACE1 function is not clear but may influence its trafficking and turnover.

**Gamma-secretase cleavage.** The other key player in the amyloidogenic processing of APP,  $\gamma$ -secretase, is a multiprotein complex comprising four transmembrane proteins, presenilin-1 or -2 (PSEN1 or PSEN2), nicastrin (NCSTN), the  $\gamma$ -secretase subunits APH1A or -1B (APH1A or APH1B) and PEN-2 (PSENEN) (reviewed in<sup>56</sup>). The presenilins contain the catalytic site of  $\gamma$ -secretase. Nicastrin acts as a substrate recruiter. PEN-2 is required for the autocatalytic maturation of presenilin and  $\gamma$ -secretase activity. APH1 forms stable subcomplexes with nicastrin that are believed to be an initial scaffold before the generation of the full presenilin complex. Around 25 interactors affecting, generally by inhibiting, the  $\gamma$ -secretase processing of APP and A $\beta$  generation have been identified. Again, for most of them, the molecular mechanisms underlying the biological effect remain elusive. However, several clues have been obtained. Many inhibitors directly interfere with the assembly of the complex. In addition to the four core components that assemble into a functional complex into the early compartments of the secretory pathway, additional proteins have been found to associate more or less transiently with the  $\gamma$ -secretase complex. Sorting nexin 27 (SNX27) associates with PSEN1 and dissociates  $\gamma$ -secretase complex,<sup>57</sup> phospholipase D1 (PLD1)<sup>58</sup> has been reported to interrupt the incorporation of PEN-2 into the complex. The protein RER1<sup>59</sup> is thought to downregulate  $\gamma$ -secretase assembly by competing with APH1 for nicastrin binding. In turn, the E3 ubiquitin–protein ligase synoviolin (SYVN1)<sup>60</sup> indirectly upregulates A $\beta$  production by targeting RER1 for degradation. The protein crumbs homolog 2 (CRB2) could compete with PSEN1 and APH1 for binding with nicastrin.<sup>61</sup>

Proteins that bind to the mature complex and interfere with A $\beta$  production have also been identified. The transmembrane protein TMP21 (TMED10), a member of p24 cargo protein, binds to the presenilin complex and reduces its activity.<sup>62</sup> The gamma-secretase-activating protein (GSAP), a 16 kDa protein derived from the proteolytic cleavage of an uncharacterized protein termed pigeon homolog protein, also binds to the presenilin complex.<sup>63</sup> It has also been claimed that basigin (BSG, alias CD147)<sup>64</sup> is a regulatory subunit of the  $\gamma$ -secretase complex.

Another set of proteins modulate A $\beta$  production by regulating the trafficking of  $\gamma$ -secretase. Of particular interest is the ARC/ARG3.1 protein (ARC), which binds PSEN1 and plays a key role in trafficking  $\gamma$ -secretase to postsynaptic vesicles that process APP in an activity-dependent manner.<sup>65</sup> In addition, several G protein-coupled receptors, in association with beta-arrestin-2 (ARRB2),<sup>66</sup> bind to APH1A and activate the  $\gamma$ -secretase complex by favoring its redistribution toward lipid rafts.

**Cholesterol homeostasis.** Both BACE1 and  $\gamma$ -secretase are enriched in cholesterol and sphingolipid-rich membrane raft microdomains (reviewed in<sup>67</sup>) where they interact with proteins such as caveolins and flotillins, which in turn are APP interactors. By modulating the fine partitioning of secretases and APP between membrane microdomains, these proteins are important modulators of APP processing and A $\beta$  generation. A $\beta$  generation is also dependent on the level of cholesterol within the lipid rafts where the proamyloidogenic cleaving machinery resides. High intracellular cholesterol levels result in enhanced release of A $\beta$ . Low intracellular cholesterol levels favor the processing of APP through the non-amyloidogenic pathway. Accordingly, proteins that regulate cholesterol homeostasis have been found to impact A $\beta$  biogenesis (reviewed in<sup>68</sup>). Almost all brain cholesterol originates from *in situ* neosynthesis. The sterol O-acyltransferase 1 (SOAT1, alias ACAT1) which transforms free

cholesterol into 24-OH-cholesterol allowing its storage or its efflux through the blood–brain barrier (BBB), the cholesterol 24-hydroxylase (CYP46A1) which mediates elimination of 24-OH-cholesterol through the BBB, or the Niemann-Pick C associated proteins (NPC1 and NPC2) which sequester intracellular cholesterol in the lysosomal compartment, all regulate the production of A $\beta$ . The ATP-binding cassette transporter 2 (ABCA2), that plays a role in intracellular sterol trafficking, regulates both the  $\beta$ - and  $\gamma$ -secretase cleavage of APP by a complex mechanism involving (i) accurate spatial segregation of APP and secretases at the compartmental and microdomain levels, and (ii) correct subcellular localization of nicastrin.<sup>69</sup>

**Internalization of APP.** The precise subcellular location(s) at which A $\beta$  is produced remains somewhat debated. Whereas some studies suggest that this processing can occur in the trans-Golgi network, it is generally agreed that the main site is the endosomal compartment where APP encounters  $\beta$ - and  $\gamma$ -secretases in a favorable membrane microenvironment. *In vivo* studies have demonstrated that around 70% of A $\beta$  released in the brain requires ongoing endocytosis<sup>70</sup> and internalization of APP present at the cell membrane. Indeed, ~70% of the surface-bound APP is internalized within minutes of arriving at the plasma membrane.<sup>71</sup> Within the intracellular domain of APP, the YENPTY motif contains the consensus sequence for clathrin-coated pit internalization. Deletion of this motif inhibits APP internalization and decreases A $\beta$  production. This motif is the intracellular binding site for several cytosolic APP partners that possess a phosphotyrosine binding domain, for example, the Fe65 family (APBB1-3), the mint/X11 family (APBA1-3), the disabled homologs 1 and 2 (DAB1 and 2) and the protein numb homolog (NUMB) (reviewed by<sup>72</sup>). Over-expression or knockout of these adaptors has been shown to affect APP processing and secreted A $\beta$  levels. Knowing that the complex formed by APP, FE65 and LRP-1 (one of the fastest proteins being internalized into a cell) promotes endocytosis of APP and that numb also plays a role in the internalization of various receptors involved in numerous cell functions, it is tempting to speculate that the effect of these adaptor proteins on APP processing is mediated at least in part by regulation of APP internalization. Of note, it has recently been shown that the disrupted in schizophrenia 1 protein (DISC1), a risk factor for several psychiatric disorders, also binds to the intracellular domain of APP and favors APP internalization.<sup>73</sup>

APP internalization is mediated by clathrin. Several members of the clathrin-mediated endocytotic machinery such as phosphatidylinositol-binding clathrin assembly protein (PICALM),<sup>74</sup> an adaptor protein that plays a key role in clathrin coat formation, the dynamins (DNM1 and DNM2)<sup>75</sup> which cleave the invaginated vesicles, or the sorting nexin-33 (SNX33),<sup>76</sup> a dynamin-binding partner, modulate A $\beta$  generation. By modifying the proportion of APP present at the surface membrane or in endosomes, these molecules regulate the amount of substrate available for amyloidogenic or non-amyloidogenic processing and are often referred to as indirect modulators of secretase activity. Note that the PICALM frequent variants were associated with AD in genome-wide association studies.<sup>14</sup>

**Endosomal sorting of APP.** Following endocytosis, APP is delivered to endosomes, and can be proteolytically processed by  $\beta$ - and  $\gamma$ -secretase, the cleavage products—including A $\beta$ —being subsequently released in the extracellular space, whereas another fraction undergoes degradation in lysosomes. Alternatively, endocytosed molecules containing mainly full-length APP can be recycled to the cell surface (reviewed by<sup>77</sup>). Intracellular trafficking through the endocytic and recycling pathway is regulated by the small GTPase Ras-related protein Rab, whose altered expression has been shown to modify A $\beta$  production.<sup>48</sup>

The sortilin-related receptor SorLA/LR11 (*SORL1*) and its four type 1 membrane homologs (sortilin, SorCS1, SorCS2 and SorCS3) act as molecular switches, allowing the alternative targeting of APP to recycling compartments or to the exocytosis pathway. By binding both APP and the vacuolar protein sorting-associated protein 35 (*VPS35*), they couple APP to the retromer, a protein complex localized at the endosomal membrane that allows retrograde transport of cargo proteins from the endosomes to the TGN (reviewed by<sup>78</sup>). Manipulating several members (*VPS35*, *VPS26*) of this complex has been shown to impact A $\beta$  production.<sup>79</sup> Interestingly, frequent non-coding variants of *SORL1* were associated with AD.<sup>14,17</sup> We previously found a high frequency of rare, predicted pathogenic, coding variants in the *SORL1* gene in a small sample of familial cases of early onset AD,<sup>80</sup> and we subsequently identified that rare predicted damaging missense and disruptive *SORL1* variants are a strong risk factor for early onset AD, with an even higher odds ratio for cases with positive family history.<sup>81</sup> Moreover, we recently documented a *VPS35* *de novo* missense mutation leading to partial loss-of-function of the protein in an AD patient.<sup>19</sup>

#### A $\beta$ clearance and degradation

After secretion, the A $\beta$  peptide binds to a neuronal membrane receptor to exert its physiological functions (see below) and is subsequently cleared. The major routes of A $\beta$  clearance include A $\beta$  transport across the BBB into the blood, extracellular proteolysis, receptor-mediated endocytosis followed by endosomal-lysosomal degradation and phagocytosis by microglia and other cells.

**Blood-brain barrier clearance.** The clearance of A $\beta$  across the BBB is mediated by a set of molecules. In the interstitial fluid, apolipoprotein-E (*APOE*) and clusterin (*CLU*, alias *APOJ*), mainly secreted by astrocytes, as well as alpha-2-macroglobulin (*A2M*), bind free A $\beta$  and cooperate to deliver it to the BBB.<sup>82</sup> Aside from *APOE*, *CLU* frequent variants are one of the strongest risk factors identified so far in AD<sup>14</sup>, and rare variants in *CLU* have also been associated with increased AD risk.<sup>83</sup> The product of the *APOE*  $\epsilon$ 4 risk allele binds A $\beta$  with lower affinity than does *APOE*  $\epsilon$ 3, suggesting it might be less efficient in mediating A $\beta$  clearance (reviewed by<sup>84</sup>). The lipidation status of *APOE*, thereby modifying its conformation, is a crucial parameter to govern its binding affinity for A $\beta$ . *APOE* is lipidated by the ATP-binding cassette transporter 1 (*ABCA1*), and gene dosage of *ABCA1* influences A $\beta$  clearance through its effect on *APOE* lipidation.<sup>85</sup> At the BBB, *APOE* and alpha-2-macroglobulin, complexed to A $\beta$ , bind to LRP-1, whereas clusterin-A $\beta$  complexes bind to LRP-2.<sup>86</sup> In mice, clusterin specifically increases the BBB clearance of A $\beta$ 42, the most toxic of A $\beta$  species.<sup>87</sup> To note, among this set of molecules acting primarily at the BBB, *APOE* and LRP-1 are also involved in an A $\beta$  neuronal reuptake pathway, which after endocytosis leads to lysosomal degradation of A $\beta$ .<sup>88</sup>

**Enzymatic degradation.** Proteolytic processing of A $\beta$  in the extracellular space is another important mechanism that controls A $\beta$  clearance. Several A $\beta$ -degrading enzymes have been characterized including neprilysin (*MME*),<sup>89</sup> insulin-degrading enzyme (*IDE*),<sup>90</sup> the matrix metalloproteinases 2 and 9 (*MMP2* and 9),<sup>91,92</sup> angiotensin-converting enzyme (*ACE*),<sup>93</sup> plasmin, generated from plasminogen (*PLG*), the endothelin-converting enzymes 1 and 2 (*ECE1* and 2),<sup>94</sup> glutamate carboxypeptidase 2 (*FOLH1*),<sup>95</sup> beta-secretase 2 (*BACE2*)<sup>96</sup> and cathepsin B (*CTSB*).<sup>97</sup> Together, they combine different endo- and exo-peptidase activities and target various forms of A $\beta$  in different states of aggregation and oligomerization, with greatest affinity for soluble A $\beta$ . With the exception of insulin-degrading enzyme which has been shown to degrade both cytosolic and extracellular A $\beta$ , most of them are either membrane-bound or secreted enzymes and are expected

to remove A $\beta$  from the extracellular milieu. Despite the lack of consistent association with common or rare variants reported so far, the genes encoding these enzymes, as well as their endogenous activator / inhibitor proteins, remain biologically relevant candidates in AD genetic risk assessment. Finally, some extracellular matrix proteins such as heparan sulfate proteoglycans, in addition to interfering with aggregation of the A $\beta$  peptide (see below), may also prevent proteolysis of the amyloid fibrils.

**Microglia phagocytosis.** Microglial cells belong to the lineage of innate immunity myeloid cells. They are able to clear the brain from danger-associated molecule patterns (DAMP) or pathogen-associated molecular patterns, through their binding to cell-surface receptors of the innate immunity. A $\beta$  can be assimilated to a DAMP as microglial cells have the ability to bind it through receptors from the same family. Microglia mediates A $\beta$  clearance by cleaving fibrillar aggregates (fA $\beta$ ) internalized through phagocytosis (reviewed by<sup>98</sup>). Microglia interact with fA $\beta$  via a set of cell-surface receptors, including toll-like receptors (*TLR1*, 2, 4, 6), the N-formyl peptide receptor 2 (*FPR2*), scavenger receptors (*SCARB1* and *MSR1*), and the advanced glycosylation end product-specific receptor (*AGER*). Upon binding fA $\beta$ , these receptors activate intracellular cascades triggering the phagocytic uptake of fA $\beta$  but also inducing inflammatory response.<sup>99,100</sup> Moreover, A $\beta$  activates the complement cascade. A complement C3-deficient transgenic AD mouse showed a significant increase in plaque burden.<sup>101</sup> Activated C3b fragment binds the membrane-associated complement receptor type 1 (*CR1*) allowing the ingestion of particles that have activated the complement. *CR1* is a known AD genetic risk factor (reviewed by<sup>102</sup>).

The balance between phagocytic and inflammatory responses is controlled by several chemokines or cytokines, for example, fractalkine (*CX3CL1*), C-C motif chemokine 2 (*CCL2*), transforming growth factor beta-1 (*TGFB1*), the interleukine-1 beta and -10 (*IL1B* and *IL10*) and other innate immune modulators, for example, the myeloid cell surface antigen CD33 (for review see<sup>103</sup>). The innate immune receptor TREM2, for which one rare variant (p.R47H) has been associated with the risk for AD,<sup>15</sup> can activate phagocytosis and create an anti-inflammatory cytokine milieu in microglia. TREM2 and its signaling adaptor protein TYROBP sense anionic lipids that interact with fA $\beta$ , and the p.R47H variant has been reported to attenuate microglial detection of these lipids.<sup>104</sup> Progranulin (*GRN*) is a secreted growth factor involved in the control of phagocytic activities. *GRN* haploinsufficiency is a cause of frontotemporal lobar degeneration with a peculiar neuropathological phenotype including TDP-43 positive inclusions.<sup>105,106</sup> However, some *GRN* mutation carriers exhibit AD pathology including neuritic plaques and NFT in neocortical regions.<sup>107</sup> Accordingly, selectively reducing microglial expression of *GRN* in AD mouse models impaired phagocytosis and increased plaque load.<sup>108</sup>

Finally, *ABCA7* is an AD risk factor for which both frequent<sup>14</sup> and rare<sup>109</sup> variants have been associated with the disease. As its closest homolog *ABCA1*, *ABCA7* has a lipid transport function. However, it is primarily expressed in microglial cells and its main role in AD pathogenesis might be the clearance of oligomeric A $\beta$ ,<sup>110</sup> although a role in APP processing has recently been suggested.<sup>111</sup>

**Proteasome and autophagy.** Besides these well-characterized mechanisms targeting extracellular A $\beta$ , two other degradative pathways are thought to be involved in intracellular A $\beta$  clearance. First, A $\beta$  can be degraded by the ubiquitin-proteasome system in neurons. During this process, A $\beta$  has been shown to inhibit the proteolytic activities of the 26S proteasome, leading to the accumulation of ubiquitinated A $\beta$  species in AD brains.<sup>112</sup> The E3 ubiquitin-protein ligase parkin (*PARK2*) mediates the ubiquitination of A $\beta$  and targets it to degradation by the proteasome.<sup>113</sup>

These results raise the question of the intriguing cytosolic localization of A $\beta$ , a place where it is not supposed to be. Leakage from endosomes or aberrant APP processing in the ER in conditions of ER stress, followed by subsequent targeting of misfolded fragments to proteasomal degradation by the ERAD system, may be involved.

In the same vein, dystrophic neurites associated with amyloid plaques have been shown to be enriched in autophagic vacuoles containing both APP and cleaved A $\beta$  peptide.<sup>114</sup> Autophagy is a cellular process by which cellular components are isolated from the rest of the cell within a double-membraned vesicle known as an autophagosome. The autophagosome then fuses with a lysosome and its content is degraded. In a transgenic APP mouse model, genetic reduction of beclin-1 (*BECN1*), a molecular platform assembling an interactome that regulates the initiation of the autophagosome, increased intraneuronal A $\beta$  accumulation, extracellular A $\beta$  deposition and neurodegeneration. However, it remains to be determined whether beclin-1 is necessary for the degradation of intracellular A $\beta$  itself or of APP by autophagy.<sup>115</sup>

### A $\beta$ aggregation

Which specific A $\beta$  assembly state is the etiological agent in AD? Although the emphasis has shifted from highly insoluble A $\beta$  fibrils, such as those present in senile plaques, to intermediate soluble oligomers (A $\beta$ o), the identity of the real culprit remains a subject of controversy.<sup>116</sup> Thus, every intermediate between the monomeric A $\beta$  physiologically released by neurons and the probably inert fibrils found in plaques is a suspect and every biological process modulating A $\beta$  aggregation should be considered.

Post-translational modifications of A $\beta$  can modify its propensity to aggregate. A $\beta$  peptide can be cross-linked by transglutaminase-2 (*TGM2*), forming multimers.<sup>117</sup> Similarly, formation of pyroglutamic acid at position 3 or 11, catalyzed by the glutamyl-peptide cyclotransferase (*QPCT*), enforces oligomer formation.<sup>118</sup> In addition, phosphorylation of serine residue 8 by extracellular soluble protein kinase A (*PKA*) also increases the formation of oligomeric A $\beta$  aggregates.<sup>119</sup> Zinc is another contributor of A $\beta$  aggregation. Synaptic Zn<sup>2+</sup> delivery is under the control of the zinc transporter 3 (*SLC30A3*). The absence of synaptic zinc by deleting *SLC30A3* in APP mice reduces plaque load and increases the ratio of soluble/insoluble A $\beta$ .<sup>120</sup>

Molecular chaperones recognize and bind misfolded proteins, prevent their aggregation and assist in their correct refolding. Several members of this class of protein, including the heat shock 70 kDa protein 4 and DnaJ homolog subfamily members (*HSPA4*, *DNAJB1*, *DNAJB6*),<sup>121</sup> the heat shock protein HSP 90-alpha (*HSP90AA1*),<sup>121</sup> HSP 27 (*HSPB1*)<sup>122</sup> and the alpha-crystallin B chain (*HSPB5*),<sup>123</sup> can interact with A $\beta$  and inhibit the formation of A $\beta$  fibrils by maintaining the peptide in an oligomeric state.

Finally, a heterogeneous set of molecules present in the extracellular milieu have been shown to modulate the kinetics of amyloid fibril formation *in vitro*. Most of these molecules colocalize with A $\beta$  in senile plaques. Some of these molecules, such as acetylcholinesterase (*ACHE*),<sup>124</sup> APOE,<sup>125</sup> agrin (*AGRN*),<sup>126</sup> and perlecan (*HSPG2*)<sup>127</sup> promote amyloid fibril formation, whereas others such as cholinesterase (*BCHE*),<sup>128</sup> transthyretin (*TTR*),<sup>129</sup> clusterin,<sup>130</sup> alpha-2-macroglobulin (*A2M*),<sup>131</sup> cystatin-C (*CST3*),<sup>132,133</sup> alpha-1-antichymotrypsin (*SERPINA3*)<sup>134</sup> and reelin (*RELN*)<sup>135</sup> inhibit fibril formation. Of note, two of the strongest AD frequent genetic risk factors, *APOE* and *CLU*, encode proteins that are clearly involved both in A $\beta$  clearance and aggregation.

**Cross seeding.** Two proteins interfering with A $\beta$  aggregation, alpha-synuclein (*SNCA*) and TDP-43 (*TARDBP*), deserve particular attention because (i) they form pathological aggregates involved in other neurodegenerative disorders and (ii) for each of them, small amounts of misfolded protein can act as seeds to initiate

templated recruitment of their soluble counterparts into fibrils. Importantly, for each of these proteins a particular oligomer seed can influence fibrillization of other aggregation-prone proteins according to a specific template. A $\beta$  is itself a peptide for which a seeding effect has been demonstrated (for review see<sup>136</sup>). Therefore, the 'cross seeding' effect by which alpha-synuclein, TDP-43 and A $\beta$  influence each other's aggregation,<sup>137–140</sup> probably explains why 30–50% of AD cases also present with pathological intraneuronal aggregates of synuclein or TDP-43, in addition to senile plaques and NFT.<sup>141,142</sup>

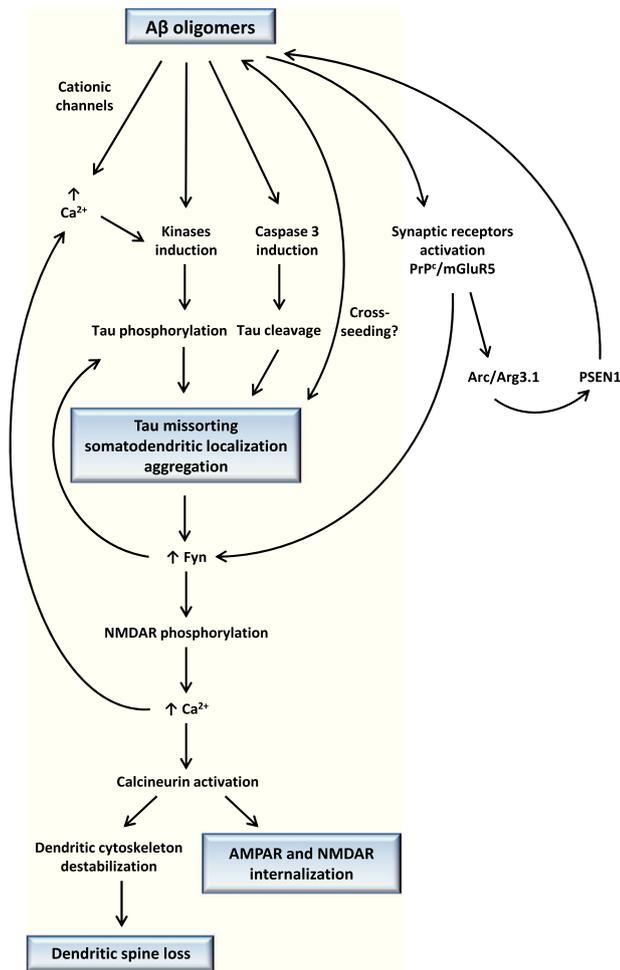
### A $\beta$ toxicity

As previously discussed, it is now generally agreed that, in AD, A $\beta$  functions upstream of tau and that tau is necessary for A $\beta$  toxicity.<sup>143,144</sup> On the other hand, tau pathology in the absence of overt A $\beta$  pathology is sufficient to cause neurodegeneration, whereas there is no example of a human neurodegenerative disease associated with A $\beta$  pathology alone. Overall, these facts support the view that, in AD, A $\beta$  is the trigger and tau the bullet.<sup>10</sup>

**A $\beta$ -tau interplay.** In FTDP-17, due to missense or splicing mutations in the *MAPT* gene, tau is detached from the microtubules, hyperphosphorylated, and redistributed from axons into the somatodendritic compartment where it undergoes conformational changes, oligomerizes and finally polymerizes into insoluble filaments. Missorting of tau also causes the targeting of the tyrosine-protein kinase Fyn, with which tau interacts, to the dendrites.<sup>145</sup> Here, Fyn phosphorylates the glutamatergic N-methyl-d-aspartate (NMDA) receptor subunit NR2B. The resulting upregulation of NMDA receptor activity may lead to a harmful enhancement of Ca<sup>2+</sup> entry, resulting in synapses and neuronal loss (discussed by<sup>146</sup>).

In AD, in the absence of *MAPT* pathogenic mutations, the exact mechanism by which A $\beta$  triggers these events is not completely understood. However, as discussed in several recent reviews<sup>9,147–149</sup> a plausible scenario has emerged (Figure 2). The three features that characterize abnormal tau in all tauopathies are hyperphosphorylation, missorting and aggregation.<sup>150</sup> An open question is the relationship between these changes. It is often assumed that in AD, hyperphosphorylation plays a key role because it can (i) promote the dissociation of tau from microtubules allowing subsequent missorting<sup>151,152</sup> and (ii) induce structural modifications that promote aggregation. In AD, tau is both hyperphosphorylated at 'physiological' sites and abnormally phosphorylated at specific sites.<sup>153,154</sup> The hyperphosphorylation of tau is catalyzed by a combination of kinases including glycogen synthase kinase-3 beta (*GSK3B*), the 5'-AMP-activated protein kinase (AMPK), the cyclin-dependent-like kinase 5 (*CDK5*), the MAP/microtubule affinity-regulating kinases (MARKs) and the dual specificity tyrosine-phosphorylation-regulated kinase 1A (*DYRK1A*). In addition, the serine/threonine-protein phosphatase 2A (PP2A) regulates the level of tau phosphorylation.<sup>155</sup> In this complex interplay, A $\beta$  can induce several kinase activities resulting in tau hyperphosphorylation.<sup>156–160</sup> In most cases, kinase dysregulation is mediated by A $\beta$ -induced disruption of calcium homeostasis. Due to its ability to form cationic channels in the plasma membrane or the activation of NMDA receptors, A $\beta$  induces a calcium influx into the neuron. A $\beta$  is also able to mobilize the ER calcium store to the cytosol through activation of ryanodine receptors (RyR),<sup>161</sup> inositol 1,4,5-trisphosphate receptor type 1 (*ITPR1*)<sup>162</sup> or calcium homeostasis modulator protein 1 (*CALHM1*).<sup>163</sup> This influx of calcium directly activates at least two phosphorylation pathways, that is, the CAMKK2-AMPK pathway and the calpain/p35/Cdk5 pathway.

Beyond its driving effect on tau phosphorylation, A $\beta$  might also interfere with tau oligomerization and aggregation. A $\beta$  can induce caspase-3 (*CASP3*) cleavage of tau, resulting in a truncated form



**Figure 2.** The amyloid cascade: schematic representation of amyloid beta (A $\beta$ )-tau interplay leading to A $\beta$ -driven neuronal toxicity. Via Ca<sup>2+</sup>-dependent processes, A $\beta$  oligomers can drive tau phosphorylation and cleavage, promoting the detachment of tau from microtubules, its misfolding and its redistribution to the somatodendritic compartment. In addition, tau and A $\beta$  could influence each other's self-aggregation. Tau binds to the Fyn kinase and targets it to a dendritic localization. A $\beta$  oligomers bound to postsynaptic prion protein lead to Fyn activation and neuron impairment via NMDAR phosphorylation, internalization and ultimately spine loss. PrP<sup>C</sup>, prion protein; NMDAR, NMDA (*N*-methyl-D-aspartate) receptor; AMPAR,  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor.

that is more aggregation-prone.<sup>164</sup> Tau is a natively unfolded protein, and conformational changes are required to bypass the rate-limiting nucleation step and accelerate polymerization of monomeric tau. Misfolded hyperphosphorylated and/or cleaved tau proteins have prion-like seeding properties: at their contact, wild-type protein can adopt an abnormal conformation. It has been reported that seeding of normal tau by pathological tau conformers rapidly recruits large amounts of soluble tau into filamentous inclusions.<sup>165</sup> In addition, A $\beta$  and tau can form soluble complexes that may promote self-aggregation of both.<sup>166</sup>

**Synaptic toxicity.** Synapses are an important site for A $\beta$  production and release. A $\beta$  can be released from both presynaptic and postsynaptic compartments and synaptic activity increases A $\beta$  release. In turn, A $\beta$  depresses excitatory synaptic transmission by driving endocytosis of AMPA glutamate receptors and reduces

spine density.<sup>167</sup> These data strongly suggest that a key physiological role of A $\beta$  is maintaining synaptic homeostasis by preventing potentially harmful excess of synaptic activity. This protective role is compromised in AD. A $\beta$  binds to multiple synaptic sites including excitatory neurotransmitter receptors modulating calcium homeostasis such as the neuronal acetylcholine receptor subunit alpha-7 (*CHRNA7*)<sup>168</sup> or signaling proteins expressed at the synaptic membrane such as the cellular prion protein (PrP<sup>C</sup>, *PRNP*)<sup>169,170</sup> the leukocyte immunoglobulin-like receptor 2 (*LILRB2*),<sup>171</sup> the low affinity neurotrophin receptor p75NTR (*NGFR*)<sup>172</sup> or the ephrin type-B receptor 2 (*EPHB2*) (reviewed by<sup>173</sup>). Notably, binding of A $\beta$  oligomers to the PrP<sup>C</sup>/mGluR5/PSD-95 complex activates Fyn, which facilitates the stabilization of the complex by recruitment of PSD-95<sup>174</sup> and increases the phosphorylation of the NMDAR subunit NR2B.<sup>146</sup> This process is associated with transient increase in NR2B at the cell surface, yielding elevated neuronal calcium influx and consequent excitotoxicity.<sup>170</sup> Eventually, exposure to A $\beta$  oligomers results in the loss of surface NMDAR.<sup>175,176</sup> Downstream of calcium influx, the Ca<sup>2+</sup>-dependent protein phosphatase calcineurin/PP2B (*PPP3CA, B and C*) is a key mediator of the A $\beta$ -induced excitotoxic processes. Upon activation, calcineurin on one hand induces a cascade of events, leading to tau hyperphosphorylation and cleavage,<sup>148,149</sup> on the other hand, it might also activate cofilin, thereby inducing depolymerization of F-actin and dendritic spine collapse.<sup>176</sup>

In addition, signaling by A $\beta$ o-PrP<sup>C</sup>-mGluR5 complexes mediates phosphorylation of the elongation factor 2 by a pathway requiring Fyn.<sup>174</sup> In turn, this allows increased expression of the activity-regulated *ARC* gene, whose product physically associates with PSEN1 and upregulates it,<sup>65</sup> thereby increasing the amount of activity-dependent A $\beta$  generated and amplifying the initial signal. It should be emphasized that if activation of Fyn, whose migration into dendritic spines has been facilitated by tau, is a key event in the A $\beta$ o-induced synaptic toxicity, the exact mechanisms by which synaptic transmission is impaired in the AD brain are far from being fully understood. More interactors than those listed here certainly play a role in these processes.

**Mitochondrial toxicity.** Another cellular compartment in which A $\beta$  and tau conspire to damage neurons is the mitochondria. It has been shown that the intracellular A $\beta$  can be translocated into mitochondria. Here, it interacts with the 3-hydroxyacyl-CoA dehydrogenase type-2 (*HSD17B10*, also known as *ABAD*), and this interaction promotes mitochondrial dysfunction and leakage of reactive oxygen species.<sup>177</sup> Another protein localized at the outer mitochondrial membrane, the dynamin-1-like protein Drp1 (*DNM1L*), can interact with both A $\beta$  and phosphorylated tau, likely leading to excessive mitochondrial fragmentation and neuronal damage.<sup>178</sup>

## CONCLUSION

Here we have explored an A $\beta$ -centered network of genes potentially involved in AD pathophysiology. A large number of biological functions and cellular compartments are involved in A $\beta$  homeostasis. To guarantee the specificity of the network, we have chosen to restrict our search to a reasonably narrow definition of the amyloid cascade and have not included genes related to comorbid pathologies that might exacerbate amyloid pathology, e.g. diabetes or vascular pathologies. Nor were the genes of apoptotic cascades that are triggered in cell models by exposure to high concentrations of A $\beta$  considered. Depending on the interest of the researchers, broader and less conservative versions of this network can easily be built by adding additional subcomponents. Of course, the delineation of the network is based on the current knowledge of A $\beta$  biology, is not exhaustive, and would greatly benefit from regular updates.

Confronting this set of gene whose function is theoretically strongly related to the amyloid cascade hypothesis with the actual pattern of mutations found in AD patients will be a major challenge in the next years. Focusing on sub-parts of the networks will offer opportunities to decipher the physiopathology of the disease with more accuracy. Will all sub-processes be equally involved or could some hierarchy be demonstrated? Only such comings and goings from expert-curated networks to genetic data will help answer these questions. When the genetic architecture of AD becomes clearer it will be interesting to compare the topology of this network with interactome networks from public databases or regulatory networks inferred from transcriptome data. Finally, the amyloid cascade hypothesis is still debated, with pro<sup>12</sup> and con<sup>179</sup> arguments. Specifically, the validity of the amyloid cascade hypothesis in sporadic AD remains a matter of discussion.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### ACKNOWLEDGMENTS

Authors would like to thank Magalie Lecourtois and Camille Charbonnier for helpful discussions and Tracey Avequin for editing the manuscript. The authors are supported by grants from the Clinical Research Hospital Program from the French Ministry of Health (GMAJ, PHRC 2008/067), the CNR-MAJ, the JPND PERADES and Inserm.

### REFERENCES

- Rosso SM, Kamphorst W, Ravid R, van Swieten JC. Coexistent tau and amyloid pathology in hereditary frontotemporal dementia with tau mutations. *Ann NY Acad Sci* 2000; **920**: 115–119.
- Ishida C, Kobayashi K, Kitamura T, Ujike H, Iwasa K, Yamada M. Frontotemporal dementia with parkinsonism linked to chromosome 17 with the MAPT R406W mutation presenting with a broad distribution of abundant senile plaques. *Neuropathology* 2015; **35**: 75–82.
- Hurtado DE, Molina-Porcel L, Iba M, Aboagye AK, Paul SM, Trojanowski JQ *et al*. A[ $\beta$ ] accelerates the spatiotemporal progression of tau pathology and augments tau amyloidosis in an Alzheimer mouse model. *Am J Pathol* 2010; **177**: 1977–1988.
- Hardy JA, Higgins GA. Alzheimer's disease: the amyloid cascade hypothesis. *Science* 1992; **256**: 184–185.
- Walsh DM, Selkoe DJ. A beta oligomers—a decade of discovery. *J Neurochem* 2007; **101**: 1172–1184.
- Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol* 1991; **82**: 239–259.
- Arriagada PV, Growdon JH, Hedley-Whyte ET, Hyman BT. Neurofibrillary tangles but not senile plaques parallel duration and severity of Alzheimer's disease. *Neurology* 1992; **42**: 631–639.
- Delacourte A, Sergeant N, Champain D, Watzet A, Maurage CA, Lebert F *et al*. Nonoverlapping but synergetic tau and APP pathologies in sporadic Alzheimer's disease. *Neurology* 2002; **59**: 398–407.
- Iltner LM, Gotz J. Amyloid-beta and tau—a toxic pas de deux in Alzheimer's disease. *Nat Rev Neurosci* 2011; **12**: 65–72.
- Bloom GS. Amyloid-beta and tau: the trigger and bullet in Alzheimer disease pathogenesis. *JAMA Neurol* 2014; **71**: 505–508.
- Duyckaerts C. Neurodegenerative lesions: seeding and spreading. *Rev Neurol* 2013; **169**: 825–833.
- Musiek ES, Holtzman DM. Three dimensions of the amyloid hypothesis: time, space and 'wingmen'. *Nat Neurosci* 2015; **18**: 800–806.
- Genin E, Hannequin D, Wallon D, Sleegers K, Hiltunen M, Combarros O *et al*. APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol Psychiatry* 2011; **16**: 903–907.
- Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C *et al*. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45**: 1452–1458.
- Jonsson T, Stefansson H, Steinberg S, Jonsson PV, Snaedal J *et al*. Variant of TREM2 associated with the risk of Alzheimer's disease. *New Engl J Med* 2013; **368**: 107–116.
- Wu G, Zhi D. Pathway-based approaches for sequencing-based genome-wide association studies. *Genet Epidemiol* 2013; **37**: 478–494.
- Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F *et al*. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat Genet* 2007; **39**: 168–177.
- Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D *et al*. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014; **515**: 216–221.
- Rovelet-Lecrux A, Charbonnier C, Wallon D, Nicolas G, Seaman MN, Pottier C *et al*. De novo deleterious genetic variations target a biological network centered on Abeta peptide in early-onset Alzheimer disease. *Mol Psychiatry* 2015; **20**: 1046–1056.
- Haass C, Kaether C, Thinakaran G, Sisodia S. Trafficking and proteolytic processing of APP. *Cold Spring Harb Perspect Med* 2012; **2**: a006270.
- Claeysen S, Cochet M, Donnegier R, Dumuis A, Bockaert J, Giannoni P. Alzheimer culprits: cellular crossroads and interplay. *Cell Signal* 2012; **24**: 1831–1840.
- De Strooper B. Proteases and proteolysis in Alzheimer disease: a multifactorial view on the disease process. *Physiol Rev* 2010; **90**: 465–494.
- Hoshino T, Nakaya T, Araki W, Suzuki K, Suzuki T, Mizushima T. Endoplasmic reticulum chaperones inhibit the production of amyloid-beta peptides. *Biochem J* 2007; **402**: 581–589.
- Huttunen HJ, Guenette SY, Peach C, Greco C, Xia W, Kim DY *et al*. HtrA2 regulates beta-amyloid precursor protein (APP) metabolism through endoplasmic reticulum-associated degradation. *J Biol Chem* 2007; **282**: 28285–28295.
- Kaneko M, Koike H, Saito R, Kitamura Y, Okuma Y, Nomura Y. Loss of HRD1-mediated protein degradation causes amyloid precursor protein accumulation and amyloid-beta generation. *J Neurosci* 2010; **30**: 3924–3932.
- Kumar P, Ambasta RK, Veereshwarayya V, Rosen KM, Kosik KS, Band H *et al*. CHIP and HSPs interact with beta-APP in a proteasome-dependent manner and influence Abeta metabolism. *Hum Mol Genet* 2007; **16**: 848–864.
- Ullrich S, Munch A, Neumann S, Kremmer E, Tatzelt J, Lichtenthaler SF. The novel membrane protein TMEM59 modulates complex glycosylation, cell surface expression, and secretion of the amyloid precursor protein. *J Biol Chem* 2010; **285**: 20664–20674.
- Parisiadou L, Bethani I, Michaki V, Krousti K, Rapti G, Efthimiopoulos S. Homer2 and Homer3 interact with amyloid precursor protein and inhibit Abeta production. *Neurobiol Dis* 2008; **30**: 353–364.
- Zhou X, Hu X, He W, Tang X, Shi Q, Zhang Z *et al*. Interaction between amyloid precursor protein and Nogo receptors regulates amyloid deposition. *FASEB J* 2011; **25**: 3146–3156.
- Wang J, Shan C, Cao W, Zhang C, Teng J, Chen J. SCG10 promotes non-amyloidogenic processing of amyloid precursor protein by facilitating its trafficking to the cell surface. *Hum Mol Genet* 2013; **22**: 4888–4900.
- Stokin GB, Lillo C, Falzone TL, Brusch RG, Rockenstein E, Mount SL *et al*. Axonopathy and transport deficits early in the pathogenesis of Alzheimer's disease. *Science* 2005; **307**: 1282–1288.
- Steuble M, Diep TM, Schatzle P, Ludwig A, Tagaya M, Kunz B *et al*. Calsynenin-1 shelters APP from proteolytic processing during anterograde axonal transport. *Biol Open* 2012; **1**: 761–774.
- Yang GZ, Yang M, Lim Y, Lu JJ, Wang TH, Qi JG *et al*. Huntingtin associated protein 1 regulates trafficking of the amyloid precursor protein and modulates amyloid beta levels in neurons. *J Neurochem* 2012; **122**: 1010–1022.
- Sisodia SS. Beta-amyloid precursor protein cleavage by a membrane-bound protease. *Proc Natl Acad Sci USA* 1992; **89**: 6075–6079.
- Kuhn PH, Wang H, Dislich B, Colombo A, Zeitschel U, Ellwart JW *et al*. ADAM10 is the physiologically relevant, constitutive alpha-secretase of the amyloid precursor protein in primary neurons. *EMBO J* 2010; **29**: 3020–3032.
- Suh J, Choi SH, Romano DM, Gannon MA, Lesinski AN, Kim DY *et al*. ADAM10 missense mutations potentiate beta-amyloid accumulation by impairing prodomain chaperone function. *Neuron* 2013; **80**: 385–401.
- Donmez G, Wang D, Cohen DE, Guarente L. SIRT1 suppresses beta-amyloid production by activating the alpha-secretase gene ADAM10. *Cell* 2010; **142**: 320–332.
- Xu D, Sharma C, Hemler ME. Tetraspanin12 regulates ADAM10-dependent cleavage of amyloid precursor protein. *FASEB J* 2009; **23**: 3674–3681.
- Ohno M, Hiraoka Y, Lichtenthaler SF, Nishi K, Saijo S, Matsuoka T *et al*. Nardilysin prevents amyloid plaque formation by enhancing alpha-secretase activity in an Alzheimer's disease mouse model. *Neurobiol Aging* 2014; **35**: 213–222.
- Marcello E, Gardoni F, Mauceri D, Romorini S, Jeromin A, Epis R *et al*. Synapse-associated protein-97 mediates alpha-secretase ADAM10 trafficking and promotes its activity. *J Neurosci* 2007; **27**: 1682–1691.
- Amour A, Knight CG, Webster A, Slocombe PM, Stephens PE, Knauper V *et al*. The *in vitro* activity of ADAM-10 is inhibited by TIMP-1 and TIMP-3. *FEBS Lett* 2000; **473**: 275–279.
- Vingtdeux V, Marambaud P. Identification and biology of alpha-secretase. *J Neurochem* 2012; **120**: 34–45.

- 43 Capell A, Steiner H, Willem M, Kaiser H, Meyer C, Walter J *et al*. Maturation and pro-peptide cleavage of beta-secretase. *J Biol Chem* 2000; **275**: 30849–30854.
- 44 He W, Lu Y, Qahwash I, Hu XY, Chang A, Yan R. Reticulon family members modulate BACE1 activity and amyloid-beta peptide generation. *Nat Med* 2004; **10**: 959–965.
- 45 Ho A, Sudhof TC. Binding of F-spondin to amyloid-beta precursor protein: a candidate amyloid-beta precursor protein ligand that modulates amyloid-beta precursor protein cleavage. *Proc Natl Acad Sci USA* 2004; **101**: 2548–2553.
- 46 Matsuda S, Giliberto L, Matsuda Y, McGowan EM, D'Adamio L. BRI2 inhibits amyloid beta-peptide precursor protein processing by interfering with the docking of secretases to the substrate. *J Neurosci* 2008; **28**: 8668–8676.
- 47 Matsuda S, Matsuda Y, D'Adamio L. BRI3 inhibits amyloid precursor protein processing in a mechanistically distinct manner from its homologue dementia gene BRI2. *J Biol Chem* 2009; **284**: 15815–15825.
- 48 Udayar V, Buggia-Prevot V, Guerreiro RL, Siegel G, Rambabu N, Soohoo AL *et al*. A paired RNAi and RabGAP overexpression screen identifies Rab11 as a regulator of beta-amyloid production. *Cell Rep* 2013; **5**: 1536–1551.
- 49 Sannerud R, Declerck I, Peric A, Raemaekers T, Menendez G, Zhou L *et al*. ADP ribosylation factor 6 (ARF6) controls amyloid precursor protein (APP) processing by mediating the endosomal sorting of BACE1. *Proc Natl Acad Sci USA* 2011; **108**: E559–E568.
- 50 Buggia-Prevot V, Fernandez CG, Udayar V, Vetrivel KS, Elie A, Roseman J *et al*. A function for EHD family proteins in unidirectional retrograde dendritic transport of BACE1 and Alzheimer's disease Abeta production. *Cell Rep* 2013; **5**: 1552–1563.
- 51 Wahle T, Thal DR, Sastre M, Rentmeister A, Bogdanovic N, Famulok M *et al*. GGA1 is expressed in the human brain and affects the generation of amyloid beta-peptide. *J Neurosci* 2006; **26**: 12838–12846.
- 52 Sarajarvi T, Haapasalo A, Viswanathan J, Makinen P, Laitinen M, Soininen H *et al*. Down-regulation of seladin-1 increases BACE1 levels and activity through enhanced GGA3 depletion during apoptosis. *J Biol Chem* 2009; **284**: 34433–34443.
- 53 Tesco G, Koh YH, Kang EL, Cameron AN, Das S, Sena-Estevés M *et al*. Depletion of GGA3 stabilizes BACE and enhances beta-secretase activity. *Neuron* 2007; **54**: 721–737.
- 54 Zhang M, Deng Y, Luo Y, Zhang S, Zou H, Cai F *et al*. Control of BACE1 degradation and APP processing by ubiquitin carboxyl-terminal hydrolase L1. *J Neurochem* 2012; **120**: 1129–1138.
- 55 Gray EH, De Vos KJ, Dingwall C, Perkinton MS, Miller CC. Deficiency of the copper chaperone for superoxide dismutase increases amyloid-beta production. *JAD* 2010; **21**: 1101–1105.
- 56 Li Y, Bohm C, Dodd R, Chen F, Qamar S, Schmitt-Ulms G *et al*. Structural biology of presenilin 1 complexes. *Mol Neurodegener* 2014; **9**: 59.
- 57 Wang X, Huang T, Zhao Y, Zheng Q, Thompson RC, Bu G *et al*. Sorting nexin 27 regulates Abeta production through modulating gamma-secretase activity. *Cell Rep* 2014; **9**: 1023–1033.
- 58 Cai D, Netzer WJ, Zhong M, Lin Y, Du G, Frohman M *et al*. Presenilin-1 uses phospholipase D1 as a negative regulator of beta-amyloid formation. *Proc Natl Acad Sci USA* 2006; **103**: 1941–1946.
- 59 Park HJ, Shabashvili D, Nekorchuk MD, Shyqriyu E, Jung JI, Ladd TB *et al*. Retention in endoplasmic reticulum 1 (RER1) modulates amyloid-beta (Abeta) production by altering trafficking of gamma-secretase and amyloid precursor protein (APP). *J Biol Chem* 2012; **287**: 40629–40640.
- 60 Tanabe C, Maeda T, Zou K, Liu J, Liu S, Nakajima T *et al*. The ubiquitin ligase synviolin up-regulates amyloid beta production by targeting a negative regulator of gamma-secretase, Rer1, for degradation. *J Biol Chem* 2012; **287**: 44203–44211.
- 61 Mitsuishi Y, Hasegawa H, Matsuo A, Araki W, Suzuki T, Tagami S *et al*. Human CRB2 inhibits gamma-secretase cleavage of amyloid precursor protein by binding to the presenilin complex. *J Biol Chem* 2010; **285**: 14920–14931.
- 62 Pardossi-Piquard R, Bohm C, Chen F, Kanemoto S, Checler F, Schmitt-Ulms G *et al*. TMAP21 transmembrane domain regulates gamma-secretase cleavage. *J Biol Chem* 2009; **284**: 28634–28641.
- 63 He G, Luo W, Li P, Remmers C, Netzer WJ, Hendrick J *et al*. Gamma-secretase activating protein is a therapeutic target for Alzheimer's disease. *Nature* 2010; **467**: 95–98.
- 64 Zhou S, Zhou H, Walian PJ, Jap BK. CD147 is a regulatory subunit of the gamma-secretase complex in Alzheimer's disease amyloid beta-peptide production. *Proc Natl Acad Sci USA* 2005; **102**: 7499–7504.
- 65 Wu J, Petralia RS, Kurushima H, Patel H, Jung MY, Volk L *et al*. Arc/Arg3.1 regulates an endosomal pathway essential for activity-dependent beta-amyloid generation. *Cell* 2011; **147**: 615–628.
- 66 Thathiah A, Horre K, Snellinx A, Vandeweyer E, Huang Y, Ciesielska M *et al*. Beta-arrestin 2 regulates Abeta generation and gamma-secretase activity in Alzheimer's disease. *Nat Med* 2013; **19**: 43–49.
- 67 Vetrivel KS, Thinakaran G. Membrane rafts in Alzheimer's disease beta-amyloid production. *Biochim Biophys Acta* 2010; **1801**: 860–867.
- 68 Wollmer MA. Cholesterol-related genes in Alzheimer's disease. *Biochim Biophys Acta* 2010; **1801**: 762–773.
- 69 Michaki V, Guix FX, Vennekens K, Munck S, Dingwall C, Davis JB *et al*. Down-regulation of the ATP-binding cassette transporter 2 (Abca2) reduces amyloid-beta production by altering nicastrin maturation and intracellular localization. *J Biol Chem* 2012; **287**: 1100–1111.
- 70 Cirrito JR, Kang JE, Lee J, Stewart FR, Verges DK, Silverio LM *et al*. Endocytosis is required for synaptic activity-dependent release of amyloid-beta *in vivo*. *Neuron* 2008; **58**: 42–51.
- 71 Cheng H, Vetrivel KS, Gong P, Meckler X, Parent A, Thinakaran G. Mechanisms of disease: new therapeutic strategies for Alzheimer's disease—targeting APP processing in lipid rafts. *Nat Clin Pract Neurol* 2007; **3**: 374–382.
- 72 King GD, Scott Turner R. Adaptor protein interactions: modulators of amyloid precursor protein metabolism and Alzheimer's disease risk? *Exp Neurol* 2004; **185**: 208–219.
- 73 Shahani N, Seshadri S, Jaaro-Peled H, Ishizuka K, Hirota-Tsuyada Y, Wang Q *et al*. DISC1 regulates trafficking and processing of APP and Abeta generation. *Mol Psychiatry* 2015; **20**: 874–879.
- 74 Xiao Q, Gil SC, Yan P, Wang Y, Han S, Gonzales E *et al*. Role of phosphatidylinositol clathrin assembly lymphoid-myeloid leukemia (PICALM) in intracellular amyloid precursor protein (APP) processing and amyloid plaque pathogenesis. *J Biol Chem* 2012; **287**: 21279–21289.
- 75 Carey RM, Balcz BA, Lopez-Coviella I, Slack BE. Inhibition of dynamin-dependent endocytosis increases shedding of the amyloid precursor protein ectodomain and reduces generation of amyloid beta protein. *BMC Cell Biol* 2005; **6**: 30.
- 76 Schobel S, Neumann S, Hertweck M, Dislich B, Kuhn PH, Kremmer E *et al*. A novel sorting nexin modulates endocytic trafficking and alpha-secretase cleavage of the amyloid precursor protein. *J Biol Chem* 2008; **283**: 14257–14268.
- 77 Small SA, Gandy S. Sorting through the cell biology of Alzheimer's disease: intracellular pathways to pathogenesis. *Neuron* 2006; **52**: 15–31.
- 78 Reitz C. The role of the retromer complex in aging-related neurodegeneration: a molecular and genomic review. *Mol Genet Genomics* 2015; **290**: 413–427.
- 79 Muhammad A, Flores I, Zhang H, Yu R, Staniszewski A, Planel E *et al*. Retromer deficiency observed in Alzheimer's disease causes hippocampal dysfunction, neurodegeneration, and Abeta accumulation. *Proc Natl Acad Sci USA* 2008; **105**: 7327–7332.
- 80 Pottier C, Hannequin D, Coutant S, Rovelet-Lecrux A, Wallon D, Rousseau S *et al*. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. *Mol Psychiatry* 2012; **17**: 875–879.
- 81 Nicolas G, Charbonnier C, Wallon D, Quenez O, Bellenguez C, Grenier-Boley B *et al*. SORL1 rare variants: a major risk factor for familial early-onset Alzheimer's disease. *Mol Psychiatry* 2015. doi: 10.1038/mp.2015.121.
- 82 DeMattos RB, Cirrito JR, Parsadanian M, May PC, O'Dell MA, Taylor JW *et al*. ApoE and clusterin cooperatively suppress Abeta levels and deposition: evidence that ApoE regulates extracellular Abeta metabolism *in vivo*. *Neuron* 2004; **41**: 193–202.
- 83 Bettens K, Brouwers N, Engelborghs S, Lambert JC, Roggeva E, Vandenberghe R *et al*. Both common variations and rare non-synonymous substitutions and small insertion/deletions in CLU are associated with increased Alzheimer risk. *Mol Neurodegener* 2012; **7**: 3.
- 84 Liu CC, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol* 2013; **9**: 106–118.
- 85 Koldamova R, Fitz NF, Lefterov I. The role of ATP-binding cassette transporter A1 in Alzheimer's disease and neurodegeneration. *Biochim Biophys Acta* 2010; **1801**: 824–830.
- 86 Deane R, Bell RD, Sagare A, Zlokovic BV. Clearance of amyloid-beta peptide across the blood-brain barrier: implication for therapies in Alzheimer's disease. *CNS Neurol Disord Drug Targets* 2009; **8**: 16–30.
- 87 Bell RD, Sagare AP, Friedman AE, Bedi GS, Holtzman DM, Deane R *et al*. Transport pathways for clearance of human Alzheimer's amyloid beta-peptide and apolipoproteins E and J in the mouse central nervous system. *J Cereb Blood Flow Metab* 2007; **27**: 909–918.
- 88 Kanekiyo T, Cirrito JR, Liu CC, Shinohara M, Li J, Schuler DR *et al*. Neuronal clearance of amyloid-beta by endocytic receptor LRP1. *J Neurosci* 2013; **33**: 19276–19283.
- 89 Iwata N, Tsubuki S, Takaki Y, Shirohata K, Lu B, Gerard NP *et al*. Metabolic regulation of brain Abeta by neprilysin. *Science* 2001; **292**: 1550–1552.
- 90 Qiu WQ, Walsh DM, Ye Z, Vekrellis K, Zhang J, Podlisny MB *et al*. Insulin-degrading enzyme regulates extracellular levels of amyloid beta-protein by degradation. *J Biol Chem* 1998; **273**: 32730–32738.
- 91 Liao MC, Van Nostrand WE. Degradation of soluble and fibrillar amyloid beta-protein by matrix metalloproteinase (MT1-MMP) *in vitro*. *Biochemistry* 2010; **49**: 1127–1136.

- 92 Backstrom JR, Lim GP, Cullen MJ, Tokes ZA. Matrix metalloproteinase-9 (MMP-9) is synthesized in neurons of the human hippocampus and is capable of degrading the amyloid-beta peptide (1-40). *J Neurosci* 1996; **16**: 7910-7919.
- 93 Zou K, Yamaguchi H, Akatsu H, Sakamoto T, Ko M, Mizoguchi K et al. Angiotensin-converting enzyme converts amyloid beta-protein 1-42 (Abeta(1-42)) to Abeta(1-40), and its inhibition enhances brain Abeta deposition. *J Neurosci* 2007; **27**: 8628-8635.
- 94 Pacheco-Quinto J, Eckman EA. Endothelin-converting enzymes degrade intracellular beta-amyloid produced within the endosomal/lysosomal pathway and autophagosomes. *J Biol Chem* 2013; **288**: 5606-5615.
- 95 Kim MJ, Chae SS, Koh YH, Lee SK, Jo SA. Glutamate carboxypeptidase II: an amyloid peptide-degrading enzyme with physiological function in the brain. *FASEB J* 2010; **24**: 4491-4502.
- 96 Abdul-Hay SO, Sahara T, McBride M, Kang D, Leissring MA. Identification of BACE2 as an avid ss-amyloid-degrading protease. *Mol Neurodegener* 2012; **7**: 46.
- 97 Mueller-Steiner S, Zhou Y, Arai H, Roberson ED, Sun B, Chen J et al. Anti-amyloidogenic and neuroprotective functions of cathepsin B: implications for Alzheimer's disease. *Neuron* 2006; **51**: 703-714.
- 98 Lee CY, Landreth GE. The role of microglia in amyloid clearance from the AD brain. *J Neural Transm* 2010; **117**: 949-960.
- 99 Salminen A, Ojala J, Kauppinen A, Kaariranta K, Suuronen T. Inflammation in Alzheimer's disease: amyloid-beta oligomers trigger innate immunity defence via pattern recognition receptors. *Prog Neurobiol* 2009; **87**: 181-194.
- 100 Heneka MT, Golenbock DT, Latz E. Innate immunity in Alzheimer's disease. *Nat Immunol* 2015; **16**: 229-236.
- 101 Maier M, Peng Y, Jiang L, Seabrook TJ, Carroll MC, Lemere CA. Complement C3 deficiency leads to accelerated amyloid beta plaque deposition and neurodegeneration and modulation of the microglia/macrophage phenotype in amyloid precursor protein transgenic mice. *J Neurosci* 2008; **28**: 6333-6341.
- 102 Crehan H, Holton P, Wray S, Pocock J, Guerreiro R, Hardy J. Complement receptor 1 (CR1) and Alzheimer's disease. *Immunobiology* 2012; **217**: 244-250.
- 103 Heneka MT, Carson MJ, El Khoury J, Landreth GE, Brosseron F, Feinstein DL et al. Neuroinflammation in Alzheimer's disease. *Lancet Neurol* 2015; **14**: 388-405.
- 104 Wang Y, Cella M, Mallinson K, Ulrich JD, Young KL, Robinette ML et al. TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. *Cell* 2015; **160**: 1061-1071.
- 105 Cruts M, Gijssels I, van der Zee J, Engelborghs S, Wils H, Pirici D et al. Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. *Nature* 2006; **442**: 920-924.
- 106 Baker M, Mackenzie IR, Pickering-Brown SM, Gass J, Rademakers R, Lindholm C et al. Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature* 2006; **442**: 916-919.
- 107 Perry DC, Lehmann M, Yokoyama JS, Karydas A, Lee JJ, Coppola G et al. Progranulin mutations as risk factors for Alzheimer disease. *JAMA Neurol* 2013; **70**: 774-778.
- 108 Minami SS, Min SW, Krabbe G, Wang C, Zhou Y, Asgarov R et al. Progranulin protects against amyloid beta deposition and toxicity in Alzheimer's disease mouse models. *Nat Med* 2014; **20**: 1157-1164.
- 109 Steinberg S, Stefansson H, Jonsson T, Johannsdottir H, Ingason A, Helgason H et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nat Genet* 2015; **47**: 445-447.
- 110 Zhao QF, Yu JT, Tan MS, Tan L. ABCA7 in Alzheimer's disease. *Mol Neurobiol* 2015; **51**: 1008-1016.
- 111 Satoh K, Abe-Dohmae S, Yokoyama S, George-Hyslop P, Fraser PE. ABCA7 loss of function alters Alzheimer amyloid processing. *J Biol Chem* 2015; **290**: 24152-24165.
- 112 Gregori L, Fuchs C, Figueiredo-Pereira ME, Van Nostrand WE, Goldgaber D. Amyloid beta-protein inhibits ubiquitin-dependent protein degradation *in vitro*. *J Biol Chem* 1995; **270**: 19702-19708.
- 113 Burns MP, Zhang L, Rebeck GW, Querfurth HW, Moussa CE. Parkin promotes intracellular Abeta1-42 clearance. *Hum Mol Genet* 2009; **18**: 3206-3216.
- 114 Nixon RA, Wegiel J, Kumar A, Yu WH, Peterhoff C, Cataldo A et al. Extensive involvement of autophagy in Alzheimer disease: an immuno-electron microscopy study. *J Neuropathol Exp Neurol* 2005; **64**: 113-122.
- 115 Pickford F, Masliah E, Britschgi M, Lucin K, Narasimhan R, Jaeger PA et al. The autophagy-related protein beclin 1 shows reduced expression in early Alzheimer disease and regulates amyloid beta accumulation in mice. *J Clin Invest* 2008; **118**: 2190-2199.
- 116 Benilova I, Karran E, De Strooper B. The toxic Abeta oligomer and Alzheimer's disease: an emperor in need of clothes. *Nat Neurosci* 2012; **15**: 349-357.
- 117 Hartley DM, Zhao C, Speier AC, Woodard GA, Li S, Li Z et al. Transglutaminase induces protofibril-like amyloid beta-protein assemblies that are protease-resistant and inhibit long-term potentiation. *J Biol Chem* 2008; **283**: 16790-16800.
- 118 Schlenzig D, Ronicke R, Cynis H, Ludwig HH, Scheel E, Reymann K et al. N-terminal pyroglutamate formation of Abeta38 and Abeta40 enforces oligomer formation and potency to disrupt hippocampal long-term potentiation. *J Neurochem* 2012; **121**: 774-784.
- 119 Kumar S, Rezaei-Ghaleh N, Terwel D, Thal DR, Richard M, Hoch M et al. Extracellular phosphorylation of the amyloid beta-peptide promotes formation of toxic aggregates during the pathogenesis of Alzheimer's disease. *EMBO J* 2011; **30**: 2255-2265.
- 120 Lee JY, Cole TB, Palmiter RD, Suh SW, Koh JY. Contribution by synaptic zinc to the gender-disparate plaque formation in human Swedish mutant APP transgenic mice. *Proc Natl Acad Sci USA* 2002; **99**: 7705-7710.
- 121 Evans CG, Wisen S, Gestwicki JE. Heat shock proteins 70 and 90 inhibit early stages of amyloid beta(1-42) aggregation *in vitro*. *J Biol Chem* 2006; **281**: 33182-33191.
- 122 Ojha J, Masilamani G, Dunlap D, Udoff RA, Cashikar AG. Sequestration of toxic oligomers by HspB1 as a cytoprotective mechanism. *Mol Cell Biol* 2011; **31**: 3146-3157.
- 123 Wilhelmus MM, Boelens WC, Otte-Holler I, Kamps B, de Waal RM, Verbeek MM. Small heat shock proteins inhibit amyloid-beta protein aggregation and cerebrovascular amyloid-beta protein toxicity. *Brain Res* 2006; **1089**: 67-78.
- 124 Inestrosa NC, Alvarez A, Perez CA, Moreno RD, Vicente M, Linker C et al. Acetylcholinesterase accelerates assembly of amyloid-beta-peptides into Alzheimer's fibrils: possible role of the peripheral site of the enzyme. *Neuron* 1996; **16**: 881-891.
- 125 Holtzman DM, Bales KR, Tenkova T, Fagan AM, Parsadanian M, Sartorius LJ et al. Apolipoprotein E isoform-dependent amyloid deposition and neuritic degeneration in a mouse model of Alzheimer's disease. *Proc Natl Acad Sci USA* 2000; **97**: 2892-2897.
- 126 Cotman SL, Halfter W, Cole GJ. Agrin binds to beta-amyloid (Abeta), accelerates abeta fibril formation, and is localized to Abeta deposits in Alzheimer's disease brain. *Mol Cell Neurosci* 2000; **15**: 183-198.
- 127 Castillo GM, Ngo C, Cummings J, Wight TN, Snow AD. Perlecan binds to the beta-amyloid proteins (A beta) of Alzheimer's disease, accelerates A beta fibril formation, and maintains A beta fibril stability. *J Neurochem* 1997; **69**: 2452-2465.
- 128 Diamant S, Podoly E, Friedler A, Ligumsky H, Livnah O, Soreq H. Butyrylcholinesterase attenuates amyloid fibril formation *in vitro*. *Proc Natl Acad Sci USA* 2006; **103**: 8628-8633.
- 129 Li X, Zhang X, Ladiwala AR, Du D, Yadav JK, Tessier PM et al. Mechanisms of transthyretin inhibition of beta-amyloid aggregation *in vitro*. *J Neurosci* 2013; **33**: 19423-19433.
- 130 Nuutinen T, Suuronen T, Kauppinen A, Salminen A. Clusterin: a forgotten player in Alzheimer's disease. *Brain Res Rev* 2009; **61**: 89-104.
- 131 Hughes SR, Khorkova O, Goyal S, Knaeblein J, Heroux J, Riedel NG et al. Alpha2-macroglobulin associates with beta-amyloid peptide and prevents fibril formation. *Proc Natl Acad Sci USA* 1998; **95**: 3275-3280.
- 132 Kaeser SA, Herzig MC, Coomaraswamy J, Kilger E, Selenica ML, Winkler DT et al. Cystatin C modulates cerebral beta-amyloidosis. *Nat Genet* 2007; **39**: 1437-1439.
- 133 Mi W, Pawlik M, Sastre M, Jung SS, Radvinsky DS, Klein AM et al. Cystatin C inhibits amyloid-beta deposition in Alzheimer's disease mouse models. *Nat Genet* 2007; **39**: 1440-1442.
- 134 Eriksson S, Janciauskiene S, Lannfelt L. Alpha 1-antichymotrypsin regulates Alzheimer beta-amyloid peptide fibril formation. *Proc Natl Acad Sci USA* 1995; **92**: 2313-2317.
- 135 Kocherhan S, Madhusudan A, Doehner J, Breu KS, Nitsch RM, Fritschy JM et al. Reduced reelin expression accelerates amyloid-beta plaque formation and tau pathology in transgenic Alzheimer's disease mice. *J Neurosci* 2010; **30**: 9228-9240.
- 136 Jucker M, Walker LC. Pathogenic protein seeding in Alzheimer disease and other neurodegenerative disorders. *Ann Neurol* 2011; **70**: 532-540.
- 137 Masliah E, Rockenstein E, Veinbergs I, Sagara Y, Mallory M, Hashimoto M et al. Beta-amyloid peptides enhance alpha-synuclein accumulation and neuronal deficits in a transgenic mouse model linking Alzheimer's disease and Parkinson's disease. *Proc Natl Acad Sci USA* 2001; **98**: 12245-12250.
- 138 Ono K, Takahashi R, Ikeda T, Yamada M. Cross-seeding effects of amyloid beta-protein and alpha-synuclein. *J Neurochem* 2012; **122**: 883-890.
- 139 Fang YS, Tsai KJ, Chang YJ, Kao P, Woods R, Kuo PH et al. Full-length TDP-43 forms toxic amyloid oligomers that are present in frontotemporal lobar dementia-TDP patients. *Nat Commun* 2014; **5**: 4824.
- 140 Guerrero-Munoz MJ, Castillo-Carranza DL, Krishnamurthy S, Paulucci-Holthausen AA, Sengupta U, Lasagna-Reeves CA et al. Amyloid-beta oligomers as a template for secondary amyloidosis in Alzheimer's disease. *Neurobiol Dis* 2014; **71**: 14-23.
- 141 Josephs KA, Whitwell JL, Weigand SD, Murray ME, Tosakulwong N, Liesinger AM et al. TDP-43 is a key player in the clinical features associated with Alzheimer's disease. *Acta Neuropathol* 2014; **127**: 811-824.

- 142 Hamilton RL. Lewy bodies in Alzheimer's disease: a neuropathological review of 145 cases using alpha-synuclein immunohistochemistry. *Brain Pathol* 2000; **10**: 378–384.
- 143 Rapoport M, Dawson HN, Binder LI, Vitke MP, Ferreira A. Tau is essential to beta-amyloid-induced neurotoxicity. *Proc Natl Acad Sci USA* 2002; **99**: 6364–6369.
- 144 Roberson ED, Scarce-Levie K, Palop JJ, Yan F, Cheng IH, Wu T *et al*. Reducing endogenous tau ameliorates amyloid beta-induced deficits in an Alzheimer's disease mouse model. *Science* 2007; **316**: 750–754.
- 145 Ittner LM, Ke YD, Delerue F, Bi M, Gladbach A, van Eersel J *et al*. Dendritic function of tau mediates amyloid-beta toxicity in Alzheimer's disease mouse models. *Cell* 2010; **142**: 387–397.
- 146 Salter MW, Kalia LV. Src kinases: a hub for NMDA receptor regulation. *Nat Rev Neurosci* 2004; **5**: 317–328.
- 147 Nisbet RM, Polanco JC, Ittner LM, Gotz J. Tau aggregation and its interplay with amyloid-beta. *Acta Neuropathol* 2015; **129**: 207–220.
- 148 Tu S, Okamoto S, Lipton SA, Xu H. Oligomeric Abeta-induced synaptic dysfunction in Alzheimer's disease. *Mol Neurodegener* 2014; **9**: 48.
- 149 Spire-Jones TL, Friedman T, Pitstick R, Polydoro M, Roe A, Carlson GA *et al*. Methylene blue does not reverse existing neurofibrillary tangle pathology in the rTg4510 mouse model of tauopathy. *Neurosci Lett* 2014; **562**: 63–68.
- 150 Spillantini MG, Goedert M. Tau pathology and neurodegeneration. *Lancet Neurol* 2013; **12**: 609–622.
- 151 Zempel H, Mandelkow E. Lost after translation: missorting of tau protein and consequences for Alzheimer disease. *Trends Neurosci* 2014; **37**: 721–732.
- 152 Frandemich ML, De Seranno S, Rush T, Borel E, Elie A, Arnal I *et al*. Activity-dependent tau protein translocation to excitatory synapse is disrupted by exposure to amyloid-beta oligomers. *J Neurosci* 2014; **34**: 6084–6097.
- 153 Noble W, Hanger DP, Miller CC, Lovestone S. The importance of tau phosphorylation for neurodegenerative diseases. *Front Neurol* 2013; **4**: 83.
- 154 Iqbal K, Gong CX, Liu F. Hyperphosphorylation-induced tau oligomers. *Front Neurol* 2013; **4**: 112.
- 155 Kins S, Cramer A, Evans DR, Hemmings BA, Nitsch RM, Gotz J. Reduced protein phosphatase 2A activity induces hyperphosphorylation and altered compartmentalization of tau in transgenic mice. *J Biol Chem* 2001; **276**: 38193–38200.
- 156 Yu W, Polepalli J, Wagh D, Rajadas J, Malenka R, Lu B. A critical role for the PAR-1/MARK-tau axis in mediating the toxic effects of Abeta on synapses and dendritic spines. *Hum Mol Genet* 2012; **21**: 1384–1390.
- 157 Mairet-Coello G, Courchet J, Pieraut S, Courchet V, Maximov A, Polleux F. The CAMKK2–AMPK kinase pathway mediates the synaptotoxic effects of Abeta oligomers through Tau phosphorylation. *Neuron* 2013; **78**: 94–108.
- 158 Lee MS, Kwon YT, Li M, Peng J, Friedlander RM, Tsai LH. Neurotoxicity induces cleavage of p35 to p25 by calpain. *Nature* 2000; **405**: 360–364.
- 159 Hartigan JA, Johnson GV. Transient increases in intracellular calcium result in prolonged site-selective increases in Tau phosphorylation through a glycogen synthase kinase 3beta-dependent pathway. *J Biol Chem* 1999; **274**: 21395–21401.
- 160 Kimura R, Kamino K, Yamamoto M, Nuripa A, Kida T, Kazui H *et al*. The DYRK1A gene, encoded in chromosome 21 Down syndrome critical region, bridges between beta-amyloid production and tau phosphorylation in Alzheimer disease. *Hum Mol Genet* 2007; **16**: 15–23.
- 161 Del Prete D, Checler F, Chami M. Ryanodine receptors: physiological function and deregulation in Alzheimer disease. *Mol Neurodegen* 2014; **9**: 21.
- 162 Ferreira E, Oliveira CR, Pereira C. Involvement of endoplasmic reticulum Ca<sup>2+</sup> release through ryanodine and inositol 1,4,5-triphosphate receptors in the neurotoxic effects induced by the amyloid-beta peptide. *J Neurosci Res* 2004; **76**: 872–880.
- 163 Dreses-Werringloer U, Lambert JC, Vingtdoux V, Zhao H, Vais H, Siebert A *et al*. A polymorphism in CALHM1 influences Ca<sup>2+</sup> homeostasis, Abeta levels, and Alzheimer's disease risk. *Cell* 2008; **133**: 1149–1161.
- 164 Gamblin TC, Chen F, Zambrano A, Abraha A, Lagalwar S, Guillozet AL *et al*. Caspase cleavage of tau: linking amyloid and neurofibrillary tangles in Alzheimer's disease. *Proc Natl Acad Sci USA* 2003; **100**: 10032–10037.
- 165 Guo JL, Lee VM. Seeding of normal tau by pathological tau conformers drives pathogenesis of Alzheimer-like tangles. *J Biol Chem* 2011; **286**: 15317–15331.
- 166 Guo JP, Arai T, Miklossy J, McGeer PL. Abeta and tau form soluble complexes that may promote self aggregation of both into the insoluble forms observed in Alzheimer's disease. *Proc Natl Acad Sci USA* 2006; **103**: 1953–1958.
- 167 Hsieh H, Boehm J, Sato C, Iwatsubo T, Tomita T, Sisodia S *et al*. AMPAR removal underlies Abeta-induced synaptic depression and dendritic spine loss. *Neuron* 2006; **52**: 831–843.
- 168 Wang HY, Lee DH, D'Andrea MR, Peterson PA, Shank RP, Reitz AB. Beta-amyloid (1–42) binds to alpha7 nicotinic acetylcholine receptor with high affinity. Implications for Alzheimer's disease pathology. *J Biol Chem* 2000; **275**: 5626–5632.
- 169 Larson M, Sherman MA, Amar F, Nuvolone M, Schneider JA, Bennett DA *et al*. The complex PrP(C)-Fyn couples human oligomeric Abeta with pathological tau changes in Alzheimer's disease. *J Neurosci* 2012; **32**: 16857–71a.
- 170 Um JW, Nygaard HB, Heiss JK, Kostylev MA, Stagi M, Vortmeyer A *et al*. Alzheimer amyloid-beta oligomer bound to postsynaptic prion protein activates Fyn to impair neurons. *Nat Neurosci* 2012; **15**: 1227–1235.
- 171 Kim T, Vidal GS, Djurisic M, William CM, Birnbaum ME, Garcia KC *et al*. Human LirB2 is a beta-amyloid receptor and its murine homolog PirB regulates synaptic plasticity in an Alzheimer's model. *Science* 2013; **341**: 1399–1404.
- 172 Knowles JK, Rajadas J, Nguyen TV, Yang T, LeMieux MC, Vander Griend L *et al*. The p75 neurotrophin receptor promotes amyloid-beta(1–42)-induced neuritic dystrophy *in vitro* and *in vivo*. *J Neurosci* 2009; **29**: 10627–10637.
- 173 Cisse M, Checler F. Eph receptors: new players in Alzheimer's disease pathogenesis. *Neurobiol Dis* 2015; **73**: 137–149.
- 174 Um JW, Kaufman AC, Kostylev M, Heiss JK, Stagi M, Takahashi H *et al*. Metabotropic glutamate receptor 5 is a coreceptor for Alzheimer amyloid beta oligomer bound to cellular prion protein. *Neuron* 2013; **79**: 887–902.
- 175 Roselli F, Tirard M, Lu J, Hutzler P, Lamberti P, Livrea P *et al*. Soluble beta-amyloid1–40 induces NMDA-dependent degradation of postsynaptic density-95 at glutamatergic synapses. *J Neurosci* 2005; **25**: 11061–11070.
- 176 Shankar GM, Bloodgood BL, Townsend M, Walsh DM, Selkoe DJ, Sabatini BL. Natural oligomers of the Alzheimer amyloid-beta protein induce reversible synapse loss by modulating an NMDA-type glutamate receptor-dependent signaling pathway. *J Neurosci* 2007; **27**: 2866–2875.
- 177 Lustbader JW, Cirilli M, Lin C, Xu HW, Takuma K, Wang N *et al*. ABAD directly links Abeta to mitochondrial toxicity in Alzheimer's disease. *Science* 2004; **304**: 448–452.
- 178 Manczak M, Calkins MJ, Reddy PH. Impaired mitochondrial dynamics and abnormal interaction of amyloid beta with mitochondrial protein Drp1 in neurons from patients with Alzheimer's disease: implications for neuronal damage. *Hum Mol Genet* 2011; **20**: 2495–2509.
- 179 Herrup K. The case for rejecting the amyloid cascade hypothesis. *Nat Neurosci* 2015; **18**: 794–799.

Supplementary Information accompanies the paper on the *Molecular Psychiatry* website (<http://www.nature.com/mp>)

### 3. Recherche de CNVs rare détectés à partir de données de séquençage d'exomes

Après l'étude des SNVs et des indels, nous avons focalisé nos analyses sur l'implication des CNVs dans la maladie d'Alzheimer du sujet jeune.

#### Un enrichissement à l'échelle d'une liste de gènes

Dans cette étude, nous avons voulu réaliser une étude d'association à partir de CNVs détectés à partir de données de séquençage d'exomes. Cette nouvelle technologie a notamment pu être développée grâce à l'amélioration et l'uniformisation des séquençages d'exomes. De la même manière que pour les études d'association sur les variants rares, nous avons sélectionné 546 individus atteints de formes précoces de MA et 597 contrôles génétiquement appariés provenant du *French Exome Project*.

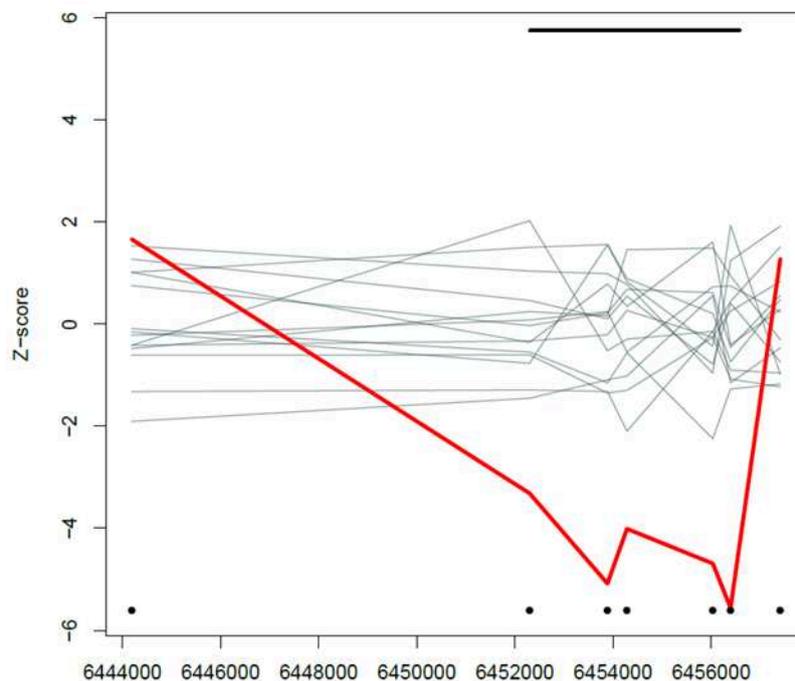
Afin de détecter les CNVs à partir de données d'exomes, nous avons utilisé le logiciel CANOES (Backenroth et al., 2014). Ce logiciel base son analyse sur les profondeurs de séquençage en créant un modèle statistique de Markov caché à partir de 30 individus séquencés dans les mêmes conditions (même kit de capture, centre et série de séquençage) que l'individu en cours d'analyse (Figure 26 page 83).

Afin d'évaluer la sensibilité du logiciel, nous avons réalisé une étude comparative grâce à des individus qui avaient fait l'objet d'un séquençage d'exome et d'une CGH (1 million de sondes : 1M). Nous avons réalisé cette comparaison en ne considérant que les CNVs de tailles supérieures à 8 kb, limite de résolution de la CGH 1M. Les CNVs situés sur des zones de duplication segmentaire ont également été retirés de l'analyse car elles ne sont pas ou peu couvertes par la technique de CGH. Les résultats nous ont permis de déterminer que CANOES identifiait 88 % (59/67 CNVs) des CNVs rares initialement détectés grâce à la CGH. Les CNVs non-détectés par CANOES sont principalement le reflet de la faible proportion de séquences codantes dans les remaniements détectés par CGH. Afin de déterminer la valeur prédictive positive, sur 93 CNVs détectés 85 (91 %) ont pu être validés par une technique de PCR multiplexe semi-quantitative (QMPSF). Le nombre plus important de CNVs détectés grâce aux données d'exomes par rapport à la CGH (6,43 vs 1,60) n'était donc pas dû à un excès de faux-positifs.

Un premier filtre basé sur la qualité a été appliquée sur les individus porteurs d'un nombre aberrant de CNVs (>50) reflétant une mauvaise qualité de séquençage d'exome. Notre analyse d'association incluait finalement 522 cas et 584 contrôles. Dans un second temps, nous avons ensuite restreint notre analyse aux CNVs rares (fréquence <1 % dans notre population) et touchant les gènes du

réseau A $\beta$  (Résultats Partie 2). Nous avons également exclus les gènes situés sur des zones de duplication segmentaire. Enfin, les CNVs restants ont été validés par QMPFSF.

Grâce à cette stratégie, nous avons pu obtenir un enrichissement nominalelement significatif de CNVs rares intersectant cette liste de gènes chez nos cas comparés à nos contrôles (OR=1,92, p=0,038). Cependant, ce résultat pourrait être amélioré si l'on pouvait prédire précisément l'effet des CNVs détectés. En effet, qu'il s'agisse de duplications ou délétions, partielles ou complètes les conséquences fonctionnelles sur la protéine seront différentes. De plus, il faut pouvoir déterminer d'un point de vue physiopathologique l'implication de ces variations en fonction du gène étudié.



**Figure 26 : Exemple de la détection d'une délétion grâce à CANOES**

Modélisation de la déviation du Z-score établie à partir des différences de la profondeur de séquençage entre les individus. Dans ce cas, le Z-score est très négatif indiquant la présence d'une délétion. La position de la délétion est indiquée en abscisse.

### La duplication du locus 17q21.31 dans la maladie d'Alzheimer

La principale découverte de cette étude est la détection chez quatre cas non-apparentés d'une duplication du locus 17q21.31. Cette duplication emporte 4 gènes : *CRHR1* (*Corticotropin Releasing Hormone Receptor 1*), *MAPT*, *STH* (*Saitohin*) et *KANSL1* (*KAT8 Regulatory NSL Complex Subunit 1*). Ce locus définit un bloc haplotypique de 900 kb correspondant à deux haplotypes principaux nommés H1 et H2.

L'association réalisée avec les contrôles de notre étude ne nous permet de détecter qu'une association nominalelement significative (p=0,049). Lorsque nous avons utilisé les données de l'étude de Coe et collaborateurs (Coe et al., 2014) qui regroupent 19 584 contrôles ayant fait l'objet d'une étude par CGH et la sous-population Européenne non-Finlandaise d'ExAC (Rudfer et al., 2016) qui

regroupent 32 850 individus indemnes de maladies psychiatriques sévères, nous obtenions des résultats très significatifs : 4/522 vs 3/19 854,  $p=1,5.10^{-5}$  ou 4/522 vs 1/32 850,  $p=2,9.10^{-7}$ .

Grâce à la technique de SNaPshot, nous avons pu déterminer que deux patients étaient porteurs d'une duplication de l'haplotype H1 et deux patients d'une duplication de l'haplotype H2. Nous avons pu déterminer que cette duplication entraînait une augmentation d'expression des transcrits de *MAPT* : 1,64 fois supérieure chez les patients porteurs de duplication H1 et 1,87 fois chez les porteurs de duplication H2.

Parmi les cas porteurs de la duplication du locus 17q21.31, deux individus étaient des cas familiaux. L'étude de la ségrégation de cette duplication au sein d'une famille pour laquelle nous disposions de l'ADN des apparentés a révélé que la duplication était également présente chez un apparenté malade et absente chez les apparentés sains. Parmi les deux cas sporadiques, nous disposions pour l'un d'entre eux de l'ADN des parents, qui ne présentaient pas de troubles cognitifs, et avons pu établir que la duplication était survenue *de novo* chez cet individu.

La présentation clinique de ces patients est typique d'une MA avec un déclin des fonctions cognitives associé à des troubles de la mémoire. Les analyses d'imagerie pratiquées ont également permis de mettre en évidence une atrophie hippocampique ainsi qu'un hypométabolisme des régions temporales et pariétales du cerveau de ces patients, en faveur d'un diagnostic de MA. L'examen des biomarqueurs du LCR appuyait également ce diagnostic de MA. Ce dosage permet de mesurer les niveaux d'Aβ40 et Aβ42 notamment en supposant que leur agrégation dans le cerveau diminue sa présence dans le LCR. Cependant, aucune plaque sénile ni aucun dépôt amyloïde n'ont pu être détectés par l'imagerie amyloïde pratiquée chez 3 de ces patients. Cette discordance entre les différentes analyses biologiques ne pouvait être expliquée que par l'analyse anatomo-pathologique. Nous ne disposions pas d'analyses anatomopathologiques pour les patients porteurs de la duplication. Cependant, nous avons pu identifier la duplication du locus 17q21.31 chez un patient atteint d'une démence précoce qui avait fait l'objet d'une étude anatomopathologique complète (Alexander et al., 2016). De manière intéressante, un apparenté malade était également porteur de cette duplication tandis qu'un autre apparenté sain ne l'était pas. L'analyse anatomopathologique avait permis de révéler que le patient était atteint d'une tauopathie pure avec des DNFs similaires à celles retrouvées dans la MA, composées des isoformes 3R et 4R de Tau, localisées au niveau de l'hippocampe, du noyau basal de Meynert et du cortex entorhinal mais sans aucun dépôts amyloïdes. En conclusion, cette étude a permis de caractériser une nouvelle pathologie causée par la duplication du locus 17q21.31 et pourrait expliquer la pathologie de certains individus inclus dans les populations de malades Alzheimer pour lesquels l'analyse anatomopathologique réalisée *post-mortem* n'a révélé aucun dépôt amyloïde.

ORIGINAL ARTICLE

# 17q21.31 duplication causes prominent tau-related dementia with increased *MAPT* expression

K Le Guennec<sup>1,2,29</sup>, O Quenez<sup>1,2,3,29</sup>, G Nicolas<sup>1,2,3,4</sup>, D Wallon<sup>1,2,3,5</sup>, S Rousseau<sup>1,2,3</sup>, A-C Richard<sup>1,2,3</sup>, J Alexander<sup>6</sup>, P Paschou<sup>6</sup>, C Charbonnier<sup>1,2,3</sup>, C Bellenguez<sup>7,8,9</sup>, B Grenier-Boley<sup>7,8,9</sup>, D Lechner<sup>10</sup>, M-T Bihoreau<sup>10</sup>, R Olaso<sup>10</sup>, A Boland<sup>10</sup>, V Meyer<sup>10</sup>, J-F Deleuze<sup>10,11</sup>, P Amouyel<sup>7,8,9</sup>, HM Munter<sup>12</sup>, G Bourque<sup>12</sup>, M Lathrop<sup>12</sup>, T Frebourg<sup>1,2,4</sup>, R Redon<sup>13,14</sup>, L Letenneur<sup>15,16</sup>, J-F Dartigues<sup>15,16</sup>, O Martinaud<sup>3,5</sup>, O Kalev<sup>17</sup>, S Mehrabian<sup>18</sup>, L Traykov<sup>18</sup>, T Ströbel<sup>19</sup>, I Le Ber<sup>20,21</sup>, P Caroppo<sup>20,21</sup>, S Epelbaum<sup>20,21</sup>, T Jonveaux<sup>22,23</sup>, F Pasquier<sup>24</sup>, A Rollin-Sillaire<sup>24</sup>, E Génin<sup>25</sup>, L Guyant-Maréchal<sup>5,26</sup>, GG Kovacs<sup>19</sup>, J-C Lambert<sup>7,8,9</sup>, D Hannequin<sup>1,2,3,4,5</sup>, D Champion<sup>1,2,3,27</sup>, A Rovelet-Lecrux<sup>1,2,3</sup> and The CNR-MAJ collaborators<sup>28</sup>

To assess the role of rare copy number variations in Alzheimer's disease (AD), we conducted a case–control study using whole-exome sequencing data from 522 early-onset cases and 584 controls. The most recurrent rearrangement was a 17q21.31 microduplication, overlapping the *CRHR1*, *MAPT*, *STH* and *KANSL1* genes that was found in four cases, including one *de novo* rearrangement, and was absent in controls. The increased *MAPT* gene dosage led to a 1.6–1.9-fold expression of the *MAPT* messenger RNA. Clinical signs, neuroimaging and cerebrospinal fluid biomarker profiles were consistent with an AD diagnosis in *MAPT* duplication carriers. However, amyloid positron emission tomography (PET) imaging, performed in three patients, was negative. Analysis of an additional case with neuropathological examination confirmed that the *MAPT* duplication causes a complex tauopathy, including prominent neurofibrillary tangle pathology in the medial temporal lobe without amyloid- $\beta$  deposits. 17q21.31 duplication is the genetic basis of a novel entity marked by prominent tauopathy, leading to early-onset dementia with an AD clinical phenotype. This entity could account for a proportion of probable AD cases with negative amyloid PET imaging recently identified in large clinical series.

*Molecular Psychiatry* advance online publication, 13 December 2016; doi:10.1038/mp.2016.226

## INTRODUCTION

Alzheimer's disease (AD) is the most common neurodegenerative dementia worldwide. Although different clinical presentations can be observed, the typical one is marked by the impairment of episodic memory related to the alteration of the hippocampal cortex, followed by the impairment of additional cognitive functions. AD is neuropathologically defined by the co-occurrence of two types of lesions: the amyloid plaques consisting of amyloid- $\beta$  (A $\beta$ ) peptides and the neurofibrillary tangles composed of hyperphosphorylated, aggregated Tau proteins. The definite diagnosis of AD requires a post mortem examination of the brain. The diagnosis of probable AD is based on clinical arguments. Different levels of evidence of an AD pathophysiological process can be provided *in vivo* by structural and metabolic neuroimaging and cerebrospinal fluid (CSF) AD biomarkers. The latter include A $\beta$ <sub>42</sub> (decreased in AD), total Tau (t-Tau increased)

and phosphorylated Tau (P-Tau, increased).<sup>1</sup> Moreover, amyloid imaging by positron emission tomography (PET) now allows identifying the presence of amyloid plaques. Importantly, using these tools, no evidence of amyloid pathology has been reported in ~12% of patients fulfilling the criteria of probable AD included in clinical series or trials.<sup>2,3</sup> A proportion of them was referred to as suspected non-AD pathophysiology (SNAP),<sup>3,4</sup> opening the question of the underlying pathology.

Besides causative mutations in *APP* (amyloid precursor protein), *PSEN1* (presenilin1), or *PSEN2* (presenilin2) and the risk conferred by rare variants in *TREM2*, *SORL1* and *ABCA7* genes,<sup>5–10</sup> rare copy number variations (CNVs) are also known to contribute to the genetic determinism of AD. First, *APP* duplications or *PSEN1* exon 9 deletions are part of the mutation spectrum causing autosomal dominant early-onset AD (EOAD, onset before 65 years).<sup>11–13</sup> Second, private CNVs have been described in extreme forms of

<sup>1</sup>Inserm, U1079, faculté de médecine, Rouen University, IRIB, Normandy University, Rouen, France; <sup>2</sup>Normandy Centre for Genomic Medicine and Personalized Medicine, Rouen, France; <sup>3</sup>CNR-MAJ, Rouen University Hospital, Rouen, France; <sup>4</sup>Department of Genetics, Rouen University Hospital, Rouen, France; <sup>5</sup>Department of Neurology, Rouen University Hospital, Rouen, France; <sup>6</sup>Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupoli, Greece; <sup>7</sup>Inserm, U1167, Lille, France; <sup>8</sup>Institut Pasteur de Lille, Lille, France; <sup>9</sup>Université Lille-Nord de France, Lille, France; <sup>10</sup>Centre National de Génotypage, Institut de Génétique, CEA, Evry, France; <sup>11</sup>Fondation Jean Dausset, Centre d'études du Polymorphisme Humain, Paris, France; <sup>12</sup>McGill University and Génome Québec Innovation Centre, Montréal, QC, Canada; <sup>13</sup>Inserm, UMR 1087, l'Institut du thorax, CHU Nantes, Nantes, France; <sup>14</sup>CNRS, UMR 6291, Université de Nantes, Nantes, France; <sup>15</sup>INSERM, U1219, Bordeaux, France; <sup>16</sup>Université de Bordeaux, Bordeaux, France; <sup>17</sup>Institute of Pathology and Neuropathology, Kepler University Hospital, Linz, Austria; <sup>18</sup>Department of Neurology, Alexandrovska University Hospital, Medical University-Sofia, Sofia, Bulgaria; <sup>19</sup>Institute of Neurology, Medical University Vienna, Vienna, Austria; <sup>20</sup>Sorbonne Universités, Inserm, CNRS, UPMC Univ Paris 06, UMR S 1127, Paris, France; <sup>21</sup>CNR-MAJ, IMMA, département des maladies du système nerveux, Hôpital Pitié-Salpêtrière, Paris, France; <sup>22</sup>Centre Mémoire de Ressources et de Recherche de Lorraine, CHRU Nancy Service de Gériatrie, Hôpital de Brabois, Vandœuvre les Nancy, France; <sup>23</sup>Laboratoire INTERPSY, EA 4432, Groupe de recherche sur les Communications (GRC), Université de Lorraine, Psychologie, Nancy, France; <sup>24</sup>CNR-MAJ Inserm U1171, Univ Lille, CHU, Lille, France; <sup>25</sup>Inserm, UMR1078, CHU Brest, Université Bretagne Occidentale, Brest, France; <sup>26</sup>Department of Neurophysiology, Rouen University Hospital, Rouen, France and <sup>27</sup>Department of Research, Rouvray Psychiatric Hospital, Sotteville-lès-Rouen, France. Correspondence: Dr A Rovelet-Lecrux or Dr D Champion, Inserm U1079, faculté de médecine, Rouen University, IRIB, Normandy University, 22 boulevard Gambetta, Rouen, 76183, France. E-mail: anne.roveletlecrux@univ-rouen.fr or dominique.champion@univ-rouen.fr

<sup>28</sup>CNR-MAJ collaborators are listed in Appendix A.

<sup>29</sup>These authors contributed equally to this work.

EOAD, some of them cosegregating with AD within pedigrees.<sup>14,15</sup> Third, a common CNV overlapping the genome-wide associated complement receptor 1 (CR1) gene has been shown to produce different CR1 isoforms; individuals carrying three copies of LCR1 (low-copy repeats 1) have an increased risk for AD compared with individuals carrying two copies.<sup>16</sup>

Here we report the results of a CNV analysis based on whole-exome sequencing (WES) data in a sample of 522 EOAD patients and 584 controls.

## MATERIALS AND METHODS

### Patients and controls

Patient ascertainment was described in detail by Nicolas *et al.*<sup>17</sup> Briefly, we included 546 unrelated EOAD patients (mean age of onset 55 years, range 44–65) from French ancestry. Patients were recruited by the French National CNR-MAJ consortium, a memory clinical network including 24 expert centers. Diagnoses of probable AD were performed according to the National Institute on Aging-Alzheimer's Association (NIA-AA) recommendations.<sup>1</sup> All patients underwent a comprehensive clinical examination, including personal medical and family history assessment, neurological examination, neuropsychological assessment, neuroimaging and 385 of them had CSF biomarkers indicative of an AD pathophysiological process, based either on three abnormal biomarkers (t-Tau, P-Tau and A $\beta$ <sub>42</sub>) or a t-Tau/A $\beta$ <sub>42</sub> ratio >0.52.<sup>17</sup> CSF samples were collected by lumbar puncture and stored in polypropylene tubes. CSF samples with >500 erythrocytes per  $\mu$ l were discarded. Samples were analyzed with the INNOTEST assays (Fujirebio, Malvern, PA, USA) according to the manufacturer protocols. No CSF biomarkers were available for the remaining patients.

Patients with *PSEN1*, *PSEN2* or *APP* mutation (detected by Sanger sequencing), *APP* duplication (detected by Quantitative Multiplex PCR of Short Fluorescent Fragments (QMPSF)<sup>13</sup>) and *C9ORF72* expansion (detected following the protocol described by DeJesus-Hernandez *et al.*<sup>18</sup>) were not included in this study. The number of *APOE*  $\epsilon$ 4-positive patients was  $n=288$  (53%).

A total of 597 controls of French ancestry were recruited in six different French cities. Among them, 396 were more than 55 years old at inclusion and had normal Mini Mental State Examination score, according to age and education level. The remaining controls were healthy blood donors. Overall, the mean age at inclusion ( $\pm$ s.d.) was  $66.4\pm 17.5$  years (range: 19–103). The number of *APOE*  $\epsilon$ 4-positive controls was  $n=139$  (23%).

All patients and controls gave informed, written consent for genetic analyses. This study was approved by the Ile de France II ethics committee.

### Whole-exome sequencing

WES was performed in three different centers: the Integragen Society (Evry, France), the Centre National de Génotypage (Evry, France) and the Génomique Québec Innovation Center (Montréal, Canada). Exomes were captured using Agilent SureSelect Human All Exon kits (V4+UTR, V5, V5 +UTR; Agilent technologies, Santa Clara, CA, USA). Final libraries were sequenced on a HiSeq2000 or 2500 (Illumina, San Diego, CA, USA) with paired ends, 76 or 100 bp reads. Exome samples were all processed through the same bioinformatics pipeline following GATK 3.3 (Broad Institute, Cambridge, MA, USA) Best Practices recommendations.<sup>19</sup> Reads were mapped to the 1000 Genomes GRCh37 build using BWA 0.7.5a.<sup>20</sup> Picard Tools 1.101 (<http://picard.sourceforge.net>) was used to flag duplicate reads. We applied GATK for short insertion and deletions (indel) realignment and base quality score recalibration.

### Quality checks

Individuals were analyzed through every recommended quality control steps for genetic case–control association studies.<sup>21</sup> Most checks were carried out with PLINK 1.9 (<https://www.cog-genomics.org/plink2>). All individuals in the sample were (i) checked for concordant sex information using Plink –sex-check,<sup>22</sup> (ii) discarded when identified with significantly higher heterozygosity rates (and contamination confirmed if relevant using the VerifyBamID software<sup>23</sup>) and (iii) discarded if Plink pi\_hat-relatedness estimation exceeded 18.5%. Out of 546 initially included cases and 597 controls, 535 cases and 593 controls passed these quality checks and were retained for CNV calling. No individual of divergent ancestry was detected by Plink–neighbor function. In particular, ethnical match between cases

and controls was confirmed by principal component analysis on common variants (minor allele frequency >5%) after exclusion of long-range linkage disequilibrium regions and variant pruning on linkage disequilibrium ( $r^2 > 0.2$ ).

### CNV calling

CNVs were detected from WES data using the CANOES software,<sup>24</sup> an algorithm dedicated to the detection of quantitative genomic variations based on read depth information. The read depth for each target was calculated using bedtools. Targets in close proximity (<30 bp distant) were fused to avoid false positives. When captured, UTR regions were removed from the analyses.

To reduce background variability, define a reference sample and build a valid statistical model, CANOES requires GC content data, and should be run on multiple exomes—ideally >30—technically as close as possible (that is, same capture kit, same sequencing center and same batch of sequencing). The GC content of each target was determined using the GATK suite. We then ran the CANOES software by batches, grouping exomes according to capture kits, sequencing centers and batches of sequencing (Supplementary Table S1).

After the calling, a 'not applicable' score was attributed to all CNVs from samples carrying >50 rearrangements. Such samples are usually characterized by higher or lower read depth and cannot be compared with the reference sample. All CNVs assigned with a not applicable score were thus removed from further analysis. At the end of this process, 13 870 CNVs were called in 1106 samples (522 cases and 584 controls). Of note, the mean number of CNV per sample was similar in the two main batches of cases and controls, which were sequenced in the Centre National de Génotypage (CNG, CEA, Evry, France) and the Genome Québec Innovation Centre (McGill University and Génomique Québec Innovation Centre, Montréal, Canada) using the V5 or V5-UTR capture kits (Supplementary Table S1).

The prediction performance of CANOES was then ascertained on a subset of patients for which array-Comparative Genomic Hybridization (aCGH) data was available.

### Validation of the false positive and false negative rates of CANOES

As CNV detection from exome data is a novel approach, we then performed a comprehensive check of our data. We compared CANOES calls with 1 M probes high-resolution aCGH data (Agilent Technologies, Santa Clara, CA, USA) in 42 individuals for whom both WES and aCGH data were available (Supplementary Table S2). Owing to the limited resolution of aCGH, we restricted this analysis to CNVs larger than 8 kb. Similarly, CNVs located in regions of segmental duplication, which are weakly or not covered by aCGH probes, were not included in the comparison analysis. Of the 67 rare (frequency <1%) CNVs intersecting genes detected by aCGH in these 42 individuals, 59 (88%) were also detected by CANOES. We noticed that CANOES specifically missed a subset of monoexonic CNVs that were successfully detected by aCGH, when embedded in large intronic or intergenic segments covered by a sufficient number of probes. Accordingly, one of the 32 CNVs larger than 50 kb was detected by aCGH, but was not detected by CANOES. Finally, when restricting our analysis to the 26 CNVs larger than 100 kb detected by aCGH, all were successfully detected by CANOES (Supplementary Table S2). Overall, CANOES detected more rare genic CNVs per individual ( $n=6.43$ ) than 1M probes-aCGH ( $n=1.60$ ). This difference is explained by a better resolution and by inclusion of CNVs located in regions of segmental duplication in CANOES raw data set.

We next undertook a validation study using QMPSF (Quantitative Multiplex of Short Fluorescent Fragments) in a set of rare CNVs identified by CANOES and including every CNV targeting either a list of genes previously reported as enriched in CNVs in AD<sup>25,26</sup> or the genes of the A $\beta$ -related gene list (Supplementary Table S3). Of the 93 events tested, 85 (91%) corresponding to 62 distinct CNVs were validated, confirming that the increased number of CNVs detected from exome data as compared with aCGH was not due to an excessive rate of false positive genic CNVs.

### Quantitative multiplex PCR of short fluorescent fragments

Each CNV from the case–control study was confirmed by QMPSF assay. In addition, two QMPSF assays were designed to search for the 17q21.31 duplication in three members of a recently reported family.<sup>27</sup> One targeted the *MAPT* gene<sup>28</sup> while the other spanned genes located inside and outside the 17q21.31 duplicated region (Supplementary Tables S4A and B). DNA was extracted from blood samples using the Flexigen DNA kit

(Qiagen, Hilden, Germany). For each QMPFSF, short genomic fragments spanning each coding exons of the *MAPT* gene (NM\_005910.5, or genes located inside (*CRHR1*) and outside (*NSF* and *ARGHAP27*) the 17q21.31 duplicated region were simultaneously amplified within a single PCR reaction, performed on 100 ng of genomic DNA. Forward primers were 6-FAM-labeled. DNA fragments were then separated on an ABI 3500 genetic analyzer (ThermoFisher Scientific, Illkirch, France). The fluorescence profiles were analyzed using the GeneMapper software 5 (ThermoFisher Scientific).

#### *MAPT* messenger RNA expression measurement by SNaPshot

Total RNA was extracted from whole blood collected into PAXgene Blood RNA tubes using the PAXgene Blood RNA kit (Qiagen). Genomic DNA was extracted using the Flexigene DNA kit (Qiagen). RNA was reverse-transcribed using the Verso cDNA synthesis kit (Fisher Scientific, Illkirch, France) with a blend of random hexamers and anchored oligo-dT (3:1). Allele-specific expression (ASE) was quantified on the five patients and eight H1/H2 heterozygous controls, using two H1/H2-tagging single-nucleotide polymorphisms (SNP): rs17652748 (c.\*2079C>T) and rs17574228 (c.\*2972T>C). ASE was measured using the SNaPshot (PE Applied Biosystems, Foster City, CA, USA) technique, following the manufacturer's recommendations. Each SNP was analyzed independently. Briefly, the complementary DNA and genomic DNA of each individual were PCR-amplified using pairs of primers specific of each SNP (Supplementary Table S5). The single-nucleotide primer extension was performed using the SNaPshot Multiplex kit (Applied Biosystems) and SNP-specific primers (Supplementary Table S5). Fluorescence was analyzed using an ABI 3500 automated sequencer (PE Applied Biosystems). The ASE ratio was calculated by normalizing the ratio between the peak area of the H1 and H2 alleles in cDNA to the same peak ratio measured on genomic DNA.

## RESULTS

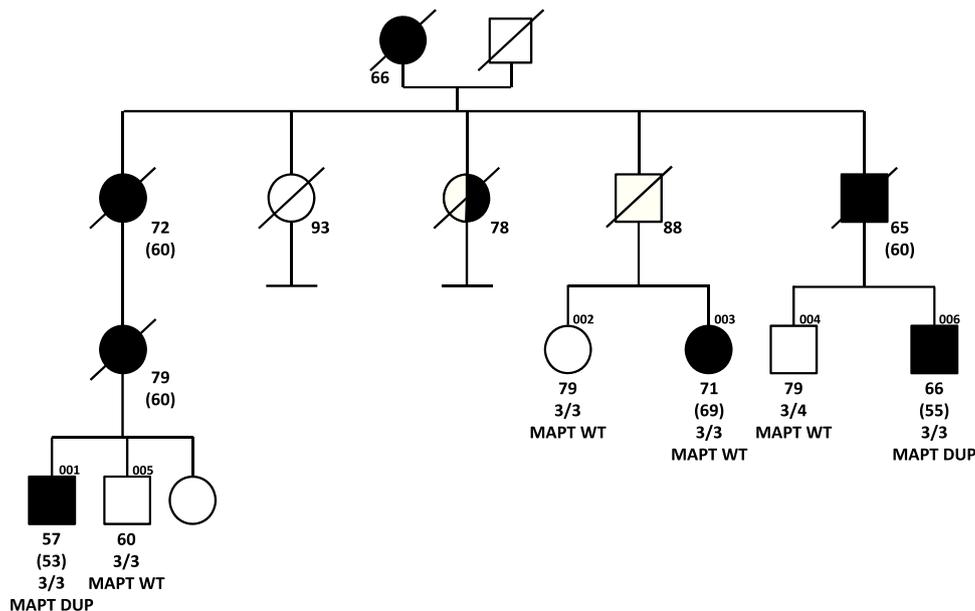
All rare CNVs (minor allele frequency < 1% in the whole sample) impacting genes were considered in case-control analyses, whatever the number of exons involved. Neither the average number of rare genic CNVs nor the average number of genes intersected by rare CNVs differed between cases and controls (6.43 vs 6.37, Mann-Whitney *U* test  $P=0.70$  and 16.2 vs 16.9, Mann-Whitney *U* test  $P=0.62$ , respectively).

We then focused on CNVs targeting an expert-curated list of genes, including 335 genes whose protein product is relevant to A $\beta$  processing, clearance, aggregation or toxicity<sup>29</sup> (Supple-

mentary Table S3). After removing events localized in known regions of segmental duplication, the burden of rare CNVs targeting the A $\beta$ -related gene list was compared between cases and controls using a Fisher exact test. We observed a significant increase in the burden of CNVs in cases compared with controls (30/522 carriers in cases vs 18/584 carriers in controls, odds ratio = 1.92, 95% confidence interval = 1.02–3.70,  $P=0.038$ ). Overall, 19 deletions (14 partial) and 30 duplications (13 partial) were found to intersect genes of this network in the whole study sample (Supplementary Tables S6A and B). The vast majority of these CNVs were singletons. We identified a novel *PSEN1* deletion encompassing both exons 9 and 10, which is most likely pathogenic. This deletion is predicted to result in an in-frame skipping of exons 9 and 10, and an amino-acid substitution at the aberrant exon 8–exon 11 junction, introducing at the beginning of exon 11 an S290W amino-acid change. Functional analysis will however be required to further ascertain the mechanisms underlying the pathogenicity of this deletion. For one gene (*CTSB*), we observed a reciprocal CNV in a patient and a control. Two genes (*SLC30A3* and *ABCA7*) were hit twice by overlapping CNVs in cases only. The deletion of the entire *ABCA7* gene is well in line with the now demonstrated excess of *ABCA7* loss-of-function variants found in AD patients of large case-control studies<sup>5,8,10</sup> and supports the claim that *ABCA7* loss of function constitutes a risk for AD.

Most importantly, four patients carried a recurrent 17q21.31 duplication including four genes, *CRHR1*, *MAPT*, *STH* and *KANSL1*, without any similar rearrangement detected in controls. Statistical analyses revealed a significant enrichment of this single rearrangement in cases compared with controls: 4/522 carriers in cases vs 0/584 carriers in controls,  $P=0.049$ , Fisher exact test. We then compared the frequency of the 17q21.31 duplication found in our patients with that observed in controls of the study of Coe *et al.*<sup>30</sup>: 4/522 carriers were observed in cases vs 3/19 584 carriers in controls,  $P=1.5 \times 10^{-5}$ , Fisher exact test. Finally, we performed the same comparison using the recently published CNV data from the ExAC database:<sup>31</sup> 4/522 carriers in cases vs 1/32 850 carriers in non-Finnish European population,  $P=2.9 \times 10^{-7}$ , Fisher exact test.

The 17q21.31 duplication is located within a ~900 kb inversion polymorphism resulting in near complete linkage disequilibrium

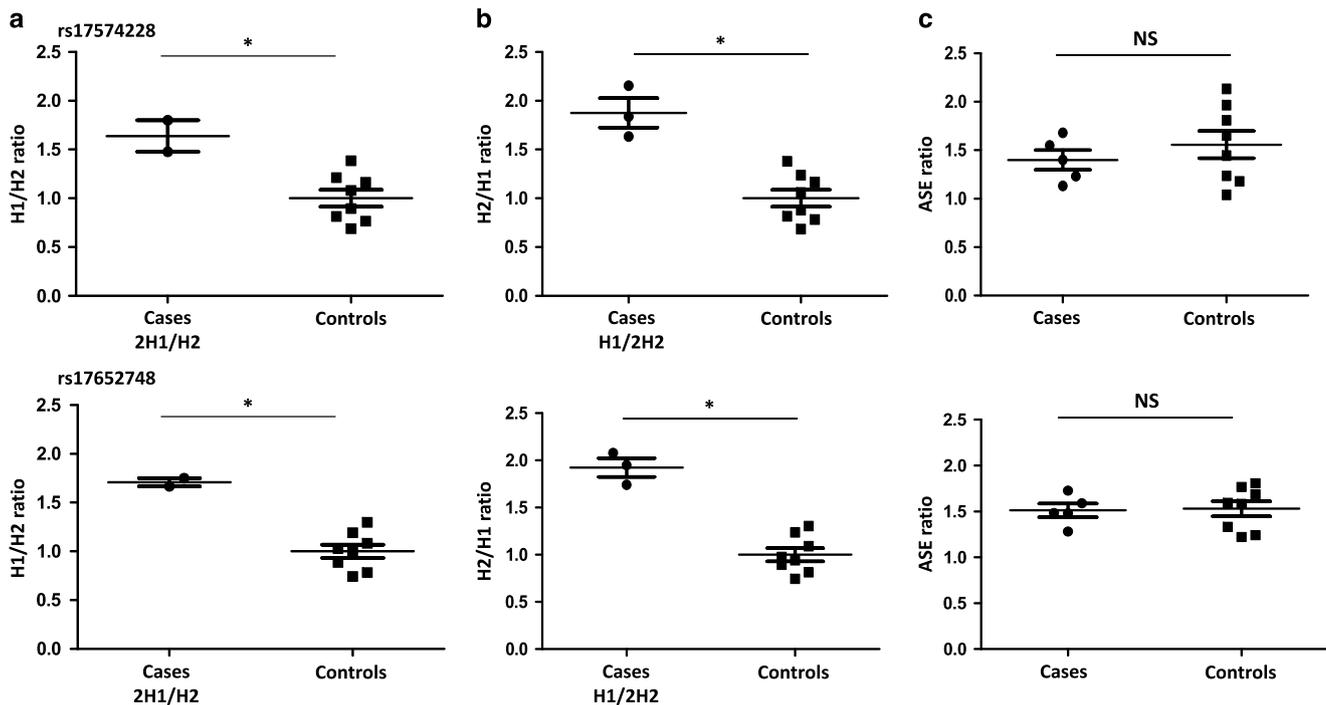


**Figure 1.** Pedigree of family ALZ 596. Filled symbols: probable AD. Semi-filled symbols: memory impairment. Current age or age at death (years) are indicated. Age of onset is in parentheses. APOE genotype and CNV status are indicated. AD, Alzheimer's disease; CNV, copy number variations; DUP, duplication, WT, wild-type.

**Table 1.** Phenotypic characteristics of patients with a 17q21.31 duplication

Family ID	ROU 1373		ALZ 596		ALZ 441	EXT 1114
Patient ID	001	001	006	005	001	
Sex	M	M	M	M	F	
Family history	Sporadic	Familial	Familial	Sporadic ( <i>de novo</i> )	Familial	
Education (years)	13	20	11	13	8	
Age at onset (years)	54	51	55	45	58	
Age at last clinical evaluation (years)	60	57	66	55	60	
MMSE score (age)	<b>25/30</b> (58) <b>23/30</b> (59) <b>21/30</b> (60)	<b>24/30</b> (55) <b>23/30</b> (57)	<b>0/30</b> (66)	<b>21/30</b> (46) <b>8/30</b> (49) <b>0/30</b> (55)	27/30 (59) <b>17/30</b> (60)	
Clinical presentation at onset	Amnestic <sup>a</sup>	Amnestic <sup>a</sup>	Amnestic <sup>a</sup>	Amnestic <sup>a</sup>	Amnestic <sup>a</sup> +behavioral changes <sup>b</sup>	
Structural imaging (age)	Bilateral hippocampal (Scheltens II) and parieto-occipital bilateral atrophy (59)	Bilateral hippocampal atrophy (Scheltens IV) (56)	Bilateral hippocampal atrophy (CT scan)	Bilateral hippocampal atrophy (Scheltens III) (46)	Bilateral hippocampal atrophy (Scheltens IV) (60)	
Functional imaging (age)	<sup>18</sup> FDG-PET: hypometabolism in medial temporal and parietal regions (59)	<sup>18</sup> FDG-PET: hypometabolism of both medial temporal regions (53)	NA	<sup>18</sup> FDG-PET: hypometabolism in prefrontal, medial temporal and posterior associative regions with right predominance (47)	<sup>99m</sup> Tc-SPECT: Left medial temporal hypoperfusion (60)	
Amyloid PET (age)	Negative (18F-flutemetamol) (58)	Negative (18F-florbetapir) (57)	NA	NA	Negative (18F-florbetapen) (61)	
Aβ <sub>42</sub> (pg mL <sup>-1</sup> )	864	<b>496</b>	NA	<b>260</b>	<b>372</b>	
t-Tau (pg mL <sup>-1</sup> )	<b>696</b>	<b>558</b>	NA	<b>507</b>	<b>866</b>	
P-Tau (pg mL <sup>-1</sup> )	<b>98</b>	<b>66</b>	NA	<b>67</b>	<b>94</b>	
t-Tau/Aβ <sub>42</sub>	<b>0.81</b>	<b>1.13</b>	NA	<b>1.95</b>	<b>2.32</b>	
APOE genotype	3-4	3-3	3-3	3-4	3-3	
MAPT haplotype	2H1/H2	2H2/H1	2H2/H1	2H1/H2	2H2/H1	

Abbreviations: Aβ, amyloid-β; CSF, cerebrospinal fluid; CT, computed tomography; NA, not available; PET, positron emission tomography; P-Tau, phosphorylated Tau; t-Tau, total Tau. <sup>a</sup>Impairment of episodic memory with inefficient cueing evaluated by Free and Cued Selective Recall Reminding Test (Supplementary Table S7). Age at examination (years) is indicated in parentheses. <sup>b</sup>Irritability and apathy. Abnormal CSF biomarker values appear in bold. Normative values for CSF biomarkers were: Aβ > 550 (pg mL<sup>-1</sup>), t-Tau < 350 (pg mL<sup>-1</sup>), P-Tau < 60 (pg mL<sup>-1</sup>), t-tau/Aβ<sub>42</sub> < 0.52. Abnormal Mini Mental State Examination (MMSE) scores according to age and education appear in bold.



**Figure 2.** Increased expression of *MAPT* in duplication carriers. SNaPshot analysis was performed on rs17574228 (upper panel) and rs17652748 (lower panel). (a) Comparison of the H1/H2 ratio in cDNA between cases carrying an H1 duplication and controls. The H1/H2 ratio was arbitrarily set at 1.0 in controls. (b) Comparison of the H2/H1 ratio in cDNA between cases carrying an H2 duplication and controls. The H2/H1 ratio was arbitrarily set at 1.0 in controls. (c) Comparison of ASE ratios between cases and controls. The ASE ratio was calculated by normalizing the ratio between the peak area of the H1 and H2 alleles in cDNA to the same peak ratio measured on genomic DNA. \*Mann–Whitney test  $P$ -value < 0.05. ASE, allele-specific expression; NS, non-significant.

across the interval and defining two haplotypes referred to as H1 and H2. All duplication carriers were heterozygous for the H1/H2 haplotype, and SNaPshot analysis on genomic DNA showed that the duplication had occurred twice on an H1 haplotype background and twice on an H2 haplotype background (Supplementary Figure S1). Two cases were sporadic. For one of them, parental DNA was available, allowing us to show that the CNV had occurred *de novo* (case ALZ 441). The two other patients had a positive family history of AD. In one family (ALZ 596), DNA was available for relatives. Segregation of the CNV is displayed in Figure 1. The CNV was present in two distantly affected patients and none of the healthy relatives. A third relative was diagnosed with probable late-onset AD and did not carry the CNV. Accordingly, her own father, who would have been an obligate carrier, died unaffected at age 88, allowing to consider this patient as a phenocopy. Phenotypic characteristics of the patients (four index cases and one affected relative) carrying the 17q21.31 duplication are displayed in detail in Table 1. Briefly, all five patients exhibited progressive cognitive decline characterized by impaired episodic memory with inefficient cueing together with impairment of at least one other cognitive domain (Supplementary Table S7). Only one patient exhibited concomitant mild behavioral disturbances (irritability and apathy). Both spatial patterns of atrophy in structural imaging (magnetic resonance imaging,  $n=4$  or computerized tomography scan,  $n=1$ ) and patterns of decreased  $^{18}\text{F}$ FDG uptake (PETscan,  $n=3$ ) or  $^{99\text{m}}\text{Tc}$ -SPECT perfusion ( $n=1$ ) were consistent with AD signature.<sup>32</sup> CSF biomarkers were available for the four index cases. Interestingly, all but one patient had the three biomarkers (t-Tau, P-Tau and  $\text{A}\beta_{42}$ ) in pathological ranges, indicating both an amyloidogenic and a Tau-related neurodegenerative process. In the fourth patient (ROU 1373), t-Tau and P-Tau were increased, but  $\text{A}\beta_{42}$  was in the normal range. However, the Tau/ $\text{A}\beta_{42}$  ratio was abnormal,<sup>33</sup>

allowing to support an AD diagnosis, albeit with a lesser evidence than for the three above-mentioned patients. Strikingly, amyloid PET imaging performed in three patients showed no evidence of amyloid plaques. None of the patients had neuropathological examination. To overcome this limitation, we analyzed DNA from three members of a recently reported additional family with a similar phenotype.<sup>27</sup> The two affected members carried the 17q21.31 duplication, whereas the third unaffected member did not (Supplementary Figures S2A and B). In this family, the neurodegenerative disorder was characterized as a pure tauopathy with neurofibrillary tangles predominating in hippocampus, entorhinal cortex and nucleus basalis Meynert without  $\text{A}\beta$  deposits (see detailed description in Alexander *et al.*<sup>27</sup>).

The *MAPT* gene encodes the Tau protein, a key player in several neurodegenerative disorders including AD.<sup>34</sup> To assess the impact of the increased gene dosage on *MAPT* expression, we performed SNaPshot experiments by genotyping two SNPs tagging the H1/H2 haplotype, rs17574228 and rs17652748 (Figure 2 and Supplementary Figure S1), on eight control individuals heterozygous for the H1/H2 haplotype and five patients, including the four index cases and one affected relative of ALZ 596-001. When focusing on the allelic ratios in cDNA, we were able to detect on average a 1.64-fold overexpression of *MAPT* transcripts in blood cells of patients carrying an H1 duplication and a 1.87-fold in patients carrying an H2 duplication, compared with controls (Figures 2a and b). The *MAPT* increased gene dosage is thus associated with increased messenger RNA expression whatever the duplicated allele. Comparison of ASE between cases and controls (Figure 2c and Supplementary Figure S1) confirmed that the H1 allele was basically more expressed than the H2 allele<sup>35</sup> in both groups. The haplotype-specific ASE was similar in WT and *MAPT* duplication carriers, confirming that the *MAPT* duplication is indeed associated with increased messenger RNA expression.

## DISCUSSION

The most striking finding of this study is the presence of a 17q21.31 duplication involving the complete coding sequence of the *MAPT* gene, as well as flanking genes in four index AD patients and one affected relative. This CNV was not present among the 584 controls. This rearrangement was initially described in rare subjects with intellectual disability (ID), but its link with this condition is highly questionable because (i) no common phenotypic feature was found among cases carrying this duplication,<sup>36</sup> (ii) some subjects either carried another candidate CNV<sup>37</sup> or were not screened for mutations in all genes relevant for their phenotype<sup>38</sup> and (iii) three out of four carriers reported in a screen including 13 070 ID patients had an intelligence quotient in the normal range (no intelligence quotient for the fourth case).<sup>39</sup> None of the subjects described here was diagnosed with ID. It is therefore likely that the identification of 17q21.31 duplications in rare cases of developmental disorders is due to a reporting bias, as demonstrated by the data of Coe *et al.*,<sup>30</sup> reporting a similar frequency of 17q21.31 duplications in their large case-control samples (4/29 085 children with developmental delay vs 3/19 584 controls). According to this large study, the established frequency of the rearrangement is thus estimated to be 0.01% in the general population.

Subsequently, we described this duplication in a patient in whom memory loss and aggressive behavior were present by age 55.<sup>28</sup> This patient, who had a positive family history of dementia, died by age 64 severely demented but without neuropathological examination. Finally, Hooli *et al.*<sup>14</sup> reported the segregation of a 170 kb duplication spanning *MAPT* in three affected sibs with a clinical diagnosis of EOAD. With this study, we add seven novel demented cases from five distinct pedigrees carrying a 17q21.31 duplication, one of them arising *de novo*. Clinical signs, standard neuroimaging and CSF biomarkers were consistent with an AD diagnosis in our patients. However, amyloid PET imaging failed to document A $\beta$  deposits in three affected subjects and neuropathological examination confirmed that, despite an AD-like clinical phenotype, no senile plaques could be detected in an additional 17q21.31 duplication carrier exhibiting Tau pathology. We thus define this clinico-pathological entity as prominent Tau-related dementia with increased Tau expression. This could be related to the biomarker-based concept referred to as SNAP.<sup>3</sup> Nevertheless, the CSF profile supporting an A $\beta$  pathology in 3/4 patients is puzzling. The decreased CSF level of soluble A $\beta$ <sub>42</sub> could be consistent with an increased level of A $\beta$  oligomers in the cerebral parenchyma, which are hardly detectable by amyloid PET imaging or immunostaining. In that case, an AD diagnosis would *in fine* be favored and this would imply a direct involvement of the *MAPT* gene in AD determinism. Such an involvement is progressively emerging. It has been suggested by a recent study reporting an association of the *MAPT* H2 haplotype with both reduced risk of late-onset AD and lower *MAPT* transcript brain levels.<sup>35</sup> In addition, the rare *MAPT* p.A152T variant, which decreases the binding of Tau to microtubules and increases the formation of Tau oligomers, has been reported to increase the risk for both fronto-temporal lobar degeneration spectrum disorders and AD.<sup>40</sup> Alternatively, these patients could be affected by a pure tauopathy such as primary age-related tauopathy (PART), a controversial neuropathological entity<sup>41</sup> defined by the presence of numerous neurofibrillary tangles in the medial temporal lobe with almost no amyloid deposit.<sup>42</sup> PART has been suggested to explain part of SNAP.<sup>3</sup> Its biomarker signature is unknown; however, PART is considered to affect the very old (>80 years),<sup>42</sup> which is not the case of our patients. It would be of great interest to analyze DNA of PART subjects to determine if prominent Tau-related dementia and PART share similar molecular bases.

The phenotype of our patients carrying this duplication is clearly different on clinical and neuropathological ground from

fronto-temporal dementia with Parkinsonism linked to chromosome 17 caused by *MAPT* missense mutations.<sup>34</sup> Moreover, to our knowledge, *MAPT* duplications have never been reported in fronto-temporal lobar degeneration patients.

Although the role of Tau in the neurodegenerative process is not doubtful, we cannot exclude that other genes included in the duplicated segment may also play a role in this disease. Indeed, a novel locus located near the *MAPT* gene and including the *KANSL1* gene, has been associated with late-onset AD risk in a recent genome-wide association study.<sup>43</sup>

We conclude that 17q21.31 microduplication associated to increased gene dosage and expression of *MAPT* leads to a neurodegenerative disorder with prominent Tau pathology in medial temporal structures. The frequency of this microduplication in the general population suggests that it could account for a significant proportion of genetically unexplained early-onset dementia cases. This novel clinico-pathological entity could represent a proportion of cases fulfilling the criteria for probable AD with no amyloid PET binding identified in large clinical series.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

This study was funded by grants from the Clinical Research Hospital Program from the French Ministry of Health (GMAJ, PHRC 2008/067), the CNR-MAJ and the JPND PERADES (ANR-13-JPRF-0001-04). This work was supported by France Génomique, Labex GENMED ANR-10-LABX-0013, the National Foundation for Alzheimer's disease and related disorders, the Institut Pasteur de Lille, the Centre National de Génotypage, Inserm, FRC (fondation pour la recherche sur le cerveau) and by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant (Development of Innovative Strategies for a Transdisciplinary approach to Alzheimer's disease). We are indebted to the Banque d'ADN et de cellules-Institut du Cerveau et de la Moelle épinière (ICM-Inserm U1127-UPMC P6 UMR-S 1127-CNRS UMR 7225).

## REFERENCES

- McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH *et al*. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement J Alzheimers Assoc* 2011; **7**: 263–269.
- Ossenkoppelle R, Jansen WJ, Rabinovici GD, Knol DL, van der Flier WM, van Berckel BNM *et al*. Prevalence of amyloid PET positivity in dementia syndromes: a meta-analysis. *JAMA* 2015; **313**: 1939–1949.
- Jack CR, Knopman DS, Chételat G, Dickson D, Fagan AM, Frisoni GB *et al*. Suspected non-Alzheimer disease pathophysiology—concept and controversy. *Nat Rev Neurol* 2016; **12**: 117–124.
- Jack CR, Knopman DS, Weigand SD, Wiste HJ, Vemuri P, Lowe V *et al*. An operational approach to National Institute on Aging-Alzheimer's Association criteria for preclinical Alzheimer disease. *Ann Neurol* 2012; **71**: 765–775.
- Cuyvers E, De Roeck A, Van den Bossche T, Van Cauwenbergh C, Bettens K, Vermeulen S *et al*. Mutations in ABCA7 in a Belgian cohort of Alzheimer's disease patients: a targeted resequencing study. *Lancet Neurol* 2015; **14**: 814–822.
- Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogava E, Majounie E *et al*. TREM2 variants in Alzheimer's disease. *N Engl J Med* 2013; **368**: 117–127.
- Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, Snaedal J *et al*. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* 2013; **368**: 107–116.
- Le Guennec K, Nicolas G, Quenez O, Charbonnier C, Wallon D, Bellenguez C *et al*. ABCA7 rare variants and Alzheimer disease risk. *Neurology* 2016; **86**: 2134–2137.
- Nicolas G, Charbonnier C, Wallon D, Quenez O, Bellenguez C, Grenier-Boley B *et al*. SORL1 rare variants: a major risk factor for familial early-onset Alzheimer's disease. *Mol Psychiatry* 2016; **21**: 831–836.
- Steinberg S, Stefansson H, Jonsson T, Johannsdottir H, Ingason A, Helgason H *et al*. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nat Genet* 2015; **47**: 445–447.

- 11 Hiltunen M, Helisalms S, Mannerman A, Alafuzoff I, Koivisto AM, Lehtovirta M *et al*. Identification of a novel 4.6-kb genomic deletion in presenilin-1 gene which results in exclusion of exon 9 in a Finnish early onset Alzheimer's disease family: an Alu core sequence-stimulated recombination? *Eur J Hum Genet EJHG* 2000; **8**: 259–266.
- 12 Prihar G, Verkkoniemi A, Perez-Tur J, Crook R, Lincoln S, Houlden H *et al*. Alzheimer disease PS-1 exon 9 deletion defined. *Nat Med* 1999; **5**: 1090.
- 13 Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerrière A, Vital A *et al*. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet* 2006; **38**: 24–26.
- 14 Hooli BV, Kovacs-Vajna ZM, Mullin K, Blumenthal MA, Mattheisen M, Zhang C *et al*. Rare autosomal copy number variations in early-onset familial Alzheimer's disease. *Mol Psychiatry* 2014; **19**: 676–681.
- 15 Rovelet-Lecrux A, Legallic S, Wallon D, Flaman J-M, Martinaud O, Bombois S *et al*. A genome-wide study reveals rare CNVs exclusive to extreme phenotypes of Alzheimer disease. *Eur J Hum Genet EJHG* 2012; **20**: 613–617.
- 16 Brouwers N, Van Cauwenberghc C, Engelborghs S, Lambert J-C, Bettens K, Le Bastard N *et al*. Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites. *Mol Psychiatry* 2012; **17**: 223–233.
- 17 Nicolas G, Wallon D, Charbonnier C, Quenez O, Rousseau S, Richard A-C *et al*. Screening of dementia genes by whole-exome sequencing in early-onset Alzheimer disease: input and lessons. *Eur J Hum Genet EJHG* 2016; **24**: 710–716.
- 18 DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ *et al*. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 2011; **72**: 245–256.
- 19 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A *et al*. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–1303.
- 20 Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl* 2010; **26**: 589–595.
- 21 Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc* 2010; **5**: 1564–1573.
- 22 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 2015; **4**: 7.
- 23 Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR *et al*. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet* 2012; **91**: 839–848.
- 24 Backenroth D, Homsy J, Murillo LR, Glessner J, Lin E, Brueckner M *et al*. CANOES: detecting rare copy number variants from whole exome sequencing data. *Nucleic Acids Res* 2014; **42**: e97.
- 25 Swaminathan S, Huentelman MJ, Corneveaux JJ, Myers AJ, Faber KM, Foroud T *et al*. Analysis of copy number variation in Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. *PLoS One* 2012; **7**: e50640.
- 26 Swaminathan S, Shen L, Kim S, Inlow M, West JD, Faber KM *et al*. Analysis of copy number variation in Alzheimer's disease: the NIALOAD/ NCRAD Family Study. *Curr Alzheimer Res* 2012; **9**: 801–814.
- 27 Alexander J, Kalev O, Mehrabian S, Traykov L, Raycheva M, Kanakis D *et al*. Familial early-onset dementia with complex neuropathologic phenotype and genomic background. *Neurobiol Aging* 2016; **42**: 199–204.
- 28 Rovelet-Lecrux A, Hannequin D, Guillin O, Legallic S, Jurici S, Wallon D *et al*. Frontotemporal dementia phenotype associated with MAPT gene duplication. *J Alzheimers Dis JAD* 2010; **21**: 897–902.
- 29 Campion D, Pottier C, Nicolas G, Le Guennec K, Rovelet-Lecrux A. Alzheimer disease: modeling an A $\beta$ -centered biological network. *Mol Psychiatry* 2016; **21**: 861–871.
- 30 Coe BP, Witherspoon K, Rosenfeld JA, van Bon BWM, Vulto-van Silfhout AT, Bosco P *et al*. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet* 2014; **46**: 1063–1071.
- 31 Ruderfer DM, Hamamsy T, Lek M, Karczewski KJ, Kavanagh D, Samocha KE *et al*. Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. *Nat Genet* 2016; **48**: 1107–1111.
- 32 Jack CR, Holtzman DM. Biomarker modeling of Alzheimer's disease. *Neuron* 2013; **80**: 1347–1358.
- 33 Duits FH, Teunissen CE, Bouwman FH, Visser P-J, Mattsson N, Zetterberg H *et al*. The cerebrospinal fluid 'Alzheimer profile': easily said, but what does it mean? *Alzheimers Dement J Alzheimers Assoc* 2014; **10**: 713–723.e2.
- 34 Spillantini MG, Goedert M. Tau pathology and neurodegeneration. *Lancet Neurol* 2013; **12**: 609–622.
- 35 Allen M, Kachadoorian M, Quicksall Z, Zou F, Chai HS, Younkin C *et al*. Association of MAPT haplotypes with Alzheimer's disease risk and MAPT brain gene expression levels. *Alzheimers Res Ther* 2014; **6**: 39.
- 36 Natacci F, Alfei E, Tararà L, D'Arrigo S, Zuffardi O, Gentilin B *et al*. Chromosome 17q21.31 duplication syndrome: Description of a new familial case and further delineation of the clinical spectrum. *Eur J Paediatr Neurol EJPN Off J Eur Paediatr Neurol Soc* 2016; **20**: 183–187.
- 37 Kitsiou-Tzeli S, Frysira H, Giannikou K, Symrou A, Kosma K, Kakourou G *et al*. Microdeletion and microduplication 17q21.31 plus an additional CNV, in patients with intellectual disability, identified by array-CGH. *Gene* 2012; **492**: 319–324.
- 38 Kirchhoff M, Bisgaard A-M, Duno M, Hansen FJ, Schwartz M. A 17q21.31 microduplication, reciprocal to the newly described 17q21.31 microdeletion, in a girl with severe psychomotor developmental delay and dysmorphic craniofacial features. *Eur J Med Genet* 2007; **50**: 256–263.
- 39 Grisart B, Willatt L, Destrée A, Fryns J-P, Rack K, de Ravel T *et al*. 17q21.31 microduplication patients are characterised by behavioural problems and poor social interaction. *J Med Genet* 2009; **46**: 524–530.
- 40 Coppola G, Chinnathambi S, Lee JJ, Dombroski BA, Baker MC, Soto-Ortolaza AI *et al*. Evidence for a role of the rare p.A152T variant in MAPT in increasing the risk for FTD-spectrum and Alzheimer's diseases. *Hum Mol Genet* 2012; **21**: 3500–3512.
- 41 Duyckaerts C, Braak H, Brion J-P, Buée L, Del Tredici K, Goedert M *et al*. PART is part of Alzheimer disease. *Acta Neuropathol (Berl)* 2015; **129**: 749–756.
- 42 Jellinger KA, Alafuzoff I, Attems J, Beach TG, Cairns NJ, Crary JF *et al*. PART, a distinct tauopathy, different from classical sporadic Alzheimer disease. *Acta Neuropathol (Berl)* 2015; **129**: 757–762.
- 43 Jun G, Ibrahim-Verbaas CA, Vronskaya M, Lambert J-C, Chung J, Naj AC *et al*. A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol Psychiatry* 2016; **21**: 108–117.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (<http://www.nature.com/mp>)

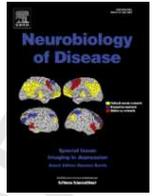
#### 4. Caractérisation fonctionnelle d'une nouvelle délétion des exons 9 et 10 de *PSEN1*

La détection de CNVs à partir de données d'exomes nous a également permis de mettre en évidence une nouvelle délétion génomique de *PSEN1* emportant les exons 9 et 10. Les mutations de *PSEN1* comptent pour la majorité des variations responsables d'ADEOAD. Parmi les 221 mutations rapportées, 6 entraînent la délétion de l'exon 9 à cause d'un remaniement génomique ou d'une mutation ponctuelle d'épissage (Crook et al., 1998; Dumanchin et al., 2006; Perez-Tur et al., 1995; Rovelet-Lecrux et al., 2015; Sato et al., 1998; Smith et al., 2001) L'analyse des points de cassure de la délétion des exons 9 et 10 a permis de déterminer que celle-ci avait une taille 10,1 kb et que le mécanisme de recombinaison était médié par des éléments Alu. L'analyse de l'ARN du patient a permis de confirmer que l'allèle mutant était correctement exprimé. D'un point de vue clinique, le patient présentait une altération des fonctions cognitives associée à une paraparésie spastique, une caractéristique retrouvée également chez les patients porteurs d'une délétion de l'exon 9 de *PSEN1*. Afin de caractériser l'effet de cette variation structurale sur l'activité du complexe  $\gamma$ -sécrétase, nous avons établi une collaboration avec l'équipe du Pr Bart De Strooper en Belgique, spécialiste de ce complexe (Chávez-Gutiérrez et al., 2012; Szaruga et al., 2015; Veugelen et al., 2016). Lors d'un stage de 2 mois dans ce laboratoire, j'ai pu étudier l'effet de cette délétion sur l'activité endopeptidase et carboxypeptidase du complexe  $\gamma$ -sécrétase à partir de fibroblastes embryonnaires murins (*Mouse Embryonic Fibroblasts*: MEFs) déficientes pour *PSEN1* et *PSEN2* et surexprimant la forme sauvage ou mutante de *PSEN1* humain.

Dans un premier temps, nous avons pu déterminer que la forme mutante de *PSEN1* est capable de reconstituer un complexe  $\gamma$ -sécrétase mature. Nous avons ensuite pu observer que la délétion des exons 9 et 10 réduit fortement l'activité endopeptidase résultant *in fine* en une baisse globale de la production de peptide A $\beta$ . Enfin, l'activité carboxypeptidase a été évaluée en dosant les différents isoformes d'A $\beta$  (38, 40, 42 et 43). La forme mutante  $\Delta$ 9-10 altère fortement le clivage carboxypeptidase, résultant en l'absence de production d'A $\beta$ 38 et en une augmentation considérable de la production d'A $\beta$ 42 et d'A $\beta$ 43. A $\beta$ 43 devient ainsi la forme majoritairement produite. Or, d'après des études *in vivo*, le peptide A $\beta$ 43 serait potentiellement plus amyloïdergique et neurotoxique que les peptides A $\beta$ 40 et A $\beta$ 42 (Saito et al., 2011).

En conclusion, nous avons pu mettre en évidence que cette nouvelle délétion des exons 9 et 10 de *PSEN1* détectée à partir de données de séquençage d'exomes est responsable de maladie d'Alzheimer du sujet jeune associée à une paraparésie spastique. Au niveau fonctionnel, cette variation structurale entraîne une dysfonction du complexe  $\gamma$ -sécrétase générant majoritairement des peptides plus amyloïdergiques et plus neurotoxiques. Ce résultat confirme les précédentes

études effectuées sur l'effet des variations de *PSEN1* qui affectent toutes le clivage carboxypeptidase (Chávez-Gutiérrez et al., 2012) et sont en faveur de la contribution d'Aβ43 dans la physiopathologie de la MA.



## Deletion of exons 9 and 10 of the Presenilin 1 gene in a patient with Early-onset Alzheimer Disease generates longer amyloid seeds

Kilan Le Guennec<sup>a</sup>, Sarah Veugelen<sup>b, c</sup>, Olivier Quenez<sup>a</sup>, Maria Szaruga<sup>b, c</sup>, Stéphane Rousseau<sup>a</sup>, Gaël Nicolas<sup>a</sup>, David Wallon<sup>d</sup>, Frédérique Fluchere<sup>e</sup>, Thierry Frébourg<sup>f</sup>, Bart De Strooper<sup>b, c, g</sup>, Dominique Campion<sup>a, h</sup>, Lucía Chávez-Gutiérrez<sup>b, c</sup>, Anne Rovelet-Lecrux<sup>a, \*</sup>

<sup>a</sup> Normandie Univ, UNIROUEN, Inserm U1245, Rouen University Hospital, Department of Genetics and CNR-MAJ, Normandy Center for Genomic and Personalized Medicine, F 76000 Rouen, France

<sup>b</sup> VIB - Center for Brain and Disease Research, University of Leuven, 3000 Leuven, Belgium

<sup>c</sup> Center for Human Genetics, Leuven Research Institute for Neuroscience & Disease (LIND), University of Leuven, 3000 Leuven, Belgium

<sup>d</sup> Normandie Univ, UNIROUEN, Inserm U1245, Rouen University Hospital, Department of Neurology and CNR-MAJ, Normandy Center for Genomic and Personalized Medicine, F 76000 Rouen, France

<sup>e</sup> Department of Neurology and Movement Disorders, APHM, La Timone, Pôle de Neurosciences cliniques, Aix-Marseille Univ, Marseille, France.

<sup>f</sup> Normandie Univ, UNIROUEN, Inserm U1245, Rouen University Hospital, Department of Genetics, Normandy Center for Genomic and Personalized Medicine, F 76000 Rouen, France

<sup>g</sup> Institute of Neurology, University College London, Queen Square, WC1N 3BG London, UK

<sup>h</sup> Department of Research, Rouvray Psychiatric Hospital, Sotteville-lès-Rouen, France

### ARTICLE INFO

#### Article history:

Received 19 January 2017

Received in revised form 27 March 2017

Accepted 27 April 2017

Available online xxx

#### Keywords:

Alzheimer Disease

Early-onset

PSEN1

Hydrophilic loop

Aβ43

Amyloid

### ABSTRACT

*Presenilin 1 (PSEN1)* mutations are the main cause of autosomal dominant Early-onset Alzheimer Disease (EOAD). Among them, deletions of exon 9 have been reported to be associated with a phenotype of spastic paraparesis.

Using exome data from a large sample of 522 EOAD cases and 584 controls to search for genomic copy-number variations (CNVs), we report here a novel partial, in-frame deletion of *PSEN1*, removing both exons 9 and 10. The patient presented with memory impairment associated with spastic paraparesis, both starting from the age of 56 years. He presented a positive family history of EOAD. We performed functional analysis to elucidate the impact of this novel deletion on PSEN1 activity as part of the  $\gamma$ -secretase complex. The deletion does not affect the assembly of a mature protease complex but has an extreme impact on its global endopeptidase activity. The mutant carboxypeptidase-like activity is also strongly impaired and the deleterious mutant effect leads to an incomplete digestion of long A $\beta$  peptides and enhances the production of A $\beta$ 43, which has been shown to be potentially amyloidogenic and neurotoxic *in vivo*.

© 2016 Published by Elsevier Ltd.

### 1. Introduction

Mutations in the *Amyloid Precursor protein (APP)*, *Presenilin 1 (PSEN1)* and *Presenilin 2 (PSEN2)* genes and duplications of the *APP* locus are the main causes of Autosomal Dominant Early-onset Alzheimer's disease (AD-EOAD). Mutations in the *PSEN1* gene account for the majority of AD-EOAD cases with 221 pathogenic mutations reported to date (<http://www.alzforum.org/>), including 6 deletions of *PSEN1* exon 9 resulting either from splicing mutations (Perez-Tur et al., 1995; Sato et al., 1998; Rovelet-Lecrux et al., 2015) or genomic deletions (Prihar et al., 1999; Hiltunen et al., 2000; Smith et al., 2001).

Presenilin is the catalytic subunit of the  $\gamma$ -secretase, an enzymatic complex involved in the processing of numerous type 1 transmem-

brane proteins, including APP and Notch (Jurisch-Yaksi et al., 2013). The sequential cleavage of APP by the  $\beta$ - and  $\gamma$ -secretase releases the A $\beta$  peptides of different lengths. The accumulation of aggregation-prone A $\beta$  peptides into senile plaques is central to AD pathophysiology. The  $\gamma$ -secretase cleavage occurs in a multi-step process: first, the endopeptidase ( $\epsilon$ ) cleavage releases the C-terminal part of APP (AICD) and, depending on the position of the  $\epsilon$ -cleavage, either the A $\beta$ 49 or A $\beta$ 48 fragment will be produced. Then, carboxypeptidase-like  $\gamma$ -cleavages sequentially shorten these A $\beta$  peptides following specific lines of production: A $\beta$ 49  $\rightarrow$  A $\beta$ 46  $\rightarrow$  A $\beta$ 43  $\rightarrow$  A $\beta$ 40  $\rightarrow$  A $\beta$ 37 and A $\beta$ 48  $\rightarrow$  A $\beta$ 45  $\rightarrow$  A $\beta$ 42  $\rightarrow$  A $\beta$ 38 (Takami et al., 2009). AD-linked *PSEN1* variations result in various effects on  $\gamma$ -secretase activity: whereas the endopeptidase cleavage is not always impaired, the carboxypeptidase efficiency is consistently reduced, resulting in qualitative shifts in A $\beta$  profile, favoring the secretion of longer, more aggregation-prone A $\beta$  peptides such as A $\beta$ 42 and A $\beta$ 43, and possibly

\* Corresponding author.

Email address: anne.roveletlecrux@univ-rouen.fr (A. Rovelet-Lecrux)

longer (Saito et al., 2011; Chávez-Gutiérrez et al., 2012; Szaruga et al., 2015; Veugelen et al., 2016).

In the present study, we used exome sequencing data from a large cohort of 522 EOAD cases and 584 controls to search for genomic copy-number variations (CNVs). We identified a novel mutation consisting of a genomic deletion of exons 9 and 10 of *PSEN1* in an EOAD patient and evaluated the consequences of this variation on  $\gamma$ -secretase activity. The deletion of exons 9-10 drastically decreases the global endopeptidase activity and results in an increased secretion of A $\beta$ 43. Following previous work demonstrating that other highly inactivating *PSEN1* mutations show similar effects on A $\beta$  secretion (Veugelen et al., 2016), the data presented here further supports the pathogenic role of A $\beta$ 43, even at low concentrations.

## 2. Materials and methods

### 2.1. Patients, whole exome sequencing, and CNV calling

We performed whole exome sequencing (WES) in 546 patients with EOAD and 597 controls from French ancestry. Inclusion of patients and controls, WES details, bioinformatics pipelines and quality checks (QC) were previously reported (Nicolas et al., 2016b). Briefly, diagnoses of probable AD were performed according to the NIA-AA recommendations (McKhann et al., 2011) and genetic diagnosis of mutation in AD-causative genes were first made prior to WES if our national criteria were met (*i.e.* two cases from the same family with AAO before 65 or sporadic case with AAO before 50) (Nicolas et al., 2016a). Screening of AD-causative CNVs was routinely performed for *APP* duplication following same criteria.

For 385 EOAD patients, cerebrospinal fluid (CSF) total-Tau, Phosphorylated-Tau, and A $\beta$ 42 biomarkers were available and were all indicative of AD pathophysiology (Le Guennec et al., 2016a). The mean age at onset was 55 years (range 44–65). Out of the 546 initially included cases and 597 controls, 535 cases and 593 controls passed the QC for genetic case-control association analyses and were retained for CNV calling.

CNV calling was performed using the CANOES software (Backenroth et al., 2014), an algorithm dedicated to the detection of quantitative genomic variations based on read depth information. A final sample of 522 patients and 584 controls passed the quality criteria for CNV calling. A total of 13,870 CNVs were called in these 1106 samples. Note that, as CNV calling by CANOES is based on read depth, only rare events are efficiently detected. Previous validation of this software using aCGH data revealed a high concordance (88%) between the two methods, and showed that CANOES specifically missed a subset of very small monoexonic CNVs with lower read depth (Le Guennec et al., 2016b).

### 2.2. Quantitative Multiplex PCR of Short Fluorescent fragments (QMPSF)

Genomic DNA was extracted from blood samples using the Flexigen DNA kit (Qiagen, Hilden, Germany). *PSEN1* exon 9/10 deletion was validated according to an already published process using QMPSF spanning exons 1, 7 and 18 of *APP* (Rovelet-Lecrux et al., 2006) to which we added two supplemental amplicons spanning exons 9 and 10 of *PSEN1*. Primers are available upon request. DNA fragments were then separated on an ABI 3500 genetic analyzer. The fluorescence profiles were analyzed using the GeneMapper software 5 (Applied Biosystems).

### 2.3. Breakpoint sequencing

After delineation of the deletion by QMPSF using amplicons located in introns 8 and 10 of *PSEN1*, the rearranged allele was PCR-amplified and Sanger sequenced using the following primers: F-5'-AGTGTATTGCCTGCCTGGTTTC-3' and R-5'-CTGGCT-GTTGCTGAGGCTTT-3'. DNA sequences were run on an ABI 3500 genetic analyzer and analyzed using the Sequencing Analysis Software 6 (Applied Biosystems).

### 2.4. RT-PCR analysis

Total RNA was extracted from whole blood collected into PAXgene Blood RNA tubes using the PAXgene Blood RNA kit (Qiagen, Hilden, Germany). RNA was reverse-transcribed using the Verso cDNA synthesis kit (Fisher Scientific, Illkirch, France) with a blend of random hexamers and anchored oligo-dT (3:1). The cDNA were then PCR-amplified using the following primers: F-5'-GTGGCT-GTTTTGTGTCCGAA-3' and R-5'-TGGTTGT-GTTCAGTCTCCA-3'. Fragments were separated on a 2% agarose gel and extracted using the Nucleospin gel and PCR clean up (Macherey-Nagel, Düren, Germany). The resulting PCR products were sequenced and run on an ABI 3500 genetic analyzer and analyzed using the Sequencing Analysis Software 6 (Applied Biosystems).

### 2.5. Antibodies

Mouse PEN-2 was detected using a polyclonal antibody (B126.2, RRID:AB\_2571520) and mouse Nicastrin was detected using the monoclonal antibody 9C3 (RRID:AB\_2571521), as previously described (Chávez-Gutiérrez et al., 2012). Human PSEN1-NTF (RRID:AB\_11215630) and PSEN1-CTF (RRID:AB\_95175) were detected using monoclonal antibodies MAB1563 (Chemicon) and MAB5232 (Millipore/Chemicon), respectively. AICD was detected using the APP C-terminus antibody B63.3.

### 2.6. Generation of stable cell lines

Psen1/Psen2 deficient (*PSEN1/2*<sup>-/-</sup>) mouse embryonic fibroblasts (MEFs) (Herreman et al., 2000) were cultured in Dulbecco's modified Eagle's medium/F-12 containing 10% fetal bovine serum. MEFs were transduced using pMSCV-puro, a replication defective recombinant retroviral expression system (Takara Bio Inc.) harboring cDNA inserts coding for human wild-type (WT) or mutant PSEN1s, carrying either the deletion of exon 9 ( $\Delta$ 9-PSEN1) or the deletion of exons 9 and 10 ( $\Delta$ 9-10-PSEN1). Stable cell lines were selected using 5  $\mu$ g/mL puromycin (Sigma-Aldrich).

### 2.7. Expression and purification of APP-C99-3xFLAG

Substrate purification was performed as previously described (Chávez-Gutiérrez et al., 2008). Purity was assessed by SDS-PAGE followed by Coomassie staining (GelCode reagent, Pierce).

### 2.8. In vitro activity assays using detergent extracted $\gamma$ -secretase complexes

*In vitro* activity assays were performed as previously described (Chávez-Gutiérrez et al., 2008) with minor modifications. Briefly, MEF's microsomal fractions were solubilized in 1% CHAPSO. *In*

*in vitro* reactions were carried out in 20 mM Pipes (pH 7.4), 0.15 M NaCl, 0.2 M Sucrose, 1 mM EGTA, 1 × EDTA-free complete protease inhibitors (Roche), 2.5% DMSO, 1% phosphatidylcholine and 0.25% CHAPSO. Reactions were incubated for 4 h at 37 °C with 1.75 μM APP-C99-3xFLAG substrate. The γ-secretase inhibitor “Inhibitor X” was purchased from EMD Millipore.

### 2.9. Cell-based assay and enzyme-linked ImmunoSorbent assay

WT and mutant *PSEN1*-rescued MEF cell lines were transduced with the recombinant adenovirus Ad5/CMV-APP bearing human WT-APP695 as previously described (Chávez-Gutiérrez et al., 2008). Medium was refreshed seven hours after transduction and extracellular media were collected 24 h later. Aβ38, Aβ40 and Aβ42 levels in extracellular media were quantified on Multi-Spot 96-well plates precoated with anti-Aβ38, -Aβ40 and -Aβ42 using MSD technologies (Haryana, India) as previously described (Szaruga et al., 2015). Aβ43 levels were measured using an Aβ43 ELISA kit (IBL, Mannedorf, Switzerland).

## 3. Results

### 3.1. Detection and validation of a *PSEN1* exon 9/10 deletion

Apart from a recurrent duplication of the *MAPT* locus and a deletion of the *ABCA7* gene (Le Guennec et al., 2016b), the analysis CNVs by CANOES in the 522 EOAD cases and 584 controls revealed a deletion of exon 9 and 10 of *PSEN1* in patient EXT-313-001. QMPSF experiments confirmed the presence of the genomic deletion (Fig. 1A). Additional QMPSF assays were designed to delineate the boundaries of the deletion, using amplicons located in introns 8 and 10 of *PSEN1*. On the basis of these results, PCR amplification and sequencing of the deleted fragment showed that the size of the deletion was 10.1 kb and that the two breakpoints were situated within a 24 bp homologous sequence located in introns 8 and 10, 1 kb distal from exon 9 and 3.4 kb proximal from exon 10, respectively (Fig. 1B). *In silico* analysis, using the Repeat Masker program (<http://www.repeatmasker.org/>), revealed that the rearrangement resulted from a homologous Alu-mediated recombination, the Alu elements spanning the breakpoints deriving from the Alu Y and Alu-Sx subfamily. Skipping of exons 9 and 10 of *PSEN1* is predicted to result in an in-frame deletion, but introduces a S290 W missense mutation at the aberrant exon 8–11 junction. RT-PCR experiments on RNA extracted from patient's blood confirmed the expression of the deleted allele (Fig. 1C), approximately to the same level as the wild-type allele (Fig. 1D). Taken together, we retained the following mutation nomenclature: Chr14(GRCh37):g.73671948\_73682054del; NM\_000021.3(PSEN1):c.869-1146\_1130-1780del; p.Ser290\_Arg377delinsTrp; r.869\_1129del, and the following Alu sequence identity: Intron8 g.73671924\_g.73671947 Intron10 g.73682031\_g.73682054.

### 3.2. Case report

The patient presented with progressive cognitive decline associated with spastic paraparesis from the age of 56 years. He had no personal medical history. His mother died after a ten-year course of dementia starting from the age of 50 years. No biomaterials were available for the relatives. Upon examination at age 59, the patient presented spasticity of the lower limbs with decreased motor strength of the left lower limb, pyramidal signs in all four limbs but no upper limb motor deficit. Dystonia as well as extrapyramidal hypertonia of

the left hand were noted. Mini Mental State Examination scored 22/30, and he presented severe episodic memory impairment characterized by failure to encode words during the assessment of the free and cued selective reminding test. Frontal assessment battery revealed executive function disabilities (score was 12/18) CSF was obtained by lumbar puncture and collected and stored in polypropylene tubes. All three CSF AD biomarkers were consistent with an AD signature: decreased Aβ42 (153 pg/mL,  $N > 700$ ), increased Total-Tau (414 pg/mL,  $N < 350$ ) and Phosphorylated-Tau (64 pg/mL,  $N < 60$ ). Cerebral MRI showed severe atrophy predominantly involving the occipital-parietal bilateral cortex in T1-weighted sequences. A similar volume loss was found in the medial temporal structures. The cerebral fluorodesoxyglucose positron emission tomography (FDG-PET) revealed a severe hypometabolism affecting the frontal and occipito-parietal bilateral cortex. No arguments were found for a cerebrovascular disease or cerebral amyloid angiopathy, as only rare non-specific periventricular hyperintensities were present in the FLAIR sequence and T2\*-weighted images did not reveal any micro or macro hemorrhage.

### 3.3. Deletion of exons 9-10 drastically impacts the endopeptidase and carboxypeptidase-like activities of *PSEN1*

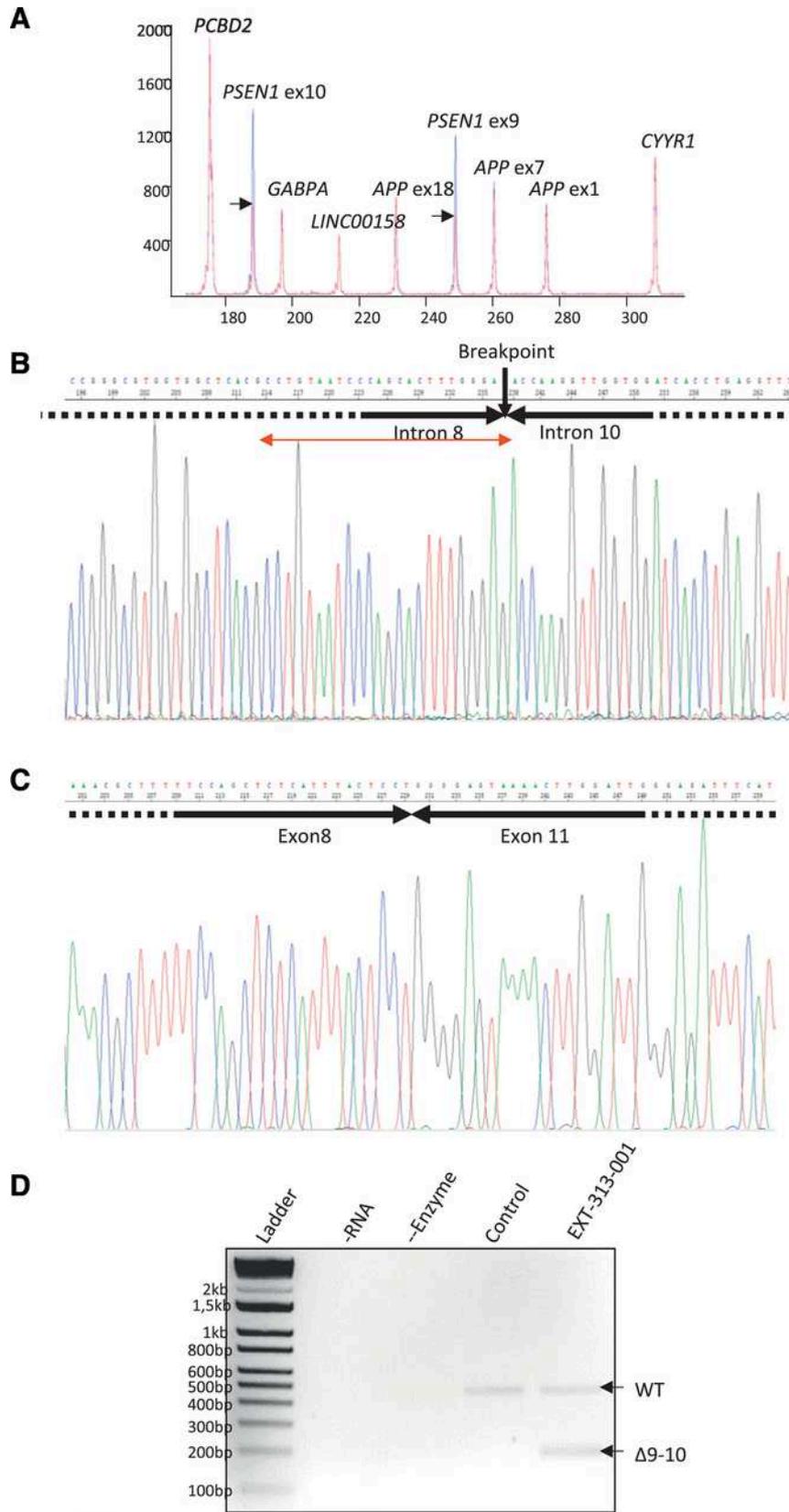
To analyze the effect(s) of the deletion of exons 9 and 10 of *PSEN1* (hereafter referred to as Δ9-10 deletion) on γ-secretase activity, *Psen1*<sup>2<sup>-/-</sup></sup> MEFs were rescued with human WT or Δ9-10 *PSEN1*. For comparison purposes, we simultaneously analyzed the well-characterized Δ9 *PSEN1* deletion (Chávez-Gutiérrez et al., 2012) and used the dKO as a negative control.

Apart from PSEN, a mature and active γ-secretase complex is made up of three additional subunits: Nicastrin (NCT), PSEN enhancer 2 (Pen-2) and anterior pharynx 1 (Aph1).

During the assembly and maturation of the γ-secretase complex, PSEN undergoes endoproteolytic cleavage within exon 9, releasing the N- and C-terminal fragments (Lee et al., 1997). As Δ9, the Δ9-10 *PSEN1* is not processed and only full-length mutant *PSEN1* could be observed on western blot (Fig. 2A). The *PSEN1*-CTF antibody did not stain the Δ9-10 mutant because it targets the *PSEN1* loop, encompassing exons 9 and 10. However, the presence of mature Nicastrin (NCT) and Pen-2 at similar levels in the WT and mutant conditions strongly indicates that the mutant Δ9-10 *PSEN1* still reconstitutes mature γ-secretase complexes.

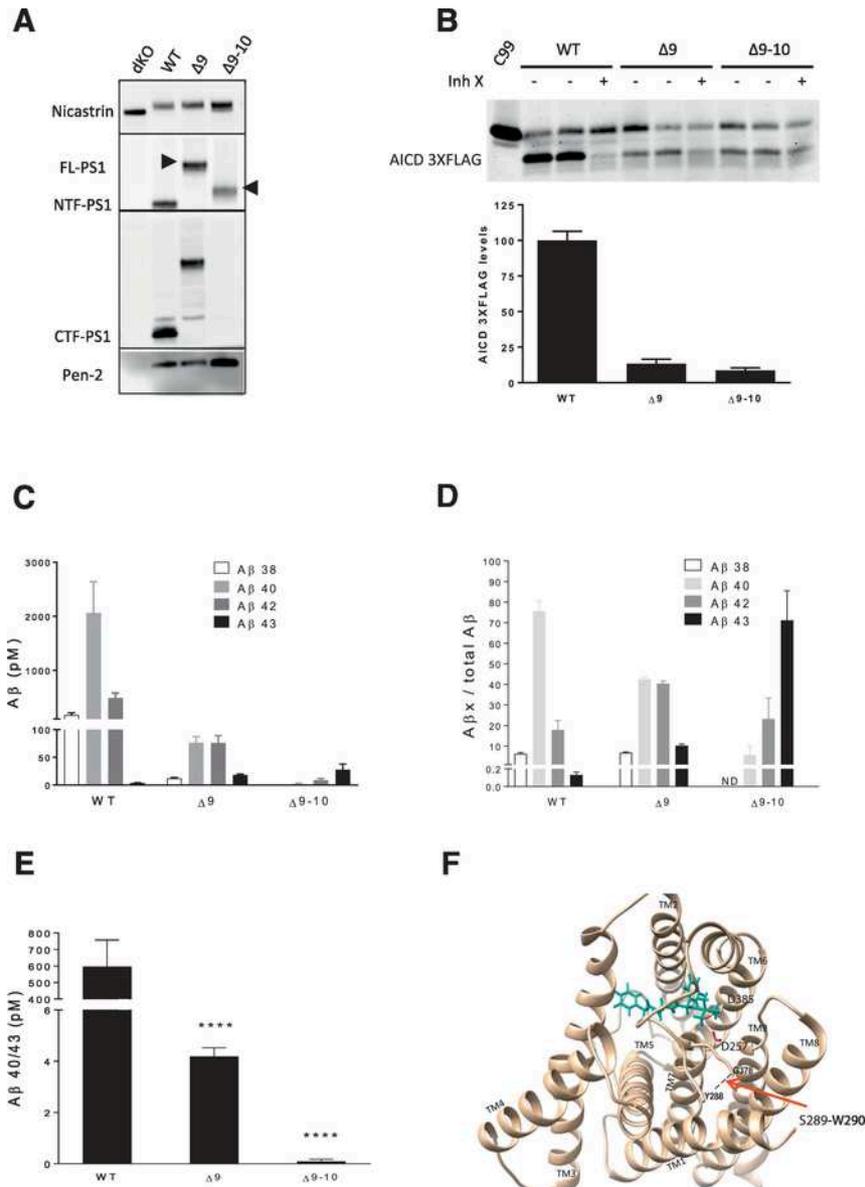
To assess the effects of the Δ9-10 deletion, the endopeptidase activity was evaluated in *in vitro* γ-secretase activity assays, using CHAPSO-solubilized membranes prepared from the WT and mutant MEF lines as source of enzyme, and purified APP-C99-3XFLAG as substrate. Western blot analysis revealed a strong decrease of AICD-3XFLAG staining with both Δ9 and Δ9-10 *PSEN1* compared to WT (Fig. 2B). These data indicate that both Δ9 and Δ9-10 deletions severely compromise the global endopeptidase protease activity.

With regard to the carboxypeptidase-like γ-secretase function, its efficiency was determined in cell-based experiments by ELISA measurement of secreted Aβ isoforms. Expression of both Δ9 and Δ9-10 *PSEN1* resulted in dramatic reductions of Aβ38, Aβ40 and Aβ42 but increased Aβ43 peptide levels compared to WT (Fig. 2C). Remarkably, the exon 9-10 deletion displayed a higher impact on the carboxypeptidase activity than the Δ9 variant, with absence of detectable Aβ38 peptide and stronger reductions in Aβ40 and Aβ42 levels. Mutant Aβ product profiles highlight Aβ43 as the major secreted peptide from the Δ9-10 *PSEN1* cell line (Fig. 2D). The extreme effect of the Δ9-10 *PSEN1* deletion on the carboxypeptidase γ-secretase efficiency is better appreciated while analyzing the Aβ40/Aβ43 ratio, which in-



**Fig. 1.** Validation of the *PSEN1* genomic deletion overlapping exon 9 and 10. **A.** Quantitative Multiplex PCR of Short Fluorescent fragments (QMPSF). The electropherogram of the patient (red) was superposed to a control subject (blue) by adjusting peak height to the control amplicon *PCBD2*. The vertical axis shows fluorescence in arbitrary units and the horizontal axis represents the size of the different amplicons in base pairs. Arrows indicate the heterozygous deletion detected by a 50% decrease of the corresponding peaks. **B.**

Sanger sequencing of the genomic deleted allele. The breakpoints are located into intron 8 and 10. Double red arrow indicates the 24 base pairs homologous sequence mediating the Alu recombination. C. Sanger sequencing of the cDNA from the deleted allele. D. Amplification of WT and  $\Delta 9$ -10 alleles of PSEN1 by RT-PCR in control individual and patient EXT-313-001. Reverse transcriptions without RNA (- RNA) or polymerase (- enzyme) were used as controls. Arrows indicate the WT and  $\Delta 9$ -10 cDNAs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 2.** Functional analysis of the endopeptidase and carboxypeptidase activities of  $\gamma$ -secretase. A. Western Blot of  $\gamma$ -secretase components. Nicastrin, PSEN1-NTF, PSEN1-CTF and Pen-2 protein levels in *PSEN1*<sup>-/-</sup> MEFs transduced to express human WT,  $\Delta 9$  and  $\Delta 9$ -10 PSEN1. Arrowheads indicate full-length forms of  $\Delta 9$  and  $\Delta 9$ -10 PSEN1. B. Analysis of endopeptidase cleavage. Measurements of AICD level for WT,  $\Delta 9$  and  $\Delta 9$ -10 PSEN1  $\gamma$ -secretase complex after *in vitro* activity assay in presence or absence of an active site  $\gamma$ -secretase Inhibitor (InhX). C. Analysis of secreted A $\beta$ 38, A $\beta$ 40, A $\beta$ 42 and A $\beta$ 43 in extracellular media by ELISA. D. The corresponding A $\beta$ <sub>x</sub>/total A $\beta$  ratio and E. the corresponding A $\beta$ 40/A $\beta$ 43 ratio, as an indication of  $\gamma$ -secretase processivity. All experiments were repeated three to eight times. Graphs show mean  $\pm$  SD and statistical significance was tested with one way ANOVA and Dunnett's post-test \*\*\*\*,  $P < 0.0001$ , \*\*\*  $P < 0.001$ . F. Presenilin 3D structure in complex with the inhibitor DAPT (blue) (PDB code: 5FN2). Side chains of active site Asp257 and Asp385 are shown and transmembrane (TM) helices are indicated. Tyr 288 (Y288) and Gly378 (G378) are connected by Ser289 and Trp290 (S289-W290) in the mutant  $\Delta 9$ -10 PSEN1 (indicated by red arrow). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

forms about the efficiency of the 4th catalytic cleavage that converts A $\beta$ 43 into A $\beta$ 40 (Fig. 2E).

#### 4. Discussion

While searching for copy-number variations using exome sequencing data from patients with EOAD, we identified a novel ge-

netic deletion overlapping exons 9 and 10 of the *PSEN1* gene, resulting from an Alu-mediated recombination. This in-frame deletion led to the production of a shorter isoform of PSEN1 lacking exons 9 and 10, which was found to be expressed in patient's blood at a similar level compared to the WT PSEN1. The  $\Delta 9$ -10 PSEN1 variant misses the largest intracellular loop that connects the transmembrane

domains VI and VII that contain the catalytic aspartates of the  $\gamma$ -secretase complex, D257 and D384.

The patient carrying the  $\Delta 9$ -10 deletion presented progressive cognitive decline characterized by impaired episodic memory associated with frontal clinical and imaging signs. Evidence of an AD pathophysiological process was provided by all three CSF AD biomarkers. Such situation highlights the importance of exploring AD CSF biomarkers for any patient with a behavioral fronto-temporal dementia according to current criteria (Rascovsky et al., 2011). Cognitive impairment was associated with spastic paraparesis, a common feature observed in patients carrying deletions of the single exon 9 of *PSEN1* and usually associated with cotton-wool plaques at neuropathological examination. The patient had a disease onset at 56 years, while his mother was affected at age 50. In this family, ages of onset are within the same ranges as the *PSEN1*  $\Delta 9$  mutation carriers (Karlstrom et al., 2008).  $\Delta 9$ -associated phenotypes are however known to be highly variable (Karlstrom et al., 2008), even between individuals from the same pedigree (Smith et al., 2001).

To assess the consequences of the novel *PSEN1* deletion on the function of the  $\gamma$ -secretase complex, we performed *in vitro* cell-free and cell-based  $\gamma$ -secretase activity assays.

Similar to the  $\Delta 9$  form, the lack of exons 9 and 10 precluded the auto-endoproteolytic cleavage of *PSEN1* into NTF and CTF. However, the presence of Pen-2 and mature NCT in both  $\Delta 9$  and  $\Delta 9$ -10 *PSEN1*-mutant  $\gamma$ -secretase complexes is a strong indication of complex assembly and maturation.

Subsequent analysis of AICD and A $\beta$  levels revealed that the  $\Delta 9$ -10 *PSEN1*-mutant strongly impairs the endopeptidase activity and drastically reduces the efficiency of the carboxypeptidase-like cleavages. The effects are indicative of a dysfunction of the  $\gamma$ -secretase complex (Chávez-Gutiérrez et al., 2012). Interestingly, the *PSEN1* -  $\Delta 9$  and  $\Delta 9$ -10 mutants impact similarly the endopeptidase activity, but differ regarding their effect on carboxypeptidase-like activity.

The  $\Delta 9$ -10 allele induces a prominent shift towards production of the longer A $\beta$ 43 peptide as demonstrated by the decreased A $\beta$ 40/A $\beta$ 43 ratio in cell-based analyses. In fact, A $\beta$ 43 is the major peptide secreted from the mutant. A $\beta$ 43 has been shown to be potently amyloidogenic and neurotoxic *in vivo* (Saito et al., 2011). Recent data functionally evaluating other extremely inactivating AD-linked *PSEN1* variations show that, as the  $\Delta 9$ -10 mutant, these pathogenic variants elevate A $\beta$ 43 secretion compared to the WT allele, supporting the idea that the A $\beta$ 43 peptide is a disease-relevant specie, even at low concentrations (Veugelen et al., 2016).

This work further highlights the involvement of deletions hitting the hydrophilic loop of *PSEN1*, mainly encoded by exons 9 and 10, in EOAD determinism.

The pathogenic  $\Delta 9$  deletion leads to decreased endopeptidase activity, associated with a change of carboxypeptidase-like activity illustrated by an increased A $\beta$ 42/A $\beta$ 38 ratio (Chávez-Gutiérrez et al., 2012), while the artificial deletion of exon 10 alone affected the carboxypeptidase-like activity of  $\gamma$ -secretase by increasing A $\beta$ 42 levels, without impairing the endopeptidase function (Wanngren et al., 2010). Thus, our data suggest a synergistic effect for the  $\Delta 9$ -10 deletion that translates into very low global endopeptidase levels and drastic changes in A $\beta$  profiles. However, our studies cannot exclude that the effects are fully or partially driven by the missense S290 W variation at the exon 8-11 junction, as previous work suggested that the effect linked to the deletion of *PSEN1* exon 9 was mostly due to the presence of the missense S290C variation at the aberrant exon 8-10 junction (Steiner et al., 1999).

Addressing the mechanistic bases of the differential effects observed for the  $\Delta 9$ ,  $\Delta 10$  and  $\Delta 9$ -10 mutants on  $\gamma$ -secretase function

may offer key insights into how the complex protease is regulated. Of note, the recent  $\gamma$ -secretase structure in complex with the inhibitor DAPT (Bai et al., 2015) shows the Tyr288 and the Gly378 at about 6 Å from each other; a linker comprised of Ser289 and Trp290 connects these two residues in the  $\Delta 9$ -10 *PSEN1* structure (see Fig. 2F). Thus, this particular structure puts closely in the space the exons 8 and 11, and may be used to understand the effects of the  $\Delta 9$ -10 mutation on the protease complex.

In conclusion, we describe in an EOAD patient a novel deletion of the *PSEN1* gene encompassing both exons 9 and 10, thus resulting in a complete removal of the largest intracellular hydrophilic loop. The large deletion, encompassing 87 amino acids, does not affect the assembly of a mature protease complex but has an extreme impact on its endopeptidase activity. Interestingly, the mutant carboxypeptidase-like activity is also strongly impaired and the deleterious mutant effect leads to an incomplete digestion of long A $\beta$  peptides and enhances the production of toxic, aggregation-prone A $\beta$ 43. In fact, A $\beta$ 43 is the main peptide released from the  $\Delta 9$ -10 *PSEN1*. These results supports the idea that, as encountered in most EOAD-linked *PSEN1* mutations (Chávez-Gutiérrez et al., 2012), AD-linked *PSEN1* mutations consistently result in impaired carboxypeptidase-like activity and most importantly underlines a potential contribution of A $\beta$ 43 to AD pathogenesis.

Genetically, our work (i) further extends the mutation spectrum of EOAD, (ii) demonstrates that exons 9 and 10, that are devoid of pathological missense mutations, are sensitive to deletions, and (iii) underlines the need to independently screen deletions of exons 9 and 10 in familial EOAD cases.

## Competing interest

We report no competing interest.

## Acknowledgements

This study was funded by grants from the Clinical Research Hospital Program from the French Ministry of Health (GMAJ, PHRC 2008/067), the CNR-MAJ and the JPND PERADES (ANR-13-JPRF-0001-04). Kilan Le Guennec benefited from an EMBO short-term fellowship. This work was also funded by the Stichting Alzheimer Onderzoek (SAO), the Fund for Scientific Research, Flanders, the KU Leuven, a Methusalem grant from the KU Leuven, the Flemish Government and Interuniversity Attraction Poles Program of the Belgian Federal Science Policy Office.

## References

- Backenroth, D., Homsy, J., Murillo, L.R., Glessner, J., Lin, E., Brueckner, M., Lifton, R., Goldmuntz, E., Chung, W.K., Shen, Y., 2014. CANOES: detecting rare copy number variants from whole exome sequencing data. *Nucleic Acids Res.* 42, e97.
- Bai, X., Rajendra, E., Yang, G., Shi, Y., Scheres, S.H.W., 2015. Sampling the conformational space of the catalytic subunit of human  $\gamma$ -secretase. *elife* 4.
- Chávez-Gutiérrez, L., Tolia, A., Maes, E., Li, T., Wong, P.C., de Strooper, B., 2008. Glu(332) in the Nicastrin ectodomain is essential for gamma-secretase complex maturation but not for its activity. *J. Biol. Chem.* 283, 20096–20105.
- Chávez-Gutiérrez, L., Bammens, L., Benilova, I., Vandersteen, A., Benurwar, M., Borgers, M., Lismont, S., Zhou, L., Van Cleynbreugel, S., Esselmann, H., Wiltfang, J., Serneels, L., et al., 2012. The mechanism of  $\gamma$ -secretase dysfunction in familial Alzheimer disease. *EMBO J.* 31, 2261–2274.
- Herreman, A., Serneels, L., Annaert, W., Collen, D., Schoonjans, L., De Strooper, B., 2000. Total inactivation of gamma-secretase activity in presenilin-deficient embryonic stem cells. *Nat. Cell Biol.* 2, 461–462.
- Hiltunen, M., Helisalmi, S., Mannermaa, A., Alafuzoff, I., Koivisto, A.M., Lehtovirta, M., Pirskanen, M., Sulkava, R., Verkkoniemi, A., Soininen, H., 2000. Identification of a novel 4.6-kb genomic deletion in presenilin-1 gene which results in exclu-

- sion of exon 9 in a Finnish early onset Alzheimer's disease family: an Alu core sequence-stimulated recombination?. *Eur. J. Hum. Genet.* 8, 259–266.
- Jurisch-Yaksi, N., Sannerud, R., Annaert, W., 2013. A fast growing spectrum of biological functions of  $\gamma$ -secretase in development and disease. *Biochim. Biophys. Acta* 1828, 2815–2827.
- Karlstrom, H., Brooks, W.S., Kwok, J.B.J., Broe, G.A., Kril, J.J., McCann, H., Halliday, G.M., Schofield, P.R., 2008. Variable phenotype of Alzheimer's disease with spastic paraparesis. *J. Neurochem.* 104, 573–583.
- Le Guennec, K., Nicolas, G., Quenez, O., Charbonnier, C., Wallon, D., Bellenguez, C., Grenier-Boley, B., Rousseau, S., Richard, A.-C., Rovelet-Lecrux, A., Bacq, D., Garnier, J.-G., et al., 2016. ABCA7 rare variants and Alzheimer disease risk. *Neurology* 86, 2134–2137.
- Le Guennec, K., Quenez, O., Nicolas, G., Wallon, D., Rousseau, S., Richard, A.-C., Alexander, J., Paschou, P., Charbonnier, C., Bellenguez, C., Grenier-Boley, B., Lechner, D., et al., 2016. 17q21.31 duplication causes prominent tau-related dementia with increased MAPT expression. *Mol. Psychiatry*
- Lee, M.K., Borchelt, D.R., Kim, G., Thinakaran, G., Slunt, H.H., Ratovitski, T., Martin, L.J., Kittur, A., Gandy, S., Levey, A.I., Jenkins, N., Copeland, N., et al., 1997. Hyperaccumulation of FAD-linked presenilin 1 variants in vivo. *Nat. Med.* 3, 756–760.
- McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., Mohs, R.C., Morris, J.C., et al., 2011. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement. J. Alzheimers Assoc.* 7, 263–269.
- Nicolas, G., Charbonnier, C., Wallon, D., Quenez, O., Bellenguez, C., Grenier-Boley, B., Rousseau, S., Richard, A.-C., Rovelet-Lecrux, A., Le Guennec, K., Bacq, D., Garnier, J.-G., et al., 2016. SORL1 rare variants: a major risk factor for familial early-onset Alzheimer's disease. *Mol. Psychiatry* 21, 831–836.
- Nicolas, G., Wallon, D., Charbonnier, C., Quenez, O., Rousseau, S., Richard, A.-C., Rovelet-Lecrux, A., Coutant, S., Le Guennec, K., Bacq, D., Garnier, J.-G., Olasso, R., et al., 2016. Screening of dementia genes by whole-exome sequencing in early-onset Alzheimer disease: input and lessons. *Eur. J. Hum. Genet.* 24, 710–716.
- Perez-Tur, J., Froelich, S., Prihar, G., Crook, R., Baker, M., Duff, K., Wragg, M., Busfield, F., Lendon, C., Clark, R.F., 1995. A mutation in Alzheimer's disease destroying a splice acceptor site in the presenilin-1 gene. *Neuroreport* 7, 297–301.
- Prihar, G., Verkoniemi, A., Perez-Tur, J., Crook, R., Lincoln, S., Houlden, H., Somer, M., Paetau, A., Kalimo, H., Grover, A., Myllykangas, L., Hutton, M., et al., 1999. Alzheimer disease PS-1 exon 9 deletion defined. *Nat. Med.* 5, 1090.
- Rascovsky, K., Hodges, J.R., Knopman, D., Mendez, M.F., Kramer, J.H., Neuhaus, J., van Swieten, J.C., Seelaar, H., Dopper, E.G.P., Onyike, C.U., Hillis, A.E., Josephs, K.A., et al., 2011. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain J. Neurol.* 134, 2456–2477.
- Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerrière, A., Vital, A., Dumanchin, C., Feuillette, S., Brice, A., Vercelletto, M., Dubas, F., Frebourg, T., et al., 2006. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat. Genet.* 38, 24–26.
- Rovelet-Lecrux, A., Charbonnier, C., Wallon, D., Nicolas, G., Seaman, M.N.J., Pottier, C., Breusegem, S.Y., Mathur, P.P., Jenardhanan, P., Le Guennec, K., Mukadam, A.S., Quenez, O., et al., 2015. De novo deleterious genetic variations target a biological network centered on A $\beta$  peptide in early-onset Alzheimer disease. *Mol. Psychiatry* 20, 1046–1056.
- Saito, T., Suemoto, T., Brouwers, N., Slegers, K., Funamoto, S., Mihira, N., Matsuba, Y., Yamada, K., Nilsson, P., Takano, J., Nishimura, M., Iwata, N., et al., 2011. Potent amyloidogenicity and pathogenicity of A $\beta$ 43. *Nat. Neurosci.* 14, 1023–1032.
- Sato, S., Kamino, K., Miki, T., Doi, A., Ii, K., St George-Hyslop, P.H., Oghihara, T., Sakaki, Y., 1998. Splicing mutation of presenilin-1 gene for early-onset familial Alzheimer's disease. *Hum. Mutat. Suppl.* 1, S91–S94.
- Smith, M.J., Kwok, J.B., McLean, C.A., Kril, J.J., Broe, G.A., Nicholson, G.A., Cappai, R., Hallupp, M., Cotton, R.G., Masters, C.L., Schofield, P.R., Brooks, W.S., 2001. Variable phenotype of Alzheimer's disease with spastic paraparesis. *Ann. Neurol.* 49, 125–129.
- Steiner, H., Romig, H., Grim, M.G., Philipp, U., Pesold, B., Citron, M., Baumeister, R., Haass, C., 1999. The biological and pathological function of the presenilin-1 Deltaexon 9 mutation is independent of its defect to undergo proteolytic processing. *J. Biol. Chem.* 274, 7615–7618.
- Szaruga, M., Veugelen, S., Benurwar, M., Lismont, S., Sepulveda-Falla, D., Lleo, A., Ryan, N.S., Lashley, T., Fox, N.C., Murayama, S., Gijzen, H., De Strooper, B., et al., 2015. Qualitative changes in human  $\gamma$ -secretase underlie familial Alzheimer's disease. *J. Exp. Med.* 212, 2003–2013.
- Takami, M., Nagashima, Y., Sano, Y., Ishihara, S., Morishima-Kawashima, M., Funamoto, S., Ihara, Y., 2009. gamma-Secretase: successive tripeptide and tetrapeptide release from the transmembrane domain of beta-carboxyl terminal fragment. *J. Neurosci.* 29, 13042–13052.
- Veugelen, S., Saito, T., Saido, T.C., Chávez-Gutiérrez, L., De Strooper, B., 2016. Familial Alzheimer's disease mutations in presenilin generate amyloidogenic A $\beta$  peptide seeds. *Neuron* 90, 410–416.
- Wanngren, J., Fränberg, J., Svensson, A.I., Laudon, H., Olsson, F., Winblad, B., Liu, F., Näslund, J., Lundkvist, J., Karlström, H., 2010. The large hydrophilic loop of presenilin 1 is important for regulating gamma-secretase complex assembly and dictating the amyloid beta peptide (A $\beta$ ) Profile without affecting Notch processing. *Biol. Chem.* 285, 8527–8536.

### III. Discussion

#### 1. Etude d'association à l'échelle d'une liste de gènes, d'un réseau, et comment prioriser certains gènes ?

Les études d'association réalisées au laboratoire nous ont permis de mettre en évidence un enrichissement de variants rares perte de fonction dans les gènes *SORL1* et *ABCA7*. Ces stratégies d'agrégation à l'échelle du gène nous ont permis d'augmenter la puissance statistique.

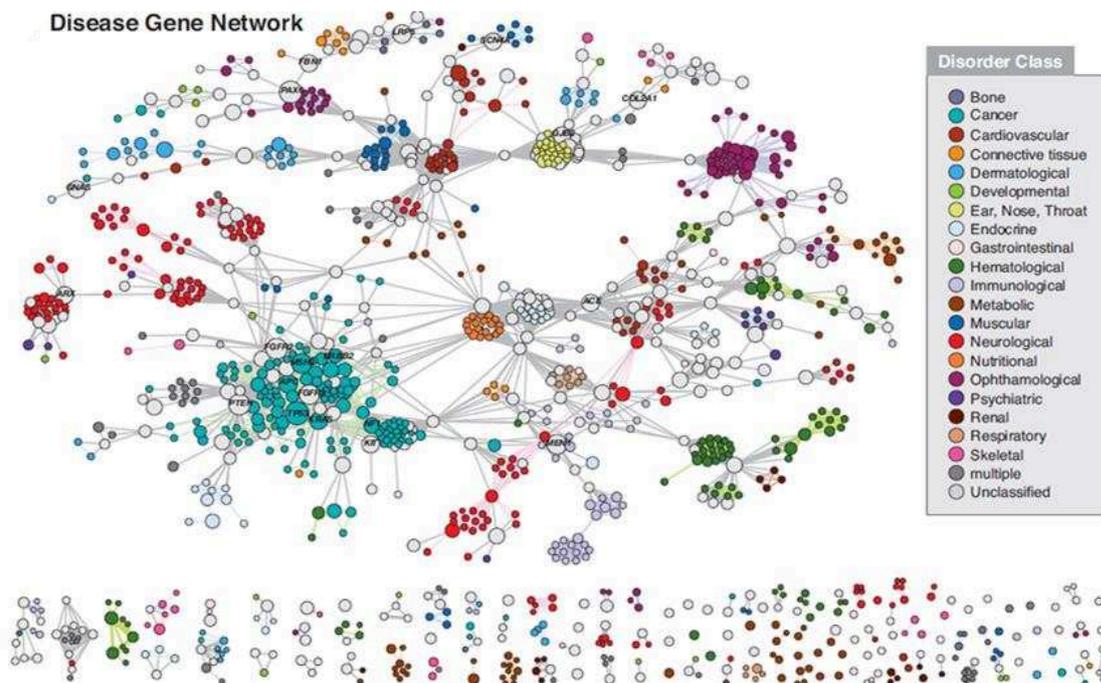
Afin de détecter de nouveaux variants impliqués dans la MA, nous avons voulu augmenter la puissance statistique en augmentant la taille de notre échantillon de cas et de contrôles. Nous avons ainsi récemment réalisé, en collaboration avec l'équipe du Docteur Jean-Charles Lambert basée à Lille, une étude d'association à partir de 1 779 cas (incluant 852 formes précoces et 927 formes tardives) et 1 273 contrôles. Nous avons alors pu mettre en évidence un enrichissement significatif d'un point de vue de l'exome chez les EOAD de variants rares PTV et faux-sens SD dans trois gènes : *SORL1*, *ABCA7* et *TREM2* (Manuscrit soumis à *Molecular Neurodegeneration*).

Pour essayer d'augmenter la puissance statistique, nous avons également essayé d'agréger les variants à l'échelle d'une liste de gènes. Pour cela, nous avons donc établi une liste de gènes à partir d'une bibliographie extensive centrée sur le peptide A $\beta$ . Grâce à cette liste, nous avons pu mettre en évidence un enrichissement de variants *de novo* chez des cas sporadiques de MA jeune (Rovelet-Lecrux et al., 2015). Nos études sur les CNVs nous a également permis de détecter un enrichissement chez les cas de remaniements intersectant cette liste de gènes. Nous avons également voulu tester l'enrichissement de variants nucléotidiques rares touchant la liste de gènes. L'analyse n'a pas permis d'obtenir un résultat significatif même lorsque nous restreignons l'analyse à des sous catégories de la liste de gènes ou quand nous ne considérons que les variants PTV et/ou SD. Cela pourrait être liée au fait que certains variants prédits comme perte de fonction peuvent être protecteurs en fonction du rôle physiologique de la protéine étudiée. À titre d'exemple, une variation perte de fonction dans le gène *SORL1*, impliquée dans la voie de recyclage d'APP pour prévenir la production d'A $\beta$ , serait considérée comme à risque alors qu'une variation perte de fonction dans le gène *BACE1*, impliqué dans le premier clivage d'APP amenant à la production d'A $\beta$ , serait *a priori* protectrice. Il est donc nécessaire de pouvoir prédire quelles sont les variations protectrices et celles à risque.

Certains tests statistiques permettent de prendre en compte ce type de données comme le test SKAT. Toutefois, l'analyse des données grâce à ce test n'a pas permis de détecter un signal d'enrichissement significatif à l'échelle de cette liste de gènes. Cette absence d'association pourrait être expliquée par l'extrême rareté des variations étudiées, celles-ci étant le plus souvent des singletons qui sont plus difficilement pris en charge par cet algorithme.

Cette liste de gènes peut être organisée sous forme d'un réseau en prenant en compte soit les interactions physiques entre les protéines impliquées (PPI : Protéine-Protéine Interaction), soit les interactions génétiques mises en évidence dans des modèles cellulaires ou animaux (Boucher and Jenna, 2013; Vidal et al., 2011). Les réseaux PPI ou de gènes sont classiquement construits à partir de nœuds, qui représentent les protéines ou les gènes, et d'arêtes représentant l'interaction physique ou génétique entre les nœuds. Un *hub* est un nœud qui est situé au centre du réseau construit. Cette centralité est calculée à partir de deux scores : le degré de centralité et l'interférence de centralité. Le degré de centralité représente le nombre d'interactions d'un nœud avec les autres nœuds. L'interférence de centralité représente le nombre de plus court chemins pour relier deux protéines passant par ce nœud. Ces deux scores permettront de placer les différents nœuds soit au centre du réseau soit en périphérie. Les *hubs* sont majoritairement des gènes définis comme essentiels dont l'inactivation est incompatible avec la survie et également des gènes impliqués dans des maladies mendéliennes (Vidal et al., 2011). De manière intéressante, une étude de réseau menée à partir de 1777 gènes impliqués dans 1284 maladies mendéliennes (*Human disease network*) a pu montrer que ces gènes interagissaient fortement entre eux (Figure 27 page 89) (Goh et al., 2007) suggérant qu'un ensemble limité de gènes en interaction sont responsables de l'ensemble des pathologies mendéliennes.

Ces approches à base de réseaux de gènes ont été utilisées pour prédire l'existence de nouveaux gènes reliés à des maladies. Par exemple, il a été montré que la plupart des gènes responsables d'ataxies cérébelleuses peuvent être inclus dans un même réseau PPI et qu'à partir de ce réseau il était possible de générer des listes de nouveaux gènes candidats (Lim et al., 2006). Dans une autre étude de séquençage d'exomes chez des patients atteints de formes héréditaires de paraplégie spastique (HPS), Novarino et collaborateurs ont identifié 18 nouveaux gènes responsables de cette pathologie sur des arguments de ségrégation intrafamiliale ou de récurrence des mutations (Novarino et al., 2014). Grâce à un réseau PPI, les auteurs ont réussi à prouver que ces gènes interagissaient entre eux mais également avec les gènes connus de HPS. Les études de réseaux sont donc importantes pour mettre en évidence les voies biologiques responsables des pathologies humaines.

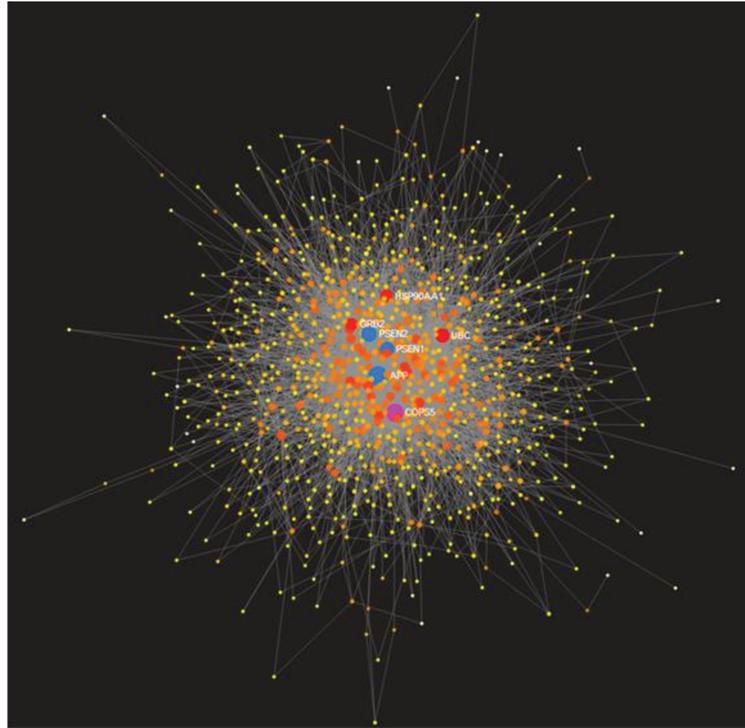
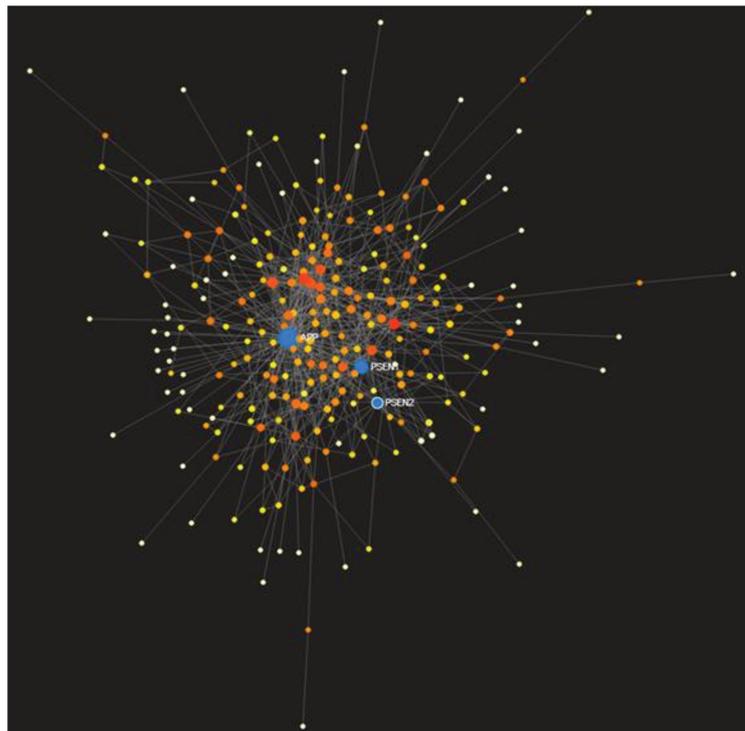


**Figure 27 : Human disease Network**

Figure issue de Goh et al., 2007.

Les gènes responsables de maladies mendéliennes ont été reliés dans un réseau lorsqu'ils étaient impliqués dans la même maladie. 1 377 sur 1 777 gènes étudiés ont ainsi pu être reliés dans plusieurs réseaux dont 903 forment un seul et même réseau.

A partir de la liste de gènes centrée sur le peptide amyloïde, nous avons essayé de créer un réseau PPI à partir du logiciel *Networkanalyst* accessible en ligne (<http://www.networkanalyst.ca/>) (Figure 28 page 90). Ce réseau, où APP occupe une place centrale, avait pour but de mettre en lumière d'autres gènes ou protéines qui avaient pu échapper à notre vigilance bibliographique. Ainsi, le gène *COP55* a été ajouté à cette liste de gènes. La protéine *COP55* faciliterait la production du peptide A $\beta$  en favorisant le clivage  $\beta$ -sécrétase. Cet ajout du gène *COP55* souligne l'imparfaite exhaustivité de la liste de gènes bien que celle-ci soit régulièrement mise à jour. Des études poussées pourraient nous permettre d'identifier des modules importants dans ce réseau qui seraient préférentiellement altérés chez des patients atteints de MA. Cependant, si nous ne travaillons que sur cette liste, nous ne pourrions pas découvrir de nouvelles voies biologiques. Nous essayons actuellement d'agrandir le réseau créé pour voir si celui-ci n'interagirait pas avec d'autres voies biologiques inconnues pour leur implication dans la MA à ce jour.

**A****B**

**Figure 28 : Réseaux A $\beta$**

Ces réseaux ont été réalisés grâce au logiciel NetworkAnalyst à partir de la liste de 342 gènes centrée sur le peptide amyloïde. La taille des nœuds représente leur score de centralité. A, réseau PPI créé en considérant l'ensemble des gènes de la liste ainsi que les interacteurs qui permettent de les relier entre eux. Parmi les 343 protéines, 325 ont ainsi pu être intégrées dans ce réseau de 916 protéines (nœuds) et 4 864 interactions (arêtes). B, réseau PPI créé en considérant uniquement les protéines de la liste. 257/343 gènes interagissent physiquement entre eux. Les protéines APP, PSEN1 et PSEN2 ont été colorées en bleu et la protéine COPS5 mise en lumière grâce à ce réseau en violet.

Dans nos études d'association, l'analyse des variations a été restreinte aux variations perte de fonction et faux-sens SD. Afin d'expliquer la part de composante génétique manquante de la maladie d'Alzheimer, l'étude des variants prédits comme « probablement pathogènes » grâce à des études fonctionnelles pourrait permettre de statuer sur le caractère à risque ou non. Le problème majeur des études fonctionnelles est qu'il faudrait pour chaque gène développer un test fonctionnel approprié afin de tester l'ensemble des variations d'intérêts. Par ailleurs, les variations synonymes ou introniques (leur détection étant limitée dans les études d'exomes) n'ont pas été étudiées. Or, nous savons que de nombreuses variations synonymes ou introniques peuvent affecter l'épissage physiologique du gène et peuvent aboutir à la production de transcrit aberrant. Ces transcrits aberrants peuvent produire des protéines anormales ou introduire une variation perte de fonction dans la séquence conduisant à une absence de production de la protéine ou à une protéine tronquée. Des algorithmes existent permettant de prédire l'effet des variations sur l'épissage ainsi que des tests fonctionnels. L'utilisation de ces outils pourrait permettre de déterminer l'effet de ces variations. Ces variations synonymes ou introniques affectant l'épissage peuvent donc avoir un effet fort sur la protéine et donc participer de manière significative à la part de la composante génétique manquante des maladies.

L'implication des CNVs dans la composante génétique de la MA est également importante. Les variations structurales peuvent (i) être des duplications, des délétions ou encore des inversions (non-délectables en exome), (ii) impliquer l'intégralité du gène ou seulement une partie et (iii) contenir plusieurs gènes. Nous avons obtenu grâce à notre étude d'association un signal nominale significatif à l'échelle de notre liste de gènes. Ce résultat pourrait être amélioré s'il était possible de distinguer les CNVs dépourvus d'effet fonctionnel ou protecteur dans la maladie, des CNVs à risque. A titre d'exemple, nous avons détecté une délétion complète de la cathepsine B (*CTSB*) chez un cas et à l'inverse une duplication complète de ce même gène chez un contrôle. *CTSB* est une enzyme de dégradation du peptide amyloïde et joue donc un rôle protecteur dans la physiopathologie de la MA. Par conséquent, la délétion est potentiellement à risque lorsque la duplication est protectrice. Ce type d'interprétation n'est pas toujours évident à réaliser pour l'ensemble des gènes. Leur interprétation est donc extrêmement complexe. Comme nous avons vu précédemment, Chiang et collaborateurs ont répondu en partie à ces questions en apportant un éclairage sur l'effet des CNVs sur l'expression des gènes (Chiang et al., 2017).

Le séquençage d'exome est limité aux séquences codantes de l'individu représentant moins de 2 % du génome humain. Nous avons donc une vision restreinte de l'ensemble des variations génétiques d'un individu. Le séquençage de génomes de patients atteints de MA semble donc la suite logique

pour identifier la composante génétique manquante. Cependant, l'interprétation des variations détectées reste encore très complexe car la plupart des analyses réalisées jusqu'à aujourd'hui l'ont été à partir des modules fonctionnels que sont les gènes. De récentes études tentent de comprendre l'impact des régions non codantes du génome et leur implication potentielle dans la composante génétique des maladies (pour revue (Zappala and Montgomery, 2016)). Les variations touchant les régions non codantes du génome peuvent notamment avoir des effets délétères lorsqu'elles sont situées sur les 3'UTR ou 5'UTR. Ainsi une variation située dans le 3'UTR d'*APP*, augmentant son expression, avait été décrite au laboratoire chez un patient atteint d'angiopathie amyloïde cérébrale (Nicolas et al., 2016c). La plupart des études de GWAS ont permis d'associer des variants intergéniques avec la maladie. Dans ces études, on suppose que les variants touchent des régions régulatrices de gènes ou sont en déséquilibre de liaison avec un variant situé dans la séquence codante d'un gène. Les variations non codantes peuvent donc également se retrouver sur des éléments régulateurs de l'expression des gènes et ainsi altérer leur fonction (Zappala and Montgomery, 2016). Par exemple, dans le cas de la rétinite pigmentaire à transmission récessive, 4 variants chez un même patient situés dans deux régions génomiques régulant l'expression de *SAMD7* ont été rapportés entraînant une diminution d'expression de ce gène (Van Schil et al., 2016). Afin d'étudier les variations non-codantes du génome sur l'expression des gènes, Melnikov et collaborateurs ont utilisé une technique appelée *Massive Parallel Reporter Assay* (Melnikov et al., 2012). Cette technique permet de mesurer l'effet des variations non codantes sur l'expression d'un gène rapporteur. Ils avaient ainsi pu observer, en utilisant deux régions promotrices, que certaines variations pouvaient diminuer jusqu'à 4 fois l'expression d'un gène (Melnikov et al., 2012). D'autres études ont également utilisé cette technique pour évaluer l'effet des variations non codantes sur certaines régions régulatrices (Patwardhan et al., 2012; Vockley et al., 2015). Ces études ont permis de mettre en évidence que seulement une faible fraction des variations non-codantes peut avoir un effet sur l'expression des gènes.

## 2. La duplication du gène *MAPT* est-elle responsable d'une maladie d'Alzheimer ou d'une tauopathie pure ?

La MA est caractérisée au niveau anatomopathologique par deux types de lésions : les plaques séniles et les dégénérescences neurofibrillaires (DNF). Chez la majorité des individus âgés, les DNFs restent confinées au cortex entorhinal et l'hippocampe. Dans le cerveau des patients atteints de MA, les DNFs suivent une progression stéréotypée entre des neurones connectés synaptiquement pour atteindre les zones corticales influencée par la présence de dépôts diffus d'A $\beta$  dans cette région (Braak and Braak, 1991; Thal et al., 2002). Les études récentes considèrent que les oligomères d'A $\beta$  plutôt que les plaques séniles sont les entités neurotoxiques dans la maladie (Walsh and Selkoe, 2007), notamment en influençant les mécanismes de LTD et LTP (Guntupalli et al., 2016).

L'étude des variations structurales à partir d'un échantillon de formes précoces de MA a permis d'identifier la récurrence de la duplication du locus 17q21.31 chez 4 cas. Les profils des biomarqueurs du LCR sont compatibles avec une pathologie amyloïde. En effet, la diminution des niveaux d'A $\beta$ 42 dans le LCR indique l'accumulation de ce peptide fibrillogène dans le cerveau des patients (Tapiola et al., 2009). Cependant, l'imagerie amyloïde s'est révélée négative. L'analyse neuroanatomopathologique d'un cas supplémentaire a confirmé que cette duplication de *MAPT* était responsable d'une tauopathie complexe avec des DNFs présentes dans l'hippocampe, le cortex entorhinal et le noyau basal de Meynert, sans dépôts amyloïdes (Alexander et al., 2016).

L'absence de détection de dépôts amyloïdes par l'analyse anatomopathologique et l'imagerie amyloïde semble donc contradictoire avec le diagnostic initial et les biomarqueurs. Cependant, les oligomères d'A $\beta$  sont difficiles à détecter par les techniques de PET amyloïde ou de marquages immunohistochimiques qui permettent de détecter principalement les plaques séniles. L'analyse du cerveau de ces patients grâce à des anticorps spécifiques des oligomères d'A $\beta$  tels que NU-1 (Lambert et al., 2007) permettra de révéler des oligomères qui ne sont pas agrégés en plaques séniles ni en dépôts amyloïdes. Dans ce cas, le mécanisme physiopathologique serait en faveur d'une MA. Néanmoins, ces patients pourraient également être atteints d'une tauopathie pure qui ressemblerait à une forme précoce de PART (Crary et al., 2014). PART est une entité neuropathologique définie par la présence de DNFs composées des isoformes 3R et 4R dans le cerveau d'individus âgés au stade I à IV de Braak (Crary et al., 2014). Les DNFs retrouvées chez ces individus âgés sont très similaires aux DNFs observées dans la MA et aux patients porteurs de duplication *MAPT* composées des mêmes isoformes de Tau (Espinoza et al., 2008).

Il faut maintenant valider ces résultats dans de plus larges cohortes et développer des modèles fonctionnels pour déterminer si la duplication 17q21.31 est responsable d'une MA jeune avec donc

l'implication des oligomères solubles d'A $\beta$  ou si elle provoque une tauopathie pure comparable à un PART précoce.

Dans un premier temps, nous rechercherons la duplication du locus 17q21.31 chez des patients atteints d'une démence similaire d'un point de vue clinique et/ou anatomopathologique à nos individus porteurs grâce à des banques de cerveaux. L'un des premiers objectifs sera de déterminer si les patients porteurs présentent des dépôts amyloïdes grâce à des analyses immunohistochimiques et des Western Blot en utilisant des anticorps spécifiques des oligomères d'A $\beta$  tels que NU-1 (Lambert et al., 2007) et celui développé par Kaye et collaborateurs (Kaye et al., 2003).

Les résultats obtenus à partir de l'analyse des oligomères d'A $\beta$  aiguilleront nos recherches au niveau fonctionnel.

Nous utiliserons ensuite des fibroblastes prélevés chez les patients et chez des apparentés sains pour générer des cellules souches pluripotentes induites (iPSC) (Yu et al., 2007). Ces cellules peuvent être différenciées en cellules neuronales et représentent les modèles d'étude les plus proches de la génétique des patients contrairement aux modèles cellulaires ou animaux (pour revue (Zhang et al., 2016b)).

En fonction de la détection ou non d'oligomères d'A $\beta$  dans le cerveau des patients, deux hypothèses seront étudiées.

En l'absence d'oligomères d'A $\beta$  dans le cerveau des patients, nous considérerons que la surexpression de *MAPT* n'est pas dépendante d'A $\beta$  pour déclencher la pathologie et n'est donc pas responsable de MA. La caractérisation de ces neurones permettra d'évaluer l'effet de surexpression de Tau sur la production, la phosphorylation, l'agrégation et la localisation subcellulaire de Tau. Ce modèle permettra également d'évaluer la toxicité synaptique de cette augmentation d'expression de Tau dans la cellule.

En présence d'oligomères, nous essayerons de déterminer si la duplication de Tau a un effet sur la production, agrégation ou toxicité d'A $\beta$  en mesurant les niveaux d'A $\beta$ 40, d'A $\beta$ 42 et des oligomères d'A $\beta$  dans le milieu de culture des cellules iPSC différenciées en neurones grâce à la technique ELISA (Englund et al., 2007).

Les hypothèses validées permettront de déterminer si le mécanisme physiopathologique des duplications du locus 17q21.31 est en faveur d'une MA atypique ou d'une tauopathie pure.

Ces résultats auront un impact sur le diagnostic de MA et la prise en charge thérapeutique des patients porteurs. La duplication du locus 17q21.31 pourrait en effet expliquer une partie des individus SNAP (*Suspected Non-Alzheimer Disease Pathophysiology*). Ces patients présentent un diagnostic clinique de MA, des biomarqueurs en faveur d'une neurodégénérescence de type Alzheimer (niveau important de Tau dans le liquide céphalo-rachidien, atrophie et hypométabolisme du cerveau) mais une imagerie amyloïde négative et représentent environ 12% des patients inclus

dans les essais thérapeutiques dans le cadre de la MA (Jack et al., 2016). La prise en charge thérapeutique de ces patients sera donc différente puisque la majorité des traitements développés à ce jour dans le cadre de la MA sont dirigées contre le peptide amyloïde (Folch et al., 2016).

En conclusion, l'étude d'association à partir de données de séquençage d'exomes nous a permis de révéler une part de la composante génétique de la MA médiée par des variations rares qu'ils s'agissent de SNVs, indels ou encore de CNVs. De nombreuses stratégies d'analyse d'exomes peuvent permettre d'augmenter la puissance statistique pour tenter de révéler de nouveaux facteurs de risque génétiques. Cependant, l'exome représente moins de 2% du génome humain. Le défi futur résidera donc dans l'interprétation des variations génétiques non-codantes détectées à partir de données de séquençage de génomes.

## **Annexes**

Annexe 1 : Liste de gènes centrée sur le peptide amyloïde

Annexe 2 : Article sur l'étude de trios Alzheimer

## Annexe 1. Liste de gènes centrée sur le peptide amyloïde

<i>Gene</i>	ID	Category	Sub category	Protein ID	PPI	PS1 PPI	Bace PPI	GWAS	Prédictions	OMIM	Transmission
<i>APP</i>	351			P05067.3	APP, Aβ	1			2	104300 / 605714	D / D
<i>CANX</i>	821	Processing / trafficking	ER	P27824.2					1	.	.
<i>HSPA5</i>	3309	Processing / trafficking	ER	P11021.2	APP				1	.	.
<i>HYOU1</i>	10525	Processing / trafficking	ER	Q9Y4L1.1					1	.	.
<i>TMCC2</i>	9911	Processing / trafficking	ER	O75069.3	APP				0	.	.
<i>HTRA2</i>	27429	Processing / trafficking	ERAD / proteasome	O43464.2	APP				2	.	.
<i>DOCK3</i>	1795	Processing / trafficking	ERAD / proteasome	Q8IZD9.1		1			1	.	.
<i>SYVN1</i>	84447	Processing / trafficking	ERAD / proteasome	Q86TM6.2	APP				1	.	.
<i>PPARG</i>	5468	Processing / trafficking	ERAD / proteasome	P37231.3					1	604367	D
<i>STUB1</i>	10273	Processing / trafficking	ERAD / proteasome	Q9UNE7.2	Aβ				1	615768	R
<i>BLMH</i>	642	Processing / trafficking	Early secretory pathway	Q13867.1	APP				1	.	.
<i>RTN4R</i>	65078	Processing / trafficking	Early secretory pathway	Q9BZR6.1	APP				2	.	.
<i>RTN4RL2</i>	349667	Processing / trafficking	Early secretory pathway	Q86UN3.1	APP				2	.	.
<i>RTN4RL1</i>	146760	Processing / trafficking	Early secretory pathway	Q86UN2.1	APP				2	.	.
<i>HOMER2</i>	9455	Processing / trafficking	Early secretory pathway	Q9NSB8.1	APP				1	.	.
<i>HOMER3</i>	9454	Processing / trafficking	Early secretory pathway	Q9NSC5.2	APP				1	.	.
<i>TMEM59</i>	9528	Processing / trafficking	Early secretory pathway	Q9BXS4.1					1	.	.
<i>STMN2</i>	11075	Processing / trafficking	Early secretory pathway	Q93045.3	APP				1	.	.
<i>VAMP1</i>	6843	Processing / trafficking	Early secretory pathway	P23763.1	APP				2	108600	D

<b>GAS1</b>	2619	Processing / trafficking	Early secretory pathway	P54826	APP				1	.	.
<b>ARCN1</b>	372	Processing / trafficking	Early secretory pathway	P48444	APP				2	617164	D
<b>WASF1</b>	8936	Processing / trafficking	Early secretory pathway	Q92558.1	APP				2	605035	.
<b>CLSTN1</b>	22883	Processing / trafficking	Axonal transport	O94985.1	A $\beta$				0	.	.
<b>KLC1</b>	3831	Processing / trafficking	Axonal transport	Q07866.2	APP				0	.	.
<b>MAPK8IP1</b>	9479	Processing / trafficking	Axonal transport	Q9UQF2.1	APP				0	125853	D
<b>HAP1</b>	9001	Processing / trafficking	Axonal transport	P54257.3	APP				1	.	.
<b>ADAM10</b>	102	Processing / trafficking	Alpha secretase cleavage	O14672.1	APP				1	615537	D
<b>ADAM17</b>	6868	Processing / trafficking	Alpha secretase cleavage	P78536.1	APP				1	614328	R
<b>ADAM9</b>	8754	Processing / trafficking	Alpha secretase cleavage	Q13443.1					1	612775	R
<b>ADAM19</b>	8728	Processing / trafficking	Alpha secretase cleavage	Q9H013.3					1	.	.
<b>FURIN</b>	5045	Processing / trafficking	Alpha secretase cleavage	P09958.2					1	.	.
<b>PCSK7</b>	9159	Processing / trafficking	Alpha secretase cleavage	Q16549.2					1	.	.
<b>TIMP3</b>	7078	Processing / trafficking	Alpha secretase cleavage	P51512.2					2	136900	D
<b>TIMP1</b>	7076	Processing / trafficking	Alpha secretase cleavage	P01033.1					2	.	.
<b>PDPK1</b>	5170	Processing / trafficking	Alpha secretase cleavage	O15530.1					2	.	.
<b>DLG1</b>	1739	Processing / trafficking	Alpha secretase cleavage	Q12959.2					1	.	.
<b>TSPAN12</b>	23554	Processing / trafficking	Alpha secretase cleavage	O95859.1					1	613310	D, R
<b>NRD1</b>	4898	Processing / trafficking	Alpha secretase cleavage	O43847.2					1	.	.
<b>SIRT1</b>	23411	Processing / trafficking	Alpha secretase cleavage	Q96EB6.2					1	.	.
<b>DNM1</b>	1759	Processing / trafficking	Clathrin/caveolin dependent endocytosis	Q05193.2					0	616346	D?
<b>DNM2</b>	1785	Processing / trafficking	Clathrin/caveolin dependent endocytosis	P50570.2					1	606482/615368/160150	D/R/D

<b>DNMBP</b>	23268	Processing / trafficking	Clathrin/caveolin dependent endocytosis	Q6XZF7.1					0	.	.
<b>SNX33</b>	257364	Processing / trafficking	Clathrin/caveolin dependent endocytosis	Q8WV41.1					1	.	.
<b>PICALM</b>	8301	Processing / trafficking	Clathrin/caveolin dependent endocytosis	Q13492.2			Y		2	.	.
<b>BIN1</b>	274	Processing / trafficking	Clathrin/caveolin dependent endocytosis	O00499.1			Y		2	255200	R
<b>SYNJ1</b>	8867	Processing / trafficking	Clathrin/caveolin dependent endocytosis	O43426.2					0	615530	R
<b>FBXL2</b>	25827	Processing / trafficking	Clathrin/caveolin dependent endocytosis	Q9UKC9.3	APP				1	.	.
<b>PIN1</b>	5300	Processing / trafficking	Clathrin/caveolin dependent endocytosis	Q13526.1	APP				1	.	.
<b>FLOT2</b>	2319	Processing / trafficking	Clathrin/caveolin dependent endocytosis	Q14254.2	APP	1			2	.	.
<b>GULP1</b>	51454	Processing / trafficking	Internalization of APP	Q9UBP9.1	APP				0	.	.
<b>DAB1</b>	1600	Processing / trafficking	Internalization of APP	O75553.3	APP				0	.	.
<b>DAB2</b>	1601	Processing / trafficking	Internalization of APP	P98082.3	APP				1	.	.
<b>NUMB</b>	8650	Processing / trafficking	Internalization of APP	P49757.2	APP				2	.	.
<b>APBA1</b>	320	Processing / trafficking	Internalization of APP	Q02410.3	APP	1			0	.	.
<b>APBA2</b>	321	Processing / trafficking	Internalization of APP	Q99767.3	APP	1			0/2	.	.
<b>APBA3</b>	9546	Processing / trafficking	Internalization of APP	O96018.1	APP				0/2	.	.
<b>APBB1</b>	322	Processing / trafficking	Internalization of APP	O00213.2	APP				0	.	.
<b>APBB2</b>	323	Processing / trafficking	Internalization of APP	Q92870.3	APP				0	.	.
<b>APBB3</b>	10307	Processing / trafficking	Internalization of APP	O95704.2	APP				0	.	.
<b>SHC3</b>	53358	Processing / trafficking	Internalization of APP	Q92529.1	APP				2	.	.
<b>SNX17</b>	9784	Processing / trafficking	Internalization of APP	Q15036.1	APP				1	.	.
<b>LRP1B</b>	53353	Processing / trafficking	Internalization of APP	Q9NZR2.2	APP				1	.	.
<b>LRP8</b>	7804	Processing / trafficking	Internalization of APP	Q14114.4	APP				0	.	.

<b>APLP1</b>	333	Processing / trafficking	Internalization of APP	P51693.3					1	.	.
<b>DISC1</b>	27185	Processing / trafficking	Internalization of APP	Q9NRI5.3	APP				0	.	.
<b>SORCS1</b>	114815	Processing / trafficking	Endosomes / retromer	Q8WY21.3	APP				1	.	.
<b>SORCS2</b>	57537	Processing / trafficking	Endosomes / retromer	Q96PQ0.3					1	.	.
<b>SORCS3</b>	22986	Processing / trafficking	Endosomes / retromer	Q9UPU3.2					1	.	.
<b>SORT1</b>	6272	Processing / trafficking	Endosomes / retromer	Q99523.3	APP				1	.	.
<b>SORL1</b>	6653	Processing / trafficking	Endosomes / retromer	Q92673.2	APP		1	Y	1	.	.
<b>LRP10</b>	26020	Processing / trafficking	Endosomes / retromer	Q724F1.2	APP				1	.	.
<b>PACS1</b>	55690	Processing / trafficking	Endosomes / retromer	Q6VY07.2					1	615009	D
<b>VPS26A</b>	9559	Processing / trafficking	Endosomes / retromer	O75436.2					1	.	.
<b>VPS35</b>	55737	Processing / trafficking	Endosomes / retromer	Q96QK1.2					1	614203	D
<b>PIK3C3</b>	5289	Processing / trafficking	Endosomes / retromer	P22543.1					1	.	.
<b>RAB5A</b>	5868	Processing / trafficking	Endosomes / retromer	P20339.2					2	.	.
<b>RAB5B</b>	5869	Processing / trafficking	Endosomes / retromer	P61020.1					2	.	.
<b>RAB5C</b>	5878	Processing / trafficking	Endosomes / retromer	P51148.2					2	.	.
<b>RAB3A</b>	5864	Processing / trafficking	Endosomes / retromer	P20336.1					2	.	.
<b>RAB17</b>	64284	Processing / trafficking	Endosomes / retromer	Q9H0T7.2					2	.	.
<b>RAB44</b>	401258	Processing / trafficking	Endosomes / retromer	Q726P3.3					2	.	.
<b>RAB6A</b>	5870	Processing / trafficking	Endosomes / retromer	P20340.3					2	.	.
<b>RAB10</b>	10890	Processing / trafficking	Endosomes / retromer	P61026.1					2	.	.
<b>RAB34</b>	83871	Processing / trafficking	Endosomes / retromer	Q9BZG1.1					2	.	.
<b>RAB8A</b>	4218	Processing / trafficking	Endosomes / retromer	P61006.1					2	.	.

<b>RAB25</b>	57111	Processing / trafficking	Endosomes / retromer	P57735.2					2	.	.
<b>RAB14</b>	51552	Processing / trafficking	Endosomes / retromer	P61106.4					2	.	.
<b>RAB2A</b>	5862	Processing / trafficking	Endosomes / retromer	P61019.1					2	.	.
<b>SNX8</b>	29886	Processing / trafficking	Endosomes / retromer	Q9Y5X2.1					2	.	.
<b>GRB2</b>	2885	Processing / trafficking	Endosomes / retromer	P62993.1	APP				0	.	.
<b>ADRA2A</b>	150	Processing / trafficking	Endosomes / retromer	P08913.3					2	.	.
<b>BACE1</b>	23621	Processing / trafficking	Beta secretase cleavage	P56817.2	APP				2	.	.
<b>BACE2</b>	25825	Processing / trafficking	Beta secretase cleavage	Q9Y5Z0.1	APP				1	.	.
<b>RANBP9</b>	10048	Processing / trafficking	Beta secretase cleavage	Q96S59.1	APP		1		2	.	.
<b>COP55</b>	10987	Processing / trafficking	Beta secretase cleavage	Q92905.4	APP		1		2	.	.
<b>RTN1</b>	6252	Processing / trafficking	Beta secretase cleavage	Q16799.1			1		1	.	.
<b>RTN2</b>	6253	Processing / trafficking	Beta secretase cleavage	O75298.1			1		1	604805	D
<b>RTN3</b>	10313	Processing / trafficking	Beta secretase cleavage	O95197.2			1		1	.	.
<b>RTN4</b>	57142	Processing / trafficking	Beta secretase cleavage	Q9NQ3.2			1		1	.	.
<b>CCS</b>	9973	Processing / trafficking	Beta secretase cleavage	O14618.1			1		1	.	.
<b>ATXN1</b>	6310	Processing / trafficking	Beta secretase cleavage	P54253.2					1	164400	D
<b>PAWR</b>	5074	Processing / trafficking	Beta secretase cleavage	Q96IZ0.1			1		2	.	.
<b>UCHL1</b>	7345	Processing / trafficking	Beta secretase cleavage	P09936.2					1	615491	R
<b>LRRTM3</b>	347731	Processing / trafficking	Beta secretase cleavage	Q86VH5.2					2	.	.
<b>EHD1</b>	10938	Processing / trafficking	Beta secretase cleavage	Q9H4M9.2					2	.	.
<b>EHD3</b>	30845	Processing / trafficking	Beta secretase cleavage	Q9NZN3.2					2	.	.
<b>ARF6</b>	382	Processing / trafficking	Beta secretase cleavage	P62330.2					1	.	.

<b>PSD</b>	5662	Processing / trafficking	Beta secretase cleavage	A5PKW4.2					1	.	.
<b>SNX12</b>	29934	Processing / trafficking	Beta secretase cleavage	Q9UMY4.3			1		1	.	.
<b>SNX6</b>	58533	Processing / trafficking	Beta secretase cleavage	Q9UNH7.1			1		1	.	.
<b>GGA1</b>	26088	Processing / trafficking	Beta secretase cleavage	Q9UJY5.1			1		1	.	.
<b>GGA2</b>	23062	Processing / trafficking	Beta secretase cleavage	Q9UJY4.3			1		1	.	.
<b>GGA3</b>	23163	Processing / trafficking	Beta secretase cleavage	Q9NZ52.1			1		1	.	.
<b>DHCR24</b>	1718	Processing / trafficking	Beta secretase cleavage	Q15392.2					1	602398	R
<b>RAB11A</b>	8766	Processing / trafficking	Beta secretase cleavage	P62491.3		1			2	.	.
<b>RAB11B</b>	9230	Processing / trafficking	Beta secretase cleavage	Q15907.4		1			2	.	.
<b>FLOT1</b>	10211	Processing / trafficking	Beta secretase cleavage	O75955.3	APP		1		2	.	.
<b>SPON1</b>	10418	Processing / trafficking	Beta secretase cleavage	Q9HCB6.2	APP				1	.	.
<b>CAV3</b>	859	Processing / trafficking	Beta secretase cleavage	P56539.1	APP				0	192600/12332/607801/614321/606072	D/D/R, D/D/D
<b>KLK6</b>	5653	Processing / trafficking	Abeta cleavage (other)	Q92876.1	APP				2	.	.
<b>ENPEP</b>	2028	Processing / trafficking	Abeta cleavage (other)	Q07075.3					0	.	.
<b>APH1B</b>	83464	Processing / trafficking	Gamma secretase cleavage	Q8WW43.3		1			0	.	.
<b>APH1A</b>	51107	Processing / trafficking	Gamma secretase cleavage	Q96B13.1		1			0	.	.
<b>PSENE1</b>	55851	Processing / trafficking	Gamma secretase cleavage	Q9NZ42.1		1			0	613736	D
<b>NCSTN</b>	23385	Processing / trafficking	Gamma secretase cleavage	Q92542.2	APP	1			0	142690	D
<b>PSEN1</b>	5663	Processing / trafficking	Gamma secretase cleavage	P49768.1	APP				0	613737/607822/613694/600274/172700	All D
<b>PSEN2</b>	5664	Processing / trafficking	Gamma secretase cleavage	P49810.1	APP				0	606889/613697	D/D
<b>BSG</b>	682	Processing / trafficking	Gamma secretase cleavage	P35613.2		1			0	.	.
<b>CSNK1A1?</b>	1452	Processing / trafficking	Gamma secretase cleavage	P48729.2					2	.	.

<b>ABCA2</b>	20	Processing / trafficking	Gamma secretase cleavage	Q9BZC7.3					2	.	.
<b>RER1</b>	11079	Processing / trafficking	Gamma secretase cleavage	O15258.1		1			1	.	.
<b>FBXW7</b>	55294	Processing / trafficking	Gamma secretase cleavage	Q969H0.1		1			2	.	.
<b>GSAP</b>	54103	Processing / trafficking	Gamma secretase cleavage	A4D1B5.2	APP	1			2	.	.
<b>TMED9</b>	54732	Processing / trafficking	Gamma secretase cleavage	Q9BVK6.2					1	.	.
<b>TMED10</b>	10972	Processing / trafficking	Gamma secretase cleavage	P49755.2		1			1	.	.
<b>ARC</b>	23237	Processing / trafficking	Gamma secretase cleavage	Q7LC44.1		1			2	.	.
<b>CRB2</b>	286204	Processing / trafficking	Gamma secretase cleavage	Q5IJ48		1			1	616220/219730	R/R
<b>VDAC1</b>	7416	Processing / trafficking	Gamma secretase cleavage	P21796.2		1			2	.	.
<b>CNTNAP1</b>	8506	Processing / trafficking	Gamma secretase cleavage	P78357.1					2	616286	R
<b>SYP</b>	6855	Processing / trafficking	Gamma secretase cleavage	P08247.3					2	300802	X-linked D
<b>CTNND1</b>	1500	Processing / trafficking	Gamma secretase cleavage	O60716.1		1			0	.	.
<b>CTNND2</b>	1501	Processing / trafficking	Gamma secretase cleavage	Q9UQB3.3		1			0	.	.
<b>HM13</b>	81502	Processing / trafficking	Gamma secretase cleavage	Q8TCT9.1		1			2	.	.
<b>KCNIP3</b>	30818	Processing / trafficking	Gamma secretase cleavage	Q9Y2W7.1		1			2	.	.
<b>PLD1</b>	5337	Processing / trafficking	Gamma secretase cleavage	Q13393.1	A $\beta$	1			1	.	.
<b>PLD3</b>	23646	Processing / trafficking	Gamma secretase cleavage	Q8IV08.1	APP				1	.	.
<b>ARRB2</b>	409	Processing / trafficking	Gamma secretase cleavage	P32121.2	APP				2	.	.
<b>GPR3</b>	2827	Processing / trafficking	Gamma secretase cleavage	P46089.1	APP	1			2	.	.
<b>TRPC6</b>	7225	Processing / trafficking	Gamma secretase cleavage	Q9Y210	APP				1	.	.
<b>ADRB2</b>	154	Processing / trafficking	Gamma secretase cleavage	P07550.3		1			2	.	.
<b>OPRD1</b>	4985	Processing / trafficking	Gamma secretase cleavage	P41143.4		1	1		2	.	.

<b>ALOX5</b>	240	Processing / trafficking	Gamma secretase cleavage	P09917.2					2	.	.
<b>CALR</b>	811	Processing / trafficking	Gamma secretase cleavage	P27797.1	Aβ	1			1	.	.
<b>ERLIN2</b>	11160	Processing / trafficking	Gamma secretase cleavage	O94905.1		1			2	611225	R
<b>STX5</b>	6811	Processing / trafficking	Gamma secretase cleavage	Q13190.2		1			1	.	.
<b>SNX27</b>	81609	Processing / trafficking	Gamma secretase cleavage	Q96L92.2		1			1	611541	.
<b>OCIAD2</b>	132299	Processing / trafficking	Gamma secretase cleavage	Q56VL3					2	.	.
<b>CAV1</b>	857	Processing / trafficking	Gamma secretase cleavage	Q03135.4	APP				1	606721	D
<b>SOAT1</b>	6646	Processing / trafficking	Cholesterol metabolism	P35610.3					2	.	.
<b>OSBP</b>	5007	Processing / trafficking	Cholesterol metabolism	P22059.1					1	.	.
<b>CH25H</b>	9023	Processing / trafficking	Cholesterol metabolism	O95992.1					0	.	.
<b>CYP46A1</b>	10858	Processing / trafficking	Cholesterol metabolism	Q9Y6A2.1					1	.	.
<b>ABCG1</b>	9619	Processing / trafficking	Cholesterol metabolism	P45844.3					0	.	.
<b>PLTP</b>	5360	Processing / trafficking	Cholesterol metabolism	P55058	APP				1	.	.
<b>ABCA1</b>	19	Processing / trafficking	Cholesterol metabolism	O95477.3					1	.	.
<b>NPC1</b>	4864	Processing / trafficking	Cholesterol metabolism	O15118.2					2	257220	R
<b>NPC2</b>	10577	Processing / trafficking	Cholesterol metabolism	P61916.1					2	607625	R
<b>GSK3A</b>	2931	Processing / trafficking	Miscellaneous	P49840.2					2	.	.
<b>UBQLN1</b>	29979	Processing / trafficking	Miscellaneous	Q9UMX0.2	APP	1			0	.	.
<b>LDLRAD3</b>	143458	Processing / trafficking	Miscellaneous	Q86YD5.3	APP				2	.	.
<b>CDH2</b>	1000	Processing / trafficking	Miscellaneous	P19022.4	APP	1			2	.	.
<b>CD74</b>	972	Processing / trafficking	Miscellaneous	P04233.3	APP				1	.	.
<b>ITM2B</b>	9445	Processing / trafficking	Miscellaneous	Q9Y287.1	APP				1	176500/17300/616079	All D

<b>ITM2C</b>	81618	Processing / trafficking	Miscellaneous	Q9NQX7.1	APP		1		1	.	.
<b>NTN1</b>	9423	Processing / trafficking	Miscellaneous	O95631.2	APP				1	.	.
<b>CTXN3</b>	613212	Processing / trafficking	Miscellaneous	Q4LDR2.1					1	.	.
<b>CD2AP</b>	23607	Processing / trafficking	Miscellaneous	Q9Y5K6.1			Y		2	607832	??
<b>LGMN</b>	5641	Processing / trafficking	Miscellaneous	Q99538.1	APP				2	.	.
<b>FERMT2</b>	10979	Processing / trafficking	Miscellaneous	Q96AC1			Y		2	.	.
<b>MMP24</b>	10893	Processing / trafficking	Miscellaneous	Q9Y5R2.1	APP				2	.	.
<b>SERPINE1</b>	5054	Clearance / degradation	Enzymatic degradation of Abeta	P05121.1					2	613329	R
<b>IDE</b>	3416	Clearance / degradation	Enzymatic degradation of Abeta	P14735.4	Aβ				1	.	.
<b>ACE</b>	1636	Clearance / degradation	Enzymatic degradation of Abeta	P12821.1	Aβ				1	267430	R
<b>CPB1</b>	1360	Clearance / degradation	Enzymatic degradation of Abeta	P15086.4					1	.	.
<b>FOLH1</b>	2346	Clearance / degradation	Enzymatic degradation of Abeta	Q04609.1					1	.	.
<b>MME</b>	4311	Clearance / degradation	Enzymatic degradation of Abeta	P08473.2					1	.	.
<b>MMEL1</b>	79258	Clearance / degradation	Enzymatic degradation of Abeta	Q495T6.2					1	.	.
<b>MMP2</b>	4313	Clearance / degradation	Enzymatic degradation of Abeta	P08253.2	APP				1	259600	R
<b>MMP9</b>	4318	Clearance / degradation	Enzymatic degradation of Abeta	P14780.3	Aβ				1	613073	R
<b>ECE2</b>	9718	Clearance / degradation	Enzymatic degradation of Abeta	O60344.4					1	.	.
<b>ECE1</b>	1889	Clearance / degradation	Enzymatic degradation of Abeta	P42892.2					1	.	.
<b>PLG</b>	5340	Clearance / degradation	Enzymatic degradation of Abeta	P00747.2					1	217090	R
<b>PLAT</b>	5327	Clearance / degradation	Enzymatic degradation of Abeta	P00750.1					1	.	.
<b>PLAU</b>	5328	Clearance / degradation	Enzymatic degradation of Abeta	P00749.2					1	601709	D
<b>HTRA1</b>	5654	Clearance / degradation	Enzymatic degradation of Abeta	Q92743					1	600142/616779	R/D

<b>CTSB</b>	1508	Clearance / degradation	Enzymatic degradation of Abeta	P07858.3					1	.	.
<b>PARK2</b>	5071	Clearance / degradation	ERAD / proteasome	O60260.2	Aβ				1	600116	R
<b>BECN1</b>	8678	Clearance / degradation	Autophagy	Q14457.2					1	.	.
<b>ATG7</b>	10533	Clearance / degradation	Autophagy	O95352.1					1	.	.
<b>ABCB1</b>	5243	Clearance / degradation	Blood brain barrier	P08183.3	Aβ				1	.	.
<b>OLR1</b>	4973	Clearance / degradation	Blood brain barrier	P78380.1					0	.	.
<b>LRP1</b>	4035	Clearance / degradation	Blood brain barrier	Q07954.2	APP				1	.	.
<b>LRP2</b>	4036	Clearance / degradation	Blood brain barrier	P98164.3					1	222448	R
<b>LDLR</b>	3949	Clearance / degradation	Blood brain barrier	P01130.1					1	143890	D
<b>SRF</b>	6722	Clearance / degradation	Blood brain barrier	P11831.1					2	.	.
<b>MEOX2</b>	4223	Clearance / degradation	Blood brain barrier	P50222.2					1	.	.
<b>MYOCD</b>	93649	Clearance / degradation	Blood brain barrier	Q8IZQ8.1					2	.	.
<b>MEGF10</b>	84466	Clearance / degradation	Microglia	Q96KG7.1	Aβ				1	614399	R
<b>LPL</b>	4023	Clearance / degradation	Microglia	P06858.1	Aβ				1	144250/238600	D/R
<b>MSR1</b>	4481	Clearance / degradation	Microglia	P21757.1	Aβ				1	176807	D
<b>SCARB1</b>	949	Clearance / degradation	Microglia	Q8WTV0.1	Aβ				1	.	.
<b>CD36</b>	948	Clearance / degradation	Microglia	P16671.2	Aβ				1	.	.
<b>CD47</b>	961	Clearance / degradation	Microglia	Q08722.1	Aβ	1			1	.	.
<b>PTPRC</b>	5788	Clearance / degradation	Microglia	P08575.2					1	608971	R
<b>CD14</b>	929	Clearance / degradation	Microglia	P08571.2	Aβ				0	.	.
<b>ITGA6</b>	3655	Clearance / degradation	Microglia	P23229.5	Aβ				0	226730	R
<b>ITGB1</b>	3688	Clearance / degradation	Microglia	P05556.2	Aβ	1			0	.	.

<b>AGER</b>	177	Clearance / degradation	Microglia	Q15109.1	Aβ				2	.	.
<b>FPR2</b>	2358	Clearance / degradation	Microglia	P25090.2	Aβ				2	.	.
<b>HMGB1</b>	3146	Clearance / degradation	Microglia	P09429.3	Aβ				2	.	.
<b>TLR2</b>	7097	Clearance / degradation	Microglia	O60603.1	Aβ				0	.	.
<b>TLR1</b>	7096	Clearance / degradation	Microglia	Q15399.3					0	.	.
<b>TLR6</b>	10333	Clearance / degradation	Microglia	Q9Y2C9.2					0	.	.
<b>TLR4</b>	7099	Clearance / degradation	Microglia	O00206.2	Aβ				0	.	.
<b>MAPK14</b>	1432	Clearance / degradation	Microglia	Q16539.3					0	.	.
<b>S100A8</b>	6279	Clearance / degradation	Microglia	P05109.1					0	.	.
<b>S100A9</b>	6280	Clearance / degradation	Microglia	P06702.1					0	.	.
<b>MYD88</b>	4615	Clearance / degradation	Microglia	Q99836.1					0	.	.
<b>NLRP3</b>	114548	Clearance / degradation	Microglia	Q96P20.3					0	120100/191900	All D
<b>TREM2</b>	54209	Clearance / degradation	Microglia	Q9NZC2.1					0	221770	R
<b>SIRPB1</b>	10326	Clearance / degradation	Microglia	Q5TFQ8.1					1	.	.
<b>TYROBP</b>	7305	Clearance / degradation	Microglia	O43914.1					0	221770	R
<b>CCR2</b>	729230	Clearance / degradation	Microglia	P41597.1					0	.	.
<b>CCL2</b>	6347	Clearance / degradation	Microglia	P13500.1					0	.	.
<b>CX3CR1</b>	1524	Clearance / degradation	Microglia	P49238.1					0	.	.
<b>CX3CL1</b>	6376	Clearance / degradation	Microglia	P78423.1					0	.	.
<b>CD33</b>	945	Clearance / degradation	Microglia	P20138.2			Y		0	.	.
<b>IL1B</b>	3553	Clearance / degradation	Microglia	P01584.2					0	.	.
<b>IL12A</b>	3592	Clearance / degradation	Microglia	P29459.2					0	.	.

<b>IL12B</b>	3593	Clearance / degradation	Microglia	P29460.1					0	.	.
<b>IL23A</b>	51561	Clearance / degradation	Microglia	Q9NPF7.1					0	.	.
<b>IL6</b>	3569	Clearance / degradation	Microglia	P05231.1					0	.	.
<b>IL4</b>	3565	Clearance / degradation	Microglia	P05112.1					0	.	.
<b>IL10</b>	3586	Clearance / degradation	Microglia	P22301.1					0	.	.
<b>TNF</b>	7124	Clearance / degradation	Microglia	P01375.1					0	.	.
<b>CSF2</b>	1437	Clearance / degradation	Microglia	P04141.1					0	.	.
<b>GRN</b>	2896	Clearance / degradation	Microglia	P28799.2					1	607485/614706	D / R
<b>PTGER2</b>	5732	Clearance / degradation	Microglia	P43116.2					0	.	.
<b>CNR2</b>	1269	Clearance / degradation	Microglia	P34972.1					1	.	.
<b>TGFB1</b>	7040	Clearance / degradation	Microglia	P01137.2	Aβ				1	131300	D
<b>IRAK4</b>	51135	Clearance / degradation	Microglia	Q9NWZ3.1					0	607676	R
<b>CR1</b>	1378	Clearance / degradation	Microglia	P17927.3				Y	0	.	.
<b>C3</b>	718	Clearance / degradation	Microglia	P01024.2					0	613779	R
<b>ABCA7</b>	10347	Clearance / degradation	Microglia	Q8IZY2.3				Y	1	.	.
<b>APCS</b>	325	Clearance / degradation	Miscellaneous	P02743.2	Aβ				0	.	.
<b>QPCT</b>	25797	Aggregation	Miscellaneous	Q16769.1					2	.	.
<b>TGM2</b>	7052	Aggregation	Miscellaneous	P21980.2	Aβ				0	.	.
<b>PRKACA</b>	5566	Aggregation	Miscellaneous	P17612.2					0	.	.
<b>PRKACB</b>	5567	Aggregation	Miscellaneous	P22694.2					0	.	.
<b>SLC30A3</b>	7781	Aggregation	Miscellaneous	Q99726.2					2	.	.
<b>SNCA</b>	6622	Aggregation	Miscellaneous	P37840.1	Aβ				0	127750/168601/605543	All D
<b>TARDBP</b>	23435	Aggregation	Miscellaneous	Q13148.1					0	612069	D
<b>TPPP</b>	11076	Aggregation	Miscellaneous	O94811.1	Aβ				0	.	.

<b>ACHE</b>	43	Aggregation	Miscellaneous	P22303.1	A $\beta$				0	.	.
<b>BCHE</b>	590	Aggregation	Miscellaneous	P06276.1	A $\beta$				0	.	.
<b>COL25A1</b>	84570	Aggregation	Miscellaneous	Q9BXS0.2	A $\beta$				0	616219	R
<b>CST3</b>	1471	Aggregation	Miscellaneous	P01034.1	A $\beta$				0	105150	D
<b>TTR</b>	7276	Aggregation	Miscellaneous	P02766.1	A $\beta$				0	105210/115430/145680	All D
<b>APOA1</b>	335	Aggregation	Miscellaneous	P02647.1	A $\beta$				0	105200/604091	All D
<b>RELN</b>	5649	Aggregation	Miscellaneous	P78509.3	A $\beta$				0	257320/616436	R/D
<b>LAMA1</b>	284217	Aggregation	Miscellaneous	P25391.2	A $\beta$				0	615960	R
<b>SERPINA3</b>	12	Aggregation	Miscellaneous	P01011.2	A $\beta$				0	.	.
<b>SERPINH1</b>	871	Aggregation	Miscellaneous	P50454.2	APP				0	613848	R
<b>CLU</b>	1191	Aggregation	Miscellaneous	P10909.1	A $\beta$			Y	1	.	.
<b>APOE</b>	348	Aggregation	Miscellaneous	P02649.1	A $\beta$				0	269600	D?
<b>A2M</b>	2	Aggregation	Miscellaneous	P01023.3	A $\beta$				1	614036	D
<b>HSPA4</b>	3308	Aggregation	Chaperone	P34932.4					0	.	.
<b>DNAJB1</b>	3337	Aggregation	Chaperone	P25685.4					0	.	.
<b>DNAJB6</b>	10049	Aggregation	Chaperone	O75190.2					0	603511	D
<b>HSP90AA1</b>	3320	Aggregation	Chaperone	P07900.5	A $\beta$				0	.	.
<b>HSPB1</b>	3315	Aggregation	Chaperone	P04792.2	A $\beta$				0	606595/608634	All D
<b>CRYAB</b>	1410	Aggregation	Chaperone	P02511.2	A $\beta$				0	615184/613763/608810/613869	D / D and R / D / R
<b>HSPB6</b>	126393	Aggregation	Chaperone	O14558.2	A $\beta$				0	.	.
<b>HSPB8</b>	26353	Aggregation	Chaperone	Q9UJY1.1	A $\beta$				0	608673/158590	All D
<b>LRPAP1</b>	4043	Aggregation	Chaperone	P30533.1	A $\beta$				0	615431	R
<b>AGRN</b>	375790	Aggregation	Heparan sulfate proteoglycan	O00468.5	A $\beta$				2	615120	R
<b>GPC1</b>	2817	Aggregation	Heparan sulfate proteoglycan	P35052.2	APP				2	.	.
<b>SDC3</b>	9672	Aggregation	Heparan sulfate proteoglycan	O75056.2					2	.	.
<b>COL18A1</b>	80781	Aggregation	Heparan sulfate proteoglycan	P39060.5	A $\beta$				2	267750	R
<b>HSPG2</b>	3339	Aggregation	Heparan sulfate proteoglycan	P98160.4	A $\beta$				2	224410/255800	R
<b>GRIA2</b>	2891	Mediator of Abeta toxicity	Synaptic toxicity	P42262.3	A $\beta$				0	.	.

<b>GRIN2A</b>	2903	Mediator of Abeta toxicity	Synaptic toxicity	Q12879.1					0	245570	D
<b>GRIN2B</b>	2904	Mediator of Abeta toxicity	Synaptic toxicity	Q13224.3					0	616139/613970	D / D
<b>CHRNA7</b>	1139	Mediator of Abeta toxicity	Synaptic toxicity	P36544.5	Aβ				0	.	.
<b>DLG4</b>	1742	Mediator of Abeta toxicity	Synaptic toxicity	P78352.3					0	.	.
<b>LILRB2</b>	10288	Mediator of Abeta toxicity	Synaptic toxicity	Q8N423.4	Aβ				0	.	.
<b>PPP3CA</b>	5530	Mediator of Abeta toxicity	Synaptic toxicity	Q08209.1					0	.	.
<b>PPP3CB</b>	5532	Mediator of Abeta toxicity	Synaptic toxicity	P16298.2					0	.	.
<b>PPP3CC</b>	5533	Mediator of Abeta toxicity	Synaptic toxicity	P48454.3					0	.	.
<b>PPP3R1</b>	5534	Mediator of Abeta toxicity	Synaptic toxicity	P63098.2					0	.	.
<b>PPP3R2</b>	5535	Mediator of Abeta toxicity	Synaptic toxicity	Q96LZ3.3					0	.	.
<b>NGFR</b>	4804	Mediator of Abeta toxicity	Synaptic toxicity	P08138.1	Aβ				0	.	.
<b>LINGO1</b>	84894	Mediator of Abeta toxicity	Synaptic toxicity	Q96FE5.2	APP				0	.	.
<b>PXN</b>	5829	Mediator of Abeta toxicity	Synaptic toxicity	P49023.3					0	.	.
<b>EPHB2</b>	2048	Mediator of Abeta toxicity	Synaptic toxicity	P29323.5	Aβ				0	.	.
<b>EPHA4</b>	2043	Mediator of Abeta toxicity	Synaptic toxicity	P54764.1					0	.	.
<b>PRNP</b>	5621	Mediator of Abeta toxicity	Synaptic toxicity	P04156.1	Aβ				0	137440/137440/123400/603218	All D
<b>GRM5</b>	2915	Mediator of Abeta toxicity	Synaptic toxicity	P41594.2	Aβ				0	.	.
<b>FYN</b>	2534	Mediator of Abeta toxicity	Synaptic toxicity	P06241.3					0	.	.
<b>MAPT</b>	4137	Mediator of Abeta toxicity	Tau Toxicity	P10636.5	Aβ				2	600274/172700/260540/601104	D / D / R / D
<b>GSK3B</b>	2932	Mediator of Abeta toxicity	Tau Toxicity	P49841.2		1			0	.	.
<b>PRKAA1</b>	5562	Mediator of Abeta toxicity	Tau Toxicity	Q13131.4					0	.	.
<b>CAMKK2</b>	10645	Mediator of Abeta toxicity	Tau Toxicity	Q96RR4.2					0	.	.

<b>CDK5R1</b>	8851	Mediator of Abeta toxicity	Tau Toxicity	Q15078.1					0	.	.
<b>CDK5</b>	1020	Mediator of Abeta toxicity	Tau Toxicity	Q00535.3					0	616342	R
<b>CAPN1</b>	823	Mediator of Abeta toxicity	Tau Toxicity	P07384.1					0	.	.
<b>CAPN2</b>	824	Mediator of Abeta toxicity	Tau Toxicity	P17655.6					0	.	.
<b>MARK4</b>	57787	Mediator of Abeta toxicity	Tau Toxicity	Q96L34.1					0	.	.
<b>DYRK1A</b>	1859	Mediator of Abeta toxicity	Tau Toxicity	Q13627.2					0	614104	D
<b>PPP2R1A</b>	5518	Mediator of Abeta toxicity	Tau Toxicity	P30153.4					0	616362	D
<b>PPP2R1B</b>	5519	Mediator of Abeta toxicity	Tau Toxicity	P30154.3					0	.	.
<b>PPP2CA</b>	5515	Mediator of Abeta toxicity	Tau Toxicity	P67775.1					0	.	.
<b>PPP2CB</b>	5516	Mediator of Abeta toxicity	Tau Toxicity	P62714.1					0	.	.
<b>PPP2R2A</b>	5520	Mediator of Abeta toxicity	Tau Toxicity	P63151.1					0	.	.
<b>PPP2R2B</b>	5521	Mediator of Abeta toxicity	Tau Toxicity	Q00005.1					0	604326	D
<b>PPP2R5D</b>	5528	Mediator of Abeta toxicity	Tau Toxicity	Q14738.1					0	616355	D
<b>PPP2R3A</b>	5523	Mediator of Abeta toxicity	Tau Toxicity	Q06190.1					0	.	.
<b>ANP32A</b>	8125	Mediator of Abeta toxicity	Tau Toxicity	P39687.1					0	.	.
<b>SET</b>	6418	Mediator of Abeta toxicity	Tau Toxicity	Q01105.3					0	.	.
<b>CASP3</b>	836	Mediator of Abeta toxicity	Tau Toxicity	P42574.2					0	.	.
<b>HSD17B10</b>	3028	Mediator of Abeta toxicity	Mitochondrial toxicity	Q99714.3	Aβ				0	300438/300220	X-linked D / X-linked R
<b>PPID</b>	5481	Mediator of Abeta toxicity	Mitochondrial toxicity	Q08752.3	Aβ				0	.	.
<b>DNM1L</b>	10059	Mediator of Abeta toxicity	Mitochondrial toxicity	O00429.2	Aβ				0	614388	D
<b>RYR1</b>	6261	Mediator of Abeta toxicity	calcium homeostasis	P21817.3					0	117000/117000/145600/145600/255320	D and R/D and R/D/D/R
<b>RYR2</b>	6262	Mediator of Abeta toxicity	calcium homeostasis	Q92736.3					0	600996/604772	All D

<b>RYR3</b>	6263	Mediator of Abeta toxicity	calcium homeostasis	Q15413.3					0	.	.
<b>CALHM1</b>	255022	Mediator of Abeta toxicity	calcium homeostasis	Q8IU99.2					0	.	.
<b>ITPR1</b>	3708	Mediator of Abeta toxicity	calcium homeostasis	Q14643.3					0	606658/117360	D/D
<b>TRPM2</b>	7226	Mediator of Abeta toxicity	calcium homeostasis	O94759.2					0	.	.
<b>TM2D1</b>	83941	Mediator of Abeta toxicity	Miscellaneous	Q9BX74.1	A $\beta$				0	.	.
<b>ADNP</b>	23394	Mediator of Abeta toxicity	Miscellaneous	Q9H2P0.1	A $\beta$				0	615873	D
<b>NPTX1</b>	4884	Mediator of Abeta toxicity	Miscellaneous	Q15818.2					0	.	.

PPI : *Protein-Protein Interaction* ; y : *yes* ;D : *Dominant* ; R : *Recessive* .

Pour chaque gène, un chiffre 1 ou 2 a été attribué dans le cas où une perte de fonction du gène serait en faveur d'un mécanisme pathologique (1) ou si un gain de fonction serait en faveur d'un mécanisme pathologique (2) d'après le rôle attribué de la protéine dans la physiopathologie. Le 0 a été utilisé lorsqu'il n'était pas possible faire de telles prédictions.

## **Annexe 2 : Article sur l'étude de trios Alzheimer**

## IMMEDIATE COMMUNICATION

# *De novo* deleterious genetic variations target a biological network centered on A $\beta$ peptide in early-onset Alzheimer disease

A Rovelet-Lecrux<sup>1,11</sup>, C Charbonnier<sup>1,2,11</sup>, D Wallon<sup>1,2,3,11</sup>, G Nicolas<sup>1,2,4,11</sup>, MNJ Seaman<sup>5</sup>, C Pottier<sup>1</sup>, SY Breusegem<sup>5</sup>, PP Mathur<sup>6,7</sup>, P Jenardhanan<sup>6</sup>, K Le Guennec<sup>1</sup>, AS Mukadam<sup>5</sup>, O Quenez<sup>1,2</sup>, S Coutant<sup>1</sup>, S Rousseau<sup>1,2</sup>, A-C Richard<sup>1,2</sup>, A Boland<sup>8</sup>, J-F Deleuze<sup>8</sup>, T Frebourg<sup>1,4</sup>, D Hannequin<sup>1,2,3,4</sup>, D Campion<sup>1,2,9</sup> and CNR-MAJ collaborators<sup>10</sup>

We hypothesized that *de novo* variants (DNV) might participate in the genetic determinism of sporadic early-onset Alzheimer disease (EOAD, onset before 65 years). We investigated 14 sporadic EOAD trios first by array-comparative genomic hybridization. Two patients carried a *de novo* copy number variation (CNV). We then performed whole-exome sequencing in the 12 remaining trios and identified 12 non-synonymous DNVs in six patients. The two *de novo* CNVs (an amyloid precursor protein (*APP*) duplication and a *BACE2* intronic deletion) and 3/12 non-synonymous DNVs (in *PSEN1*, *VPS35* and *MARK4*) targeted genes from a biological network centered on the Amyloid beta (A $\beta$ ) peptide. We showed that this *a priori*-defined genetic network was significantly enriched in amino acid-altering DNV, compared with the rest of the exome. The causality of the *APP de novo* duplication (which is the first reported one) was obvious. In addition, we provided evidence of the functional impact of the following three non-synonymous DNVs targeting this network: the novel *PSEN1* variant resulted in exon 9 skipping in patient's RNA, leading to a pathogenic missense at exons 8–10 junction; the *VPS35* missense variant led to partial loss of retromer function, which may impact neuronal APP trafficking and A $\beta$  secretion; and the *MARK4* multiple nucleotide variant resulted into increased Tau phosphorylation, which may trigger enhanced A $\beta$ -induced toxicity. Despite the difficulty to recruit Alzheimer disease (AD) trios owing to age structures of the pedigrees and the genetic heterogeneity of the disease, this strategy allowed us to highlight the role of *de novo* pathogenic events, the putative involvement of new genes in AD genetics and the key role of A $\beta$  network alteration in AD.

*Molecular Psychiatry* advance online publication, 21 July 2015; doi:10.1038/mp.2015.100

## INTRODUCTION

The genetic architecture of Alzheimer disease (AD) is still incompletely characterized. Rare autosomal dominant forms of early-onset Alzheimer disease (EOAD) are caused by mutations within three genes (*APP* (amyloid precursor protein), *PSEN1* and *PSEN2*). In addition, genome-wide association studies have implicated common variants at ~20 genetic loci.<sup>1</sup> Notably, with the exception of the  $\epsilon 4$  allele of the *APOE* gene, these variants have a modest effect on AD risk (OR ranging from 0.77 to 1.22). With the advent of whole-exome sequencing, which consists in the massive parallel sequencing of all protein-coding exons of a genome, it has become possible to interrogate low-frequency variants in case-control studies, and rare missense variants in *TREM2* and *PLD3* have been reported to increase risk for AD.<sup>2,3</sup> However, despite these findings, it is acknowledged that a substantial part of the genetic component of the disease remains unidentified.

Recently, the impact of *de novo* mutations in common neuropsychiatric diseases including mental retardation, autism or schizophrenia has been highlighted by several studies. Together, these studies have provided arguments that germline mutations

with large effect could account for a fraction of sporadic cases affected by these conditions. Consequently, a '*de novo*' experimental design has become an attractive paradigm for the discovery of relevant genes in complex diseases.<sup>4</sup> Different types of *de novo* genomic events have been studied. Although earlier studies focused on copy number variations (CNVs), more recent whole-exome sequencing studies have allowed for the detection of *de novo* single- or multiple-nucleotide variants (MNV) and indels. It is now estimated that CNVs larger than 100 kb occur *de novo* in ~1 of every 50 individuals, whereas each individual carries an average of one exonic *de novo* single-nucleotide variants.<sup>5</sup> These studies are based on the simultaneous analysis of the whole-exome sequence of a patient and of his/her unaffected parents in family trios to detect mutation (s) present in the patient but not in his/her parents. Owing to the age structure of AD pedigrees, such a design is difficult to apply when dealing with diseases affecting elderly subjects. We nevertheless succeeded in sampling 14 EOAD patients (age of onset younger than 65) with no history of the disease in first- or second-degree relatives, and their living parents (Table 1).

<sup>1</sup>Inserm U1079, Institute for Research and Innovation in Biomedicine, University of Rouen, Rouen, France; <sup>2</sup>CNR-MAJ, Rouen University Hospital, Rouen, France; <sup>3</sup>Department of Neurology, Rouen University Hospital, Rouen, France; <sup>4</sup>Department of Genetics, Rouen University Hospital, Rouen, France; <sup>5</sup>Cambridge Institute for Medical Research/Dept of Clinical Biochemistry, University of Cambridge, Addenbrookes Hospital, Cambridge, UK; <sup>6</sup>Centre of Bioinformatics and Department of Biochemistry & Molecular Biology, School of Life Sciences, Pondicherry University, Pondicherry, India; <sup>7</sup>KIIT University, Bhubaneswar, Odisha, India; <sup>8</sup>Centre National de Génotypage, Evry, France and <sup>9</sup>Department of Research, Centre Hospitalier du Rouvray, Sotteville-Les-Rouen, France. Correspondence: D Campion, Inserm U1079, Faculté de Médecine, 22 boulevard Gambetta, Rouen 76183 France.

E-mail: dominique.campion@univ-rouen.fr

<sup>11</sup>These authors contributed equally to this work.

<sup>10</sup>CNR-MAJ collaborators are listed in Appendix A.

Received 20 March 2015; revised 27 May 2015; accepted 16 June 2015

**Table 1.** Demographic and clinical data of the study participants

Patient code	Sex	AOO	APOE	Neuropathology	CSF A $\beta$ 42		CSF Tau		CSF P-Tau		Age of father (years)		Age of mother (years)			
					pg ml <sup>-1</sup>		pg ml <sup>-1</sup>		pg ml <sup>-1</sup>		42		At		At	
					N > 550	N < 350	N < 60	N < 0.52	Inclusion	Childbirth	Inclusion	Childbirth				
EXT 804	M	50	3/4	AD+ CAA	678	1028	96	1.52	81	26	80	25				
EXT 773	M	44	3/3		85	30	76	21								
EXT 235	F	46	2/3	AD	481	777	96	1.62	87	32	78	24				
ROU 682	F	54	3/3		92	27	85	20								
ROU 1331	M	58	3/4		366	612	73	1.67	92	27	86	21				
ROU 1322	F	50	3/3		573	624	100	1.09	94	35	80	21				
EXT 395	F	50	3/4		248	525	124	2.12	81	27	81	27				
EXT 488	F	52	3/3		588	627	80	1.07	86	26	85	25				
EXT 687	M	49	3/3		396	628	123	1.59	79	28	76	25				
EXT 306	F	49	3/4		430	744	101	1.73	76	23	76	23				
EXT 186	F	58	3/3		463	1200	145	2.59	91	26	89	24				
EXT 344	M	46	3/4		608	900	82	1.48	76	25	77	26				
EXT 573	M	46	3/3	390	658	27	1.69	72	24	68	20					
EXT 576	F	45	3/4	497	1013	143	2.04	80	30	73	23					

Abbreviations: A $\beta$ , Amyloid beta; AD, Alzheimer disease; AOO, age of onset; APOE, APOE genotype; CAA, cerebral amyloid angiopathy; CSF, cerebrospinal fluid.

## MATERIALS AND METHODS

For detailed methods see Supplementary Information.

### Subjects

Patients were recruited by a French memory clinical network. Diagnoses (neuropathologically proven AD,  $n=2$ ; probable AD with evidence of the AD pathophysiological process,  $n=12$ ) were performed according to the NINCDS-ADRDA<sup>6</sup> criteria. Inclusion criteria are described in Supplementary Methods and the demographic characteristics and APOE genotypes of the sample are displayed in Table 1. Only patients with no affected relative among first- or second-degree relatives were included in the study.

### CNV analysis

High-resolution array-comparative genomic hybridization analysis was performed using the Human High-Resolution Discovery Microarray Kit 1  $\times$  1 M (Agilent Technologies, Santa Clara, CA, USA) using standard protocols. All CNVs encompassing genes and not previously reported in large array-comparative genomic hybridization studies in the database of genomic variants (<http://dgv.tcag.ca>) were confirmed in patients and their inheritance was assessed by quantitative multiplex PCR of short fluorescent fragments (Supplementary Figure S1).

### Exome sequencing

Exomes were captured using Agilent SureSelect Human All Exon kits. Final libraries were then sequenced on a HiSeq2000. Bioinformatic pipelines, detection of *de novo* mutations and confirmation by Sanger sequencing are described in Supplementary Methods.

### Building the A $\beta$ network

We performed extensive literature search using Pubmed (January 2014) to collect information linking a set of genes with Amyloid beta (A $\beta$ ) processing, aggregation, clearance or toxicity using as key words APP, A $\beta$  and each of the subheaders of the gene list (Supplementary Table S1). As a result, 1988 references in English language were initially selected based on titles and abstracts. Irrelevant genes were filtered out and additional candidates were identified from the body text of these references. Routine

literature searches were subsequently used to maintain the data set and identify new interactors. We only retained genes for which experimental evidence based on cell or animal models had been published. No further exclusion criteria were used except for the A $\beta$  toxicity category for which genes involved in non-specific late stages of cell death were not retained. Direct protein-protein interaction was not a prerequisite but was noted, when documented. For that purpose, data retrieved during this search were compared with the extensive list of APP interactors previously compiled by Perreau *et al.*<sup>7</sup>

### Testing for enrichment in *de novo* mutations of the A $\beta$ network

The objective was to test whether the A $\beta$  network was enriched in *de novo* mutations among our 12 EOAD patients, compared with the expected relative mutation rate of this network in the general population, conditional on the total number of *de novo* mutations observed. We considered several candidates of the network relative baseline mutation rate, either based on the proportion of bases jointly covered at 20 $\times$  that fell within the A $\beta$  network, or based on the proportion of variants (synonymous or missense, rare or of any minor allele frequency) carried by the unaffected parents that fell within the A $\beta$  network.

We tested for the enrichment of the network in *de novo* mutations among AD patients using a one-sided binomial test and adopting the highest baseline estimation as reference to generate a *P*-value as conservative as possible.

### Analysis of PSEN1-splicing mutation

Total RNA was extracted from patient's peripheral mononuclear cells collected in PAXgene Blood RNA tubes (Qiagen, Hilden, Germany) and reverse-transcribed using the Verso cDNA synthesis kit (Fisher Scientific, Illkirch, France). WT and mutant alleles were amplified using the following primers: forward: 5'-TGGAATTTTGGTGGTGGG-3' and reverse: 5'-CCAATCCAAGTTTACTCCCCTT-3'.

### Analysis of VPS35 L625P mutant

The p.Leu625Pro (L625P) mutation was engineered into the GFP-VPS35 construct<sup>8</sup> using the QuikChange kit (Stratagene, Santa Clara, CA, USA) according to manufacturer's instructions. Two independent clones that contained the L625P mutation were used for subsequent experiments.

For immunoprecipitation experiments, the GFP-VPS35 constructs were transiently transfected into HeLa cells, and WT and mutant forms of GFP-VPS35 were immunoprecipitated, analyzed by western blotting and detected using <sup>125</sup>I-protein A as previously described.<sup>9–11</sup>

For immunofluorescence staining, HeLa cells transiently transfected with GFP-VPS35 constructs were fixed and stained 24 h post transfection as previously described.<sup>12</sup> Imaging was performed using either a Zeiss Axioplan epifluorescence microscope, or the Cellomics ArrayScan V automated microscope for automated microscopy.<sup>10</sup>

#### Conformational dynamics of MARK4 WT and mutant

The microtubule affinity-regulating kinase (MARK)4 sequence (Uniprot ID: Q96L34), was mutated ( $\Delta$ G316, E317D) and three-dimensional structure for both WT and mutant were predicted using a comparative modeling tool Modeler 9v11 embedded in Discovery Studio Suite. The crystal structure of MARK2 in active state (PDB ID: 3IEC)<sup>13</sup> was used as template, and predicted models were refined and validated, with ATP and Magnesium ion docked into the catalytic cleft using Glide module in Schrödinger Suite<sup>14</sup> to represent the functional state of protein. GROMACS v4.5.5 (ref. 15) molecular dynamics simulation suite was used to assess their stability and conformational dynamics, where root means square deviation and root means square fluctuation plots were taken as a measure of stability. Both mutant and WT-predicted models hold conventional DFG-in/ $\alpha$ C-in mode with outward folding of T-loop favorable for binding downstream substrate proteins with due changes as a result of their mutation. Adenosine triphosphate binding was stabilized by coordinate bonds with Mg<sup>2+</sup> and hydrogen bonds with K88 of P-loop and A138 of hinge region in the MARK4 catalytic cleft. The predicted models have maintained their structural stability (Supplementary Figure S6) in the given time scale and hence used to study the structural changes between mutant and WT MARK4.

Analysis of Tau Ser262 phosphorylation by WT and mutant MARK4 Plasmid encoding the full-length human MARK4 cDNA cloned into the pcDNA3 expression vector is a kind gift from Dr Alessandro Beghini. The mutant MARK4 was generated using the Site-Directed Mutagenesis kit XL-II (Agilent) according to manufacturer's instructions.

Plasmids encoding Tau and the MARK4 constructs were transiently transfected into HEK293 cells. Twenty-four hours after transfection, proteins were extracted and analyzed by western blotting. Blots were stained with antibodies directed against MARK4 (1:2000; Abcam, Cambridge, UK), total Tau (1:50000; Dako, Santa Clara, CA, USA), Tau-PhosphoSer262 (1:2000; Invitrogen, Illkirch, France) and Actin-2066 (1:10000; Sigma, Lyon, France). Differential Ser262 staining by WT or mutant MARK4 was compared using a paired *t*-test after log transformation of normalized staining values.

## RESULTS

We searched for *de novo* variations in the index cases of these trios using a two-step procedure. First, these patients were screened for CNVs by high-resolution array-comparative genomic hybridization allowing for a resolution of about 10 kb. An average of 47 CNVs (range 25–59) was detected per patient, three in average corresponding to rare CNVs encompassing genes. Subsequent analysis of rare CNVs intersecting genes in patients and their parents by quantitative multiplex PCR of short fluorescent fragments showed that all but two of these variants were inherited. In patient EXT 773, we observed a 7.6 Mb *de novo* 21q21 duplication, which encompassed *APP* and 15 other genes (Figure 1a and Supplementary Figure S1). Patient EXT 804 bore a 12 kb *de novo* deletion involving the intron 1 of *BACE2* (Figure 1b

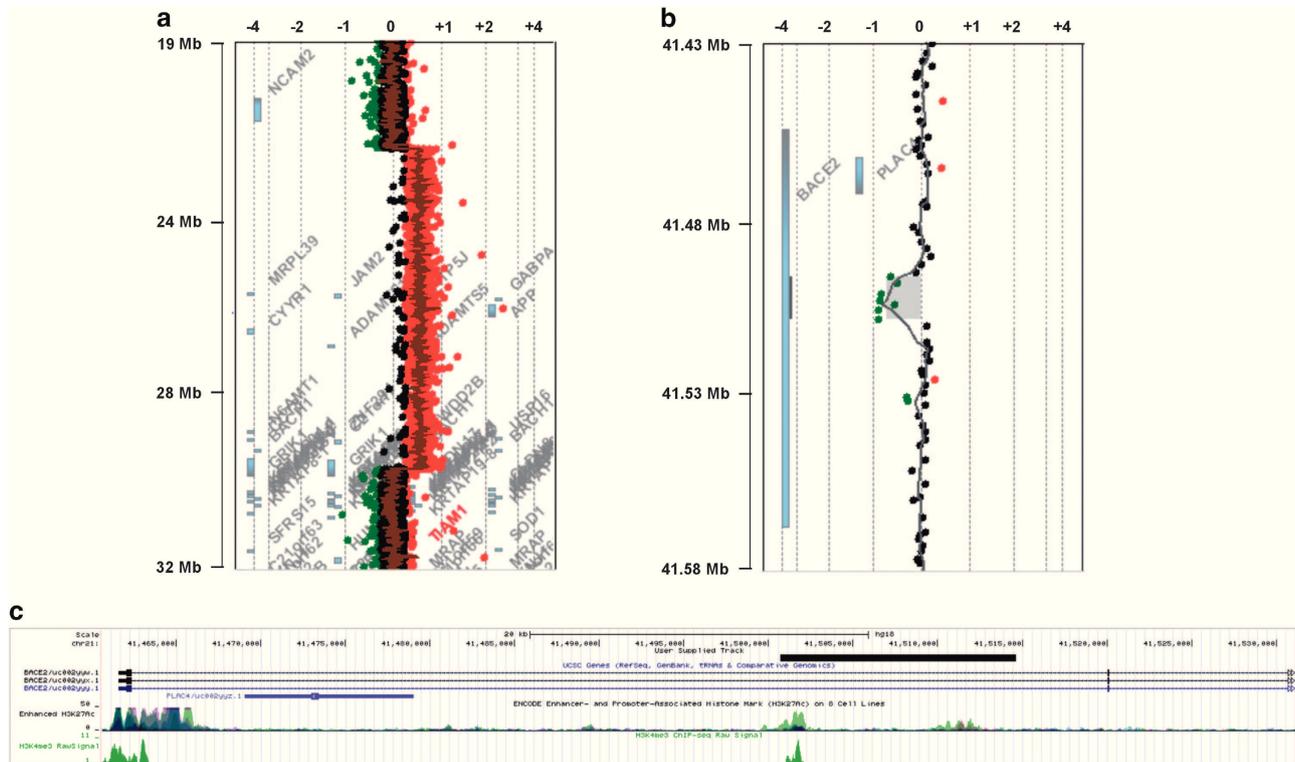
and Supplementary Figure S1), which does not overlap with any CNV reported in the database of genomic variants (URL: <http://dgv.tcag.ca/>).

In the 12 remaining trios, we then conducted a whole-exome sequencing screening. We generated ~4.5 Gb of sequence per individual with a mean depth of 96.88-fold on RefSeq coding exons and canonical splice sites ( $\pm$  2 bp). At this depth of coverage, an average of 81 and 89.8% of RefSeq coding exons and splice sites were jointly covered by at least 20 high-quality reads among the three trio members for exomes captured with Agilent SureSelect Human All Exon V4+UTRs and V5 kits, respectively (Supplementary Figure S2). Among the 20 001 exonic and splicing variants observed in average per exome, we identified 15 *de novo* variants (DNVs) in seven individuals (Table 2). All were confirmed by Sanger sequencing. The observed rate of *de novo* events per individual in the RefSeq protein coding exons was 1.25, corresponding to an exonic point mutation rate of  $2.14 \times 10^{-8}$ , in accordance with previous estimates.<sup>16</sup>

The distribution of the number of DNVs *per* trio did not differ from the Poisson distribution ( $P=0.86$ , with one expected DNV per individual). Among the 15 DNVs, we observed 3 synonymous and 12 amino-acid-altering DNVs, including 10 missense variants, 1 splice site variant and one MNV. With the exception of the *PHKA2* p.Arg290His variant (found once in the 1000 Genomes cohort), none of them was present in variants databases. The non-synonymous to synonymous ratio was 3.33. Five of the 10 missense variants were classified as damaging by at least 2 of 3 prediction programs (SIFT, Polyphen2 and Mutation Taster) (Table 2).

#### An A $\beta$ -centered network

It is widely recognized that the A $\beta$  peptide, resulting from the sequential cleavage of the APP, has a key role in the pathogenesis of AD.<sup>17</sup> Therefore, to interpret the biological significance of the DNVs identified in our trios, we registered all genes whose protein product was known to interact with A $\beta$  production, aggregation, clearance or synaptic toxicity. We undertook a manual curation to collate, via a comprehensive literature search, all relevant information. As a result, a list of 323 genes defining an A $\beta$ -centered network was defined (Supplementary Table S1). Then, for each trio analyzed by whole-exome sequencing, we calculated the number of coding nucleotide positions/canonical splice sites successfully read at at least 20 $\times$  in all trio members, summed these numbers across the 12 trios and dichotomized them between positions corresponding to genes included or not within the network. Three non-synonymous/splice site DNVs (all predicted to be damaging) affecting the *PSEN1*, *VPS35* and *MARK4* genes were identified among the 7 483 907 base pairs jointly read at 20 $\times$  within the network, and nine (four predicted to be damaging) among the 351 165 260 base pairs outside the network. This corresponds to a significant enrichment of non-synonymous/splice site DNVs within the network ( $P=0.0018$ , odds ratio = 15.3 with 95% confidence interval (2.7;61.3), Fisher's exact test). When only considering variants predicted to be damaging, with three of seven non-synonymous/splice site DNVs hitting the network, the *P*-value was 0.0003, odds ratio = 34.5 with 95% confidence interval (5.1;203.3)). To ensure that this enrichment was not the consequence of a non-specific high rate of *de novo* mutations within the genes of the network, we applied the same procedure to synonymous DNVs. No enrichment in synonymous DNVs within the network's genetic surface was found ( $P=1$ ). We also verified that the proportion of jointly covered base pairs did not significantly differ in genes of the network compared with other genes ( $P=0.7108$ , Mann–Whitney *U*-test). To further ensure that the enrichment was not artifactual, we finally computed several estimates for the expected relative *de novo* mutation rate of the network with respect to the entire exome in the general population (see methods and Supplementary Table S2). We



**Figure 1.** Detailed view of the rearrangements detected by 1M array-CGH. **(a)** The patient EXT 773 carries a 7.6 Mb duplication (chr21:23,561,982-31,154,110, hg19) that encompasses the *APP* gene and 15 other genes. **(b)** The patient EXT 804 carries a 12 kb intragenic deletion of *BACE2*. **(c)** Snapshot of genomic region 41.46–41.53 Mb of chr21 displayed using the UCSC genome browser (GRCh37/hg19) showing the intronic location of the *BACE2* partial deletion (black line). The RefSeq Genes track shows the exon/intron structure of *BACE2* transcripts and the *PLAC4* intronless gene. The H3K4Me3 Raw Signal and ENCODE H3K27Ac tracks show the location and levels of enrichment of H3K4Me3 and H3K27Ac histone marks, which are molecular correlates of promoters and enhancers, respectively. See also Supplementary Figure S1.

estimated the expected relative *de novo* mutation rate of the network by the ratio of the variant burdens found within the network and in the whole-exome sequence among the parents. Four classes of variants, which could be affected differently by selective pressure and genetic drift, namely (i) synonymous, (ii) missense and splice variants, (iii) rare synonymous and (iv) rare missense and splice variants, were considered to obtain a variety of estimates. Compared with the highest estimation of the relative *de novo* mutation rate of the network, three hits among 12 within the network corresponded to a significant enrichment in our EOAD patients ( $P=0.0032$ , binomial test). Omitting the trio bearing the *PSEN1* mutation, with two out of nine mutations hitting the network, the effect remained significant ( $P=0.0212$ ). We conclude that the excess of DNVs among a set of genes *a priori* classified as etiologically relevant in our EOAD patients supports the hypothesis that rare mutations spread among genes of the A $\beta$  network are involved in the genetic determinism of sporadic EOAD. Overall, when taking into account the *APP* duplication and the *BACE2* intronic deletion, 9 out of 14 patients carried at least one *de novo* genetic variation, among which five of them affected a gene of the A $\beta$  network (Table 2). Note that we did not find arguments in favor of mosaicism for all *de novo* mutations (Supplementary Figure S1), suggesting that they all are germline events. We next focused on the biological consequences of these five variations.

#### Biological consequences

***APP* duplication.** The interpretation of the 21q21 duplication is straightforward. We had previously established that *APP*

duplication, leading to a 1.5-fold increase in *APP* expression,<sup>18</sup> is a cause of AD with cerebral amyloid angiopathy.<sup>19</sup> Consistently, the neuropathological examination of this patient showed Braak stage VI AD lesions with prominent cerebral amyloid angiopathy. Although 15 other genes were involved in this duplication (Figure 1), it is now well-established that the duplication of *APP* alone is sufficient to cause AD.<sup>20</sup>

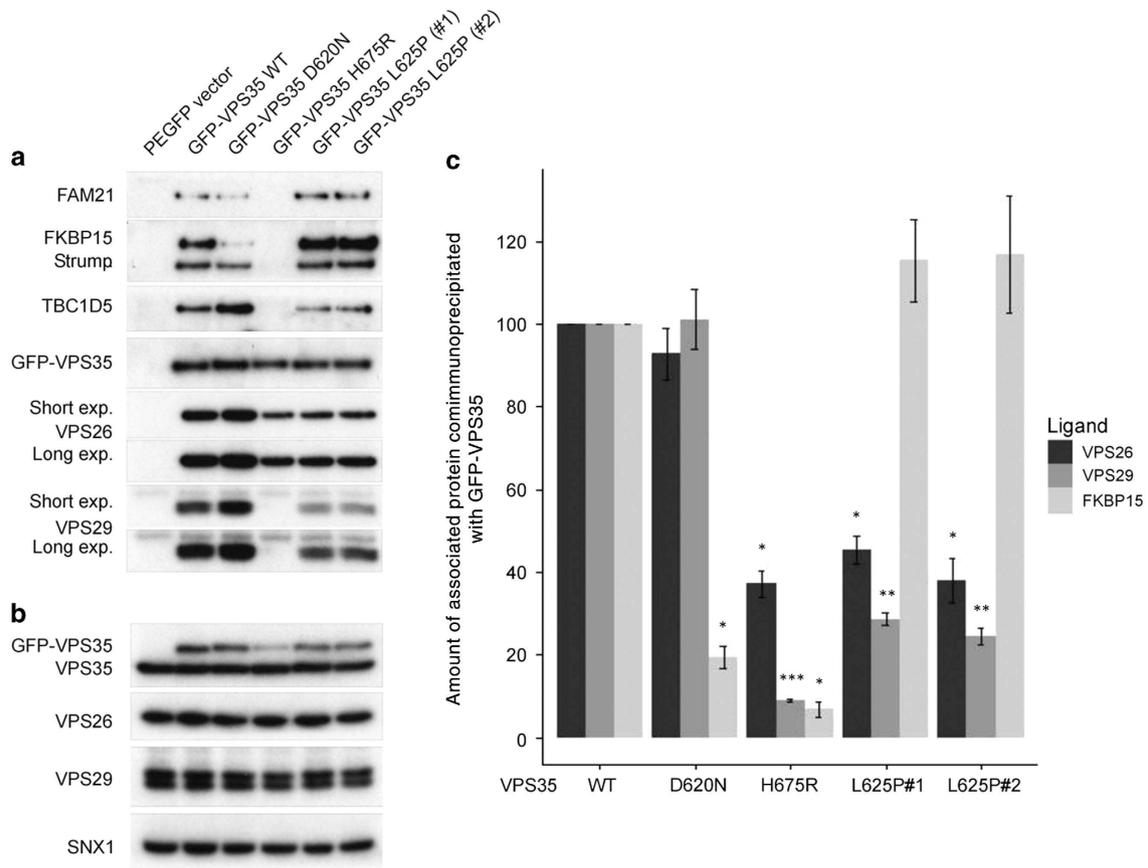
***BACE2* deletion.** The *BACE2* deletion is located at the end of intron 1, and overlaps enriched H3K4Me3 and H3K27Ac histone marks (Figure 1c and Supplementary Figure S1), which are molecular correlates of promoters and enhancers, respectively.<sup>21,22</sup> The *BACE2* gene encodes a protease known both to degrade A $\beta$  intracellularly and to lower A $\beta$  level via a  $\theta$ -secretase cleavage of APP leading to non-amyloidogenic processing.<sup>23,24</sup> Therefore, reduced *BACE2* expression is predicted to result in a proamyloidogenic effect. To investigate whether this deletion could have an effect on *BACE2* expression, we performed reverse transcription-quantitative multiplex PCR of short fluorescent fragments experiments on cultured fibroblasts of the patient and 10 control individuals (see Supplementary Methods). Although *BACE2* mRNA expression in the patient's fibroblasts was among the lowest measured in this sample, it was not possible to firmly conclude on the putative proamyloidogenic effect of this mutation because of the large inter-individual variability among controls (data not shown).

***PSEN1* c.869-2A>G mutation.** The c.869-2A>G *PSEN1* mutation is located on the splice acceptor site of exon 9, and is thus predicted to lead to the exclusion of the exon. Reverse

**Table 2.** *De novo* variants identified in the 14 EOAD trios

Trio ID		De novo CNV					
EXT 804		12-kb deletion within BACE2 intron 1					
EXT 773		7.6-Mb duplication encompassing APP gene					
<i>De novo</i> variants identified by whole-exome sequencing							
	Gene (NM ID)	Type	c.DNA level change	Protein level change	Mutation taster	Polyphen2	SIFT
EXT 235	<b>PSEN1</b> (NM_00021.3)	Splice	c.869-2A>G	Exclusion of exon 9 and Ser290Cys change at aberrant exon 8–10 junction	Disease causing (1)	NA	NA
	CASC5 (NM_170589.4)	Missense	c.2533C>T	p.Leu845Phe	Pol (1)	Benign (0)	Tolerated (0.4)
	ZDHHC13 (NM_019028.2)	Missense	c.632T>C	p.Val211Ala	Pol (0.738)	Benign (0)	Tolerated (0.61)
ROU 682	<b>VPS35</b> (NM_018206.4)	Missense	c.1874T>C	p.Leu625Pro	Disease causing (1)	Probably damaging (1)	Del (0.01)
	OXTR (NM_000916.3)	Synonymous	c.714G>T	p.Ala238Ala	NA	NA	NA
ROU 1331	<b>MARK4</b> (NM_001199867.1)	MNV	c.946_951delGGTGAGinsGAT	p.Gly316_Glu317delinsAsp	Disease causing (0.96)	NA	NA
	ABCA12 (NM_173076.2)	Missense	c.2067C>A	p.Asp689Glu	Pol (0.924)	Benign (0.012)	Tolerated (0.19)
	MYCT1 (NM_025107.2)	Missense	c.580C>T	p.Pro194Ser	Pol (1)	Benign (0.001)	Tolerated (0.88)
	ACO2 (NM_001098.2)	Synonymous	c.1443C>T	p.Asp481Asp	NA	NA	NA
ROU 1322	ADH7 (NM_000673.4)	Synonymous	c.627C>T	p.Val209Val	NA	NA	NA
EXT 395	CPM (NM_001874.4)	Missense	c.352C>T	p.Arg118Trp	Disease causing (1)	Probably damaging (1)	Del (0)
	RNF213 (NM_001256071.2)	Missense	c.5114C>T	p.Thr1705Met	Disease causing (0.79)	Probably damaging (0.987)	Del (0)
EXT 488	SLC25A27 (NM_004277.4)	Missense	c.811A>G	p.Lys271Glu	Disease causing (0.998)	Possibly damaging (0.835)	Tolerated (0.06)
	PHKA2 (NM_000292.2)	Missense	c.869G>A	p.Arg290His	Disease causing (1)	Possibly damaging (0.878)	Del (0.02)
EXT 687	RGS13 (NM_144766.2)	Missense	c.227G>A	p.Arg76Gln	Pol (1)	Benign (0.002)	Tolerated (0.23)
EXT 306	–	–	–	–	–	–	–
EXT 186	–	–	–	–	–	–	–
EXT 344	–	–	–	–	–	–	–
EXT 573	–	–	–	–	–	–	–
EXT 576	–	–	–	–	–	–	–

Abbreviations: CNV, copy number variation; Del, deleterious; NA, not applicable; Pol, polymorphism. *De novo* CNV identified by array-CGH are represented at the top of the table. Trios with no *de novo* CNV were analyzed by WES. *De novo* single-nucleotide variants or MNV are listed in the lower part of the table. Genes in bold are those belonging to the A $\beta$  network. Predicted effect on protein structure is assessed by three prediction software programs: Mutation Taster, Polyphen2 and SIFT.

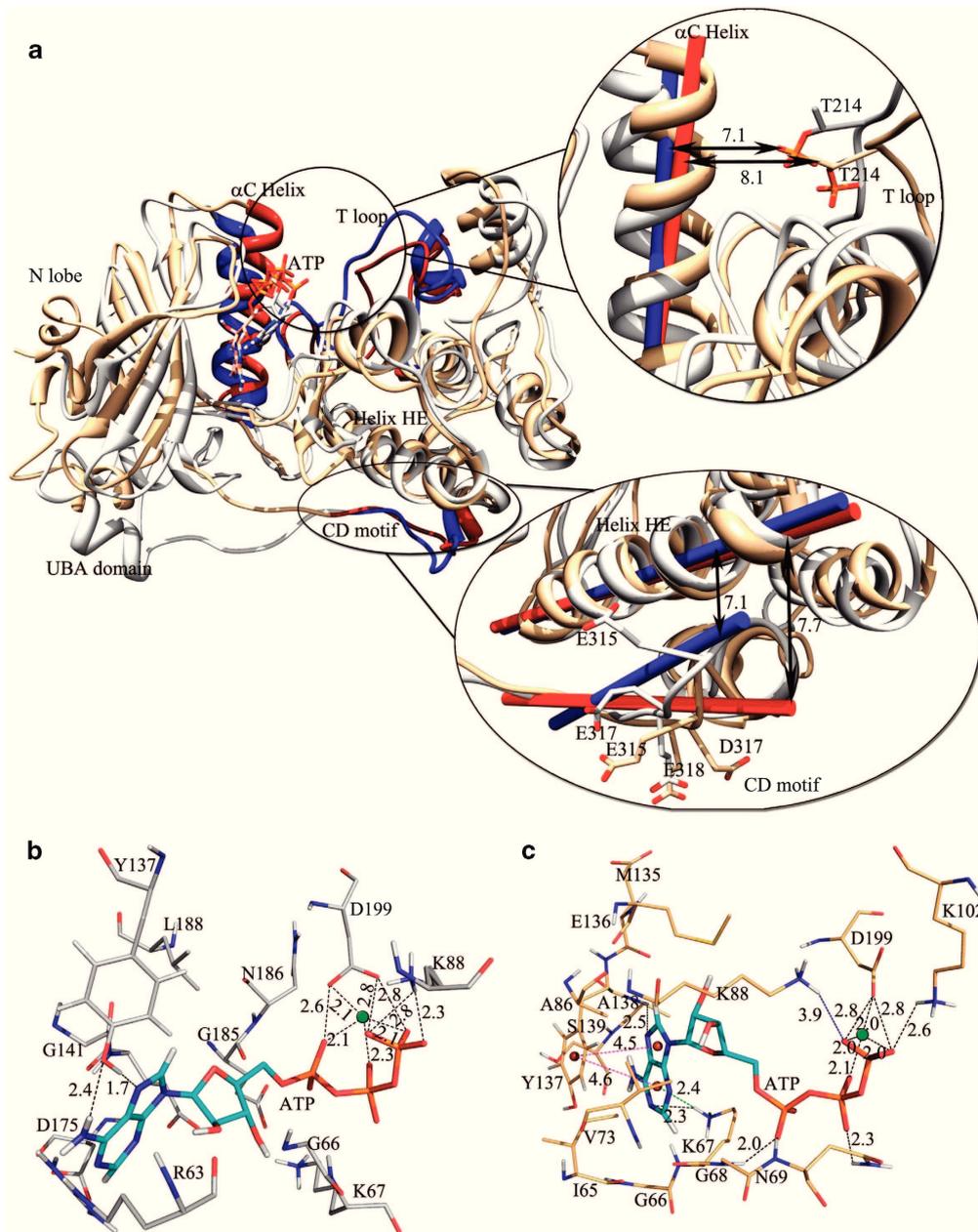


**Figure 2.** VPS35 L625P coimmunoprecipitation with retromer components and associated proteins. HeLa cells were transiently transfected with GFP-tagged VPS35 constructs. Two independent VPS35 L625P clones (#1 and #2) were analyzed. Lysates were immunoprecipitated with anti-GFP. **(a)** Detection of retromer and associated proteins by western blotting. **(b)** Analysis of lysates (1% total) by SDS-PAGE and western blotting, confirming the presence of the transfected construct and other retromer components in the initial lysates. **(c)** Quantification of the amount of proteins coimmunoprecipitated with VPS35, normalized to VPS35 signal and expressed as a % compared with WT. Results were compared using a paired *t*-test after log transformation. Error bar denotes s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . See also Supplementary Figure S5.

transcription-PCR experiments confirmed the skipping of exon 9, resulting in the introduction of a missense mutation (p.Ser290Cys) at the aberrant exon 8/10 splice junction (Supplementary Figure S4). The effect on protein structure is thus similar to that of the well-known c.869-1 G>T substitution,<sup>25</sup> whose pathological consequences on A $\beta$ 2 production have been examined in detail by Steiner *et al.*<sup>26</sup>

**VPS35 p.Leu625Pro mutation.** The VPS35 protein assembles with the other vacuolar protein-sorting VPS29 and VPS26 to form the cargo-selective subcomplex of the retromer complex, which mediates endosomal protein sorting into the endosome-to-Golgi and endosome-to-cell surface recycling pathways. Notably, the retromer complex controls the localization of the membrane protein SorLA, which has a crucial role in APP recycling and A $\beta$  lysosomal degradation. SorLA is encoded by the *SORL1* gene, for which both rare and common variants have been associated with AD risk.<sup>27,28</sup> In addition, the retromer cargo-selective subcomplex also functions as a hub for recruiting other machinery involved in endosomal protein sorting, including the WASH complex (comprising WASH1, strumpellin, FAM21, KIAA1033 and CCDC53), FKBP15 and TBC1D5.<sup>29</sup> A rare missense mutation of VPS35, p.Asp620Asn (D620N), has been found to cause Parkinson disease and results in impaired association of the retromer cargo-selective subcomplex with the WASH complex.<sup>30,31</sup>

The L625P mutation is located in the VPS35 C-terminal domain involved in VPS29 binding, similar to the D620N missense mutation. When HeLa cells were transiently transfected with a GFP-tagged VPS35 L625P mutant, the mutation reduced the amount of VPS29 that co-immunoprecipitates with the tagged VPS35 (Figure 2), although the effect was not as pronounced as that of the H675R mutation, a mutation initially characterized in *Saccharomyces Cerevisiae* that totally abolishes the binding to VPS29.<sup>9</sup> The ability of the L625P mutant to co-immunoprecipitates VPS26 also appeared compromised relative to WT VPS35 (Figure 2). Interestingly, the L625P mutation exhibited strikingly different properties than the Parkinson disease-causing D620N mutation that binds VPS26 and VPS29 normally but associates weakly with FKBP15 and the WASH complex proteins FAM21 and strumpellin. By contrast, the L625P mutant could co-immunoprecipitate FKBP15 and the WASH complex similarly to WT VPS35. However, as the association of TBC1D5 with the retromer cargo-selective subcomplex depends on binding to VPS29, the reduced binding of VPS29 to the VPS35 L625P mutant also reduced the amount of TBC1D5 that co-immunoprecipitates with the GFP-tagged VPS35 L625P mutant (Figure 2). Consistent with these findings, when the cellular localization of the L625P mutant was examined, we observed that the association of VPS35 L625P to endosomes was markedly reduced compared with WT VPS35 or the D620N mutant (Supplementary Figure S5). We

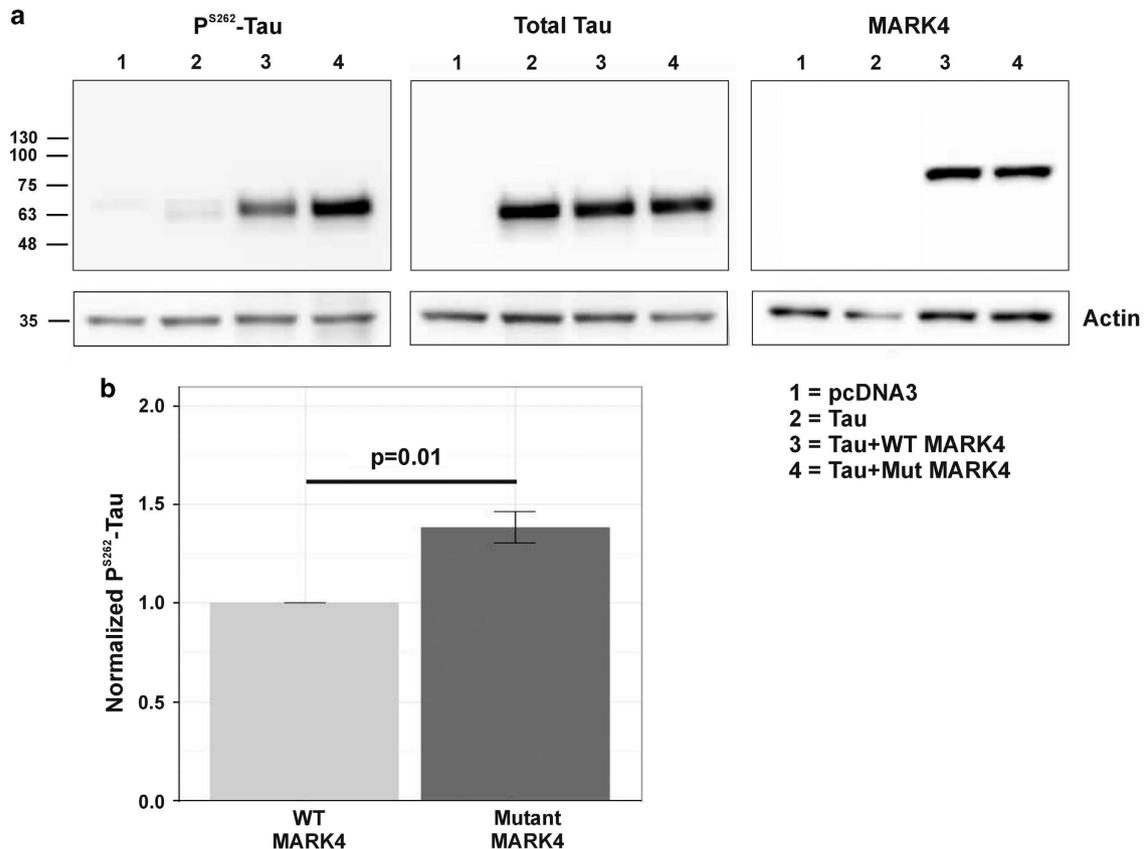


**Figure 3.** Structural changes observed as a result of  $\Delta$ G316, E317D changes in the CD motif of MARK4. **(a)** Superimposition of predicted model of wild-type (in gray) and mutant (in beige) MARK4, and their denoted structural features. Movement of T-loop in wild-type and mutant models; the distance was calculated between the helical axis of  $\alpha$ C helix and p-Thr214 from the T-loop in C lobe. Displacement of CD motif in the linker region between kinase and ubiquitin-associated (UBA) domain; the distance is calculated between the helical axis of helix HE in C lobe and the CD motif. The change in the exposure of residues E317 (WT), D317 (mutated) and E315, E318 of CD motif in both the wild-type and mutated forms are also displayed. Blue regions represent wild-type MARK4, whereas red regions represent MARK4 mutant. **(b)** and **(c)** Representation of molecular interactions between ATP and the residues in MARK4 Kinase-UBA domain of wild-type and mutant, respectively. (Black-dotted line: hydrogen bond formation, pink dotted lines: P-Pi interactions, green lines: Pi-Cation, blue-dotted lines: charge interactions. Distances represented in Å units. See also Supplementary Figure S6.

therefore conclude that the L625P mutant, acting by a mechanism different from the Parkinson disease-causing D620N mutation, behaves as a partial loss-of-function mutant that shows reduced ability to form functional complexes with both VPS29 and VPS26. To interpret this loss-of-function mechanism in the context of AD, we should remember that reduction of VPS35 expression has been shown to modulate APP processing in transgenic animals or cell cultures. Among the 11 studies addressing the role of loss of

VPS35 function,<sup>27,32–41</sup> all but one concluded that a reduction in VPS35 expression resulted in increased A $\beta$  production.

**MARK4 p.Gly316\_Glu317delinsAsp MNV.** MARK4 is a member of the MARK protein family, consisting of four highly conserved members (MARK1–4) and is involved in the early phosphorylation of Tau at Ser262, inducing detachment of Tau from microtubules and priming it for further phosphorylation by other kinases such



**Figure 4.** Tau Ser262 phosphorylation by wild-type and mutant MARK4. **(a)** HEK293 cells were transiently transfected with the empty pcDNA3 vector (lane 1), 2N4R Tau (lane 2) or co-transfected with 2N4R Tau and either the wild-type (lane 3) or mutant p.Gly316\_Glu317delinsAsp MARK4 L (lane 4). Phospho<sup>Ser262</sup> Tau, total Tau and MARK4 were detected by western blotting (upper panels). Band intensities were quantified and normalized to the actin staining (lower panels). All experiments were performed in four biological replicates. **(b)** The phospho<sup>Ser262</sup>-Tau staining was normalized to the levels of overexpressed Tau and MARK4. Differential Ser262 staining following WT or mutant MARK4 overexpression was compared using a paired *t*-test after log transformation of normalized staining values. Error bar denotes s.e.m.

as GSK3 $\beta$  and Cdk5.<sup>42</sup> MARK4 appears to have a more pronounced role than other MARKs in this process.<sup>43</sup> The overexpression of MARK4 in primary cultures of rat hippocampal neurons leads to Tau hyperphosphorylation, reduced expression of synaptic markers and dendritic spine loss, a phenotype reminiscent of that observed after A $\beta$  oligomer treatment. Furthermore, pharmacological inhibition of MARK kinases abrogated the toxic effects of A $\beta$  oligomers on dendritic spines and synapses.<sup>44</sup> Accordingly, in *Drosophila* models, Tau Ser262 phosphorylation is critical for A $\beta$ 42-induced Tau toxicity,<sup>45</sup> and PAR-1 kinase (the fly ortholog of MARKs) has an initiator role in the temporally ordered phosphorylation process that confers Tau toxicity.<sup>46</sup> Altogether, these data demonstrate that phosphorylation of Tau by PAR-1/MARK family kinases is a primary downstream event by which A $\beta$  exerts its toxicity on synapses and dendritic spines.

The MARK4 double amino-acid change ( $\Delta$ G316, E317D) caused by the MNV is located in a negatively charged short motif linker that resembles the common docking site in MAP kinases<sup>47</sup> and is known to tether the kinase to partners, that is, substrates or regulators.<sup>48</sup> To gain insight on the structural modifications induced by these changes, we predicted the three-dimensional structure of WT and mutant form of MARK4 kinase-ubiquitin-associated (UBA) domain in active form (see methods).

The two amino-acid changes did not modify the conformation of the common docking motif. However, they altered the structural arrangement of the common docking motif with respect to the above-lying helix HE of the C lobe. They increased

the distance from the original 7.1 Å observed in WT form to ~7.7 Å, thereby exposing the charged residues E315, E317 and E318, which would aid in strengthening the association of MARK4 with partner proteins (Figure 3a). Superimposition of the kinase-UBA regions of mutant form with the WT active form displays drift in the position of N and C lobes, wherein the distance between the lobes is about 17.8 Å when mutated compared with 22.4 Å in WT form, signifying that they move toward the ATP-binding cleft. As a result, the interaction between ATP and the residues of the catalytic cleft is enhanced by the addition of several new interactions. In the WT, ATP forms hydrogen bonds with K88 and Y137, and in mutant form, in addition to these interactions, it forms hydrogen bonds with K67, G68, N69 of P-Loop, and with K102 of  $\alpha$ C helix (Figure 3b and Supplementary Figure S6). Moreover, the T-Loop in the mutant model shows considerable displacement with reference to its native position observed in the WT model (Figure 3a), which further increases the area of the substrate-binding region, facilitating a strong association of MARK4 with its associated proteins. These observations indicate that the mutation in the common docking motif tends to constrict the space between the N and C lobes, strengthens the association of ATP, and increases the volume of the substrate-binding cleft. These features are predicted to result in a stronger association between MARK4 and substrates, and an increase in phosphorylation activity.

To directly assess the consequence of the MARK4 MNV on Tau phosphorylation, we co-transfected HEK293 cells with Tau and

either the WT or mutant MARK4 constructs and analyzed Tau Ser262 phosphorylation by western blotting (Figure 4a). We observed a significant increase of Tau Ser262 phosphorylation ( $P=0.01$ ; paired *t*-test on log-transformed staining values) in cells transfected with the mutant compared to the WT construct (Figure 4b). We conclude that the MARK4 MNV results in a gain-of-function regarding the ability of MARK4 to phosphorylate Tau on Ser262.

## DISCUSSION

In complex diseases such as autism or schizophrenia, despite the wealth of data recently generated, interpreting the role of *de novo* mutations remains challenging. These mutations are often retrieved as singletons and inferring causality necessitates the arduous identification of genes recurrently hit by disruptive mutations.<sup>4</sup> However, this requires hundreds or thousands of trios as illustrated by the few genes so far identified with certainty in huge studies (including ~4500 trios) in autism spectrum disorder.<sup>49,50</sup> Given the difficulty of identifying recurrently hit genes, these studies have focused on the demonstration that *de novo* mutations aggregate within particular biological pathways.<sup>49–52</sup> In the context of EOAD, we took advantage of the large body of pathophysiological data accumulated in the past 20 years to define an *a priori* biological network containing potentially etiologically relevant genes. We show that this network is indeed significantly enriched in *de novo* mutations. Note that, as the delineation of the network is based on the current knowledge of A $\beta$  biology, we cannot exclude that some genes hit by *de novo* mutations and listed as 'out of the network' are in fact genuine A $\beta$  genetic interactors, which would lead to an even higher enrichment.

By sampling 14 EOAD sporadic cases, we detected two clearly pathogenic *de novo* events, known to be responsible for autosomal-dominant forms of the disease. It should be noted that this is the first *de novo* APP duplication reported so far and that only three proven PSEN1 *de novo* missense mutations had previously been reported in AD patients.<sup>53–55</sup> Mutations in sporadic AD cases have already been reported. However, it is necessary to confirm both parenthood and the absence of mutation in unaffected parents before claiming the *de novo* occurrence of a mutation. The parents are often deceased at the time of genetic analyses of their offspring, so that their DNA cannot be obtained. The role of *de novo* APP and PSENs causative events in sporadic EOAD is therefore probably underestimated. It should be noted that, given our limited sample size, no frequency data of *de novo* causative EOAD events should be inferred from this study.

For the APP duplication and the PSEN1 mutation, the functional consequences are well-known. For the other genes in the network, we undertook functional studies to demonstrate a significant effect of these mutations on relevant biological parameters. We succeeded for two genes. With VPS35 variation, our data support the role of the endosomal trafficking pathway in EOAD and underline the importance of the retromer genes in the genetic architecture of the disease. More importantly, in the context of current efforts to link A $\beta$  pathology to Tau pathology in AD, we provide the first genetic evidence that MARK4, a gene coding for a kinase of Tau, is involved in AD pathogenesis. By increasing Tau phosphorylation at position 262, this mutation is predicted to favor detachment of Tau from the microtubule network. Whether this mutation acts synergistically with A $\beta$  exposition to promote Tau pathology or totally bypasses it remains to be established.

Despite these functional arguments and contrary to the APP duplication and the PSEN1 mutation, which are currently considered as fully penetrant mutations, no definite conclusion should be drawn regarding the penetrance of the VPS35 and the MARK4 mutations, in the absence of segregation data in the next

generations. Indeed, as the onset of AD appears after the reproduction age, we may expect transmission of these rare variants. Thus, searching for the potential implication of rare variants in the genes hit by DNVs in this work in unexplained familial cases could be promising but will require huge sample sizes in a case–control setting or large pedigrees.

Owing to the age structure of AD pedigrees it was not possible to include late-onset AD cases in our sample. The question of whether *de novo* mutations also affect a subset of late-onset AD cases or will by necessity manifest with earlier onset disease cannot be addressed. We can only speculate that some *de novo* mutations might result in a late expression of AD. Indeed, some late-onset AD cases are sporadic and have parents deceased in an advanced age without AD. This could be consistent with the contribution of a *de novo* mutation, but also of incompletely penetrant inherited variants or of environmental factors.

In conclusion, our data support the validity of the family trio strategy. They reinforce the conclusions drawn in mental retardation, another condition for which sufficient knowledge of the underlying biological pathways allowed to establish causality for a large subset of *de novo* mutations identified in a family trio screen.<sup>56</sup> More generally, as the understanding of the pathophysiological grounds of complex diseases improves, this study design combining a network-based analysis with biological validation of variants will most likely become a choice strategy to identify etiologically relevant rare variants spread among a large collection of genes.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

Sponsored by CNR-MAJ and supported by grants from the Clinical Research Hospital Program from the French Ministry of Health (GMAJ, PHRC 2008/067). MNJS and SYB are funded through grants (G0701444 and MR/L007533/1) from the Medical Research Council awarded to MNJS. PJ is supported by grant BT/PR15118/BID/07/357/2011 from the Department of Biotechnology, Government of India. We thank the Integragen company for exome sequencing. We thank Pascal Chambon, Géraldine Joly, Soumeya Bekri and Stéphanie Torre for cultured fibroblasts, and Alessandro Beghini for the MARK4 plasmid. We thank Magalie Lecourtois and Emmanuelle Genin for helpful discussions and Tracey Avequin for editing the manuscript.

## REFERENCES

- Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C *et al*. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45**: 1452–1458.
- Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, Snaedal J *et al*. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* 2013; **368**: 107–116.
- Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, Guerreiro R *et al*. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* 2014; **505**: 550–554.
- Gratten J, Visscher PM, Mowry BJ, Wray NR. Interpreting the role of *de novo* protein-coding mutations in neuropsychiatric disease. *Nat Genet* 2013; **45**: 234–238.
- Veltman JA, Brunner HG. *De novo* mutations in human genetic disease. *Nat Rev Genet* 2012; **13**: 565–575.
- McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr., Kawas CH *et al*. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011; **7**: 263–269.
- Perreau VM, Orchard S, Adlard PA, Bellingham SA, Cappai R, Ciccotosto GD *et al*. A domain level interaction network of amyloid precursor protein and Abeta of Alzheimer's disease. *Proteomics* 2010; **10**: 2377–95.
- Gokool S, Tattersall D, Seaman MN. EHD1 interacts with retromer to stabilize SNX1 tubules and facilitate endosome-to-Golgi retrieval. *Traffic* 2007; **8**: 1873–1886.

- 9 Helfer E, Harbour ME, Henriot V, Lakisic G, Sousa-Blin C, Volceanov L et al. Endosomal recruitment of the WASH complex: active sequences and mutations impairing interaction with the retromer. *Biol Cell* 2013; **105**: 191–207.
- 10 Breusegem SY, Seaman MN. Image-based and biochemical assays to investigate endosomal protein sorting. *Methods Enzymol* 2014; **534**: 155–178.
- 11 Harbour ME, Breusegem SY, Antrobus R, Freeman C, Reid E, Seaman MN. The cargo-selective retromer complex is a recruiting hub for protein complexes that regulate endosomal tubule dynamics. *J Cell Sci* 2010; **123**: 3703–3717.
- 12 Seaman MN, Harbour ME, Tattersall D, Read E, Bright N. Membrane recruitment of the cargo-selective retromer subcomplex is catalysed by the small GTPase Rab7 and inhibited by the Rab-GAP TBC1D5. *J Cell Sci* 2009; **122**: 2371–2382.
- 13 Nesic D, Miller MC, Quinkert ZT, Stein M, Chait BT, Stebbins CE. Helicobacter pylori CagA inhibits PAR1-MARK family kinases by mimicking host substrates. *Nat Struct Mol Biol* 2010; **17**: 130–132.
- 14 Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* 2004; **47**: 1750–1759.
- 15 Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 2010; **78**: 1950–1958.
- 16 Campbell CD, Eichler EE. Properties and rates of germline mutations in humans. *Trends Genet* 2013; **29**: 575–584.
- 17 Hardy J, Selkoe DJ. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* 2002; **297**: 353–356.
- 18 Pottier C, Wallon D, Lecrux AR, Maltete D, Bombois S, Jurici S et al. Amyloid-beta protein precursor gene expression in alzheimer's disease and other conditions. *J Alzheimers Dis* 2012; **28**: 561–566.
- 19 Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerriere A, Vital A et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nature Genet* 2006; **38**: 24–26.
- 20 Slegers K, Brouwers N, Gijssels I, Theuns J, Goossens D, Wauters J et al. APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. *Brain* 2006; **129**: 2977–2983.
- 21 Maunakea AK, Nagarajan RP, Bilenny M, Ballinger TJ, D'Souza C, Fouse SD et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 2010; **466**: 253–257.
- 22 Encode Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**: 57–74.
- 23 Abdul-Hay SO, Sahara T, McBride M, Kang D, Leissring MA. Identification of BACE2 as an avid  $\beta$ -amyloid-degrading protease. *Mol Neurodegener* 2012; **7**: 46.
- 24 Sun X, He G, Song W. BACE2 as a novel APP theta-secretase, is not responsible for the pathogenesis of Alzheimer's disease in Down syndrome. *FASEB J* 2006; **20**: 1369–1376.
- 25 Perez-Tur J, Froelich S, Prihar G, Crook R, Baker M, Duff K et al. A mutation in Alzheimer's disease destroying a splice acceptor site in the presenilin-1 gene. *Neuroreport* 1995; **7**: 297–301.
- 26 Steiner H, Romig H, Grim MG, Philipp U, Pesold B, Citron M et al. The biological and pathological function of the presenilin-1 Deltaexon 9 mutation is independent of its defect to undergo proteolytic processing. *J Biol Chem* 1999; **274**: 7615–7618.
- 27 Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F et al. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat Genet* 2007; **39**: 168–177.
- 28 Pottier C, Hannequin D, Coutant S, Rovelet-Lecrux A, Wallon D, Rousseau S et al. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. *Mol Psychiatry* 2012; **17**: 875–879.
- 29 Seaman MN, Gautreau A, Billadeau DD. Retromer-mediated endosomal protein sorting: all WASHed up! *Trends Cell Biol* 2013; **23**: 522–528.
- 30 McGough IJ, Steinberg F, Jia D, Barbuti PA, McMillan KJ, Heesom KJ et al. Retromer binding to FAM21 and the WASH complex is perturbed by the Parkinson disease-linked VPS35(D620N) mutation. *Curr Biol* 2014; **24**: 1670–1676.
- 31 Zavodszky E, Seaman MN, Moreau K, Jimenez-Sanchez M, Breusegem SY, Harbour ME et al. Mutation in VPS35 associated with Parkinson's disease impairs WASH complex association and inhibits autophagy. *Nat Commun* 2014; **5**: 3828.
- 32 Small SA, Kent K, Pierce A, Leung C, Kang MS, Okada H et al. Model-guided microarray implicates the retromer complex in Alzheimer's disease. *Ann Neurol* 2005; **58**: 909–919.
- 33 Muhammad A, Flores I, Zhang H, Yu R, Staniszewski A, Planel E et al. Retromer deficiency observed in Alzheimer's disease causes hippocampal dysfunction, neurodegeneration, and Abeta accumulation. *Proc Natl Acad Sci USA* 2008; **105**: 7327–7332.
- 34 Sullivan CP, Jay AG, Stack EC, Pakaluk M, Wadlinger E, Fine RE et al. Retromer disruption promotes amyloidogenic APP processing. *Neurobiol Dis* 2011; **43**: 338–345.
- 35 Wen L, Tang FL, Hong Y, Luo SW, Wang CL, He W et al. VPS35 haploinsufficiency increases Alzheimer's disease neuropathology. *J Cell Biol* 2011; **195**: 765–779.
- 36 Bhalla A, Vetanovetz CP, Morel E, Chamoun Z, Di Paolo G, Small SA. The location and trafficking routes of the neuronal retromer and its role in amyloid precursor protein transport. *Neurobiol Dis* 2012; **47**: 126–134.
- 37 Choy RW, Cheng Z, Schekman R. Amyloid precursor protein (APP) traffics from the cell surface via endosomes for amyloid beta (Abeta) production in the trans-Golgi network. *Proc Natl Acad Sci USA* 2012; **109**: E2077–E2082.
- 38 Lane RF, Raines SM, Steele JW, Ehrlich ME, Lah JA, Small SA et al. Diabetes-associated SorCS1 regulates Alzheimer's amyloid-beta metabolism: evidence for involvement of SorL1 and the retromer complex. *J Neurosci* 2010; **30**: 13110–13115.
- 39 Vieira SI, Rebelo S, Esselmann H, Wiltfang J, Lah J, Lane R et al. Retrieval of the Alzheimer's amyloid precursor protein from the endosome to the TGN is S655 phosphorylation state-dependent and retromer-mediated. *Mol Neurodegener* 2010; **5**: 40.
- 40 Fjorback AW, Seaman M, Gustafsen C, Mehmedbasic A, Gokool S, Wu C et al. Retromer binds the FANSHY sorting motif in SorLA to regulate amyloid precursor protein sorting and processing. *J Neurosci* 2012; **32**: 1467–1480.
- 41 Mecozzi VJ, Berman DE, Simoes S, Vetanovetz C, Awal MR, Patel VM et al. Pharmacological chaperones stabilize retromer to limit APP processing. *Nat Chem Biol* 2014; **10**: 443–449.
- 42 Drewes G. MARKing tau for tangles and toxicity. *Trends Biochem Sci* 2004; **29**: 548–555.
- 43 Gu GJ, Lund H, Wu D, Blokzijl A, Classon C, von Euler G et al. Role of individual MARK isoforms in phosphorylation of tau at Ser(2)(6)(2) in Alzheimer's disease. *Neuromol Med* 2013; **15**: 458–469.
- 44 Yu W, Polepalli J, Wagh D, Rajadas J, Malenka R, Lu B. A critical role for the PAR-1/MARK-tau axis in mediating the toxic effects of Abeta on synapses and dendritic spines. *Hum Mol Genet* 2012; **21**: 1384–1390.
- 45 Iijima K, Gatt A, Iijima-Ando K. Tau Ser262 phosphorylation is critical for Abeta42-induced tau toxicity in a transgenic Drosophila model of Alzheimer's disease. *Hum Mol Genet* 2010; **19**: 2947–2957.
- 46 Nishimura I, Yang Y, Lu B. PAR-1 kinase plays an initiator role in a temporally ordered phosphorylation process that confers tau toxicity in Drosophila. *Cell* 2004; **116**: 617–682.
- 47 Naz F, Anjum F, Islam A, Ahmad F, Hassan MI. Microtubule affinity-regulating kinase 4: structure, function, and regulation. *Cell Biochem Biophys* 2013; **67**: 485–499.
- 48 Tanoue T, Adachi M, Moriguchi T, Nishida E. A conserved docking motif in MAP kinases common to substrates, activators and regulators. *Nat Cell Biol* 2000; **2**: 110–116.
- 49 Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014; **515**: 216–221.
- 50 De Rubeis S, He X, Goldberg AP, Poultnery CS, Samocha K, Cicek AE et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014; **515**: 209–215.
- 51 Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J et al. De novo gene disruptions in children on the autistic spectrum. *Neuron* 2012; **74**: 285–299.
- 52 Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012; **485**: 242–245.
- 53 Dumanchin C, Brice A, Campion D, Hannequin D, Martin C, Moreau V et al. De novo presenilin 1 mutations are rare in clinically sporadic, early onset Alzheimer's disease cases. French Alzheimer's Disease Study Group. *J Med Genet* 1998; **35**: 672–673.
- 54 Portet F, Dauvilliers Y, Campion D, Raux G, Hauw JJ, Lyon-Caen O et al. Very early onset AD with a de novo mutation in the presenilin 1 gene (Met 233 Leu). *Neurology* 2003; **61**: 1136–1137.
- 55 Golan MP, Styczyńska M, Józwiak K, Walecki J, Maruszak A, Pniewski J et al. Early-onset Alzheimer's disease with a de novo mutation in the presenilin 1 gene. *Exp Neurol* 2007; **208**: 264–268.
- 56 Visser LE, de Ligt J, Gilissen C, Janssen I, Stehouwer M, de Vries P et al. A de novo paradigm for mental retardation. *Nat Genet* 2010; **42**: 1109–1112.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (<http://www.nature.com/mp>)

## APPENDIX A

### Appendix A CNR-MAJ Collaborators

Claire Boutoleau-Bretonnière<sup>1</sup>, Mathieu Ceccaldi<sup>2</sup>, Julien Cogez<sup>3</sup>, Bernard Croisile<sup>4</sup>, Olivier Felician<sup>2</sup>, Annie Laquerrière<sup>5</sup>, Isabelle Le Ber<sup>6,7</sup>, Andre Maues De Paula<sup>8</sup>, Olivier Moreaud<sup>9</sup>, Jérémie Pariente<sup>10</sup>, Florence Pasquier<sup>11,12</sup>, Adeline Rollin-Sillaire<sup>11,12</sup>, Olivier Rouaud<sup>13</sup>, François Sellal<sup>14</sup>, Christel Thauvin<sup>15</sup>

#### Collaborators affiliations

<sup>1</sup>Department of Neurology, CMRR, Nantes University Hospital, Nantes, France.

<sup>2</sup>Neurology and Neuropsychology department, AP-HM La Timône Marseille University Hospital & INSERM UMR1106, University of Aix-Marseille, Marseille, France.

<sup>3</sup>Department of Neurology, CMRR, Caen University Hospital, Caen, France.

<sup>4</sup>Department of Neuropsychology, CMRR, Groupe Hospitalier Est, University Hospital, Bron, France.

<sup>5</sup>Neuropathology Laboratory, Rouen University Hospital, Rouen, France.

<sup>6</sup>CNR-MAJ, University Hospital of Paris La Pitié-Salpêtrière, Paris, France.

<sup>7</sup>CRCICM, IM2A, UMR-S975 AP-HP, University Hospital of Paris La Pitié-Salpêtrière, Paris, France.

<sup>8</sup>Neuropathology Laboratory, AP-HM La Timône Marseille University Hospital, Marseille, France.

<sup>9</sup>Department of Neurology, CMRR, Grenoble University Hospital, Grenoble, France.

<sup>10</sup>Neurology Department, CMRR, Purpan University Hospital, Toulouse, France.

<sup>11</sup>CNR-MAJ, University Hospital of Lille, Lille, France.

<sup>12</sup>CMRR, Lille University Hospital & EA1040 Université Lille Nord de France, Lille, France.

<sup>13</sup>Memory Centre, Department of Neurology, Centre Georges François Leclerc, University Hospital, Dijon, France.

<sup>14</sup>Department of Neurology, CMRR, Hôpitaux Civils de Colmar, Colmar, France.

<sup>15</sup>Department of Genetics, Dijon University Hospital, Dijon, France.

## Références

- 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65.
- 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74.
- Acuna-Hidalgo, R., Veltman, J.A., and Hoischen, A. (2016). New insights into the generation and role of de novo mutations in health and disease. *Genome Biol.* 17, 241.
- Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* Chapter 7, Unit7.20.
- Alexander, J., Kalev, O., Mehrabian, S., Traykov, L., Raycheva, M., Kanakis, D., Drineas, P., Lutz, M.I., Ströbel, T., Penz, T., et al. (2016). Familial early-onset dementia with complex neuropathologic phenotype and genomic background. *Neurobiol. Aging* 42, 199–204.
- Alkan, C., Coe, B.P., and Eichler, E.E. (2011). Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* 12, 363–376.
- Alzheimer, A., Stelzmann, R.A., Schnitzlein, H.N., and Murtagh, F.R. (1995). An English translation of Alzheimer's 1907 paper, "Über eine eigenartige Erkrankung der Hirnrinde." *Clin. Anat. N. Y.* N 8, 429–431.
- Andersen, O.M., Reiche, J., Schmidt, V., Gotthardt, M., Spoelgen, R., Behlke, J., von Arnim, C.A.F., Breiderhoff, T., Jansen, P., Wu, X., et al. (2005). Neuronal sorting protein-related receptor sorLA/LR11 regulates processing of the amyloid precursor protein. *Proc. Natl. Acad. Sci. U. S. A.* 102, 13461–13466.
- Andersen, O.M., Rudolph, I.-M., and Willnow, T.E. (2016). Risk factor SORL1: from genetic association to functional validation in Alzheimer's disease. *Acta Neuropathol. (Berl.)* 132, 653–665.
- Backenroth, D., Homsy, J., Murillo, L.R., Glessner, J., Lin, E., Brueckner, M., Lifton, R., Goldmuntz, E., Chung, W.K., and Shen, Y. (2014). CANOES: detecting rare copy number variants from whole exome sequencing data. *Nucleic Acids Res.* 42, e97.
- Bai, X., Yan, C., Yang, G., Lu, P., Ma, D., Sun, L., Zhou, R., Scheres, S.H.W., and Shi, Y. (2015). An atomic structure of human  $\gamma$ -secretase. *Nature* 525, 212–217.
- Barnett, I.J., Lee, S., and Lin, X. (2013). Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. *Genet. Epidemiol.* 37, 142–151.
- Beattie, E.C., Carroll, R.C., Yu, X., Morishita, W., Yasuda, H., von Zastrow, M., and Malenka, R.C. (2000). Regulation of AMPA receptor endocytosis by a signaling mechanism shared with LTD. *Nat. Neurosci.* 3, 1291–1300.
- Benilova, I., Karran, E., and De Strooper, B. (2012). The toxic A $\beta$  oligomer and Alzheimer's disease: an emperor in need of clothes. *Nat. Neurosci.* 15, 349–357.

- Bennett, S.T., Barnes, C., Cox, A., Davies, L., and Brown, C. (2005). Toward the 1,000 dollars human genome. *Pharmacogenomics* 6, 373–382.
- Bloom, G.S. (2014). Amyloid- $\beta$  and tau: the trigger and bullet in Alzheimer disease pathogenesis. *JAMA Neurol.* 71, 505–508.
- Boucher, B., and Jenna, S. (2013). Genetic interaction networks: better understand to better predict. *Front. Genet.* 4, 290.
- Braak, H., and Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol. (Berl.)* 82, 239–259.
- Brouwers, N., Van Cauwenberghe, C., Engelborghs, S., Lambert, J.-C., Bettens, K., Le Bastard, N., Pasquier, F., Montoya, A.G., Peeters, K., Mattheijssens, M., et al. (2012). Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites. *Mol. Psychiatry* 17, 223–233.
- Cacace, R., Van den Bossche, T., Engelborghs, S., Geerts, N., Laureys, A., Dillen, L., Graff, C., Thonberg, H., Chiang, H.-H., Pastor, P., et al. (2015). Rare Variants in PLD3 Do Not Affect Risk for Early-Onset Alzheimer Disease in a European Consortium Cohort. *Hum. Mutat.* 36, 1226–1235.
- Cacace, R., Sleegers, K., and Van Broeckhoven, C. (2016). Molecular genetics of early-onset Alzheimer's disease revisited. *Alzheimers Dement. J. Alzheimers Assoc.* 12, 733–748.
- Caglayan, S., Takagi-Niidome, S., Liao, F., Carlo, A.-S., Schmidt, V., Burgert, T., Kitago, Y., Füchtbauer, E.-M., Füchtbauer, A., Holtzman, D.M., et al. (2014). Lysosomal sorting of amyloid- $\beta$  by the SORLA receptor is impaired by a familial Alzheimer's disease mutation. *Sci. Transl. Med.* 6, 223ra20.
- Cai, X.D., Golde, T.E., and Younkin, S.G. (1993). Release of excess amyloid beta protein from a mutant amyloid beta protein precursor. *Science* 259, 514–516.
- de Calignon, A., Spires-Jones, T.L., Pitstick, R., Carlson, G.A., and Hyman, B.T. (2009). Tangle-bearing neurons survive despite disruption of membrane integrity in a mouse model of tauopathy. *J. Neuropathol. Exp. Neurol.* 68, 757–761.
- Campbell, C.D., and Eichler, E.E. (2013). Properties and rates of germline mutations in humans. *Trends Genet. TIG* 29, 575–584.
- Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., et al. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 22, 231–238.
- Carvalho, C.M.B., and Lupski, J.R. (2016). Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* 17, 224–238.
- Castellana, S., and Mazza, T. (2013). Congruency in the prediction of pathogenic missense mutations: state-of-the-art web-based tools. *Brief. Bioinform.* 14, 448–459.
- Chaisson, M.J.P., Wilson, R.K., and Eichler, E.E. (2015). Genetic variation and the de novo assembly of human genomes. *Nat. Rev. Genet.* 16, 627–640.

- Chapman, J., Rees, E., Harold, D., Ivanov, D., Gerrish, A., Sims, R., Hollingworth, P., Stretton, A., GERAD1 Consortium, Holmans, P., et al. (2013). A genome-wide study shows a limited contribution of rare copy number variants to Alzheimer's disease risk. *Hum. Mol. Genet.* *22*, 816–824.
- Chapuis, J., Flaig, A., Grenier-Boley, B., Eysert, F., Pottiez, V., Deloison, G., Vandeputte, A., Ayral, A.-M., Mendes, T., Desai, S., et al. (2016). Genome-wide, high-content siRNA screening identifies the Alzheimer's genetic risk factor FERMT2 as a major modulator of APP metabolism. *Acta Neuropathol.* (Berl.).
- Chávez-Gutiérrez, L., Bammens, L., Benilova, I., Vandersteen, A., Benurwar, M., Borgers, M., Lismont, S., Zhou, L., Van Cleynenbreugel, S., Esselmann, H., et al. (2012). The mechanism of  $\gamma$ -Secretase dysfunction in familial Alzheimer disease. *EMBO J.* *31*, 2261–2274.
- Chiang, C., Scott, A.J., Davis, J.R., Tsang, E.K., Li, X., Kim, Y., Hadzic, T., Damani, F.N., Ganel, L., GTEx Consortium, et al. (2017). The impact of structural variation on human gene expression. *Nat. Genet.*
- Citron, M., Oltersdorf, T., Haass, C., McConlogue, L., Hung, A.Y., Seubert, P., Vigo-Pelfrey, C., Lieberburg, I., and Selkoe, D.J. (1992). Mutation of the beta-amyloid precursor protein in familial Alzheimer's disease increases beta-protein production. *Nature* *360*, 672–674.
- Citron, M., Westaway, D., Xia, W., Carlson, G., Diehl, T., Levesque, G., Johnson-Wood, K., Lee, M., Seubert, P., Davis, A., et al. (1997). Mutant presenilins of Alzheimer's disease increase production of 42-residue amyloid beta-protein in both transfected cells and transgenic mice. *Nat. Med.* *3*, 67–72.
- Cleveland, D.W., Hwo, S.Y., and Kirschner, M.W. (1977). Purification of tau, a microtubule-associated protein that induces assembly of microtubules from purified tubulin. *J. Mol. Biol.* *116*, 207–225.
- Coe, B.P., Witherspoon, K., Rosenfeld, J.A., van Bon, B.W.M., Vulto-van Silfhout, A.T., Bosco, P., Friend, K.L., Baker, C., Buono, S., Vissers, L.E.L.M., et al. (2014). Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat. Genet.* *46*, 1063–1071.
- Condello, C., Yuan, P., Schain, A., and Grutzendler, J. (2015). Microglia constitute a barrier that prevents neurotoxic protofibrillar A $\beta$ 42 hotspots around plaques. *Nat. Commun.* *6*, 6176.
- Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., et al. (2010). Origins and functional impact of copy number variation in the human genome. *Nature* *464*, 704–712.
- Crary, J.F., Trojanowski, J.Q., Schneider, J.A., Abisambra, J.F., Abner, E.L., Alafuzoff, I., Arnold, S.E., Attems, J., Beach, T.G., Bigio, E.H., et al. (2014). Primary age-related tauopathy (PART): a common pathology associated with human aging. *Acta Neuropathol.* (Berl.) *128*, 755–766.
- Crook, R., Verkkoniemi, A., Perez-Tur, J., Mehta, N., Baker, M., Houlden, H., Farrer, M., Hutton, M., Lincoln, S., Hardy, J., et al. (1998). A variant of Alzheimer's disease with spastic paraparesis and unusual plaques due to deletion of exon 9 of presenilin 1. *Nat. Med.* *4*, 452–455.
- Cruchaga, C., Karch, C.M., Jin, S.C., Benitez, B.A., Cai, Y., Guerreiro, R., Harari, O., Norton, J., Budde, J., Bertelsen, S., et al. (2014). Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* *505*, 550–554.
- Cuyvers, E., and Sleegers, K. (2016). Genetic variations underlying Alzheimer's disease: evidence from genome-wide association studies and beyond. *Lancet Neurol.* *15*, 857–868.

- Cuyvers, E., De Roeck, A., Van den Bossche, T., Van Cauwenberghe, C., Bettens, K., Vermeulen, S., Mattheijssens, M., Peeters, K., Engelborghs, S., Vandenbulcke, M., et al. (2015). Mutations in ABCA7 in a Belgian cohort of Alzheimer's disease patients: a targeted resequencing study. *Lancet Neurol.* *14*, 814–822.
- Dawson, H.N., Ferreira, A., Eyster, M.V., Ghoshal, N., Binder, L.I., and Vitek, M.P. (2001). Inhibition of neuronal maturation in primary hippocampal neurons from tau deficient mice. *J. Cell Sci.* *114*, 1179–1187.
- De Jonghe, C., Esselens, C., Kumar-Singh, S., Craessaerts, K., Serneels, S., Checler, F., Annaert, W., Van Broeckhoven, C., and De Strooper, B. (2001). Pathogenic APP mutations near the gamma-secretase cleavage site differentially affect Abeta secretion and APP C-terminal fragment stability. *Hum. Mol. Genet.* *10*, 1665–1671.
- De Strooper, B., Saftig, P., Craessaerts, K., Vanderstichele, H., Guhde, G., Annaert, W., Von Figura, K., and Van Leuven, F. (1998). Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. *Nature* *391*, 387–390.
- Delacourte, A., David, J.P., Sergeant, N., Buée, L., Wattez, A., Vermersch, P., Ghozali, F., Fallet-Bianco, C., Pasquier, F., Lebert, F., et al. (1999). The biochemical pathway of neurofibrillary degeneration in aging and Alzheimer's disease. *Neurology* *52*, 1158–1165.
- Delay, C., Grenier-Boley, B., Amouyel, P., Dumont, J., and Lambert, J.-C. (2016). miRNA-dependent target regulation: functional characterization of single-nucleotide polymorphisms identified in genome-wide association studies of Alzheimer's disease. *Alzheimers Res. Ther.* *8*, 20.
- Dewey, F.E., Murray, M.F., Overton, J.D., Habegger, L., Leader, J.B., Fetterolf, S.N., O'Dushlaine, C., Van Hout, C.V., Staples, J., Gonzaga-Jauregui, C., et al. (2016). Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* *354*.
- Dodson, S.E., Andersen, O.M., Karmali, V., Fritz, J.J., Cheng, D., Peng, J., Levey, A.I., Willnow, T.E., and Lah, J.J. (2008). Loss of LR11/SORLA enhances early pathology in a mouse model of amyloidosis: evidence for a proximal role in Alzheimer's disease. *J. Neurosci. Off. J. Soc. Neurosci.* *28*, 12877–12886.
- Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., and Liu, X. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* *24*, 2125–2137.
- Dourlen, P., Fernandez-Gomez, F.J., Dupont, C., Grenier-Boley, B., Bellenguez, C., Obriot, H., Caillierez, R., Sottejeau, Y., Chapuis, J., Bretteville, A., et al. (2016). Functional screening of Alzheimer risk loci identifies PTK2B as an in vivo modulator and early marker of Tau pathology. *Mol. Psychiatry*.
- Dumanchin, C., Tournier, I., Martin, C., Didic, M., Belliard, S., Carlander, B., Rouhart, F., Duyckaerts, C., Pellissier, J.-F., Latouche, J.B., et al. (2006). Biological effects of four PSEN1 gene mutations causing Alzheimer disease with spastic paraparesis and cotton wool plaques. *Hum. Mutat.* *27*, 1063.
- Dumanis, S.B., Burgert, T., Caglayan, S., Füchtbauer, A., Füchtbauer, E.-M., Schmidt, V., and Willnow, T.E. (2015). Distinct Functions for Anterograde and Retrograde Sorting of SORLA in Amyloidogenic Processes in the Brain. *J. Neurosci. Off. J. Soc. Neurosci.* *35*, 12703–12713.

- Duyckaerts, C., Hauw, J.J., Piette, F., Rainsard, C., Poulain, V., Berthaux, P., and Escourolle, R. (1985). Cortical atrophy in senile dementia of the Alzheimer type is mainly due to a decrease in cortical length. *Acta Neuropathol. (Berl.)* 66, 72–74.
- Duyckaerts, C., Colle, M.A., Delatour, B., and Hauw, J.J. (1999). [Alzheimer's disease: lesions and their progression]. *Rev. Neurol. (Paris)* 155 Suppl 4, S17-27.
- Ehlers, M.D. (2000). Reinsertion or degradation of AMPA receptors determined by activity-dependent endocytic sorting. *Neuron* 28, 511–525.
- Englund, H., Sehlin, D., Johansson, A.-S., Nilsson, L.N.G., Gellerfors, P., Paulie, S., Lannfelt, L., and Pettersson, F.E. (2007). Sensitive ELISA detection of amyloid-beta protofibrils in biological samples. *J. Neurochem.* 103, 334–345.
- Espinoza, M., de Silva, R., Dickson, D.W., and Davies, P. (2008). Differential incorporation of tau isoforms in Alzheimer's disease. *J. Alzheimers Dis. JAD* 14, 1–16.
- Fazzari, P., Horre, K., Arranz, A.M., Frigerio, C.S., Saito, T., Saido, T.C., and De Strooper, B. (2017). PLD3 gene and processing of APP. *Nature* 541, E1–E2.
- Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* 84, 524–533.
- Folch, J., Petrov, D., Ettcheto, M., Abad, S., Sánchez-López, E., García, M.L., Olloquequi, J., Beas-Zarate, C., Auladell, C., and Camins, A. (2016). Current Research Therapeutic Strategies for Alzheimer's Disease Treatment. *Neural Plast.* 2016, 8501693.
- Frandemiche, M.L., De Seranno, S., Rush, T., Borel, E., Elie, A., Arnal, I., Lanté, F., and Buisson, A. (2014). Activity-dependent tau protein translocation to excitatory synapse is disrupted by exposure to amyloid-beta oligomers. *J. Neurosci. Off. J. Soc. Neurosci.* 34, 6084–6097.
- Fromer, M., Moran, J.L., Chambert, K., Banks, E., Bergen, S.E., Ruderfer, D.M., Handsaker, R.E., McCarroll, S.A., O'Donovan, M.C., Owen, M.J., et al. (2012). Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am. J. Hum. Genet.* 91, 597–607.
- Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D.M., et al. (2014). De novo mutations in schizophrenia implicate synaptic networks. *Nature* 506, 179–184.
- Fu, Y., Hsiao, J.-H.T., Paxinos, G., Halliday, G.M., and Kim, W.S. (2016). ABCA7 Mediates Phagocytic Clearance of Amyloid- $\beta$  in the Brain. *J. Alzheimers Dis. JAD* 54, 569–584.
- Fujio, K., Sato, M., Uemura, T., Sato, T., Sato-Harada, R., and Harada, A. (2007). 14-3-3 proteins and protein phosphatases are not reduced in tau-deficient mice. *Neuroreport* 18, 1049–1052.
- Garagnani, P., Giuliani, C., Pirazzini, C., Olivieri, F., Bacalini, M.G., Ostan, R., Mari, D., Passarino, G., Monti, D., Bonfigli, A.R., et al. (2013). Centenarians as super-controls to assess the biological relevance of genetic risk factors for common age-related diseases: a proof of principle on type 2 diabetes. *Aging* 5, 373–385.

- Genin, E., Hannequin, D., Wallon, D., Slegers, K., Hiltunen, M., Combarros, O., Bullido, M.J., Engelborghs, S., De Deyn, P., Berr, C., et al. (2011). APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol. Psychiatry* *16*, 903–907.
- Genovese, G., Fromer, M., Stahl, E.A., Ruderfer, D.M., Chambert, K., Landén, M., Moran, J.L., Purcell, S.M., Sklar, P., Sullivan, P.F., et al. (2016). Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* *19*, 1433–1441.
- Gibson, G. (2012). Rare and common variants: twenty arguments. *Nat. Rev. Genet.* *13*, 135–145.
- Glenner, G.G., and Wong, C.W. (1984). Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochem. Biophys. Res. Commun.* *120*, 885–890.
- Goh, K.-I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., and Barabási, A.-L. (2007). The human disease network. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 8685–8690.
- González-Pérez, A., and López-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am. J. Hum. Genet.* *88*, 440–449.
- Greenfield, J.P., Tsai, J., Gouras, G.K., Hai, B., Thinakaran, G., Checler, F., Sisodia, S.S., Greengard, P., and Xu, H. (1999). Endoplasmic reticulum and trans-Golgi network generate distinct populations of Alzheimer beta-amyloid peptides. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 742–747.
- Guerreiro, R., Wojtas, A., Bras, J., Carrasquillo, M., Rogaeva, E., Majounie, E., Cruchaga, C., Sassi, C., Kauwe, J.S.K., Younkin, S., et al. (2013). TREM2 variants in Alzheimer's disease. *N. Engl. J. Med.* *368*, 117–127.
- Guerreiro, R.J., Baquero, M., Blesa, R., Boada, M., Brás, J.M., Bullido, M.J., Calado, A., Crook, R., Ferreira, C., Frank, A., et al. (2010). Genetic screening of Alzheimer's disease genes in Iberian and African samples yields novel mutations in presenilins and APP. *Neurobiol. Aging* *31*, 725–731.
- Guffanti, G., Torri, F., Rasmussen, J., Clark, A.P., Lakatos, A., Turner, J.A., Fallon, J.H., Saykin, A.J., Weiner, M., ADNI the Alzheimer's Disease Neuroimaging Initiative, et al. (2013). Increased CNV-region deletions in mild cognitive impairment (MCI) and Alzheimer's disease (AD) subjects in the ADNI sample. *Genomics* *102*, 112–122.
- Guntupalli, S., Widagdo, J., and Anggono, V. (2016). Amyloid- $\beta$ -Induced Dysregulation of AMPA Receptor Trafficking. *Neural Plast.* *2016*, 3204519.
- Guo, Y., Sheng, Q., Samuels, D.C., Lehmann, B., Bauer, J.A., Pietenpol, J., and Shyr, Y. (2013). Comparative study of exome copy number variation estimation tools using array comparative genomic hybridization as control. *BioMed Res. Int.* *2013*, 915636.
- Haass, C., Kaether, C., Thinakaran, G., and Sisodia, S. (2012). Trafficking and proteolytic processing of APP. *Cold Spring Harb. Perspect. Med.* *2*, a006270.
- Harada, A., Oguchi, K., Okabe, S., Kuno, J., Terada, S., Ohshima, T., Sato-Yoshitake, R., Takei, Y., Noda, T., and Hirokawa, N. (1994). Altered microtubule organization in small-calibre axons of mice lacking tau protein. *Nature* *369*, 488–491.

- Hardy, J.A., and Higgins, G.A. (1992). Alzheimer's disease: the amyloid cascade hypothesis. *Science* 256, 184–185.
- Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M.L., Pahwa, J.S., Moskva, V., Dowzell, K., Williams, A., et al. (2009). Genome-wide association study identifies variants at *CLU* and *PICALM* associated with Alzheimer's disease. *Nat. Genet.* 41, 1088–1093.
- Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al. (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* 163, 1515–1526.
- Hartigan, J.A., and Johnson, G.V. (1999). Transient increases in intracellular calcium result in prolonged site-selective increases in Tau phosphorylation through a glycogen synthase kinase 3 $\beta$ -dependent pathway. *J. Biol. Chem.* 274, 21395–21401.
- Hartmann, T., Bieger, S.C., Brühl, B., Tienari, P.J., Ida, N., Allsop, D., Roberts, G.W., Masters, C.L., Dotti, C.G., Unsicker, K., et al. (1997). Distinct sites of intracellular production for Alzheimer's disease A $\beta$ 40/42 amyloid peptides. *Nat. Med.* 3, 1016–1020.
- Hastings, P.J., Lupski, J.R., Rosenberg, S.M., and Ira, G. (2009). Mechanisms of change in gene copy number. *Nat. Rev. Genet.* 10, 551–564.
- He, K., Song, L., Cummings, L.W., Goldman, J., Hagan, R.L., and Lee, H.-K. (2009). Stabilization of Ca<sup>2+</sup>-permeable AMPA receptors at perisynaptic sites by GluR1-S845 phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* 106, 20033–20038.
- Hehir-Kwa, J.Y., Pfundt, R., and Veltman, J.A. (2015). Exome sequencing and whole genome sequencing for the detection of copy number variation. *Expert Rev. Mol. Diagn.* 15, 1023–1032.
- Heilmann, S., Driemel, D., Clarimon, J., Fernández, V., Lacour, A., Wagner, H., Thelen, M., Hernández, I., Fortea, J., Alegret, M., et al. (2015). *PLD3* in non-familial Alzheimer's disease. *Nature* 520, E3-5.
- Henjum, K., Almdahl, I.S., Årskog, V., Minthon, L., Hansson, O., Fladby, T., and Nilsson, L.N.G. (2016). Cerebrospinal fluid soluble TREM2 in aging and Alzheimer's disease. *Alzheimers Res. Ther.* 8, 17.
- Hert, D.G., Fredlake, C.P., and Barron, A.E. (2008). Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis* 29, 4618–4626.
- Heslegrave, A., Heywood, W., Paterson, R., Magdalinos, N., Svensson, J., Johansson, P., Öhrfelt, A., Blennow, K., Hardy, J., Schott, J., et al. (2016). Increased cerebrospinal fluid soluble TREM2 concentration in Alzheimer's disease. *Mol. Neurodegener.* 11, 3.
- Hirtz, D., Thurman, D.J., Gwinn-Hardy, K., Mohamed, M., Chaudhuri, A.R., and Zalutsky, R. (2007). How common are the “common” neurologic disorders? *Neurology* 68, 326–337.
- Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J.-C., Carrasquillo, M.M., Abraham, R., Hamshere, M.L., Pahwa, J.S., Moskva, V., et al. (2011). Common variants at *ABCA7*, *MS4A6A/MS4A4E*, *EPHA1*, *CD33* and *CD2AP* are associated with Alzheimer's disease. *Nat. Genet.* 43, 429–435.
- Holmes, B.B., and Diamond, M.I. (2014). Prion-like properties of Tau protein: the importance of extracellular Tau as a therapeutic target. *J. Biol. Chem.* 289, 19855–19861.

- Hooli, B.V., Kovacs-Vajna, Z.M., Mullin, K., Blumenthal, M.A., Mattheisen, M., Zhang, C., Lange, C., Mohapatra, G., Bertram, L., and Tanzi, R.E. (2014). Rare autosomal copy number variations in early-onset familial Alzheimer's disease. *Mol. Psychiatry* 19, 676–681.
- Hooli, B.V., Lill, C.M., Mullin, K., Qiao, D., Lange, C., Bertram, L., and Tanzi, R.E. (2015). PLD3 gene variants and Alzheimer's disease. *Nature* 520, E7-8.
- Houdayer, C., Caux-Moncoutier, V., Krieger, S., Barrois, M., Bonnet, F., Bourdon, V., Bronner, M., Buisson, M., Coulet, F., Gaildrat, P., et al. (2012). Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum. Mutat.* 33, 1228–1238.
- Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet.* 6, e1001154.
- Huang, Y.-W.A., Zhou, B., Wernig, M., and Südhof, T.C. (2017). ApoE2, ApoE3, and ApoE4 Differentially Stimulate APP Transcription and A $\beta$  Secretion. *Cell* 168, 427–441.e21.
- Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. *Nat. Genet.* 36, 949–951.
- International HapMap 3 Consortium, Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., et al. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58.
- International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945.
- Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515, 216–221.
- Jack, C.R., Knopman, D.S., Chételat, G., Dickson, D., Fagan, A.M., Frisoni, G.B., Jagust, W., Mormino, E.C., Petersen, R.C., Sperling, R.A., et al. (2016). Suspected non-Alzheimer disease pathophysiology--concept and controversy. *Nat. Rev. Neurol.* 12, 117–124.
- Jay, T.R., Miller, C.M., Cheng, P.J., Graham, L.C., Bemiller, S., Broihier, M.L., Xu, G., Margevicius, D., Karlo, J.C., Sousa, G.L., et al. (2015). TREM2 deficiency eliminates TREM2+ inflammatory macrophages and ameliorates pathology in Alzheimer's disease mouse models. *J. Exp. Med.* 212, 287–295.
- Jian, X., Boerwinkle, E., and Liu, X. (2014). In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res.* 42, 13534–13544.
- Jin, S.C., Benitez, B.A., Karch, C.M., Cooper, B., Skorupa, T., Carrell, D., Norton, J.B., Hsu, S., Harari, O., Cai, Y., et al. (2014). Coding variants in TREM2 increase risk for Alzheimer's disease. *Hum. Mol. Genet.* 23, 5838–5846.
- Jonsson, T., Atwal, J.K., Steinberg, S., Snaedal, J., Jonsson, P.V., Bjornsson, S., Stefansson, H., Sulem, P., Gudbjartsson, D., Maloney, J., et al. (2012). A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 488, 96–99.

- Jonsson, T., Stefansson, H., Steinberg, S., Jonsdottir, I., Jonsson, P.V., Snaedal, J., Bjornsson, S., Huttenlocher, J., Levey, A.I., Lah, J.J., et al. (2013). Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* *368*, 107–116.
- Jun, G., Ibrahim-Verbaas, C.A., Vronskaya, M., Lambert, J.-C., Chung, J., Naj, A.C., Kunkle, B.W., Wang, L.-S., Bis, J.C., Bellenguez, C., et al. (2016). A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol. Psychiatry* *21*, 108–117.
- Kadalayil, L., Rafiq, S., Rose-Zerilli, M.J.J., Pengelly, R.J., Parker, H., Oscier, D., Strefford, J.C., Tapper, W.J., Gibson, J., Ennis, S., et al. (2015). Exome sequence read depth methods for identifying copy number changes. *Brief. Bioinform.* *16*, 380–392.
- Kawarabayashi, T., Shoji, M., Younkin, L.H., Wen-Lang, L., Dickson, D.W., Murakami, T., Matsubara, E., Abe, K., Ashe, K.H., and Younkin, S.G. (2004). Dimeric amyloid beta protein rapidly accumulates in lipid rafts followed by apolipoprotein E and phosphorylated tau accumulation in the Tg2576 mouse model of Alzheimer's disease. *J. Neurosci. Off. J. Soc. Neurosci.* *24*, 3801–3809.
- Kayed, R., Head, E., Thompson, J.L., McIntire, T.M., Milton, S.C., Cotman, C.W., and Glabe, C.G. (2003). Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* *300*, 486–489.
- Kelleher, R.J., and Shen, J. (2010). Genetics. Gamma-secretase and human disease. *Science* *330*, 1055–1056.
- Kim, W.S., Guillemin, G.J., Glaros, E.N., Lim, C.K., and Garner, B. (2006). Quantitation of ATP-binding cassette subfamily-A transporter gene expression in primary human brain cells. *Neuroreport* *17*, 891–896.
- Kim, W.S., Li, H., Ruberu, K., Chan, S., Elliott, D.A., Low, J.K., Cheng, D., Karl, T., and Garner, B. (2013). Deletion of *Abca7* increases cerebral amyloid- $\beta$  accumulation in the J20 mouse model of Alzheimer's disease. *J. Neurosci. Off. J. Soc. Neurosci.* *33*, 4387–4394.
- Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* *46*, 310–315.
- Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L., et al. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* *17*, 142–153.
- Kleinberger, G., Yamanishi, Y., Suárez-Calvet, M., Czirr, E., Lohmann, E., Cuyvers, E., Struyfs, H., Pettkus, N., Wenninger-Weinzierl, A., Mazaheri, F., et al. (2014). TREM2 mutations implicated in neurodegeneration impair cell surface transport and phagocytosis. *Sci. Transl. Med.* *6*, 243ra86.
- Koldamova, R., Fitz, N.F., and Lefterov, I. (2010). The role of ATP-binding cassette transporter A1 in Alzheimer's disease and neurodegeneration. *Biochim. Biophys. Acta* *1801*, 824–830.
- Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonsdottir, A., Jonsdottir, A., et al. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature* *488*, 471–475.

- Koolen, D.A., Vissers, L.E.L.M., Pfundt, R., de Leeuw, N., Knight, S.J.L., Regan, R., Kooy, R.F., Reyniers, E., Romano, C., Fichera, M., et al. (2006). A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat. Genet.* *38*, 999–1001.
- Koolen, D.A., Kramer, J.M., Neveling, K., Nillesen, W.M., Moore-Barton, H.L., Elmslie, F.V., Toutain, A., Amiel, J., Malan, V., Tsai, A.C.-H., et al. (2012). Mutations in the chromatin modifier gene *KANSL1* cause the 17q21.31 microdeletion syndrome. *Nat. Genet.* *44*, 639–641.
- Kovacs, G.G. (2015). Invited review: Neuropathology of tauopathies: principles and practice. *Neuropathol. Appl. Neurobiol.* *41*, 3–23.
- Krumm, N., Sudmant, P.H., Ko, A., O’Roak, B.J., Malig, M., Coe, B.P., NHLBI Exome Sequencing Project, Quinlan, A.R., Nickerson, D.A., and Eichler, E.E. (2012). Copy number variation detection and genotyping from exome sequence data. *Genome Res.* *22*, 1525–1532.
- Kuhn, P.-H., Wang, H., Dislich, B., Colombo, A., Zeitschel, U., Ellwart, J.W., Kremmer, E., Rossner, S., and Lichtenthaler, S.F. (2010). *ADAM10* is the physiologically relevant, constitutive alpha-secretase of the amyloid precursor protein in primary neurons. *EMBO J.* *29*, 3020–3032.
- LaFramboise, T. (2009). Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res.* *37*, 4181–4193.
- Lambert, J.-C., Heath, S., Even, G., Campion, D., Sleegers, K., Hiltunen, M., Combarros, O., Zelenika, D., Bullido, M.J., Tavernier, B., et al. (2009). Genome-wide association study identifies variants at *CLU* and *CR1* associated with Alzheimer’s disease. *Nat. Genet.* *41*, 1094–1099.
- Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., et al. (2013a). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease. *Nat. Genet.* *45*, 1452–1458.
- Lambert, J.-C., Grenier-Boley, B., Harold, D., Zelenika, D., Chouraki, V., Kamatani, Y., Sleegers, K., Ikram, M.A., Hiltunen, M., Reitz, C., et al. (2013b). Genome-wide haplotype association study identifies the *FRMD4A* gene as a risk locus for Alzheimer’s disease. *Mol. Psychiatry* *18*, 461–470.
- Lambert, J.-C., Grenier-Boley, B., Bellenguez, C., Pasquier, F., Campion, D., Dartigues, J.-F., Berr, C., Tzourio, C., and Amouyel, P. (2015). *PLD3* and sporadic Alzheimer’s disease risk. *Nature* *520*, E1.
- Lambert, M.P., Velasco, P.T., Chang, L., Viola, K.L., Fernandez, S., Lacor, P.N., Khuon, D., Gong, Y., Bigio, E.H., Shaw, P., et al. (2007). Monoclonal antibodies that target pathological assemblies of Aβeta. *J. Neurochem.* *100*, 23–35.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* *409*, 860–921.
- Lanoiselée, H.-M., Nicolas, G., Wallon, D., Rovelet-Lecrux, A., Lacour, M., Rousseau, S., Richard, A.-C., Pasquier, F., Rollin-Sillaire, A., Martinaud, O., et al. (2017). *APP*, *PSEN1*, and *PSEN2* mutations in early-onset Alzheimer disease: A genetic screening study of familial and sporadic cases. *PLoS Med.* *14*, e1002270.
- Laurén, J., Gimbel, D.A., Nygaard, H.B., Gilbert, J.W., and Strittmatter, S.M. (2009). Cellular prion protein mediates impairment of synaptic plasticity by amyloid-beta oligomers. *Nature* *457*, 1128–1132.

Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M. (2014). LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* *15*, R84.

Le Guennec, K., Quenez, O., Nicolas, G., Wallon, D., Rousseau, S., Richard, A.-C., Alexander, J., Paschou, P., Charbonnier, C., Bellenguez, C., et al. (2016). 17q21.31 duplication causes prominent tau-related dementia with increased MAPT expression. *Mol. Psychiatry.*

Lee, M.S., Kwon, Y.T., Li, M., Peng, J., Friedlander, R.M., and Tsai, L.H. (2000). Neurotoxicity induces cleavage of p35 to p25 by calpain. *Nature* *405*, 360–364.

Lee, S., Abecasis, G.R., Boehnke, M., and Lin, X. (2014). Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.* *95*, 5–23.

van der Lee, S.J., Holstege, H., Wong, T.H., Jakobsdottir, J., Bis, J.C., Chouraki, V., van Rooij, J.G.J., Grove, M.L., Smith, A.V., Amin, N., et al. (2015). PLD3 variants in population studies. *Nature* *520*, E2–3.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* *536*, 285–291.

Levinson, D.F., Duan, J., Oh, S., Wang, K., Sanders, A.R., Shi, J., Zhang, N., Mowry, B.J., Olincy, A., Amin, F., et al. (2011). Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am. J. Psychiatry* *168*, 302–316.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* *25*, 1754–1760.

Li, J., Lupat, R., Amarasinghe, K.C., Thompson, E.R., Doyle, M.A., Ryland, G.L., Tothill, R.W., Halgamuge, S.K., Campbell, I.G., and Goringe, K.L. (2012a). CONTRA: copy number analysis for targeted resequencing. *Bioinforma. Oxf. Engl.* *28*, 1307–1313.

Li, M.-X., Gui, H.-S., Kwan, J.S.H., Bao, S.-Y., and Sham, P.C. (2012b). A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. *Nucleic Acids Res.* *40*, e53.

Li, S., Jin, M., Koeglsperger, T., Shepardson, N.E., Shankar, G.M., and Selkoe, D.J. (2011). Soluble A $\beta$  oligomers inhibit long-term potentiation through a mechanism involving excessive activation of extrasynaptic NR2B-containing NMDA receptors. *J. Neurosci. Off. J. Soc. Neurosci.* *31*, 6627–6638.

Li, Y., Shaw, C.A., Sheffer, I., Sule, N., Powell, S.Z., Dawson, B., Zaidi, S.N.Y., Bucayas, K.L., Lupski, J.R., Wilhelmsen, K.C., et al. (2012c). Integrated copy number and gene expression analysis detects a CREB1 association with Alzheimer's disease. *Transl. Psychiatry* *2*, e192.

Lim, J., Hao, T., Shaw, C., Patel, A.J., Szabó, G., Rual, J.-F., Fisk, C.J., Li, N., Smolyar, A., Hill, D.E., et al. (2006). A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* *125*, 801–814.

Liu, X., Jian, X., and Boerwinkle, E. (2013). dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum. Mutat.* *34*, E2393-2402.

Logeat, F., Bessia, C., Brou, C., LeBail, O., Jarriault, S., Seidah, N.G., and Israël, A. (1998). The Notch1 receptor is cleaved constitutively by a furin-like convertase. *Proc. Natl. Acad. Sci. U. S. A.* *95*, 8108–8112.

Louwersheimer, E., Cohn-Hokke, P.E., Pijnenburg, Y.A.L., Weiss, M.M., Sistermans, E.A., Rozemuller, A.J., Hulsman, M., van Swieten, J.C., van Duijn, C.M., Barkhof, F., et al. (2017). Rare Genetic Variant in SORL1 May Increase Penetrance of Alzheimer's Disease in a Family with Several Generations of APOE-ε4 Homozygosity. *J. Alzheimers Dis. JAD* 56, 63–74.

MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335, 823–828.

MacDonald, J.R., Ziman, R., Yuen, R.K.C., Feuk, L., and Scherer, S.W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* 42, D986-992.

Mairet-Coello, G., Courchet, J., Pieraut, S., Courchet, V., Maximov, A., and Polleux, F. (2013). The CAMKK2-AMPK kinase pathway mediates the synaptotoxic effects of Aβ oligomers through Tau phosphorylation. *Neuron* 78, 94–108.

Malekpour, S.A., Pezeshk, H., and Sadeghi, M. (2016). MGP-HMM: Detecting genome-wide CNVs using an HMM for modeling mate pair insertion sizes and read counts. *Math. Biosci.* 279, 53–62.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. *Nature* 461, 747–753.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., and Stadlan, E.M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 34, 939–944.

McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement. J. Alzheimers Assoc.* 7, 263–269.

Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G., Kinney, J.B., et al. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* 30, 271–277.

Michaelson, J.J., Shi, Y., Gujral, M., Zheng, H., Malhotra, D., Jin, X., Jian, M., Liu, G., Greer, D., Bhandari, A., et al. (2012). Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* 151, 1431–1442.

Mooney, M.A., and Wilmot, B. (2015). Gene set analysis: A step-by-step guide. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.* 168, 517–527.

Morris, G.P., Clark, I.A., and Vissel, B. (2014). Inconsistencies and controversies surrounding the amyloid hypothesis of Alzheimer's disease. *Acta Neuropathol. Commun.* 2, 135.

Morsch, R., Simon, W., and Coleman, P.D. (1999). Neurons may live for decades with neurofibrillary tangles. *J. Neuropathol. Exp. Neurol.* 58, 188–197.

Naj, A.C., Jun, G., Beecham, G.W., Wang, L.-S., Vardarajan, B.N., Buross, J., Gallins, P.J., Buxbaum, J.D., Jarvik, G.P., Crane, P.K., et al. (2011). Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.* *43*, 436–441.

Naj, A.C., Schellenberg, G.D., and Alzheimer's Disease Genetics Consortium (ADGC) (2017). Genomic variants, genes, and pathways of Alzheimer's disease: An overview. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.* *174*, 5–26.

Nam, J.-Y., Kim, N.K.D., Kim, S.C., Joung, J.-G., Xi, R., Lee, S., Park, P.J., and Park, W.-Y. (2016). Evaluation of somatic copy number estimation tools for whole-exome sequencing data. *Bioinform.* *17*, 185–192.

Ng, P.C., and Henikoff, S. (2002). Accounting for human polymorphisms predicted to affect protein function. *Genome Res.* *12*, 436–446.

Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* *31*, 3812–3814.

Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* *461*, 272–276.

Nicolas, G., Wallon, D., Charbonnier, C., Quenez, O., Rousseau, S., Richard, A.-C., Rovelet-Lecrux, A., Coutant, S., Le Guennec, K., Bacq, D., et al. (2016a). Screening of dementia genes by whole-exome sequencing in early-onset Alzheimer disease: input and lessons. *Eur. J. Hum. Genet. EJHG* *24*, 710–716.

Nicolas, G., Charbonnier, C., Wallon, D., Quenez, O., Bellenguez, C., Grenier-Boley, B., Rousseau, S., Richard, A.-C., Rovelet-Lecrux, A., Le Guennec, K., et al. (2016b). SORL1 rare variants: a major risk factor for familial early-onset Alzheimer's disease. *Mol. Psychiatry* *21*, 831–836.

Nicolas, G., Wallon, D., Goupil, C., Richard, A.-C., Pottier, C., Dorval, V., Sarov-Rivière, M., Riant, F., Hervé, D., Amouyel, P., et al. (2016c). Mutation in the 3'untranslated region of APP as a genetic determinant of cerebral amyloid angiopathy. *Eur. J. Hum. Genet. EJHG* *24*, 92–98.

Nilsberth, C., Westlind-Danielsson, A., Eckman, C.B., Condron, M.M., Axelman, K., Forsell, C., Sten, C., Luthman, J., Teplow, D.B., Younkin, S.G., et al. (2001). The "Arctic" APP mutation (E693G) causes Alzheimer's disease by enhanced Aβ protofibril formation. *Nat. Neurosci.* *4*, 887–893.

Nisbet, R.M., Polanco, J.-C., Ittner, L.M., and Götz, J. (2015). Tau aggregation and its interplay with amyloid-β. *Acta Neuropathol. (Berl.)* *129*, 207–220.

Novarino, G., Fenstermaker, A.G., Zaki, M.S., Hofree, M., Silhavy, J.L., Heiberg, A.D., Abdellateef, M., Rosti, B., Scott, E., Mansour, L., et al. (2014). Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. *Science* *343*, 506–511.

Oh, M.C., Derkach, V.A., Guire, E.S., and Soderling, T.R. (2006). Extrasynaptic membrane trafficking regulated by GluR1 serine 845 phosphorylation primes AMPA receptors for long-term potentiation. *J. Biol. Chem.* *281*, 752–758.

Oyama, F., Sawamura, N., Kobayashi, K., Morishima-Kawashima, M., Kuramochi, T., Ito, M., Tomita, T., Maruyama, K., Saido, T.C., Iwatsubo, T., et al. (1998). Mutant presenilin 2 transgenic mouse: effect on an age-dependent increase of amyloid beta-protein 42 in the brain. *J. Neurochem.* *71*, 313–322.

Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.-I., Cooper, G.M., et al. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* *30*, 265–270.

Perez-Tur, J., Froelich, S., Prihar, G., Crook, R., Baker, M., Duff, K., Wragg, M., Busfield, F., Lendon, C., and Clark, R.F. (1995). A mutation in Alzheimer's disease destroying a splice acceptor site in the presenilin-1 gene. *Neuroreport* *7*, 297–301.

Piccio, L., Deming, Y., Del-Águila, J.L., Ghezzi, L., Holtzman, D.M., Fagan, A.M., Fenoglio, C., Galimberti, D., Borroni, B., and Cruchaga, C. (2016). Cerebrospinal fluid soluble TREM2 is higher in Alzheimer disease and associated with mutation status. *Acta Neuropathol. (Berl.)* *131*, 925–933.

Pinkel, D., and Albertson, D.G. (2005). Comparative genomic hybridization. *Annu. Rev. Genomics Hum. Genet.* *6*, 331–354.

Plagnol, V., Curtis, J., Epstein, M., Mok, K.Y., Stebbings, E., Grigoriadou, S., Wood, N.W., Hambleton, S., Burns, S.O., Thrasher, A.J., et al. (2012). A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinforma. Oxf. Engl.* *28*, 2747–2754.

Pocklington, A.J., Rees, E., Walters, J.T.R., Han, J., Kavanagh, D.H., Chambert, K.D., Holmans, P., Moran, J.L., McCarroll, S.A., Kirov, G., et al. (2015). Novel Findings from CNVs Implicate Inhibitory and Excitatory Signaling Complexes in Schizophrenia. *Neuron* *86*, 1203–1214.

Pottier, C., Wallon, D., Lecrux, A.R., Maltete, D., Bombois, S., Jurici, S., Frebourg, T., Hannequin, D., and Campion, D. (2012a). Amyloid- $\beta$  protein precursor gene expression in alzheimer's disease and other conditions. *J. Alzheimers Dis. JAD* *28*, 561–566.

Pottier, C., Hannequin, D., Coutant, S., Rovelet-Lecrux, A., Wallon, D., Rousseau, S., Legallic, S., Paquet, C., Bombois, S., Pariente, J., et al. (2012b). High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. *Mol. Psychiatry* *17*, 875–879.

Rapoport, M., Dawson, H.N., Binder, L.I., Vitek, M.P., and Ferreira, A. (2002). Tau is essential to beta - amyloid-induced neurotoxicity. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 6364–6369.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., et al. (2006). Global variation in copy number in the human genome. *Nature* *444*, 444–454.

Reitz, C. (2015). The role of the retromer complex in aging-related neurodegeneration: a molecular and genomic review. *Mol. Genet. Genomics MGG* *290*, 413–427.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med. Off. J. Am. Coll. Med. Genet.* *17*, 405–424.

Rosenthal, S.L., Bamne, M.N., Wang, X., Berman, S., Snitz, B.E., Klunk, W.E., Sweet, R.A., Demirci, F.Y., Lopez, O.L., and Kamboh, M.I. (2015). More evidence for association of a rare TREM2 mutation (R47H) with Alzheimer's disease risk. *Neurobiol. Aging* *36*, 2443.e21-26.

Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerrière, A., Vital, A., Dumanchin, C., Feuillette, S., Brice, A., Vercelletto, M., et al. (2006). APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat. Genet.* *38*, 24–26.

Rovelet-Lecrux, A., Legallic, S., Wallon, D., Flaman, J.-M., Martinaud, O., Bombois, S., Rollin-Sillaire, A., Michon, A., Le Ber, I., Pariente, J., et al. (2012). A genome-wide study reveals rare CNVs exclusive to extreme phenotypes of Alzheimer disease. *Eur. J. Hum. Genet. EJHG* *20*, 613–617.

Rovelet-Lecrux, A., Charbonnier, C., Wallon, D., Nicolas, G., Seaman, M.N.J., Pottier, C., Breusegem, S.Y., Mathur, P.P., Jenardhanan, P., Le Guennec, K., et al. (2015). De novo deleterious genetic variations target a biological network centered on A $\beta$  peptide in early-onset Alzheimer disease. *Mol. Psychiatry* *20*, 1046–1056.

Ruderfer, D.M., Hamamsy, T., Lek, M., Karczewski, K.J., Kavanagh, D., Samocha, K.E., Exome Aggregation Consortium, Daly, M.J., MacArthur, D.G., Fromer, M., et al. (2016). Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. *Nat. Genet.* *48*, 1107–1111.

Saito, T., Suemoto, T., Brouwers, N., Slegers, K., Funamoto, S., Mihira, N., Matsuba, Y., Yamada, K., Nilsson, P., Takano, J., et al. (2011). Potent amyloidogenicity and pathogenicity of A $\beta$ 43. *Nat. Neurosci.* *14*, 1023–1032.

Samarakoon, P.S., Sorte, H.S., Stray-Pedersen, A., Rødningen, O.K., Rognes, T., and Lyle, R. (2016). cnvScan: a CNV screening and annotation tool to improve the clinical utility of computational CNV prediction from exome sequencing data. *BMC Genomics* *17*, 51.

Sanger, F., and Coulson, A.R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* *94*, 441–448.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* *74*, 5463–5467.

Sannerud, R., Esselens, C., Ejsmont, P., Mattera, R., Rochin, L., Tharkeshwar, A.K., De Baets, G., De Wever, V., Habets, R., Baert, V., et al. (2016). Restricted Location of PSEN2/ $\gamma$ -Secretase Determines Substrate Specificity and Generates an Intracellular A $\beta$  Pool. *Cell* *166*, 193–208.

Sato, S., Kamino, K., Miki, T., Doi, A., Ii, K., St George-Hyslop, P.H., Ogiwara, T., and Sakaki, Y. (1998). Splicing mutation of presenilin-1 gene for early-onset familial Alzheimer's disease. *Hum. Mutat. Suppl* *1*, S91-94.

Satoh, K., Abe-Dohmae, S., Yokoyama, S., St George-Hyslop, P., and Fraser, P.E. (2015). ATP-binding cassette transporter A7 (ABCA7) loss of function alters Alzheimer amyloid processing. *J. Biol. Chem.* *290*, 24152–24165.

Sawamura, N., Morishima-Kawashima, M., Waki, H., Kobayashi, K., Kuramochi, T., Frosch, M.P., Ding, K., Ito, M., Kim, T.W., Tanzi, R.E., et al. (2000). Mutant presenilin 2 transgenic mice. A large increase in the levels of Abeta 42 is presumably associated with the low density membrane domain that contains decreased levels of glycerophospholipids and sphingomyelin. *J. Biol. Chem.* *275*, 27901–27908.

Schellenberg, G.D., Deeb, S.S., Boehnke, M., Bryant, E.M., Martin, G.M., Lampe, T.H., and Bird, T.D. (1987). Association of an apolipoprotein CII allele with familial dementia of the Alzheimer type. *J. Neurogenet.* *4*, 97–108.

Scheuner, D., Eckman, C., Jensen, M., Song, X., Citron, M., Suzuki, N., Bird, T.D., Hardy, J., Hutton, M., Kukull, W., et al. (1996). Secreted amyloid beta-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nat. Med.* *2*, 864–870.

- Schlattl, A., Anders, S., Waszak, S.M., Huber, W., and Korbel, J.O. (2011). Relating CNVs to transcriptome data at fine resolution: assessment of the effect of variant size, type, and overlap with functional regions. *Genome Res.* *21*, 2004–2013.
- Schulte, E.C., Kurz, A., Alexopoulos, P., Hampel, H., Peters, A., Gieger, C., Rujescu, D., Diehl-Schmid, J., and Winkelmann, J. (2015). Excess of rare coding variants in PLD3 in late- but not early-onset Alzheimer's disease. *Hum. Genome Var.* *2*, 14028.
- Schwarz, J.M., Rödelberger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* *7*, 575–576.
- Schwarz, J.M., Cooper, D.N., Schuelke, M., and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* *11*, 361–362.
- Seaman, M.N.J. (2012). The retromer complex - endosomal protein recycling and beyond. *J. Cell Sci.* *125*, 4693–4702.
- Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., et al. (2004). Large-scale copy number polymorphism in the human genome. *Science* *305*, 525–528.
- Seshadri, S., Fitzpatrick, A.L., Ikram, M.A., DeStefano, A.L., Gudnason, V., Boada, M., Bis, J.C., Smith, A.V., Carassquillo, M.M., Lambert, J.C., et al. (2010). Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* *303*, 1832–1840.
- Shankar, G.M., Bloodgood, B.L., Townsend, M., Walsh, D.M., Selkoe, D.J., and Sabatini, B.L. (2007). Natural oligomers of the Alzheimer amyloid-beta protein induce reversible synapse loss by modulating an NMDA-type glutamate receptor-dependent signaling pathway. *J. Neurosci. Off. J. Soc. Neurosci.* *27*, 2866–2875.
- Shankar, G.M., Li, S., Mehta, T.H., Garcia-Munoz, A., Shepardson, N.E., Smith, I., Brett, F.M., Farrell, M.A., Rowan, M.J., Lemere, C.A., et al. (2008). Amyloid-beta protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. *Nat. Med.* *14*, 837–842.
- Shaw-Smith, C., Pittman, A.M., Willatt, L., Martin, H., Rickman, L., Gribble, S., Curley, R., Cumming, S., Dunn, C., Kalaitzopoulos, D., et al. (2006). Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat. Genet.* *38*, 1032–1037.
- Shen, J., and Kelleher, R.J. (2007). The presenilin hypothesis of Alzheimer's disease: evidence for a loss-of-function pathogenic mechanism. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 403–409.
- Sisodia, S.S., Koo, E.H., Beyreuther, K., Unterbeck, A., and Price, D.L. (1990). Evidence that beta-amyloid protein in Alzheimer's disease is not derived by normal processing. *Science* *248*, 492–495.
- Smith, M.J., Kwok, J.B., McLean, C.A., Kril, J.J., Broe, G.A., Nicholson, G.A., Cappai, R., Hallupp, M., Cotton, R.G., Masters, C.L., et al. (2001). Variable phenotype of Alzheimer's disease with spastic paraparesis. *Ann. Neurol.* *49*, 125–129.
- Snyder, E.M., Nong, Y., Almeida, C.G., Paul, S., Moran, T., Choi, E.Y., Nairn, A.C., Salter, M.W., Lombroso, P.J., Gouras, G.K., et al. (2005). Regulation of NMDA receptor trafficking by amyloid-beta. *Nat. Neurosci.* *8*, 1051–1058.

- Sottejeau, Y., Bretteville, A., Cantrelle, F.-X., Malmanche, N., Demiaute, F., Mendes, T., Delay, C., Alves Dos Alves, H., Flaig, A., Davies, P., et al. (2015). Tau phosphorylation regulates the interaction between BIN1's SH3 domain and Tau's proline-rich domain. *Acta Neuropathol. Commun.* 3, 58.
- Spires-Jones, T.L., de Calignon, A., Matsui, T., Zehr, C., Pitstick, R., Wu, H.-Y., Osetek, J.D., Jones, P.B., Bacskai, B.J., Feany, M.B., et al. (2008). In vivo imaging reveals dissociation between caspase activation and acute neuronal death in tangle-bearing neurons. *J. Neurosci. Off. J. Soc. Neurosci.* 28, 862–867.
- St George-Hyslop, P.H., Tanzi, R.E., Polinsky, R.J., Haines, J.L., Nee, L., Watkins, P.C., Myers, R.H., Feldman, R.G., Pollen, D., and Drachman, D. (1987). The genetic defect causing familial Alzheimer's disease maps on chromosome 21. *Science* 235, 885–890.
- Steinberg, S., Stefansson, H., Jonsson, T., Johannsdottir, H., Ingason, A., Helgason, H., Sulem, P., Magnusson, O.T., Gudjonsson, S.A., Unnsteinsdottir, U., et al. (2015). Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nat. Genet.* 47, 445–447.
- Suárez-Calvet, M., Kleinberger, G., Araque Caballero, M.Á., Brendel, M., Rominger, A., Alcolea, D., Fortea, J., Lleó, A., Blesa, R., Gispert, J.D., et al. (2016). sTREM2 cerebrospinal fluid levels are a potential biomarker for microglia activity in early-stage Alzheimer's disease and associate with neuronal injury markers. *EMBO Mol. Med.* 8, 466–476.
- Swaminathan, S., Huentelman, M.J., Corneveaux, J.J., Myers, A.J., Faber, K.M., Foroud, T., Mayeux, R., Shen, L., Kim, S., Turk, M., et al. (2012a). Analysis of copy number variation in Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. *PloS One* 7, e50640.
- Swaminathan, S., Shen, L., Kim, S., Inlow, M., West, J.D., Faber, K.M., Foroud, T., Mayeux, R., Saykin, A.J., Alzheimer's Disease Neuroimaging Initiative, et al. (2012b). Analysis of copy number variation in Alzheimer's disease: the NIALOAD/ NCRAD Family Study. *Curr. Alzheimer Res.* 9, 801–814.
- Szaruga, M., Veugelen, S., Benurwar, M., Lismont, S., Sepulveda-Falla, D., Lleo, A., Ryan, N.S., Lashley, T., Fox, N.C., Murayama, S., et al. (2015). Qualitative changes in human  $\gamma$ -secretase underlie familial Alzheimer's disease. *J. Exp. Med.* 212, 2003–2013.
- Szatkiewicz, J.P., O'Dushlaine, C., Chen, G., Chambert, K., Moran, J.L., Neale, B.M., Fromer, M., Ruderfer, D., Akterin, S., Bergen, S.E., et al. (2014). Copy number variation in schizophrenia in Sweden. *Mol. Psychiatry* 19, 762–773.
- Szigeti, K., Lal, D., Li, Y., Doody, R.S., Wilhelmsen, K., Yan, L., Liu, S., Ma, C., and Texas Alzheimer Research and Care Consortium (2013). Genome-wide scan for copy number variation association with age at onset of Alzheimer's disease. *J. Alzheimers Dis. JAD* 33, 517–523.
- Takami, M., Nagashima, Y., Sano, Y., Ishihara, S., Morishima-Kawashima, M., Funamoto, S., and Ihara, Y. (2009).  $\gamma$ -Secretase: successive tripeptide and tetrapeptide release from the transmembrane domain of beta-carboxyl terminal fragment. *J. Neurosci. Off. J. Soc. Neurosci.* 29, 13042–13052.
- Tan, R., Wang, Y., Kleinstein, S.E., Liu, Y., Zhu, X., Guo, H., Jiang, Q., Allen, A.S., and Zhu, M. (2014). An evaluation of copy number variation detection tools from whole-exome sequencing data. *Hum. Mutat.* 35, 899–907.
- Tapiola, T., Alafuzoff, I., Herukka, S.-K., Parkkinen, L., Hartikainen, P., Soininen, H., and Pirttilä, T. (2009). Cerebrospinal fluid {beta}-amyloid 42 and tau proteins as biomarkers of Alzheimer-type pathological changes in the brain. *Arch. Neurol.* 66, 382–389.

- Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., et al. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337, 64–69.
- Thal, D.R., Rüb, U., Orantes, M., and Braak, H. (2002). Phases of A beta-deposition in the human brain and its relevance for the development of AD. *Neurology* 58, 1791–1800.
- Thinakaran, G., Borchelt, D.R., Lee, M.K., Slunt, H.H., Spitzer, L., Kim, G., Ratovitsky, T., Davenport, F., Nordstedt, C., Seeger, M., et al. (1996). Endoproteolysis of presenilin 1 and accumulation of processed derivatives in vivo. *Neuron* 17, 181–190.
- Thusberg, J., Olatubosun, A., and Vihinen, M. (2011). Performance of mutation pathogenicity prediction methods on missense variants. *Hum. Mutat.* 32, 358–368.
- Tosto, G., Fu, H., Vardarajan, B.N., Lee, J.H., Cheng, R., Reyes-Dumeyer, D., Lantigua, R., Medrano, M., Jimenez-Velazquez, I.Z., Elkind, M.S.V., et al. (2015). F-box/LRR-repeat protein 7 is genetically associated with Alzheimer's disease. *Ann. Clin. Transl. Neurol.* 2, 810–820.
- Tucker, K.L., Meyer, M., and Barde, Y.A. (2001). Neurotrophins are required for nerve growth during development. *Nat. Neurosci.* 4, 29–37.
- Ubelmann, F., Burrinha, T., Salavessa, L., Gomes, R., Ferreira, C., Moreno, N., and Guimas Almeida, C. (2017). Bin1 and CD2AP polarise the endocytic generation of beta-amyloid. *EMBO Rep.* 18, 102–122.
- UK10K Consortium, Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J.R.B., Xu, C., Futema, M., et al. (2015). The UK10K project identifies rare variants in health and disease. *Nature* 526, 82–90.
- Ulrich, J.D., Finn, M.B., Wang, Y., Shen, A., Mahan, T.E., Jiang, H., Stewart, F.R., Piccio, L., Colonna, M., and Holtzman, D.M. (2014). Altered microglial response to A $\beta$  plaques in APPPS1-21 mice heterozygous for TREM2. *Mol. Neurodegener.* 9, 20.
- Van Cauwenberghe, C., Van Broeckhoven, C., and Sleegers, K. (2016). The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet. Med. Off. J. Am. Coll. Med. Genet.* 18, 421–430.
- Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* 43, 11.10.1-33.
- Van Schil, K., Karlstetter, M., Aslanidis, A., Dannhausen, K., Azam, M., Qamar, R., Leroy, B.P., Depasse, F., Langmann, T., and De Baere, E. (2016). Autosomal recessive retinitis pigmentosa with homozygous rhodopsin mutation E150K and non-coding cis-regulatory variants in CRX-binding regions of SAMD7. *Sci. Rep.* 6, 21307.
- Vardarajan, B.N., Zhang, Y., Lee, J.H., Cheng, R., Bohm, C., Ghani, M., Reitz, C., Reyes-Dumeyer, D., Shen, Y., Rogaeva, E., et al. (2015). Coding mutations in SORL1 and Alzheimer disease. *Ann. Neurol.* 77, 215–227.
- Vasiliou, V., Vasiliou, K., and Nebert, D.W. (2009). Human ATP-binding cassette (ABC) transporter family. *Hum. Genomics* 3, 281–290.

- Vassar, R. (2004). BACE1: the beta-secretase enzyme in Alzheimer's disease. *J. Mol. Neurosci.* *MN 23*, 105–114.
- Vassar, R., Bennett, B.D., Babu-Khan, S., Kahn, S., Mendiaz, E.A., Denis, P., Teplow, D.B., Ross, S., Amarante, P., Loeloff, R., et al. (1999). Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science* *286*, 735–741.
- Veltman, J.A., and Brunner, H.G. (2012). De novo mutations in human genetic disease. *Nat. Rev. Genet.* *13*, 565–575.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. *Science* *291*, 1304–1351.
- Vetrivel, K.S., and Thinakaran, G. (2006). Amyloidogenic processing of beta-amyloid precursor protein in intracellular compartments. *Neurology* *66*, S69-73.
- Veugelen, S., Saito, T., Saido, T.C., Chávez-Gutiérrez, L., and De Strooper, B. (2016). Familial Alzheimer's Disease Mutations in Presenilin Generate Amyloidogenic A $\beta$  Peptide Seeds. *Neuron* *90*, 410–416.
- Vidal, M., Cusick, M.E., and Barabási, A.-L. (2011). Interactome networks and human disease. *Cell* *144*, 986–998.
- Vockley, C.M., Guo, C., Majoros, W.H., Nodzenski, M., Scholtens, D.M., Hayes, M.G., Lowe, W.L., and Reddy, T.E. (2015). Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. *Genome Res.* *25*, 1206–1214.
- Vogelsberg-Ragaglia, V., Bruce, J., Richter-Landsberg, C., Zhang, B., Hong, M., Trojanowski, J.Q., and Lee, V.M. (2000). Distinct FTDP-17 missense mutations in tau produce tau aggregates and other pathological phenotypes in transfected CHO cells. *Mol. Biol. Cell* *11*, 4093–4104.
- Wallon, D., Rousseau, S., Rovelet-Lecrux, A., Quillard-Muraine, M., Guyant-Maréchal, L., Martinaud, O., Pariente, J., Puel, M., Rollin-Sillaire, A., Pasquier, F., et al. (2012). The French series of autosomal dominant early onset Alzheimer's disease cases: mutation spectrum and cerebrospinal fluid biomarkers. *J. Alzheimers Dis. JAD* *30*, 847–856.
- Walsh, D.M., and Selkoe, D.J. (2007). A beta oligomers - a decade of discovery. *J. Neurochem.* *101*, 1172–1184.
- Wang, Y., and Mandelkow, E. (2016). Tau in physiology and pathology. *Nat. Rev. Neurosci.* *17*, 5–21.
- Wang, B., Yang, W., Wen, W., Sun, J., Su, B., Liu, B., Ma, D., Lv, D., Wen, Y., Qu, T., et al. (2010). Gamma-secretase gene mutations in familial acne inversa. *Science* *330*, 1065.
- Wang, H.Y., Lee, D.H., Davis, C.B., and Shank, R.P. (2000). Amyloid peptide A $\beta$ (1-42) binds selectively and with picomolar affinity to  $\alpha$ 7 nicotinic acetylcholine receptors. *J. Neurochem.* *75*, 1155–1161.
- Wang, Q., Lu, Q., and Zhao, H. (2015a). A review of study designs and statistical methods for genomic epidemiology studies using next generation sequencing. *Front. Genet.* *6*, 149.

Wang, Y., Cella, M., Mallinson, K., Ulrich, J.D., Young, K.L., Robinette, M.L., Gilfillan, S., Krishnan, G.M., Sudhakar, S., Zinselmeyer, B.H., et al. (2015b). TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. *Cell* *160*, 1061–1071.

Wang, Y., Ulland, T.K., Ulrich, J.D., Song, W., Tzaferis, J.A., Hole, J.T., Yuan, P., Mahan, T.E., Shi, Y., Gilfillan, S., et al. (2016). TREM2-mediated early microglial response limits diffusion and toxicity of amyloid plaques. *J. Exp. Med.* *213*, 667–675.

Weingarten, M.D., Lockwood, A.H., Hwo, S.Y., and Kirschner, M.W. (1975). A protein factor essential for microtubule assembly. *Proc. Natl. Acad. Sci. U. S. A.* *72*, 1858–1862.

Whitehouse, P.J., Price, D.L., Clark, A.W., Coyle, J.T., and DeLong, M.R. (1981). Alzheimer disease: evidence for selective loss of cholinergic neurons in the nucleus basalis. *Ann. Neurol.* *10*, 122–126.

Wolfe, M.S., Xia, W., Ostaszewski, B.L., Diehl, T.S., Kimberly, W.T., and Selkoe, D.J. (1999). Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and gamma-secretase activity. *Nature* *398*, 513–517.

Xia, D., Watanabe, H., Wu, B., Lee, S.H., Li, Y., Tsvetkov, E., Bolshakov, V.Y., Shen, J., and Kelleher, R.J. (2015). Presenilin-1 knockin mice reveal loss-of-function mechanism for familial Alzheimer's disease. *Neuron* *85*, 967–981.

Xia, D., Kelleher, R.J., and Shen, J. (2016). Loss of A $\beta$ 43 Production Caused by Presenilin-1 Mutations in the Knockin Mouse Brain. *Neuron* *90*, 417–422.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* *318*, 1917–1920.

Yu, J.-T., Tan, L., and Hardy, J. (2014). Apolipoprotein E in Alzheimer's disease: an update. *Annu. Rev. Neurosci.* *37*, 79–100.

Yu, W., Polepalli, J., Wagh, D., Rajadas, J., Malenka, R., and Lu, B. (2012). A critical role for the PAR-1/MARK-tau axis in mediating the toxic effects of A $\beta$  on synapses and dendritic spines. *Hum. Mol. Genet.* *21*, 1384–1390.

Yuan, P., Condello, C., Keene, C.D., Wang, Y., Bird, T.D., Paul, S.M., Luo, W., Colonna, M., Baddeley, D., and Grutzendler, J. (2016). TREM2 Haplodeficiency in Mice and Humans Impairs the Microglia Barrier Function Leading to Decreased Amyloid Compaction and Severe Axonal Dystrophy. *Neuron* *92*, 252–264.

Zappala, Z., and Montgomery, S.B. (2016). Non-Coding Loss-of-Function Variation in Human Genomes. *Hum. Hered.* *81*, 78–87.

Zarrei, M., MacDonald, J.R., Merico, D., and Scherer, S.W. (2015). A copy number variation map of the human genome. *Nat. Rev. Genet.* *16*, 172–183.

Zhang, D.-F., Fan, Y., Wang, D., Bi, R., Zhang, C., Fang, Y., and Yao, Y.-G. (2016a). PLD3 in Alzheimer's Disease: a Modest Effect as Revealed by Updated Association and Expression Analyses. *Mol. Neurobiol.* *53*, 4034–4045.

Zhang, W., Jiao, B., Zhou, M., Zhou, T., and Shen, L. (2016b). Modeling Alzheimer's Disease with Induced Pluripotent Stem Cells: Current Challenges and Future Concerns. *Stem Cells Int.* 2016, 7828049.

Zhao, M., Wang, Q., Wang, Q., Jia, P., and Zhao, Z. (2013). Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics* 14 Suppl 11, S1.

Zhong, L., Chen, X.-F., Wang, T., Wang, Z., Liao, C., Wang, Z., Huang, R., Wang, D., Li, X., Wu, L., et al. (2017). Soluble TREM2 induces inflammatory responses and enhances microglial survival. *J. Exp. Med.* 214, 597–607.

Zhu, X.-C., Yu, J.-T., Jiang, T., Wang, P., Cao, L., and Tan, L. (2015). CR1 in Alzheimer's disease. *Mol. Neurobiol.* 51, 753–765.

Zollino, M., Orteschi, D., Murdolo, M., Lattante, S., Battaglia, D., Stefanini, C., Mercuri, E., Chiurazzi, P., Neri, G., and Marangi, G. (2012). Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype. *Nat. Genet.* 44, 636–638.