

# Intron retention as a productive mechanism in human *MAPT*: RNA species generated by retention of intron 3



Daniel Ruiz-Gabarre,<sup>a,b,c</sup> Laura Vallés-Saiz,<sup>b</sup> Almudena Carnero-Espejo,<sup>a,c</sup> Isidro Ferrer,<sup>d,e,f,g</sup> Félix Hernández,<sup>b</sup> Ramon García-Escudero,<sup>h,i,j</sup> Jesús Ávila,<sup>b,d,\*\*</sup> and Vega García-Escudero<sup>a,c,d,k,\*</sup>



<sup>a</sup>Anatomy, Histology and Neuroscience Department, School of Medicine, Universidad Autónoma de Madrid (UAM), 28029, Madrid, Spain

<sup>b</sup>Centro de Biología Molecular Severo Ochoa (UAM-CSIC), 28049, Madrid, Spain

<sup>c</sup>Graduate Programa in Neuroscience, Universidad Autónoma de Madrid (UAM), 28029, Madrid, Spain

<sup>d</sup>Networking Research Centre on Neurodegenerative Diseases (CIBERNED), Instituto de Salud Carlos III, 28029, Madrid, Spain

<sup>e</sup>Department of Pathology and Experimental Therapeutics, University of Barcelona, 08907, Barcelona, Spain

<sup>f</sup>Bellvitge University Hospital, IDIBELL (Bellvitge Biomedical Research Centre), 08908, Barcelona, Spain

<sup>g</sup>Institute of Neurosciences, University of Barcelona, 08035, Barcelona, Spain

<sup>h</sup>Biomedical Oncology Unit, CIEMAT, 28040, Madrid, Spain

<sup>i</sup>Research Institute Hospital 12 de Octubre (imas12), 28041, Madrid, Spain

<sup>j</sup>Networking Research Centre on Cancer (CIBERONC), Instituto de Salud Carlos III, 28029, Madrid, Spain

<sup>k</sup>Institute for Molecular Biology-IUBM (Universidad Autónoma de Madrid), 28049, Madrid, Spain

## Summary

**Background** Tau is a microtubule-binding protein encoded by the *MAPT* gene. Tau is essential for several physiological functions and associated with pathological processes, including Alzheimer's disease (AD). Six tau isoforms are typically described in the central nervous system, but current research paints a more diverse landscape and a more nuanced balance between isoforms. Recent work has described tau isoforms generated by intron 11 and intron 12 retention. This work adds to that evidence, proving the existence of *MAPT* transcripts retaining intron 3. Our aim is to demonstrate the existence of mature *MAPT* RNA species that retain intron 3 in human brain samples and to study its correlation with Alzheimer's disease across different regions.

**Methods** Initial evidence of intron-3-retaining *MAPT* species come from *in silico* analysis of RNA-seq databases. We further demonstrate the existence of these mature RNA species in a human neuroepithelioma cell line and human brain samples by quantitative PCR. We also use digital droplet PCR to demonstrate the existence of RNA species that retain either intron 3, intron 12 or both introns.

**Findings** Intron-3-retaining species are even more prominently present than intron-12-retaining ones. We show the presence of *MAPT* transcripts that retain both introns 3 and 12. These intron-retaining species are diminished in brain samples of patients with Alzheimer's disease with respect to individuals without dementia. Conversely, relative abundance of intron-3- or intron-12-retaining *MAPT* species with respect to double-retaining species as well as their percentage of expression with respect to total *MAPT* are increased in patients with Alzheimer's disease, especially in hippocampal samples. Among these TIR-*MAPT* species, TIR<sup>3+12</sup> double truncation allows better classification potential of Alzheimer's disease samples. Moreover, we find a significant increase in intron-3- or intron-12-retaining species and its relative abundance with respect to double-retaining *MAPT* species in cerebellum in contrast to frontal lateral cortex and hippocampus in individuals with no signs of dementia.

**Interpretation** Intron retention constitutes a potential mechanism to generate Tau isoforms whose mature RNA expression levels correlate with Alzheimer's pathology showing its potential as a biomarker associated to the disease.

**Funding** This research was funded by the Spanish Ministry of Science, Innovation and Universities: PGC2018-096177-B-I00 (J.A.); Spanish Ministry of Science and Innovation (MCIN): PID2020-113204GB-I00 (F.H.) and PID2021-123859OB-I00 from MCIN/AEI/10.13039/501100011033/FEDER, UE (J.A.). It was also supported by CSIC through an intramural grant (201920E104) (J.A.) and the Centre for Networked Biomedical Research on Neurodegenerative Diseases (J.A.). The Centro de Biología Molecular Severo Ochoa (CBMSO) is a Severo Ochoa Center of Excellence (MICIN, award CEX2021-001154-S).

eBioMedicine

2024;100: 104953

Published Online 5 January

2024

[https://doi.org/10.](https://doi.org/10.1016/j.ebiom.2023.104953)

[1016/j.ebiom.2023.](https://doi.org/10.1016/j.ebiom.2023.104953)

104953

\*Corresponding author. Anatomy, Histology and Neuroscience Department, School of Medicine, Universidad Autónoma de Madrid (UAM), 28029, Madrid, Spain.

\*\*Corresponding author. Centro de Biología Molecular Severo Ochoa (UAM-CSIC), 28049, Madrid, Spain.

E-mail addresses: [v.garcia-escudero@uam.es](mailto:v.garcia-escudero@uam.es) (V. García-Escudero), [jesus.avila@csic.es](mailto:jesus.avila@csic.es) (J. Ávila).

Copyright © 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Tau; Alzheimer's disease; Alternative splicing; Intron retention; Tau isoforms; Tauopathies

### Research in context

#### Evidence before this study

Alzheimer's disease (AD) is the most prevalent cause of dementia worldwide, which postulates it as one of the most prominent public health challenges of our time, especially in the context of a progressively aged population. Despite having been first described more than 100 years ago, the precise mechanisms that govern the disease remain unclear. However, decades of intensive research have clearly outlined the role of certain actors, including the protein tau. Tau is a predominantly neuronal microtubule-associated protein with varied physiological functions that include, but are not limited to, microtubule stabilisation and polymerisation and axonal transport regulation. In addition, tau has been linked to a number of pathologies termed tauopathies, which include AD. Tau is expressed from a single-copy gene, *MAPT*, comprised by 16 exons of which exons 2, 3, 4a, 6, 8, 10 and 14 are subjected to alternative splicing. In the central nervous system (CNS) of adult humans, mainly 6 isoforms are expressed from the combination of inclusion and exclusion of exons 2, 3 and 10. The balance between these physiological and pathological functions of tau is mediated by the relative proportion of these isoforms. In addition, tau functions can be regulated by means of several post-translational modifications, with (hyper) phosphorylation and truncation having been consistently linked to Alzheimer's disease and other tauopathies. Recently, our lab has described the existence of an intron retention phenomenon between exons 12 and 13 (intron 12). This intron retention leads to the expression of mature transcripts that lack the terminal part of canonical *MAPT* transcripts, thus giving rise to a set of truncated isoforms that lack the C-terminal portion of canonical tau. Instead, it displays a 16-residue fragment that includes two tryptophan residues (W), which is curiously the only amino acid that cannot be found elsewhere within canonical tau. For this reason, we named it W-tau. Contrary to what has been described for other truncated tau isoforms, W-tau isoforms seem to be more soluble and exhibit less tendency to self-aggregate when compared to full-length and other truncated tau isoforms, while maintaining physiological functions of canonical tau. Strikingly, W-tau isoforms were found to be diminished in patients with AD when compared to control

participants without dementia, pointing towards a potential role of these in preventing the development of the disease. In contrast, other colleagues have reported the existence of other truncated isoforms generated upon intron 11 retention (Tau11i), which are more prone to aggregation than full-length canonical tau.

#### Added value of this study

In the present work we describe yet another set of isoforms generated by intron retention, in this case by retaining intron 3; a process that can occur by itself or simultaneously with intron 12 retention, highlighting intron retention as a protein-productive mechanism of tau alternative splicing that seems to be subjected to intensive regulation. We observed a decrease of these intron retaining species in brain samples of patients with AD. Conversely, the relative abundance of intron-3- or intron-12-retaining *MAPT* species with respect to double-retaining species and their percentage of expression with respect to total-*MAPT* levels are increased in patients with Alzheimer's disease specially in hippocampal samples. Moreover, we found a significant increase in intron-3- or intron-12-retaining species and its relative abundance with respect to double-retaining *MAPT* species in cerebellum in contrast to frontal lateral cortex and hippocampus in individuals with no signs of dementia. Our results suggest a correlation between intron retaining species and susceptibility to Alzheimer's degeneration and that double-intron retaining species have the strongest classification potential.

#### Implications of all the available evidence

Alzheimer's disease is exclusively a human disease, not well reproduced in animal models. This is particularly relevant in the context of this line of research, since introns are not conserved as well as exons across species, which implies that the set of isoforms described above generated by different intron retention events are exclusive to humans. Thus, the levels of these transcripts and their corresponding isoforms could present potentially interesting therapeutic targets or diagnostic and prognostic biomarkers, something that is direly needed in the clinical research applied to this terrible disease.

### Introduction

Tau is a microtubule-associated protein first described in 1975.<sup>1</sup> It is involved in several physiological processes (including microtubule stabilisation and axonal transport)<sup>2,3</sup> and it has also been associated with a series of pathologies, named tauopathies.<sup>3,4</sup>

Although frequently referred to as a single protein, tau is an ensemble of remarkably diverse isoforms generated by different alternative splicing events, yielding a current estimate of more than 50 different possible isoforms.<sup>5,6</sup>

Alternative splicing is the process by which introns of a pre-mRNA are severed co- or post-transcriptionally

in at least two different manners, so that different proteins are produced, which may not only differ structurally but can also be functionally distinct.<sup>7,8</sup> Alternative splicing encompasses a number of different events that can occur simultaneously, including exon skipping, alternative 3' and 5' splice sites, mutually exclusive exons, alternative polyadenylation, and intron retention.<sup>5,8</sup>

In the case of tau, all the isoforms of human tau are generated from a single gene, *MAPT*, which is composed of 16 exons.<sup>5,9,10</sup> From these, exons 2, 3, 4a, 6, 8 and 10 are typically considered to be subjected to alternative splicing, while the rest would remain constitutive.<sup>5,6</sup>

Six isoforms arising from the differential inclusion and exclusion of exons 2, 3 and 10 are the main isoforms of adult human Central Nervous System<sup>2,5,11</sup> while isoforms including exons 4a and 6 are present in the brain in limited amounts, being more circumscribed to peripheral nervous tissues.<sup>5,6</sup> Exon 6 can generate at least three different variants, using two additional 3' splicing sites.<sup>5,12</sup> As for exon 8, no isoform including it has been reported in humans, but it has been described in certain primates and other mammals.<sup>5,13–15</sup> Very recently, the splicing landscape of human *MAPT* has been further complicated, with our description of a human-specific isoform named W-Tau, originated by means of intron 12 retention.<sup>16</sup> This results in a truncated protein that lacks exon 13 but includes 16 amino acids arising from the translation of a small fragment of intron 12. This discovery poses the question as to whether other tau isoforms can be found arising from understudied mechanisms such as intron retention. Moreover, recent evidence has shown other *MAPT* intron-retention events such as the retention of intron 11 that generates a truncated tau isoform that is related with AD tau aggregates and is increased in the disease brain.<sup>17</sup>

Indeed, in the present work, we describe yet another set of potential isoforms of Tau generated by intron retention, in this case retaining intron 3, which would result in a truncation of the amino-terminal portion of the protein. These intron-retaining species would have something striking in common with intron-12-retaining

isoforms: they contain tryptophan (W), an amino acid that is absent in the rest of the protein. Together, these intron-retaining species that contain a W residue would comprise a previously undescribed *family* of W-Tau isoforms.

## Methods

### Nomenclature

Several *MAPT* RNA species and tau isoforms are mentioned in this work. To avoid confusion or having to describe the one we were referring to every time, we have decided to use a self-explanatory naming approach (Table 1).

For RNA species we kept the general name TIR-*MAPT* (Truncated by Intron Retention *MAPT*) and added the number of the intron retained as a superindex after TIR. Thus, we differentiate between TIR<sup>3</sup>-*MAPT*, retaining intron 3, and TIR<sup>12</sup>-*MAPT*, retaining intron 12. An additive approach was followed for those species retaining both introns (TIR<sup>3+12</sup>-*MAPT*). In any case, the exons included or excluded from the rest of *MAPT* can be indicated in the conventional way as a superindex after *MAPT*, with either a + symbol (inclusion) or a – symbol (exclusion).

On the protein level, all these isoforms conform to what we have named the *W-Tau family*, because they would contain an amino acid sequence that includes at least one tryptophan (W) residue on either the N- or the C-terminus. Isoforms containing only the sequence with a tryptophan on the N-terminal end have been named NW-Tau (N-terminal W-Tau). In contrast, those including the sequence on the C-terminus have been referred to as CW-Tau (C-terminal W-Tau). Those isoforms that would include the sequences at both termini have been named DW-Tau (Double intron retention W-Tau). Should any other intron retention events be discovered in the future, a superindex strategy can also be implemented.

### RNA-seq data

RNA-seq data was obtained and analysed as previously described.<sup>16</sup> Raw data was obtained from the Genotype-Tissue Expression project (GTEx).<sup>18</sup> A total of 363

Name	MAPT alternatively spliced exons					Retained introns	
	Exon 2	Exon 3	Exon 4a	Exon 6	Exon 10	Intron 3	Intron 12
TIR <sup>12</sup> -MAPT <sup>2-3-10-</sup>	–	–	–	–	–	–	+
TIR <sup>12</sup> -MAPT <sup>2+3-10+</sup>	+	–	–	–	+	–	+
TIR <sup>12</sup> -MAPT <sup>2+3-4a+6+10+</sup>	+	+	+	+	+	–	+
TIR <sup>3</sup> -MAPT	NA	NA	–	–	–	+	–
TIR <sup>3+12</sup> -MAPT	NA	NA	–	–	–	+	+

NA = not applicable.

**Table 1:** Nomenclature of *MAPT* alternatively spliced exons and retained introns.

samples of frontal cortex, dorsolateral prefrontal cortex, and hippocampus from 180 human brain donors without dementia were analysed (Supplementary Table S1). For donors with more than one sample in the same brain region, only the one with the highest levels of *MAPT* was analysed. FASTQ files were obtained from the SRA files and reads were re-mapped to the human genome GRCh38 using STAR 2.5.2a.<sup>19</sup> Gene expression was quantified using RSEM 1.3.1,<sup>20</sup> as Transcripts per Million (TPM). The annotation file was obtained from GENCODE v23 and was modified to include the TIR<sup>12</sup>-*MAPT* gene (coordinates chr17:45894382–46018851), which contains part of intron 12 (coordinates chr17:46018731–46018851) as the 3' end of the gene.

### Brain samples

No specific statistical method was used to predetermine sample size because there were no previous data on the expression of intron-3-retaining *MAPT* species. However, our cohort sizes are similar or larger than those reported in earlier publications in which statistically significant differences in *MAPT* and intron-retaining *MAPT* species have been demonstrated.<sup>16,21</sup> Human brain samples from hippocampus and frontal lateral cortex, including both control, non-demented participants and patients with sporadic Alzheimer's disease were obtained from Banco de Tejidos (Fundación CIEN, Instituto de Salud Carlos III, Madrid, Spain). Cerebellum and frontal cortex samples from both control and patients with Alzheimer's disease were provided by Dr. Isidro Ferrer, from Bellvitge University Hospital, IDIBELL (Bellvitge Biomedical Research Centre), Barcelona, Spain.

Upon examination of quantitative pathological features, brain specimens were classified as control subjects without AD-related pathology ( $n = 14$ ) or patients with Alzheimer's disease, classified according to neurofibrillary tangle Braak stages: I ( $n = 3$ ), II ( $n = 9$ ), III ( $n = 4$ ), IV ( $n = 1$ ), V ( $n = 14$ ) or VI ( $n = 10$ ). Efforts were made to try and ensure a well-balanced population in terms of subject's sex (60% of the samples were from male subjects and 40% from female subjects). See Supplementary Table S2 for a detailed description of each subject and the regions of brain available from each individual. Written informed consent was obtained premortem from all patients and relatives.

### Cell culture

Human neuroepithelioma SK-N-MC cells (HTB-10, ATCC. RRID:CVCL\_0530) were cultured in MEM supplemented with 10% foetal bovine serum, 2 mM glutamine, non-essential amino acids, 10 U/ml penicillin and 10 µg/ml streptomycin, at 37 °C and 5% CO<sub>2</sub>. Mycoplasma test was periodically performed using PlasmoTest (InvivoGen, #rep-pt) to rule out contamination. Validation of SK-N-MC was recently performed by STR

profiling in the Genomic service of Instituto de Investigaciones Biomédicas Sols-Morreale, Madrid, Spain.

### RNA extraction

Cytoplasmic RNA was extracted from SK-N-MC cell pellets immediately after harvesting and centrifugation using RNeasy Mini Kit (74104) purchased from Qiagen (RRID:SCR\_008539), according to manufacturer's protocols with the following modifications. Pellets were carefully resuspended in 175 µl of precooled (4 °C) buffer: 50 mM Tris-HCl pH 8, 140 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 0.5% (v/v) Nonidet P40 (1.06 g/ml) plus 1 mM 1,4-dithiothreitol just before use and incubated on ice for 5 min. Lysates were centrifuged at 4 °C for 2 min at 300×g. Supernatants were kept as cytoplasmic enriched fraction. To each fraction, 600 µl of Buffer RLT was added. After vortexing, 430 µl of ethanol 100% was added to the homogenized lysate. Samples were transferred to RNeasy spin columns, centrifuged for 15 s at 9000×g and the flow-through was discarded. Samples were treated with 10 µl of DNase in 70 µl of buffer RDD (RNase-free DNase Set 79254, Qiagen) for 15 min at room temperature. The rest of the extraction was performed following the protocol.

For brain samples, frontal lateral cortex, hippocampus, frontal cortex, or cerebellum samples were extracted using Qiazol lysis reagent (79306) from Qiagen. One stainless steel bead (Qiagen, 69989) was introduced in a clean tube with a small portion of each brain sample and was homogenised through agitation for 4 min at a frequency of 30/s on a Qiagen Retsch MM300 TissueLyser. RNA integrity numbers (RIN) were calculated using the Agilent 2100 Bioanalyzer system (Agilent Technologies RRID:SCR\_018043).

### Silencing experiments

For shRNA cloning, specific oligos containing the palindromic sequence were designed against the region of the *MAPT* sequence that encodes the fourth tandem repeat on the microtubule-binding repeats region, which is a constitutive region, shared by all tau isoforms (shMAPT-F1: CCGGACAGAGTCCAGTCCGAAGATTGCTCGAGCAATCTTCGACTGGACTCTGTTTTTTG and shMAPT-R1: AATTCAAAAAACAGAGTCCAGTCCGAAGATTGCTCGAGCAATCTTCGACTGGACTCTGT). These sequences were cloned into pLKO.1 TCR-cloning vector (RRID:Addgene\_10878) following the supplier's instructions (<http://www.addgene.org/plko>). As a control, pLKO.1-puro Non-Target shRNA Control Plasmid vector was used (SHC016 Sigma-Aldrich, RRID:SCR\_008988).

Pseudotyped lentivectors were produced in HEK293T cells transiently co-transfected with 10 µg of the lentivector plasmid described above, 6 µg of the packaging plasmid pCMVdr8.74 (RRID:Addgene\_22036) and 5 µg of the VSV G envelope protein plasmid pMD2G (RRID:Addgene\_12259) using

Lipofectamine and Plus reagents according to manufacturer's instructions (18324 and 11514, respectively, Invitrogen Life Technologies RRID:SCR\_008817, Thermo Fisher Scientific RRID:SCR\_008452).

SK-N-MC cells were transduced with lentiviral vectors encoding either *MAPT* shRNA or a scramble shRNA used as control. 48 h post-transduction, cells were pelleted, and cytoplasmic RNA was extracted as described above. Mature RNA extracts were analysed by qPCR.

### Reverse transcription and quantitative PCR

cDNA was obtained from previously extracted RNA samples with the PrimeScript 1st strand cDNA Synthesis Kit from Takara (6110A) using oligo dT primer, during 45 min at 42 °C, as per the manufacturer's instructions. A pool of samples without retrotranscriptase was included as an RT-control.

mRNA levels were then determined using quantitative PCR with fast SYBR Green reagent from Applied Biosystems (RRID:SCR\_005039, 4385612) on a CFX384 Touch Real-Time PCR Detection System (Bio-Rad, RRID:SCR\_018057). Specific oligonucleotides to detect TIR<sup>3</sup>-*MAPT* species were designed, including one forward (TIRI3-Fw) and two reverse (TIRI3-Rv1 and TIRI3-Rv2) oligonucleotides (Supplementary Table S3). Specific oligonucleotides were also employed to detect normalising genes *GAPDH*, *BACT*, *TUB* and 18 S. Finally, total *MAPT* levels were assessed, using another specific oligonucleotide pair. All of these are described in detail in Supplementary Table S3. All of them were previously tested to ensure one single, sharp peak was obtained, pointing to a single specific amplification product.

For cDNA amplification, 25 ng of cDNA template per sample were added to each well, with a mix of oligonucleotides pair for each gene to a final concentration of 5 µM and Fast SYBR Green reagent, on a 1:5 proportion. The final reaction volume was 10 µl and each sample was replicated thrice. Amplification was carried out for 40 cycles of 1s at 95 °C followed by 20s at 60 °C. No amplification was observed for non-template nor RT-controls.

### Digital droplet PCR

After a process of optimisation of reaction conditions (testing different concentrations of primers, probes and cDNA) the final reaction mix per well used was 11 µl ddPCR supermix for probes (Bio-Rad, RRID:SCR\_008426, 1863010), 250 nM of TIR3 Probe labelled with FAM fluorophore, 450 nM of TIR3 Fw and Rv primers, 500 nM of TIR12 Probe labelled with Hex fluorophore, and 900 nM of TIR12 Fw and Rv primers or Total *MAPT* Fw and Rv primers, and 10 ng of cDNA for TIR detection in a total volume of 22 µl that was manually set up in ddPCR 96-well PCR plates (Bio-Rad, 12001925). In the case of total-*MAPT* detection, similar conditions

were used but 500 nM of total-*MAPT* Probe labelled with FAM fluorophore and 1 ng cDNA. In the case of  $\beta$ -tubulin (TubB) an intercalating agent was used, and the reaction conditions were 11 µl of ddPCR EvaGreen Supermix (Bio-Rad, 1864034) and 500 nM of TubB Fw and Rv primers and 1 ng of cDNA. The sequence of the probes and primers is provided in Supplementary Table S4. No template control (NTC) in which all the reagents are added except the samples was used to exclude potential contamination of reagents and primer-dimer artifacts. Droplet generation and transfer of the 40 µl of emulsified samples to 96 well PCR plates was performed using the QX200 droplet generator (Bio-Rad 1864002) according to manufacturer's instructions. Briefly, 96-well plate with 22 µl of reaction mix was sealed with a Pierceable Foil Heat seal (Bio-Rad 1814040) using a PX1 PCR Plate Sealer (Bio-Rad 1814000), centrifuged for 2 min at 1200xg and loaded in Automated Droplet Generator. In the equipment, the 22 µl of reaction mix and 20 µl of Automated Droplet Generation Oil for Probes (Bio-Rad 1864110) were combined within the microchannel of DG32 AutoDG Cartridges (Bio-Rad 1864108) to generate the droplet emulsion that was transferred to a new 96-well plate. The Droplet emulsion plate was sealed (as a reaction mix plate) and processed in C1000 Touch Thermal Cycler (Bio-Rad). Optimized PCR thermal conditions for TIR-*MAPT* detection were 10'x 95 °C + (30'x 95 °C + 2'x 55 °C) x 50 + 10'x 98 °C. In the case of total-*MAPT* annealing temperature was 54,8 °C and the rest of the steps were unmodified. All the steps were performed with a ramp rate of 1 °C/s. In the case of TubB detection PCR thermal conditions were 5'x 95 °C + (30'x 95 °C + 1'x 60 °C) x 40 + 5'x 4 °C + 5'x 90 °C, using a with a ramp rate of 2 °C/s for all steps (per TubB reactions). After that, a 96-well plate was loaded into the QX200 Droplet Reader (Bio-Rad, 18640303). The droplet reader sips each sample, partitions the droplets, and streams them in a single file past a two-colour detector. The detector reads each droplet and determines which contain a target TIR3 (Fam +, Hex -), which contain target TIR12 (Fam -, Hex +), which contain both targets (Fam +, Hex +) and which do not contain any target (Fam -, Hex -). In the case of TubB and total-*MAPT*, the detector reads each droplet and determines which contain the target (Fam +) and which do not contain any target (Fam -). The droplet reader requires ddPCR Droplet Reader Oil (Bio-Rad, 1863004). Collection of Fam positive, Hex positive, double positive and double negative droplets and calculation of the target concentration (copies/µl) was performed by Quantasoft software v1.7.4.0917 (Bio-Rad).

For these analyses, 64 cDNAs (50 ng/µl) were obtained from human frontal cortex, hippocampus and cerebellum samples from control individuals and patients with Alzheimer's disease (Supplementary Table S2).



Then, the samples were evaluated manually to determine the threshold value to difference between the positive droplets and the negative droplets. The absolute quantification of target cDNA copies was performed by Quantasoft Software (Bio-Rad) measuring the number of positive and negative droplets for each sample. It then fits the fraction of positive droplets to a Poisson algorithm to determine the starting concentration of the target per  $\mu\text{l}$  of ddPCR reaction. To calculate the number of copies per well 22  $\mu\text{l}$  of mastermix were analysed. To increase the number of total droplets, improving the reliability of the absolute quantification, positive and negative droplets from all the wells of each sample were combined according to Droplet Digital PCR applications guide (Bio-Rad). Graphs show relative copy number by calculating the ratio between different isoforms signals.

#### Classification of AD brain using MAPT isoform quantification

Digital droplet PCR values (copies/ng) for the *MAPT* isoforms were obtained using hippocampal and frontal lateral cortex brain samples from control participants and individuals with AD. Receiver operating characteristic (ROC) curves were calculated using GraphPad Prism representing the area under the curve (AUC), 95% confidence intervals (CI), and *P*-values. ROC curves were also analysed upon 5-fold cross-validation for each isoform in both regions as well as the combination of two isoforms. For this analysis, AUC, 95% CI, and accuracy data were given (Supplementary Table S5). Accuracy estimates how many observations, both positive and negative, were accurately classified and was calculated as True Positive + True Negative/Positive + Negative.

#### Ethics

The collection of brain tissue samples was coordinated by the local Brain Bank (Banco de Tejidos CIEN Foundation, Madrid, Spain, B.000741 project PLA\_003\_ES) and Bellvitge University Hospital IDIBELL (Bellvitge Biomedical Research Centre, Barcelona, Spain, project BB001-2022), following national laws and international ethical and technical guidelines on the use of human samples for biomedical research purposes. The study was approved by the Ethical Committees of the Universidad Autónoma de Madrid (CEI-98- 18131) and the Instituto de Salud Carlos III (CEI PI 74\_2019-v2). Informed consent was obtained from all participating patients.

#### Statistics

Statistical analyses were performed using GraphPad Prism 9 software. For every analysis, outlier values for each experimental group were detected by the GraphPad outlier detector tool (<https://www.graphpad.com/quickcalcs/Grubbs1.cfm>) using a significance level  $\alpha = 0.05$  and those values were eliminated from the

analysis. The Shapiro–Wilk test or Kolmogorov–Smirnov test (when the number of samples is above 50 in RNA-seq datasets) and Q–Q plot graphical methods were used to check the normality of sample distribution (all analyses can be found at <https://doi.org/10.6084/m9.figshare.24331189> for qPCR and at <https://doi.org/10.6084/m9.figshare.24298141.v1> for ddPCR). Homoscedasticity of data was assessed by F-test for t-test analyses, using Brown–Forsythe test for ANOVA tests. In those cases in which the effect of a single variable was analysed in more than two conditions, the data were compared using one-way ANOVA tests. In cases with Gaussian data distribution, ordinary one-way ANOVA with Tukey for multiple comparison *post-hoc* analyses were used to compare the differences between individual groups. When the data distribution was not Gaussian a Kruskal–Wallis nonparametric test followed by Dunn’s *post-hoc* test for multiple comparisons was used. When unequal variances, Brown–Forsythe and Welch ANOVA test with Dunnett T3 correction for multiple comparisons was performed. A two-way ANOVA using Tukey’s *post hoc* test for multiple comparisons was used when the effects of more than one variable were analysed to search for possible interactions between them such as brain region and control or AD. When variables were independent, an unpaired t-test was used to compare two groups and Welch’s correction was applied in the case of unequal variances. When the data distribution was not Gaussian, the Mann–Whitney test was used when groups showed equal variances or the two-sample Kolmogorov–Smirnov test when the groups showed unequal variances. Graphs show scatter dot plots with bars representing means and 95% confidence intervals (CI) of data. ROC curves were calculated using GraphPad Prism 9 and RStudio using Wilson/Brown method.<sup>22</sup> The differences are given with their corresponding statistical significance or *P*-value, which is an index of compatibility between data and the model, including the test hypothesis and background assumptions (absence of bias in selection, performance, detection, attrition, and reporting) and the confidence interval or compatibility interval (CI), which is the range of parameter values that are compatible with the data under background assumptions,<sup>23</sup> using exact CI of difference between medians when non parametric test are used.

#### Role of funders

Funders had no role in study design, data collection, data analyses, interpretation, writing of the report, or the decision to submit it for publication.

#### Results

##### Intron-3-retaining splicing variants of MAPT can be found through RNA-seq of human brain samples

Recently, we have described the existence of a splicing variant of human *MAPT* generated by intron 12

retention,<sup>16</sup> in which the retention causes the lack of the translated fragment corresponding to exon 13 and presents in its place a 16-residue sequence from the translation of a portion of intron 12. Notably, this sequence includes two tryptophan residues (W), an amino acid that cannot be found at any other point of the Tau molecule. Thus, this isoform was named W-Tau.<sup>16</sup>

Intron-12-retaining species were firstly described in SH-SY5Y mature RNA and later confirmed in human tissue and contrasted in RNA-seq data of a larger cohort of 363 human samples from the frontal cortex, dorsolateral prefrontal cortex, and hippocampus of 180 healthy donors<sup>16</sup> obtained from the Genotype-Tissue Expression project (GTEx). These intron-retaining species that would give rise to W-Tau isoforms were then named TIR-MAPT (Truncated by Intron Retention MAPT) but will from hereon be referred to as TIR<sup>12</sup>-MAPT, to make a specific reference to the intron that is retained.

Further *a posteriori* examination of this dataset revealed that, although some of the samples solely displayed this intron 12 retention, most also retained intron 3 of the MAPT gene (TIR<sup>3+12</sup>-MAPT, Fig. 1). Anecdotally, among the species retaining only intron 12, the majority excluded exons 2, 3 and 10 (TIR<sup>12</sup>-MAPT<sup>2-3-10</sup>, Table 1), but a few data points show other RNA species, namely: transcripts including exons 2 and 10 simultaneously with intron 12 retention (TIR<sup>12</sup>-MAPT<sup>2+3-10+</sup>, Fig. 1a and Table 1) and transcripts that include exons 2, 3, 4a, 6 and 10 together with intron 12 (TIR<sup>12</sup>-MAPT<sup>2+3-4a+6+10+</sup>, Fig. 1a and Table 1).

Intron-3-retaining RNA species were also found without simultaneous intron 12 retention (analogously termed TIR<sup>3</sup>-MAPT, Fig. 1a). Interestingly, this *in silico* analysis revealed that TIR<sup>3</sup>-MAPT and TIR<sup>3+12</sup>-MAPT RNA species were significantly more abundant than those that retained only intron 12 ( $P < 0.0001$  Kruskal–Wallis test with Dunn's correction for multiple comparisons, Fig. 1a) in this dataset, which may suggest a more prominent biological relevance of the former or, at least, an increased potential to interact with other isoforms and other proteins within the cell. Also notably, between these more abundant species, TIR<sup>3+12</sup>-MAPT RNA is much more abundant than TIR<sup>3</sup>-MAPT ( $P = 0.0003$  Kruskal–Wallis test with Dunn's correction for multiple comparisons) in this dataset, which could hint towards shared mechanisms for both intron retention mechanisms that lead to doubly truncated species to be the majority (Fig. 1a).

Analysing the levels of all these transcripts in absolute terms, we found that TIR<sup>3+12</sup>-MAPT and TIR<sup>3</sup>-MAPT were expressed at RNA-seq detectable levels in roughly 60% and 42% of these samples, respectively, while the rest of intron-12-retaining species were expressed in a notably decreased proportion (11% of these samples for TIR<sup>12</sup>-MAPT<sup>2-3-10</sup>, and barely 3% for

TIR<sup>12</sup>-MAPT<sup>2+3-4a+6+10+</sup> and 1% for TIR<sup>12</sup>-MAPT<sup>2+3-10+</sup>) (Fig. 1b and Table 1).

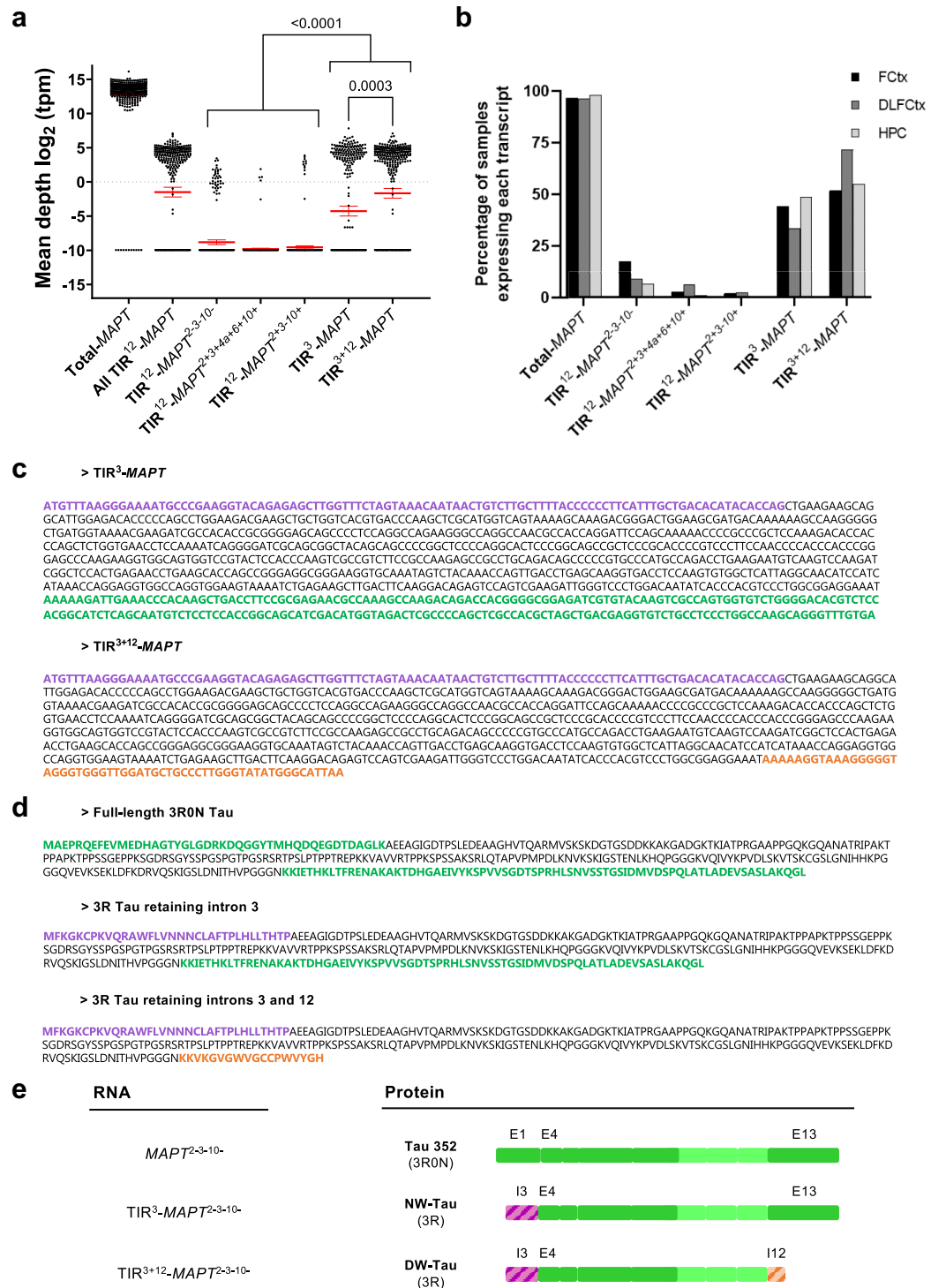
These intron-3-retaining transcripts included a specific sequence on their 5' side that differs from canonical MAPT RNA (sequence coloured in violet in Fig. 1c), but while TIR<sup>3</sup>-MAPT preserves the canonical 3' end (green in Fig. 1c), TIR<sup>3+12</sup>-MAPT displays the intron-12-retention sequence described in our previous work<sup>16</sup> (orange in Fig. 1c). The sequences displayed in Fig. 1c correspond to the RNA species that do not include exon 10. Hereon, we will use TIR<sup>3</sup>-MAPT to describe all MAPT species that retain only intron 3 regardless of the combination of exons displayed in the rest of the molecule. Similarly, we will use TIR<sup>12</sup>-MAPT or TIR<sup>3+12</sup>-MAPT to describe all MAPT species that retain only intron 12 or both introns 3 and 12, respectively, regardless of the combination of exons displayed in the rest of the molecule.

Should these mRNA species be translated, they would give rise to isoforms that differ from canonical tau in a specific 33 amino acid sequence on the N-terminal region of the molecule (MFKGKCPKVQRAWFLVNNCLAFPLHLLTHTP), that is different from the N-terminus of the analogous 3R0N full-length tau (Fig. 1d and e). In the case of transcripts that retained introns 3 and 12, the canonical tau C-terminus would also be substituted, in this case by an 18-residue sequence (KKVKGVGWVGCCPWVYGH).

Strikingly, within the 33-residue sequence resulting from the translation of the fragment of intron 3 that is retained, there is another tryptophan residue, which may hint to potential similarities in function to isoforms retaining intron 12. So as to be able to differentiate between these potential isoforms, we decided to name them according to where they display the tryptophan-containing sequence. Thus, we have named NW-Tau (N-terminal W-Tau) for those arising from TIR<sup>3</sup>-MAPT transcripts, while DW-Tau (Double W-Tau) refers to the isoforms resulting from TIR<sup>3+12</sup>-MAPT that retain both introns 3 and 12 (Fig. 1e). Analogously, we have decided to re-name the previously described W-Tau generated by retention of only intron 12<sup>16</sup> as CW-Tau (C-terminal W-Tau).

### Intron-3-retaining MAPT mRNA can be found in human neuroepithelioma cells and brain tissue

To confirm the presence of these intron-3-retaining species in human tissue and that we were able to retrieve them by means of qPCR, we designed specific oligonucleotides (named TIRI3, after Truncated by Intron Retention on Intron 3). We designed one forward oligo (TIRI3-Fw) and two slightly different reverse oligonucleotides (TIRI3-Rv1 and TIRI3-Rv2) (Supplementary Table S3, Fig. 2a). Efficiency and results obtained from these latter two were practically identical, so the results on this work will be shown using only TIRI3-Rv1 but were confirmed by using both.



**Fig. 1: RNA-seq study of Intron-3-retaining MAPT in human brain donors without dementia.** (a) RNA-seq data from GTEx was analysed. A total of 363 samples of frontal cortex, dorsolateral prefrontal cortex, and hippocampus from 180 human brain donors without dementia were used. mRNA level values [ $\log_2$  (TPM)] of the canonical MAPT gene (total-MAPT), MAPT gene with intron 12 retention (All TIR<sup>12</sup>-MAPT), and additional specific MAPT-derived transcripts, including intron retention for introns 12 or 3 are shown. Red lines represent mean mRNA levels and 95% confidence intervals (CI); and black dots mark single values. Black dots accumulating near  $\log_2$  (tpm) = -10 as a black line represent the number of samples that do not express those species. All TIR<sup>12</sup>-MAPT are all transcripts that retain intron 12. TIR<sup>12</sup>-MAPT<sup>2-3-10-</sup> are RNA species



Using these oligonucleotides, we were able to detect the existence of RNA species retaining intron 3 of human *MAPT* in mature (poly-adenylated) cytoplasmic RNA extracts from a human neuroepithelioma cell line (SK-N-MC cells) and frontal cortex samples from human brains (Fig. 2b). In every case, only samples that showed a single, sharp peak were analysed. Non-template controls and RT-controls showed no amplification of the specific product.

In comparison to the results obtained for TIR<sup>12</sup>-*MAPT*,<sup>16</sup> TIR<sup>3</sup>-*MAPT* species appeared on earlier qPCR cycles, suggesting a comparatively higher amount of the latter, which would confirm the previous RNA-seq data (Fig. 1a), pointing towards a potentially more important biological role. However, it is worth noting that the oligonucleotides used do not allow us to differentiate between TIR<sup>3</sup>-*MAPT* and TIR<sup>3+12</sup>-*MAPT* species.

In order to confirm the identity of these intron-3-retaining species, we transduced SK-N-MC cells with lentiviral vectors encoding shRNAs against total *MAPT* or a scramble, control shRNA. RNA extracts from these cells were analysed by qPCR, confirming that our shRNA can effectively silence *MAPT* (Fig. 2c). When TIR<sup>3</sup>-*MAPT* species were analysed, we could confirm that these species were indeed *MAPT* RNA, since the use of the shRNA against total *MAPT* can also decrease the number of species that are detected with TIR<sup>3</sup>-*MAPT*-specific oligonucleotides, irrespectively of whether we used TIRI3-Rv1 or TIRI3-Rv2 (Fig. 2d).

### Intron-3-retaining RNA species are expressed in healthy subjects and patients with Alzheimer's disease

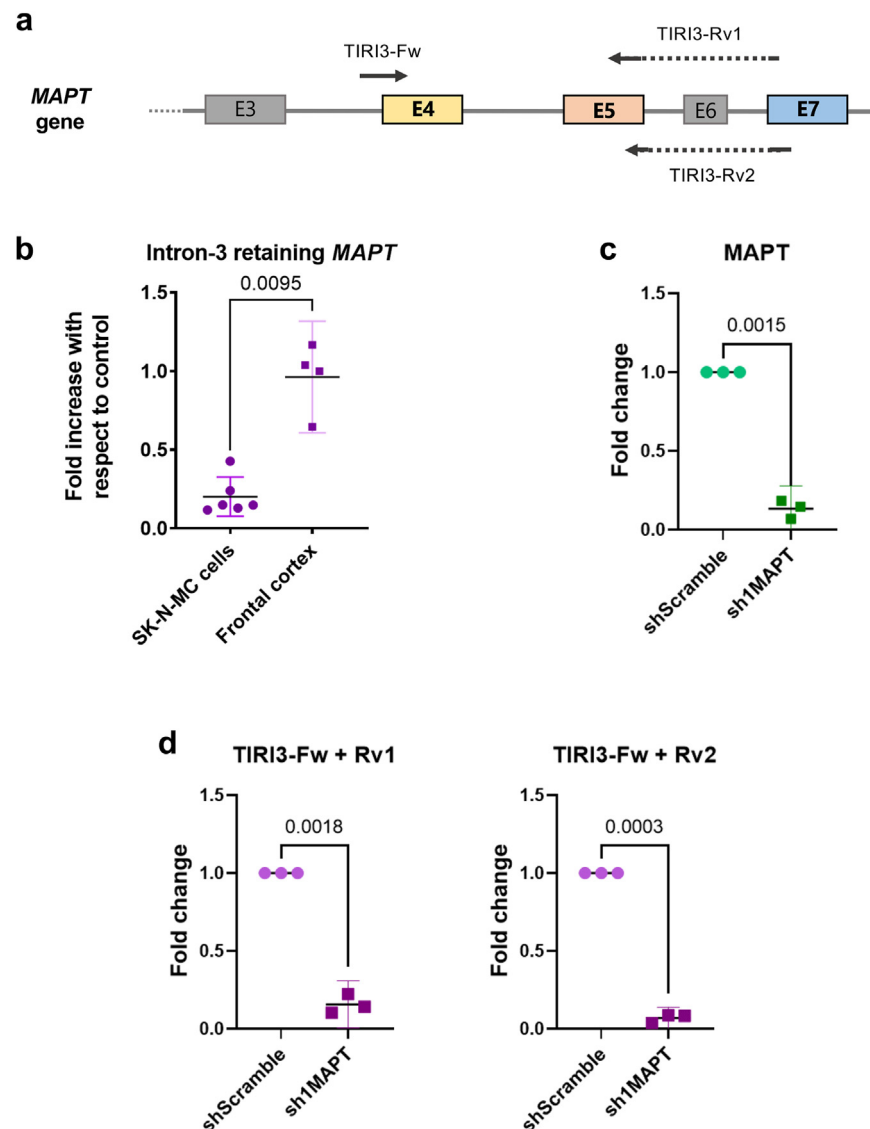
Although CW-Tau isoforms protein levels have been proven to be specifically decreased in patients with Alzheimer's disease with respect to control, participants without AD, the corresponding TIR<sup>12</sup>-*MAPT* RNA species were similarly maintained across both groups.<sup>16</sup>

With that in mind, we wanted to test TIR<sup>3</sup>-*MAPT* RNA levels by comparing control and patients with AD. Thus, we performed qPCR with the same oligonucleotides described above, comparing three different regions: frontal lateral cortex, hippocampus, and cerebellum.

Our results showed that TIR<sup>3</sup>-*MAPT* levels were not significantly different between our brain samples of control participants and patients with AD in none of the regions (Fig. 3a). However, comparing the three evaluated regions within the group of patients with AD, we observed an increase in RNA species retaining intron 3 in frontal lateral cortex (95% CI, -0.4287 to -0.01260; *P* = 0.035 one-way ANOVA with Tukey's *post-hoc* test for multiple comparisons) and cerebellum (95% CI, -0.8974 to -0.1801; *P* = 0.0019 one-way ANOVA with Tukey's *post-hoc* test for multiple comparisons), with respect to hippocampal samples (Fig. 3a). These data were also validated with TIRI3-Rv2 instead of TIRI3-Rv1 (data not shown).

In contrast, total *MAPT* levels decreased in our samples of patients with Alzheimer's disease with respect to control participants without AD in the hippocampal region (95.26% CI of difference between medians, -0.4797 to -0.04418; *P* = 0.0093 Mann-Whitney U test, Fig. 3b) and frontal lateral cortex (95% CI, -0.2397 to -0.01202; *P* = 0.031 unpaired t-test, Fig. 3b) which may be explained by neuronal death, especially in more vulnerable regions and in advanced stages of the disease. The fact that intron-retaining isoforms kept steady values while total-*MAPT* decreased suggested that intron-retained species are subjected to different regulatory mechanisms. An alternative explanation is that the steady levels maintained even in the presence of neuronal death indicate an increase in intron-retaining RNA species. When comparing different regions within each group for *MAPT*, we observed a significantly lower level in the hippocampus

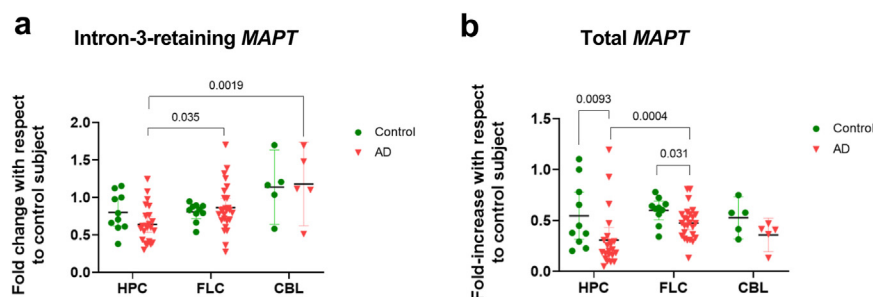
including intron 12 but excluding exons 2, 3 and 10, as reported in.<sup>16</sup> TIR<sup>12</sup>-*MAPT*<sup>2+3+4A+6+10+</sup> includes exons 2, 3, 4a, 6 and 10 in addition to the constitutive ones and intron 12 retention, while TIR<sup>12</sup>-*MAPT*<sup>2+3+10+</sup> includes exons 2 and 10 but excludes exon 3 while retaining intron 12. TIR<sup>3</sup>-*MAPT* refers to RNA species that retain only intron 3, while TIR<sup>3+12</sup>-*MAPT* represents all species that retain both intron 3 and intron 12. Kruskal-Wallis test with Dunn's correction for multiple comparison analysis was used to compare the differences between groups and *P*-values are represented. (b) Percentage of samples expressing each of the transcripts detected in (a). (c) DNA nucleotide sequence of intron-3-retaining *MAPT* species, excluding exons 2, 3 and 10, with (TIR<sup>3+12</sup>-*MAPT*) and without (TIR<sup>3</sup>-*MAPT*) simultaneous retention of intron 12. The sequence highlighted in violet represents the part of intron 3 that is retained, while the one highlighted in orange represents the part of intron 12 that is retained in TIR<sup>3+12</sup>-*MAPT*. The green-coloured sequence is the one that can be found in canonical *MAPT* transcripts not retaining intron 12. (d) Amino acid sequence of the hypothetical tau isoform resulting from intron 3 or both introns 3 and 12 retention with only three microtubule-binding repetitions (3 R), with respect to its full-length counterpart (Tau 3R0N or Tau 352). Again, green-coloured sequences represent canonical tau, while violet and orange represent the translation of the retained portion of introns 3 and 12, respectively. (e) Schematic representation of the protein isoforms that would arise from intron-retaining transcripts. The violet dashed area represents the portion of the protein arising from the translation of the portion of intron 3 that is retained, while the dashed orange one represents the equivalent portion from intron 12. All the protein representations include a label for the fragment of protein that would be generated by the translation of exon 1 (E1), exon 4 (E4), exon 13 (E13), and the part of introns 3 and 12 that are retained (I3 and I12) to highlight the differences between isoforms. NW-Tau: N-terminal W-Tau. DW-Tau: Double W-Tau.



**Fig. 2: Intron-3-retaining MAPT RNA species are confirmed to exist and are more prominent in the human brain than in cultured cells.** (a) Schematic representation of the MAPT gene sequence from exon 3 (E3) to exon 7 (E7) and the hybridisation sites for oligonucleotides detecting intron-3-retaining MAPT species: TIRI3-Fw, TIRI3-Rv1 and TIRI3-Rv2. Dashed arrows represent intron-spanning sequences on oligonucleotides. (b) qPCR results showing detection of TIR<sup>3</sup>-MAPT cDNA in SK-N-MC human neuroepithelioma cells (n = 6) with respect to frontal cortex samples from control, participants without dementia (n = 4), using the reverse oligonucleotide TIRI3-Rv1. (c) Confirmation of silencing efficiency for shRNA against MAPT in comparison to scramble shRNA (n = 3), proving MAPT can be effectively silenced. (d) Confirmation of TIR<sup>3</sup>-MAPT silencing using the same shRNA against MAPT (n = 3), confirming TIR<sup>3</sup>-MAPT's identity as MAPT species. These results are shown using both TIRI3-Rv1 and TIRI3-Rv2. Unpaired t-tests were performed to compare among groups using Welch's correction for unequal variances or the Mann-Whitney test when the data distribution was not Gaussian and equal variances. Scatter plot graphs show error-bars representing 95% confidence intervals (CI) and the corresponding P-values when P < 0.05.

with respect to frontal cortex in our samples of patients with Alzheimer's disease ( $P = 0.0004$  Kruskal-Wallis test with Dunn's correction for multiple comparisons, Fig. 3b), which seems to be a direct consequence of MAPT levels decreasing in the hippocampus in AD. Of note, cerebellum levels of total MAPT were not different

from hippocampal nor frontal lateral cortex level in our samples, but TIR<sup>3+</sup>-MAPT levels were (Fig. 3). This may imply a relative enrichment of this region with regards to RNA species retaining intron 3, which may be relevant given the cerebellum's apparent preservation to AD related pathology.



**Fig. 3: Intron-3-retaining RNA species in healthy controls and patients with AD.** (a) Comparison of intron-3-retaining cDNA levels between control participants without dementia, and patients with AD in hippocampus, frontal lateral cortex, and cerebellum. Oligonucleotide TIR3-Rv1 was used. (b) Analogous analysis to that carried out on (a) for total MAPT, comparing control and AD subjects across three different regions. HPC: hippocampus (n = 10 control subjects and 23 patients with Alzheimer's disease). FLC: frontal lateral cortex (n = 10 control subjects and 26 patients with Alzheimer's disease). CBL: cerebellum (n = 5 samples used as control and 5 patients with advanced Alzheimer's disease). Two-way ANOVA test was done to rule out interaction between variables and one-way ANOVA with Tukey's correction for multiple comparison was performed to examine the differences between the means of three brain regions when there was a Gaussian distribution of the data or Kruskal-Wallis's test with Dunn's correction for multiple comparison when the distribution of data was non-Gaussian. Unpaired t-test was performed to compare between control and AD using the Mann-Whitney test when the data distribution was not Gaussian. Scatter plot graphs show error bars representing 95% confidence intervals (CI) and the corresponding P-values between control and AD and between different brain areas when  $P < 0.05$ .

In addition, further analysis of the data showed no difference between females and males in our cohort when segregating TIR<sup>3</sup>-MAPT levels by sex whatsoever with neither TIR3-Rv1 nor TIR3-Rv2 ([Supplementary Material 1](#)).

### Abundance of different TIR<sup>3</sup>-MAPT species is related to AD pathology

The fact that intron-12 MAPT species that give rise to CW-Tau have been demonstrated to be decreased in AD<sup>16</sup> led us to wonder whether the levels of single intron-3 retention MAPT species would be altered with respect to species that include intron 12 such as TIR<sup>12+</sup>-MAPT or TIR<sup>3+12+</sup>-MAPT in health and disease.

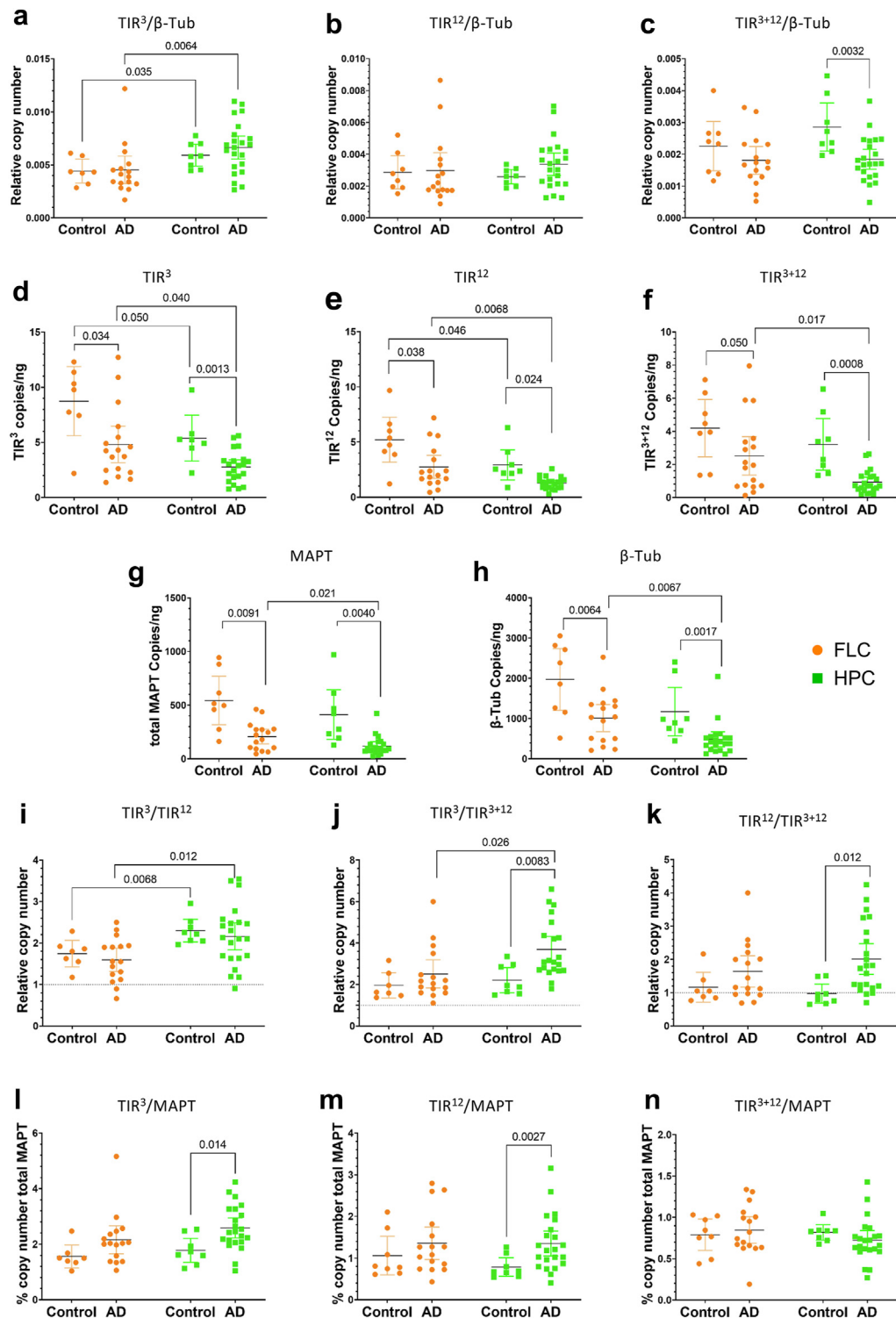
As mentioned before, our qPCR analyses were unable to differentiate between intron-3-retaining MAPT species and species that retain both intron-3 and -12. Although our RNA-seq analysis suggested the existence of species that retain both introns, this is an *in silico* analysis based on the use of STAR aligner and RSEM quantifier tools from well-known RNA-seq samples within the GTEx consortium. However, those predictions are based in the use of short reads that are not expected to include both exons in the same molecule. Therefore, we cannot fully assure that all our predictions were correct.

To unequivocally contrast these predictions, we designed a Droplet Digital PCR (ddPCR) analysis in which a pair of oligos combined with a fluorescent probe were designed against intron-3-retention and another set of oligos and a probe using other fluorophore were designed against intron-12-retention ([Supplementary Table S4](#)). This experimental design allowed us to differentiate the presence or absence of

both retentions in each single cDNA molecule, differentiating between TIR<sup>3</sup>-MAPT, TIR<sup>12</sup>-MAPT and TIR<sup>3+12</sup>-MAPT.

In order to minimise experimental errors, we selected a housekeeping gene for normalisation. With this aim, we used previous qPCR analyses in which four normalising genes were tested (18S,  $\beta$ -actin, GAPDH and  $\beta$ -tubulin). Using the *Reference Gene Selection Tool* of CFX Maestro 2.3 Software (Bio-Rad, 12013758) we detected that only  $\beta$ -actin, GAPDH and  $\beta$ -tubulin genes obtained acceptable scores in the stability plot ( $\text{Ln}(1/\text{AvgM})$ ) in all brain samples datasets, with  $\beta$ -tubulin (TubB) being the one that showed the highest stability among them. Therefore, TubB was selected for normalisation.

Using ddPCR we were able to study the relative RNA number of copies for each intron-retention with respect to TubB in frontal lateral cortex and hippocampal samples of non-demented subjects with respect to patients with AD ([Fig. 4a–c](#)). We could not find significant differences in single retention TIR<sup>3</sup>-MAPT between our cohort of patients with AD and controls. However, there were slightly higher levels in the hippocampus with respect to frontal lateral cortex in both non-demented subjects and patients with AD (95% CI, 0.00013–0.0029;  $P = 0.035$  unpaired t-test and 95.16% CI, 0.00074–0.0037;  $P = 0.0064$  Mann-Whitney U test, respectively, [Fig. 4a](#)). Similarly, we could not find differences in single retention TIR<sup>12</sup>-MAPT between regions between health and disease in our samples ([Fig. 4b](#)). Conversely, when we studied double retention TIR<sup>3+12</sup>-MAPT we could observe a diminishment in our brain samples of patients with AD with respect to control participants that was significant in the case of



**Fig. 4:** Changes in different intron-retaining MAPT RNA species between individuals without dementia and patients with AD. (a-c) Relative RNA copy number of different intron-retaining MAPT species, TIR<sup>3</sup>-MAPT (a), TIR<sup>12</sup>-MAPT (b), and TIR<sup>3+12</sup>-MAPT (c), measured by ddPCR normalised with β-tubulin (TubB). (d-h) Absolute RNA levels measured in copies per nanogram of RNA of TIR<sup>3</sup>-MAPT (d), TIR<sup>12</sup>-MAPT (e), TIR<sup>3+12</sup>-MAPT (f), total-MAPT (g), and β-tubulin (h). (i-j) Relative RNA copy number ratios of different intron-retaining MAPT species, TIR<sup>3</sup>-MAPT/TIR<sup>12</sup>-MAPT (i), TIR<sup>3</sup>-MAPT/TIR<sup>3+12</sup>-MAPT (j), and TIR<sup>12</sup>-MAPT/TIR<sup>3+12</sup>-MAPT (k), measured by ddPCR. (l-n) Percentage of different intron-retaining MAPT

hippocampal samples (95% CI,  $-0.0017$  to  $-0.00037$ ;  $P = 0.0032$  unpaired t-test, Fig. 4c).

It is well known that postmortem brain tissue, especially when affected by neurodegenerative processes, hinders the selection of optimum housekeeping genes because most of them decrease in pathology. Since ddPCR is a full quantitative technique, we studied the levels of each isoform without any normalising gene. We could observe that TIR<sup>3</sup>-MAPT copies/ng were significantly reduced in our samples from patients with AD in both frontal lateral cortex and hippocampus (Fig. 4d and 95.28% CI,  $-7.51$  to  $-0.54$ ;  $P = 0.034$  Mann-Whitney U test and 95% CI  $-4.10$  to  $-1.12$ ;  $P = 0.0013$  unpaired t-test, respectively). Additionally, the levels of TIR<sup>3</sup>-MAPT were significantly lower in our hippocampal samples with respect to our frontal lateral cortex in healthy individuals and patients with AD (Fig. 4d and 95% CI,  $-6.70$  to  $-0.0077$ ;  $P = 0.050$  unpaired t-test and  $P = 0.040$  Kolmogorov-Smirnov test, respectively). A similar pattern was found for TIR<sup>12</sup>-MAPT species (Fig. 4e). When studying the levels of double retention TIR<sup>3+12</sup>-MAPT we show a decrease in the brains of patients with AD that is extremely marked in hippocampal samples ( $P = 0.0008$ , Kolmogorov-Smirnov test). According to our previous data,<sup>21</sup> we observed a marked reduction of total-MAPT levels in our brains of patients with AD either in frontal lateral cortex or hippocampus (Fig. 4g, 95% CI,  $-564.7$  to  $-107.4$ ;  $P = 0.0091$  Welch's t-test, and  $P = 0.0040$  Kolmogorov-Smirnov test, respectively). Additionally, we observed a significant reduction in the levels of  $\beta$ -tubulin in the brains of the patients with AD (Fig. 4h, frontal lateral cortex, 95% CI,  $-1617$  to  $-297.7$ ;  $P = 0.0064$  unpaired t-test, and hippocampus  $P = 0.0017$  Kolmogorov-Smirnov test) confirming that normalising against  $\beta$ -tubulin may undermine the results of the diminishing of TIR-MAPT species.

To avoid misleading results due to a general drop in mRNA levels in the brain samples of patients with AD we analysed the ratios among different TIR-MAPT species and the percentage of expression with respect to total-MAPT levels (Fig. 4i–n).

According to our previous descriptions, ddPCR revealed that TIR<sup>3</sup>-MAPT levels were approximately twice those of TIR<sup>12</sup>-MAPT, and this ratio increased in our hippocampal samples, although no significant differences were found in our participants without AD with respect to the disease (Fig. 4i). In contrast to our *in silico* prediction from the RNA-seq dataset, we could

observe two-fold relative abundance of TIR<sup>3</sup>-MAPT species with respect to the double truncated isoforms TIR<sup>3+12</sup>-MAPT, and this ratio tended to increase in our patients with AD being significant in the hippocampus (Fig. 4j, 95.34% CI,  $0.3100$ – $2.54$ ;  $P = 0.0083$  Mann-Whitney U test) and in this area with respect to frontal lateral cortex in the disease. On the other hand, the levels of TIR<sup>12</sup>-MAPT and TIR<sup>3+12</sup>-MAPT were similar in our control subjects but were increased in our patients with Alzheimer's, being significant in our hippocampal samples (Fig. 4k  $P = 0.012$ , Kolmogorov-Smirnov test). These data suggests that increased ratio of either TIR<sup>3</sup>-or TIR<sup>12</sup>-MAPT with respect to double truncated isoforms TIR<sup>3+12</sup>-MAPT may also serve as a marker of disease neurodegeneration.

Finally, the percentage of each transcript with respect to total amount of MAPT transcribed was also studied showing that TIR<sup>3</sup>-MAPT levels were around 2%, TIR<sup>12</sup>-MAPT around 1%, and TIR<sup>3+12</sup>-MAPT slightly less than 1% of total MAPT levels in our samples (Fig. 4l–n, respectively). The percentage of single retention expression tended to increase in our patients with AD being significant in hippocampal samples (Fig. 4l–j, TIR<sup>3</sup>-MAPT, 95% CI  $0.17$ – $1.43$ ;  $P = 0.014$ , unpaired t-test and TIR<sup>12</sup>-MAPT, 95% CI,  $0.21$ – $0.91$ ;  $P = 0.0027$ , Welch's test), while double retention remained constant in our samples (Fig. 4n).

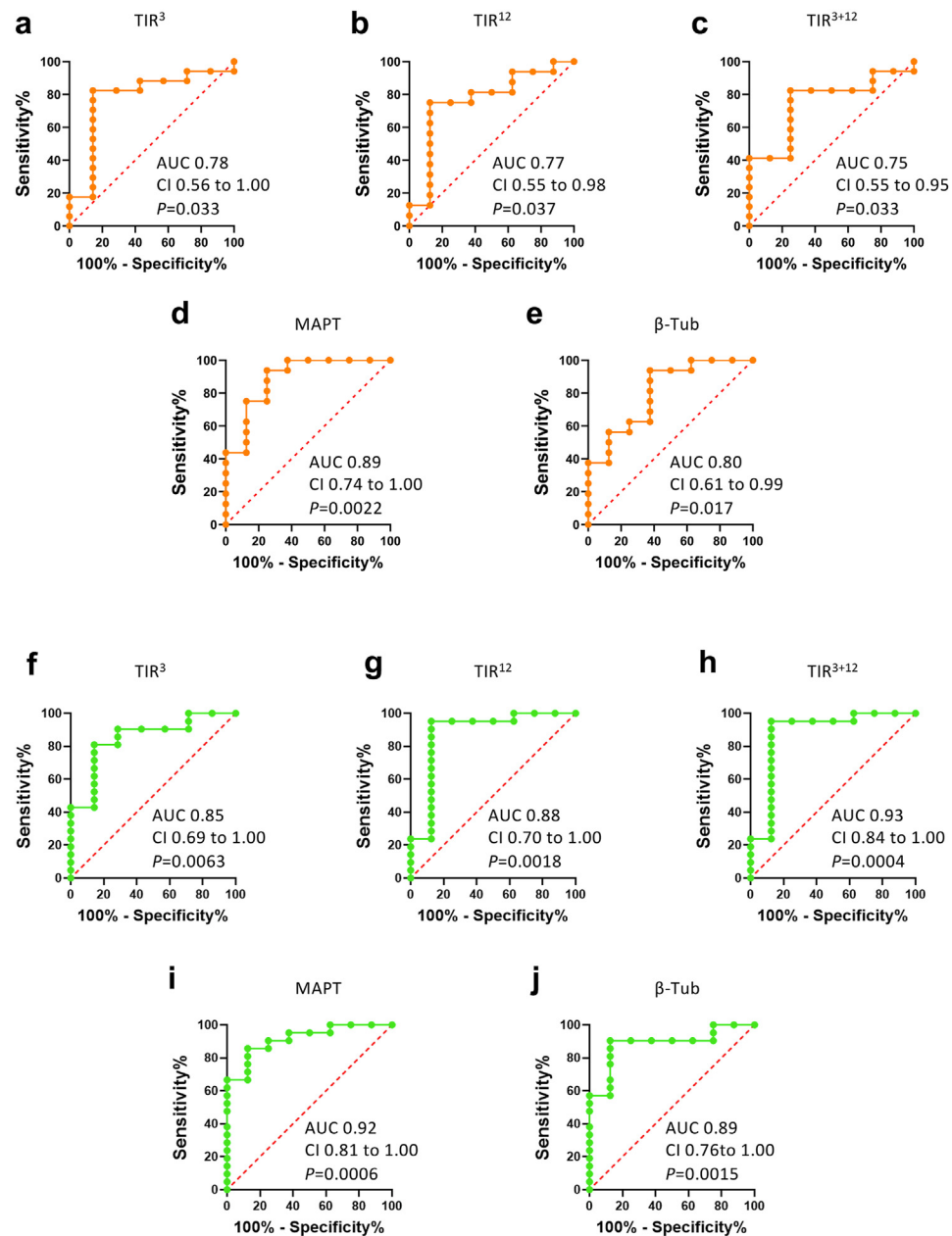
We could not observe differences between male and female subjects in health or disease, even when pooled together, in any of the parameters analysed in our samples (Supplementary Material 2).

### AD classification potential of TIR-MAPT RNA species

Due to the significant differences in expression values of TIR<sup>3</sup>-, TIR<sup>12</sup>-, TIR<sup>3+12</sup>-MAPT, total MAPT and  $\beta$ -tubulin between brains of patients with AD and control subjects, we analysed whether the isoforms can correctly classify AD and control samples. Therefore, we calculated Receiver Operating Characteristic (ROC) curves for each isoform in both brain regions, and we obtained the corresponding area under the ROC curve (AUC) values. Good classification activity is associated with AUC close to 1 and above 0.5. As represented in Fig. 5, the values were generally higher in hippocampal samples (in green) when compared to frontal lateral cortex region (in orange). This may be expected, with the hippocampus being more affected by Alzheimer's

species, TIR<sup>3</sup>-MAPT (l), TIR<sup>12</sup>-MAPT (m) and TIR<sup>3+12</sup>-MAPT (n), with respect to total-MAPT RNA copy number. Graphs show results of samples from frontal lateral cortex in orange (FLC: n = 10 control subjects and 26 patients with Alzheimer's disease, AD) and hippocampus in green (HPC: n = 10 control subjects and 23 patients with Alzheimer's disease, AD). Two-way ANOVA test was done to rule out interaction between variables and t-test analysis was performed to examine the differences between the means of two groups using Welch's correction when unequal variances or Mann-Whitney test when data distribution was non-Gaussian, and equal variance or Kolmogorov-Smirnov test when unequal variances. Scatter plot graphs show error-bars representing 95% confidence intervals (CI) and the corresponding P-values when  $P < 0.05$ .





**Fig. 5: Diagnostic relevance of TIR-MAPT species for Alzheimer's disease.** ROC curves of the classification potential of patients with AD and individuals without dementia using ddPCR values (copies/ng) of each MAPT isoform, total-MAPT and  $\beta$ -tubulin measured in brain samples of frontal lateral cortex (FLC) in orange (a-e) and hippocampus (HPC) in green (f-j). Graphs also show values of area under the ROC curve (AUC), 95% confidence intervals (CI), and P-values of each ROC curve using Wilson/Brown method.

disease pathology than the cortex. Importantly, TIR<sup>3+12</sup>-MAPT displayed an AUC = 0.93 (95% CI, 0.84–1.00;  $P = 0.0004$ ) in the hippocampus (Fig. 5h and Supplementary Table S5) showing the higher classification potential (see complete analysis at <https://doi.org/10.6084/m9.figshare.24298141.v1>). Therefore, TIR<sup>3+12</sup>-MAPT can potentially be used to distinguish AD from control brain samples in future validation

studies (Fig. 5h). This classification potential cannot be improved by any combination of the parameters analysed (Supplementary Table S5).

#### Relative abundance of different TIR-MAPT species across different brain regions

Due to the differences observed in TIR-MAPT species expression in Alzheimer's disease, we sought to

explore whether the level of each isoform may correlate with sensitivity to neurodegeneration of different regions of the brain. With this aim, we studied RNA expression levels by ddPCR of each TIR-MAPT species in frontal lateral cortex, hippocampus, and cerebellum samples of individuals with no signs of dementia including controls and Braak I and II (Fig. 6). We could observe in our set of samples that regions that are most vulnerable to AD degeneration such as frontal cortex and hippocampus, exhibited lower levels of TIR<sup>3</sup>-MAPT and TIR<sup>12</sup>-MAPT than the cerebellum, which is not affected in the disease; regardless of whether we normalised with  $\beta$ -tubulin levels (Fig. 6a and b) or evaluated absolute levels (Fig. 6d and e). However, double retention TIR<sup>3+12</sup>-MAPT is diminished in cerebellum samples with respect to hippocampus when normalised with respect to  $\beta$ -tubulin levels (Fig. 6c, 95% CI, 0.00003–0.002260,  $P = 0.043$ , ordinary one-way ANOVA with Tukey's correction for multiple comparison), but no significant differences are found in absolute levels in our brain samples (Fig. 5f). This is due to the fact that  $\beta$ -tubulin levels are increased in cerebellum with respect to hippocampal samples (Fig. 6h). Then we studied the ratios between TIR-MAPT species, and we could observe that, although there are slight differences in the TIR<sup>3</sup>/TIR<sup>12</sup>-MAPT species ratio across different regions (Fig. 6i), both TIR<sup>3</sup>- and TIR<sup>12</sup>-MAPT species ratio with respect to double truncated species TIR<sup>3+12</sup>-MAPT were highly increased in cerebellum with respect to AD susceptible regions in our samples (Fig. 6j and k). We could not find significant differences in total MAPT levels across different regions in our samples (Fig. 6g). Of note, our data showed a significant increase in the percentage of expression of TIR<sup>3</sup>-MAPT levels with respect to total MAPT levels in the cerebellum in contrast to frontal cortex and hippocampus (Fig. 6l, 95% CI, –2.209 to –0.4090;  $P = 0.0035$ ; and –1.913 to –0.1125  $P = 0.025$ ; respectively, ordinary one-way ANOVA with Tukey's correction for multiple comparisons). However, we did not observe significant differences in the percentage of TIR<sup>12</sup>-MAPT expression with respect to total MAPT in our brain samples (Fig. 6m). Conversely, our data showed that the percentage of double truncated isoforms TIR<sup>3+12</sup>-MAPT is significantly reduced in cerebellum with respect to more vulnerable areas to AD (Fig. 6n, 95% CI, 0.1620–0.6850;  $P = 0.0012$ ; and 0.09443–0.6235;  $P = 0.0063$ ; respectively, ordinary one-way ANOVA with Tukey's correction for multiple comparisons).

While more studies are necessary to fully understand the relevance of intron-3-MAPT retaining species at the protein level, our results point out that there might be a relationship between the levels of these transcript and their ratios and susceptibility to neurodegeneration, especially in Alzheimer's disease.

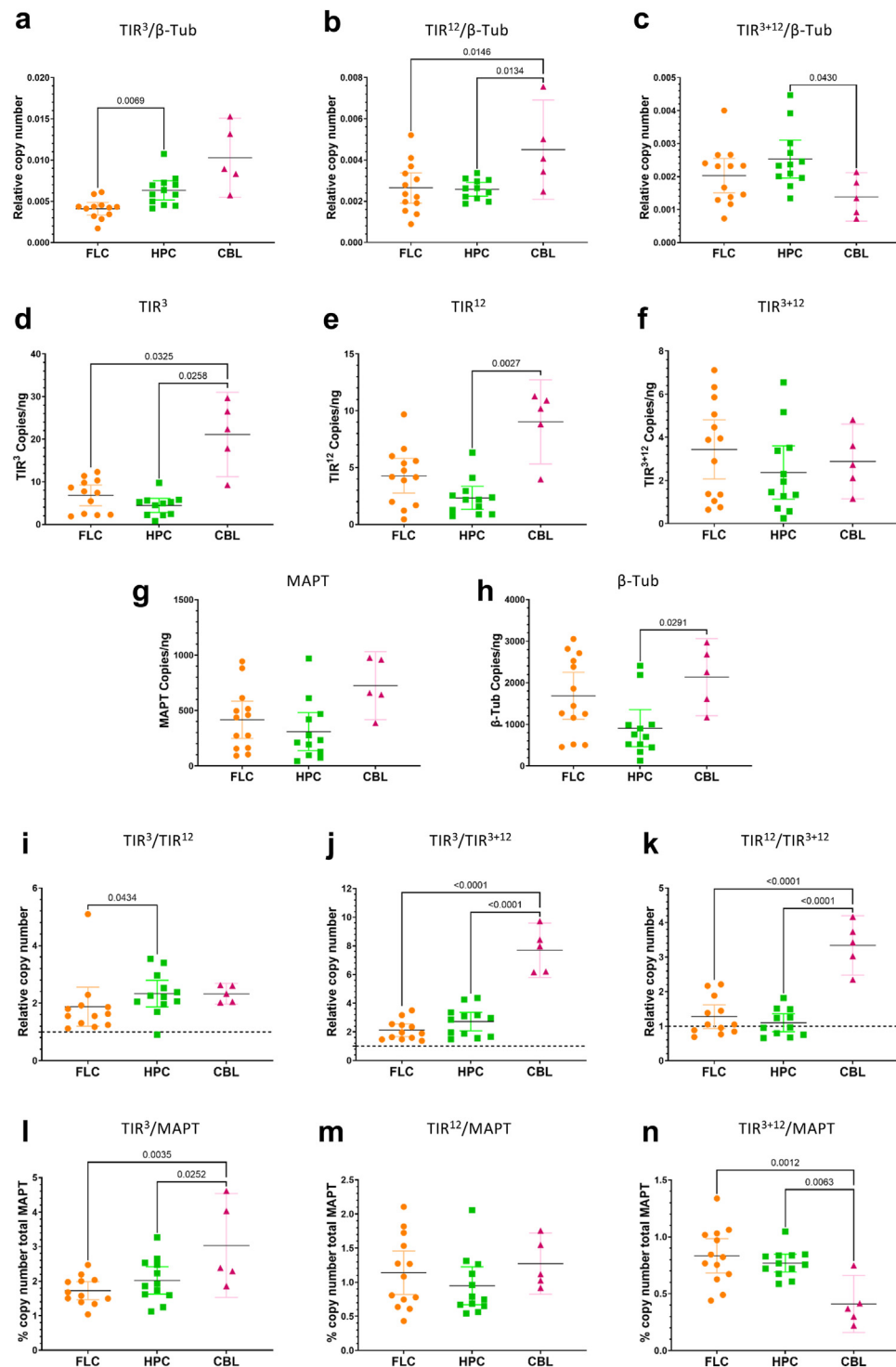
## Discussion

Although some evidence points out that alternative splicing might not be the main driver of proteome complexity as once was proposed,<sup>24</sup> some genes, including MAPT, are genuinely subjected to the alternative splicing process.<sup>24,25</sup>

The retention of intron 3 implies the addition of a specific nucleotide sequence that, if translated, generates a 33-residue fragment that would appear instead of exons 1, 2 and 3 (MFKGKCPKVQRAWFLVNN-CLAFPLHLLHTHP) (Fig. 7a). This addition does not cause a shift in the reading frame, so the rest of the tau molecule would remain intact and potentially be susceptible to undergo any pre-translational and post-translational modifications it usually goes through (Fig. 7b).

Quantitative RT-PCR analysis of our brain samples revealed that total intron-3-retaining MAPT RNA species are consistently maintained on similar levels across regions in control individuals but show notable differences in patients with AD. Further study of each isoform separately using ddPCR revealed that the absolute levels of TIR<sup>3</sup>-MAPT and TIR<sup>12</sup>-MAPT significantly decreased in our Alzheimer's disease samples. However, their percentage with respect to total MAPT tends to increase, being significant in the most vulnerable regions, such as the hippocampus. Accordingly, the absolute levels of double-retention TIR<sup>3+12</sup>-MAPT significantly decrease in the disease although their percentage of expression with respect to total-MAPT remains invariable in our brain samples. Moreover, relative abundance of TIR<sup>3</sup>-MAPT or TIR<sup>12</sup>-MAPT species with respect to double-truncated ones (TIR<sup>3+12</sup>-MAPT) are increased in our patients with AD compared to controls, being also especially significant in the hippocampus.

Notably, when studying different brain regions in patients with AD by qPCR, we could observe that the levels of total intron-3-retaining MAPT species are lower in brain regions more susceptible to this degeneration such as the hippocampus with respect to frontal lateral cortex and especially the cerebellum. Similarly, the study of each intron-retaining species separately by ddPCR in our cohort of individuals with no sign of dementia revealed that single-intron-retention TIR<sup>3</sup>-MAPT or TIR<sup>12</sup>-MAPT species are increased in cerebellum. At the same time, double-retention TIR<sup>3+12</sup>-MAPT levels are diminished or invariable. Moreover, the percentage of TIR<sup>3</sup>-MAPT to total-MAPT species is augmented, while TIR<sup>3+12</sup>-MAPT is reduced in our samples of cerebellum. Therefore, the ratios of single intron-3 or intron-12 retention with respect to double retention MAPT species are highly increased in our samples of cerebellum. Such results may be relevant, considering that the cerebellum shows less vulnerability to AD-related pathology, even if it cannot be considered



**Fig. 6: Relative abundance of  $TIR^3$ -MAPT species in different brain regions of asymptomatic patients.** (a-c) Relative RNA copy number of different intron-retaining MAPT species,  $TIR^3$ -MAPT (a),  $TIR^{12}$ -MAPT (b), and  $TIR^{3+12}$ -MAPT (c), measured by ddPCR normalized with respect to  $\beta$ -tubulin. (d-h) Absolute RNA levels measured in copies per nanogram of RNA of  $TIR^3$ -MAPT (d),  $TIR^{12}$ -MAPT (e),  $TIR^{3+12}$ -MAPT (f), total-MAPT (g), and  $\beta$ -tubulin (h). (i-j) Relative RNA copy number ratios of different intron-retaining MAPT species,  $TIR^3$ -MAPT/ $TIR^{12}$ -MAPT (i),  $TIR^3$ -MAPT/ $TIR^{3+12}$ -MAPT (j), and  $TIR^{12}$ -MAPT/ $TIR^{3+12}$ -MAPT (k), measured by ddPCR. (l-n) Percentage of different intron-retaining MAPT species,  $TIR^3$ -MAPT (l),  $TIR^{12}$ -MAPT (m), and  $TIR^{3+12}$ -MAPT (n), with respect to total MAPT RNA copy number. Graphs show results of samples from frontal

completely immune to it.<sup>26</sup> Therefore, our data point out that the ratios single-to double-intron-retention of *MAPT* significantly correlate with the sensitivity of the brain area to Alzheimer's degeneration.

It is important to note that our observations indicate a compatibility between the analysis data and the assumption of our analysis procedures. However, there is no guarantee that the background assumption embedded in our model are correct, because we cannot rule out bias in sample selection, performance, detection, attrition, and reporting as well as confounding factors.<sup>23</sup> Further analysis will be necessary, using larger cohorts and including age-matched individuals, sex-balanced cohorts, counting different ethnicities, involving different countries, and people from different cultural and socioeconomic status, comprising analysis of all comorbidities, complete medical records and patient treatments, GWAS analysis, etc, to restate our findings.

Nevertheless, our findings point out that increased levels of single-intron-retention with respect to double-intron-retention correlate with the resilience of brain tissue to Alzheimer's neurodegeneration and the increase of this ratio in patients with AD may be linked to resilience mechanisms of surviving neurons. These results suggest that single-to-double intron retention rate may be used with diagnostic proposes and further prospective studies in biological fluids will be necessary to ascertain whether it could be used to predict Alzheimer's neurodegeneration. Of note, in our brain sample cohort, absolute levels of double-intron retention show higher classification potential between individuals without dementia and patients with AD.

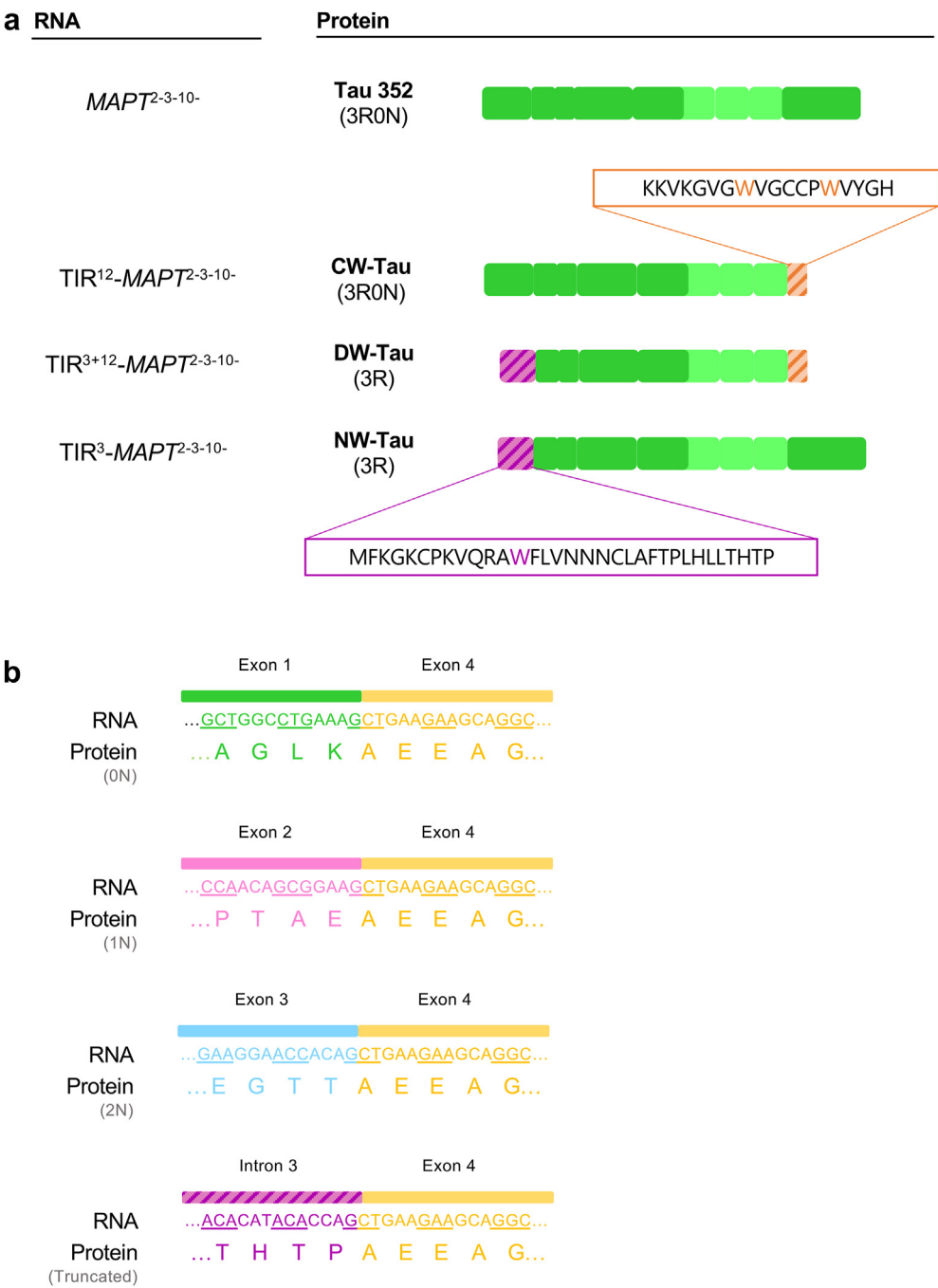
The fact that RNA levels of total intron-3-retaining *MAPT* remain steady while total *MAPT* tends to reduce in the brain of patients with AD echoes the results obtained for total intron-12- retaining *MAPT* species,<sup>16</sup> with TIR<sup>12</sup>-*MAPT* remaining stable between control and patients with AD, while total *MAPT* decreases. In fact, the percentage of number of copies of single intron retention with respect to total-*MAPT* significantly increase in our hippocampal samples of patients with AD. That may indicate that intron-retaining RNA species are subjected to specific regulatory mechanisms that keep their levels steady, even during advanced stages of Alzheimer's disease. Importantly, tau tends to accumulate in AD, especially in advanced stages of the disease,<sup>27–30</sup> usually leading to aggregation that generates neurofibrillary tangles, one of

the hallmarks of Alzheimer's disease.<sup>3,28,31</sup> This accumulation is often linked to neuronal death.<sup>2,32,33</sup> If the decreasing levels of total *MAPT* we observe are caused by such neuronal death, the fact that the ratio of intron retaining RNA species to total *MAPT* is increased even in the presence of such neurodegeneration would indicate that these RNA species are actually upregulated which would in turn correlate with what was observed for TIR<sup>12</sup>-*MAPT* RNA levels.<sup>16</sup>

On the protein level, the 33-residue specific sequence that would arise from the translation of the fragment of intron 3 that is retained contains a tryptophan (W). Tryptophan is an amino acid that can only be found in the portion of W-Tau isoforms corresponding to the translation of a fragment of intron 12,<sup>16</sup> but that is not present in any other region of the protein in the canonical sequence of tau. It is worth mentioning that there is a mutated form of Tau R406W that is linked to frontotemporal dementia, with one tryptophan residue in its sequence arising from a single mutation, which impairs membrane binding capacity.<sup>34</sup> TIR-Tau isoforms would constitute a *W-Tau family*, which share the intron retention splicing mechanism of generation and the presence of at least one tryptophan residue on either the amino- or the carboxy-terminal end (Fig. 7a). Of note, another tau isoform generated by intron-11 retention also generates two tryptophan residues in its sequence resulting from the translation of the intron (GSPVEGEGWDGRVQGVVEESW).<sup>17</sup> Therefore, it could be considered another *W-Tau* isoform generated by TIR-*MAPT* transcripts. However, the existence of tau isoforms resulting from mRNA retaining intron 3 *in vivo* as a human protein has not yet been confirmed. Future research will need to generate a specific antibody targeting the protein region resulting from the translation of intron 3 to perform the appropriate protein assays.

If any such intron-3-retaining isoforms are indeed generated, intron 3 retention would imply the loss of protein regions encoded by exons 1, 2, and 3. This region is an integral part of the N-terminal region, which is typically considered to span from exon 1 to the first amino acids of exon 7.<sup>6</sup> The N-terminal region is fundamental to tau's interaction with other proteins within the cell and cytoplasmatic membrane proteins and organelles' membranes.<sup>31,35</sup> In addition, the region encoded by exons 2 and 3 is the most conserved evolutionarily within the N-terminal region, which may indicate the existence of well-conserved interactions with specific ligands.<sup>5,31</sup> In the case of NW-Tau isoforms,

lateral cortex in orange (FLC: n = 12), hippocampus in green (HPC: n = 12) and cerebellum in pink (CBL: n = 5) of patients with no clinical signs of dementia, including controls and Braak stages I and II. One-way ANOVA test with Tukey's correction for multiple comparisons was performed to examine the differences between the means of the groups. When non-Gaussian distribution of data, Kruskal-Wallis test for multiple comparison with Dunn's correction was done. When unequal variances, Brown-Forsythe and Welch ANOVA test was performed. Scatter plot graphs show error-bars representing 95% confidence intervals (CI) and the corresponding P-values when P < 0.05.



**Fig. 7: Schematic representation of the W-Tau family.** (a) Representation of the protein sequence of isoforms retaining introns 3, 12 or both, in comparison with full-length tau. Tau 352 represents full-length tau, excluding exons 2, 3, and 10 (3R0N). 3 R (and 0 N) variants are represented for the members of the W-Tau family, but we cannot rule out that 4 R equivalents would exist, potentially. Dashed orange fragments represent the region of intron 12 translated upon intron 12's retention, whose amino acid sequence is detailed above the drawing of CW-Tau. Dashed violet fragments represent the region that would be translated if RNA species retaining intron 3 synthesise proteins, with its 33-residue specific sequence detailed below NW-Tau's illustration. (b) Representation of different exons 1, 2, or 3-exon 4 junctions depending on the splicing variants or intron 3-exon 4 junction and their translations into protein. None of the combinations alters the reading frame of exon 4.



these and any other interaction established by this region would be affected, if not wholly lost. Conversely, it is likely that the 33-residue sequence encoded by the part of intron 3 that is retained can interact with its own set of ligands.

However, as for DW-Tau, it is important to highlight that the Western blot analysis carried out when CW-Tau was described showed at least two different immunoreactive bands when probed with the antibody that recognizes the small sequence originated upon translation of intron 12's retained fragment.<sup>16</sup> The difference between these bands might be explained by partial proteolysis, but it certainly cannot be ruled out that they could be CW-Tau and DW-Tau, respectively, since the latter would be significantly smaller than the former (Fig. 7a) and both would contain intron 12's translated immunoreactive fragment.

Tau is an example of an intrinsically disordered protein, meaning that tau's conformation cannot be described as a single state.<sup>36</sup> Instead, we can point towards different potential states of the protein that together constitute a *conformational ensemble*; one of which—the *paperclip* conformation—has been thoroughly studied and proposed to be a usual conformation of soluble tau, at least *in vitro*.<sup>36,37</sup> This *paperclip* conformation entails the approach and interaction of the N-terminus and C-terminus of the protein.<sup>37</sup> Remarkably, the members of this proposed W-tau family carry a modification of either the N- or the C-terminus. This modification would make it impossible for these isoforms to adopt the *paperclip* conformation, which means they would either have a different conformation or lack one entirely. Protein conformation can affect epitope availability, post-translational modification capability, or degradation susceptibility, which would, in turn, modify the protein's functions and its ability to interact with the rest of the elements in the cell.<sup>36,38,39</sup> Ultimately, protein conformation and its properties are intimately linked.<sup>38</sup> This, together with the fact that unusual, non-aggregating properties have been described for CW-Tau, lead us to think that NW-Tau and DW-Tau isoforms would also have differential properties linked to a different conformation or the presence of tryptophan residues.

Furthermore, our results suggest that RNA species retaining intron 3 are two times more abundant than those retaining intron 12 or harbouring double intron retention. This could offer a more relevant biological function of TIR<sup>3</sup>-MAPT species and their potential protein products. It is important to note that the relative abundance of each TIR-MAPT isoform in our first analysis using the RNA-seq dataset does not match exactly our ddPCR results. It is worth mentioning that in our RNA-seq analysis, many samples showed no TIR-MAPT isoforms, which can be true biological zeros or might have undetectable levels for the sensitivity of the technique and the analysis. Those quantifications have

been done *in silico* using a STAR aligner and RSEM quantifier from high-quality RNA-seq samples within the GTEx consortium. However, we cannot fully assure that the predictions for each isoform are correct because RNA-seq does not include reads long enough to span all nucleotides from intron 3 to intron 12.

For this reason, we used ddPCR that unequivocally detects both intron sequences in the same cDNA molecule. Both techniques use oligo-dT to select only mature polyadenylated mRNA transcripts. However, the posterior processing of the samples is different. Therefore, we cannot rule out that changes in the relative abundance of each TIR-MAPT species are due to intrinsic technical variances, sensitivities, sample processing or even RNA quality.

The existence and functions of their protein products are, however, a speculative matter for the time being. In this work, we have described the presence of RNA mature species of MAPT transcripts retaining intron 3 alone or in combination with intron-12 retention, which has a relevance on its own. Namely, it confirms that intron retention is a productive mechanism of mature transcripts susceptible to being translated for the MAPT gene. Intron retention has been studied in-depth in plants and yeast, organisms in which it is considered one of the major gene expression regulation mechanisms, while it has been frequently ignored or overlooked in other organisms, including humans.<sup>40</sup> Our discovery builds on the accumulating evidence suggesting that intron retention constitutes an important regulatory mechanism of gene expression in mammals in general and humans in particular.<sup>40–42</sup> Indeed, it has even been linked to specific physiological and pathological situations such as ageing, cancer or Alzheimer's disease.<sup>15,40,43,44</sup>

As for the possible mechanism leading to intron 3 retention and the subsequent production of shorter MAPT transcripts, it is possible that these shortened species arise from the usage of an alternative promoter of MAPT, such as the one described by Huin et al.<sup>45</sup> In that work, the authors proved the existence of an alternative promoter in the intron between exons 1 and 2, which was responsible for the generation of shorter MAPT transcripts that were selectively increased in pathological conditions, such as Alzheimer's disease and progressive supranuclear palsy.<sup>45</sup> Using an alternative promoter is frequently the reason behind tissue-specific expression of different genes and has also been proposed as a suitable explanation for MAPT.<sup>5,45</sup> Together, tissue and pathology-dependent specificity may help explain the differences reported here for transcripts retaining intron 3 between hippocampus, frontal lateral cortex, and cerebellum (Figs. 3 and 6). At the moment, we cannot precisely surmise whether it is likely that intron 3 retention can be mediated by this same alternative promoter or if another such promoter exists closer to the retained region, but it is important to

consider that several CCTCCCI elements can be found within this intron. These elements, called CT boxes, constitute evolutionarily conserved sequences that have been linked to promoter-like regions.<sup>46–48</sup>

In any case, our results show that these shortened transcripts are present in non-depreciable amounts and the mechanisms by which they are generated should be investigated in the future. Interestingly, the retained 100-nucleotide sequence of intron 3 ends with a spare guanine, as do exons 1, 2 and 3, so this G comprises the first nucleotide on the first codon of exon 4 (Fig. 7b). This is what allows intron-3-retaining species to not cause a reading frameshift and generate a shortened MAPT transcript instead of a nonsense transcript that would be immediately processed.

That should, however, not be understood as a caveat that any other intron of the MAPT gene may undergo similar retention. Indeed, since three intron retention processes have been recently described for the MAPT gene<sup>16,17</sup> and present work), we cannot rule out that other such events may occur with other introns of the same gene, contributing to enrich (and complicate) its splicing landscape.

Further research is needed, however, to determine not only the mechanisms that may lead to these intron-retaining transcripts but also whether they are translated into the corresponding proteins and if these might have specific functions in physiological or pathological processes. If that would be the case, TIR<sup>3</sup>-MAPT transcripts should escape nonsense mediated decay (NMD), which is an mRNA surveillance pathway that recognizes and selectively degrades transcripts containing premature translation-termination codons (PTCs), thereby preventing the production of truncated proteins. In our case, intron-3-retention does not generate a premature stop codon, so it is not expected to be subjected to NMD. In the case of intron-12-retention, a premature stop codon is generated. However, this is followed by a canonical poly(A) site in the same intron. Therefore, there is no exon junction downstream of this stop codon, so it is not predicted to be subjected to Exon Junction Complex (EJC)-dependent NMD.<sup>49</sup> Additionally, the polyadenylation site is quite close (270 nucleotides) to the stop codon, thus it is not predicted to be driven to EJC-independent NMD. Finally, the mechanism that generates NMD relies on the recruitment of decapping enzymes and the deadenylation complex which results in the degradation of the cap and the poly(A) tail, respectively.<sup>49</sup> The fact that we have used oligo d(T) for all retrotranscription reactions in the present work to positively select polyadenylated transcripts demonstrates that measured RNA levels have escaped NMD.

Here, we provide some insight into MAPT alternative splicing and contribute to the mounting evidence showing the extraordinary diversity in the generation of transcripts from the MAPT gene, consequently opening

exciting research avenues that we will most likely see develop in the future.

#### Contributors

Conceptualisation, D.R.G., R.G.E., J.A., and V.G.E.; Methodology, D.R.G., L.V-S, R.G.E., J.A., and V.G.E.; Investigation, D.R.G., L.V-S, A.C-E, R.G-E, and V.G-E; Validation, D.R.G., L.V-S, and A.C-E; Formal analysis, D.R.G., L.V-S, A.C-E, and R.G-E; Writing—Original Draft, D.R.G.; Writing—Review & Editing, D.R.G., L.V-S, A.C-E, R.G-E, J.A., I.F., and V.G-E; Visualisation, D.R.G., L.V-S, A.C-E, R.G-E, I.F., and V.G-E; Funding Acquisition, F.H. and J.A.; Resources, I.F., F.H. and J.A.; Supervision, J.A. and V.G-E. All authors have read and approved the final version of the manuscript. D.R.G., J.A., and V.G-E. have checked and verified the underlying data.

#### Data sharing statement

All qPCR data can be found at <https://doi.org/10.6084/m9.figshare.24331189> and ddPCR raw data and further analysis are available at <https://doi.org/10.6084/m9.figshare.24298141.v1>.

#### Declaration of interests

The authors declare no competing interests.

#### Acknowledgements

This research was funded by the Spanish Ministry of Science, Innovation and Universities: PGC2018-096177-B-I00 (J.A.); Spanish Ministry of Science and Innovation (MCIN): PID2020-113204GB-I00 (F.H.) and PID2021-123859OB-I00 from MCIN/AEI/10.13039/501100011033/FEDER, UE (J.A.). It was also supported by CSIC through an intramural grant (201920E104) (J.A.) and the Centre for Networked Biomedical Research on Neurodegenerative Diseases (J.A.). The Centro de Biología Molecular Severo Ochoa (CBMSO) is a Severo Ochoa Center of Excellence (MICIN, award CEX2021-001154-S). The ddPCR experimental development and data analysis were provided by the Genomics and NGS Core Facility (GENGS) at the Centro de Biología Molecular Severo Ochoa (CBMSO, CSIC-UAM) Madrid, Spain- <http://www.cbm.uam.es/genomica>. We specially acknowledge the work of Laura Tabera (GENGS-CBMSO) in ddPCR experiments as well as María Santos-Galindo (GENGS-CBMSO) and Marta Caamaño-Moreno (Hospital Universitario 12 de Octubre, Madrid, Spain) for their advice in statistical analysis of data. The authors would like to thank CSIC Interdisciplinary Thematic Platform (PTI+) NEURO-AGING+ (PTI-NEURO-AGING+). This work was partially supported by the computing facilities of Extremadura Research Centre for Advanced Technologies (CETA-CIE-MAT), funded by the European Regional Development Fund (ERDF). CETA-CIE-MAT belongs to CIE-MAT and the Government of Spain.

We worked to ensure gender balance in the recruitment of human subjects. One or more of the authors of this paper self-identifies as a member of the LGBTQ + community. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

#### Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ebiom.2023.104953>.

#### References

- Weingarten MD, Lockwood AH, Hwo S-Y, Kirschner MW. A protein factor essential for microtubule assembly. *Proc Natl Acad Sci U S A*. 1975;72(5):1858–1862.
- Wang Y, Mandelkow E. Tau in physiology and pathology. *Nat Rev Neurosci*. 2016;17(1):22–35.
- Avila J, Lucas JJ, Perez M, Hernandez F. Role of tau protein in both physiological and pathological conditions. *Physiol Rev*. 2004;84(2):361–384.
- Buée L, Bussi  re T, Bu  e-Scherrer V, Delacourte A, Hof PR. Tau protein isoforms, phosphorylation and role in neurodegenerative disorders. *Brain Res Rev*. 2000;33(1):95–130.

- 5 Andreadis A. Tau gene alternative splicing: expression patterns, regulation and modulation of function in normal brain and neurodegenerative diseases. *Biochim Biophys Acta*. 2005;1739(2-3):91–103.
- 6 Ruiz-Gabarre D, Carnero-Espejo A, Ávila J, García-Escudero V. What's in a gene? The outstanding diversity of MAPT. *Cells*. 2022;11(5):840.
- 7 Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*. 2003;72(1):291–336.
- 8 Blencowe BJ. Alternative splicing: new insights from global analyses. *Cell*. 2006;126(1):37–47.
- 9 Neve RL, Harris P, Kosik KS, Kurnit DM, Donlon TA. Identification of cDNA clones for the human microtubule-associated protein tau and chromosomal localization of the genes for tau and microtubule-associated protein 2. *Mol Brain Res*. 1986;1(3):271–280.
- 10 Andreadis A, Brown WM, Kosik KS. Structure and novel exons of the human tau gene. *Biochemistry*. 1992;31(43):10626–10633.
- 11 Goedert M. Tau filaments in neurodegenerative diseases. *FEBS Lett*. 2018;592(14):2383–2391.
- 12 Wei ML, Andreadis A. Splicing of a regulated exon reveals additional complexity in the axonal microtubule-associated protein tau. *J Neurochem*. 1998;70(4):1346–1356.
- 13 Holzer M, Craxton M, Jakes R, Arendt T, Goedert M. Tau gene (MAPT) sequence variation among primates. *Gene*. 2004;341:313–322.
- 14 Himmler A. Structure of the bovine tau gene: alternatively spliced transcripts generate a protein family. *Mol Cell Biol*. 1989;9(4):1389–1396.
- 15 Andreadis A. Tau splicing and the intricacies of dementia. *J Cell Physiol*. 2012;227(3):1220–1225.
- 16 García-Escudero V, Ruiz-Gabarre D, Gargini R, et al. A new non-aggregative splicing isoform of human Tau is decreased in Alzheimer's disease. *Acta Neuropathol*. 2021;142(1):159–177.
- 17 Ngian Z-K, Tan Y-Y, Choo C-T, et al. Truncated Tau caused by intron retention is enriched in Alzheimer's disease cortex and exhibits altered biochemical properties. *Proc Natl Acad Sci U S A*. 2022;119(37):e2204179119.
- 18 Lonsdale J, Thomas J, Salvatore M, et al. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013;45(6):580–585.
- 19 Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
- 20 Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12(1):1–16.
- 21 García-Escudero V, Gargini R, Martín-Maestro P, García E, García-Escudero R, Ávila J. Tau mRNA 3' UTR-to-CDS ratio is increased in Alzheimer disease. *Neurosci Lett*. 2017;655:101–108.
- 22 Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. *Stat Sci*. 2001;16(2):101–133.
- 23 Mansournia MA, Nazemipour M, Etminan M. P-value, compatibility, and S-value. *Glob Epidemiol*. 2022;4:100085.
- 24 Tress ML, Abascal F, Valencia A. Alternative splicing may not be the key to proteome complexity. *Trends Biochem Sci*. 2017;42(2):98–110.
- 25 Abascal F, Ezkurdia I, Rodríguez-Rivas J, et al. Alternatively spliced homologous exons have ancient origins and are highly expressed at the protein level. *PLoS Comput Biol*. 2015;11(6):e1004325.
- 26 Gellersen HM, Guell X, Sami S. Differential vulnerability of the cerebellum in healthy ageing and Alzheimer's disease. *Neuroimage Clin*. 2021;30:102605.
- 27 Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol*. 1991;82(4):239–259.
- 28 Nelson PT, Alafuzoff I, Bigio EH, et al. Correlation of Alzheimer disease neuropathologic changes with cognitive status: a review of the literature. *J Neuropathol Exp Neurol*. 2012;71(5):362–381.
- 29 Goedert M, Spillantini M, Jakes R, Rutherford D, Crowther R. Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease. *Neuron*. 1989;3(4):519–526.
- 30 Naseri NN, Wang H, Guo J, Sharma M, Luo W. The complexity of tau in Alzheimer's disease. *Neurosci Lett*. 2019;705:183–194.
- 31 Brandt R, Trushina NI, Bakota L. Much more than a cytoskeletal protein: physiological and pathological functions of the non-microtubule binding region of tau. *Front Neurol*. 2020;11:1269.
- 32 Cotman CW, Su JH. Mechanisms of neuronal death in Alzheimer's disease. *Brain Pathol*. 1996;6(4):493–506.
- 33 Yanagisawa K. 3. Neuronal death in Alzheimer's disease. *Intern Med*. 2000;39(4):328–330.
- 34 Gauthier-Kemper A, Weissmann C, Golovayshkina N, et al. The frontotemporal dementia mutation R406W blocks tau's interaction with the membrane in an annexin A2-dependent manner. *J Cell Biol*. 2011;192(4):647–661.
- 35 Brandt R, Léger J, Lee G. Interaction of tau with the neural plasma membrane mediated by tau's amino-terminal projection domain. *J Cell Biol*. 1995;131(5):1327–1340.
- 36 Novak P, Cehlar O, Skrabana R, Novak M. Tau conformation as a target for disease-modifying therapy: the role of truncation. *J Alzheimers Dis*. 2018;64(s1):S535–S546.
- 37 Jeganathan S, Chinnathambi S, Mandelkow E-M, Mandelkow E. Conformations of microtubule-associated protein Tau mapped by fluorescence resonance energy transfer. *Methods Mol Biol*. 2012;849:85–99.
- 38 Stollar EJ, Smith DP. Uncovering protein structure. *Essays Biochem*. 2020;64(4):649–680.
- 39 Dagliyan O, Hahn KM. Controlling protein conformation with light. *Curr Opin Struct Biol*. 2019;57:17–22.
- 40 Jacob AG, Smith CW. Intron retention as a component of regulated gene expression programs. *Hum Genet*. 2017;136(9):1043–1057.
- 41 Sakabe NJ, De Souza SJ. Sequence features responsible for intron retention in human. *BMC Genomics*. 2007;8(1):1–14.
- 42 Song R, Tikoo S, Jain R, et al. Dynamic intron retention modulates gene expression in the monocytic differentiation pathway. *Immunology*. 2021;165(2):274–286.
- 43 Adusumalli S, Ngian ZK, Lin WQ, Benoukraf T, Ong CT. Increased intron retention is a post-transcriptional signature associated with progressive aging and Alzheimer's disease. *Aging Cell*. 2019;18(3):e12928.
- 44 Inoue D, Polaski JT, Taylor J, et al. Minor intron retention drives clonal hematopoietic disorders and diverse cancer predisposition. *Nat Genet*. 2021;53(5):707–718.
- 45 Huin V, Buée L, Behal H, Labreuche J, Sablonnière B, Dhaenens C-M. Alternative promoter usage generates novel shorter MAPT mRNA transcripts in Alzheimer's disease and progressive supranuclear palsy brains. *Sci Rep*. 2017;7(1):1–10.
- 46 Zilberman A, Dave V, Miano J, Olson E, Periasamy M. Evolutionarily conserved promoter region containing CARG\*-like elements is crucial for smooth muscle myosin heavy chain gene expression. *Circ Res*. 1998;82(5):566–575.
- 47 Brown AR, Simmen RC, Raj VR, Van TT, MacLeod SL, Simmen FA. Krüppel-like factor 9 (KLF9) prevents colorectal cancer through inhibition of interferon-related signaling. *Carcinogenesis*. 2015;36(9):946–955.
- 48 Lutz W, Schwab M. In vivo regulation of single copy and amplified N-myc in human neuroblastoma cells. *Oncogene*. 1997;15(3):303–315.
- 49 García-Moreno JF, Romão L. Perspective in alternative splicing coupled to nonsense-mediated mRNA decay. *Int J Mol Sci*. 2020;21(24):9424.