Alternative promoter usage generates novel shorter MAPT mRNA transcripts in Alzheimer’s disease and progressive supranuclear palsy brains

Vincent Huin1, Luc Buée1, Hélène Behal2, Julien Labreuche2, Bernard Sablonnière2 & Claire-Marie Dhaenens1

Alternative promoter usage is an important mechanism for transcriptome diversity and the regulation of gene expression. Indeed, this alternative usage may influence tissue/subcellular specificity, protein translation and function of the proteins. The existence of an alternative promoter for MAPT gene was considered for a long time to explain differential tissue specificity and differential response to transcription and growth factors between mRNA transcripts. The alternative promoter usage could explain partly the different tau proteins expression patterns observed in tauopathies. Here, we report on our discovery of a functional alternative promoter for MAPT, located upstream of the gene’s second exon (exon 1). By analyzing genome databases and brain tissue from control individuals and patients with Alzheimer’s disease or progressive supranuclear palsy, we identified novel shorter transcripts derived from this alternative promoter. These transcripts are increased in patients’ brain tissue as assessed by 5′RACE-PCR and qPCR. We suggest that these new MAPT isoforms can be translated into normal or amino-terminal-truncated tau proteins. We further suggest that activation of MAPT’s alternative promoter under pathological conditions leads to the production of truncated proteins, changes in protein localization and function, and thus neurodegeneration.

Along with alternative splicing and alternative transcription, alternative promoter usage (APU) is an important mechanism for transcriptome diversity and flexibility, and the control of gene expression (for a review, see 1). Indeed, APU may modulate (i) alternative splicing, (ii) the tissue specificity2, regional specificity3,4 and/or subcellular specificity5 of gene expression, and (iii) gene activation during development6. Moreover, APU may alter the N-terminal primary structure of the expressed protein, which in turn may affect the latter's functional activity or subcellular localization7. For a growing number of genes, APU has been evidenced - especially in the brain, where up to 60% of transcribed genes are concerned4. For instance, the existence of several different promoters has already been described for genes linked to Alzheimer’s disease (AD), such as the apolipoprotein E gene APOE (the main AD genetic risk factor8) and the presenilin-2 gene PSEN2 (one of the catalytic moieties in the gamma secretase complex involved in amyloid precursor protein (APP) cleavage and generation of amyloid beta peptide (Aβ), which accumulates outside the cell9,10). After Aβ deposition, the second hallmark of AD is the intraneuronal aggregation of hyperphosphorylated tau proteins. Abnormal tau aggregation is also associated with other neurodegenerative diseases, referred to collectively as tauopathies.

Tau proteins are encoded by the microtubule-associated protein tau gene (MAPT) on chromosome 17. MAPT contains 15 exons, which are usually numbered from 0 to 13. MAPT mRNA transcripts undergo alternative splicing as a function of the developmental stage, tissue and species, with the inclusion or exclusion of six alternative exons (E2, E3, E4a, E6, E8 and E10). Exons 2, 3 and 10 are specific to the adult brain11, and their alternative

1Univ. Lille, Inserm, CHU Lille, UMR-S 1172 - JP Arc - Centre de Recherche Jean-Pierre AUBERT Neurosciences et Cancer, F-59000, Lille, France. 2Univ. Lille, CHU Lille, EA 2694 - Santé publique: épidémiologie et qualité des soins, Unité de Biostatistiques, F-59000, Lille, France. Correspondence and requests for materials should be addressed to V.H. (email: vincent.huin@inserm.fr)
splicing generates six different mRNA transcripts and thus six different proteins after translation (referred to as 2N3R, 1N3R, 0N3R, 2N4R, 1N4R and 0N4R). The designations 0N, 1N and 2N respectively indicate the absence of both E2 and E3, the presence of E2 but not E3, and the presence of both exons E2 and E3. The designations 3R and 4R indicate the absence or presence of E10. Some of these isoforms aggregate with specific patterns, which define five classes of tauopathies (for a review, see12). These tauopathies also vary with respect to their specific clinical and neuropathological features. However, the mechanisms underlying the expression of specific MAPT transcripts and proteins have yet to be characterized.

To date, only one promoter (located around exon 0) has been described for MAPT. The existence of an additional promoter has been suggested by Andreidis11, with a view to explaining tissue-specific differences in MAPT mRNA profiles and the MAPT gene's transcriptional response to transcription and growth factors. Interestingly, our study of tissue from the frontal cortex of patients with progressive supranuclear palsy (PSP) previously highlighted significant hypomethylation of a CpG site located in the CpG island shore 13kb upstream of E113. This region contains several putative regulatory features, such as a CpG island, DNase I clusters, binding sites for the transcription factor CCCTC-binding factor, and the histone mark H3K27Ac. The presence of these features is compatible with the existence of an enhancer linked to another promoter downstream. Furthermore, the suggestion of an alternative promoter is strengthened by reports of shorter MAPT mRNAs (starting from exon 1) in genome databases14.

Our starting hypothesis was that MAPT contains an alternative promoter within intron 0, and that expression of this promoter is deregulated in a context of disease. Moreover, APU might lead to the transcription of shorter transcripts, which would encode proteins with different tissue specificities and functions (as has been reported for PSEN2 products). Shorter translation products might then be involved in the pathogenesis of tauopathies. Our current results are consistent with this hypothesis. We described its sequence and the transcription start sites for shorter transcripts (starting at exon 1). Moreover, we found that levels of these shorter transcripts are elevated in brain tissues from patients suffering from tauopathies AD or PSP.

Results
Bioinformatics analysis of the human MAPT alternative promoter. The UCSC (genome assembly GRCh37/Hg19), Ensembl (GRCh37) database contained reports of five shorter MAPT transcripts beginning at exon 1 (Fig. 1A). These MAPT transcripts contain the same coding sequence and the same alternative cassettes as the longest MAPT transcripts but differ in their 5′ untranslated region (UTR) sequence. This observation suggested the presence of an additional mechanism for transcriptional regulation. When looking at the human genome issue Hg19 with the ECR browser (https://www.ecrbrowser.dcode.org), we found that the region of the MAPT gene encompassing the 5′ intronic region upstream of exon 1 and then exon 1 itself is conserved among mammals (Fig. 1B). Moreover, the Ensembl database revealed additional transcription start sites throughout the gene but especially in exon 1. Initiation of transcription at several sites is a characteristic of TATA-less promoters and a sequence analysis did not identify any TATA or CAAT boxes for exon 0 or exon 1. Analysis with the Gene Promoter Miner server (Institute of Bioinformatics, National Chiao Tung University, Hsinking-Chu, Taiwan) (http://gpmminer.mbc.nctu.edu.tw/about.htm) revealed a specific feature of promoters for Sp family transcription factors: a GC box 141 bp upstream of exon 1. The Promoter 2.0 Prediction server (Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark) (http://www.cbs.dtu.dk/services/Promoter/) predicted a promoter at 900 bp (score: 1.127) and 500 bp upstream of exon 1 (score: 0.53), and at the first codon of exon 1 (score: 0.55). The Neural Network Promoter Prediction server (version 2.2, Lawrence Berkeley Laboratory, Berkeley, CA, USA) (http://www.fruitfly.org/seq_tools/promoter.html) indicated the presence of promoters at 1382 bp (score: 0.84), 568 bp (score: 0.95) and 153 bp (score: 0.92) upstream of exon 1. Lastly, the Regulatory Sequence Analysis Tools (Metazoa, Plateforme ABIMS Roscoff, France) (http://rsat.eu), Nsite (version 5.2013, Softberry, Inc., Mount Kisco, NY, USA) (http://www.softberry.com/berry.phtml?topic=rsite&group=help&subgroup=promoter) and P-Match-1.0 (Helmholtz Centre for Infection Research, TRANSFAC/BIOBASE GmbH, Braunschweig, Germany) (http://gene-regulation.com/pub/programs.html) software tools predicted a large number of cis-regulatory elements and transcription factor binding sites (e.g. YY1 and E2F) in the 2000 bp upstream of exon 1 (Fig. 1C).

Functional assay of promoter B (upstream exon 1). In order to demonstrate the existence of an alternative promoter, we tested the ability of a 2 kb fragment of the genomic region upstream of MAPT exon 1 (promoter B) to drive the expression of the luciferase gene in transiently transfected HeLa and SH-SY5Y cells (Fig. 2A). Promoter B’s activity was compared with that obtained with an empty pGL3-basic vector. In HeLa and SH-SY5Y cells, use of the promoter B construct was associated with a statistically significant, two-fold increase in luciferase gene expression (p = 0.028 and p = 0.019 vs. the empty vector for HeLa and SH-SY5Y, respectively) – thus confirming promoter B’s functionality (Fig. 2B). As expected, the pGL3-promoter A construct (used as a positive control in our transient transfection studies) displayed a high level of luciferase activity (Fig. 2B).

To define the minimal essential region for the promoter B, we generated constructs with different 5′ deletions of the promoter B-1879 vector and tested them in a HeLa cell luciferase activity. Promoter B’s functional activity was lost in the B-606, B-293, B-106 and B + 30 vectors (Fig. 2C). Hence, the results of this deletion analysis did not enable us to define a classical core for promoter B – suggesting that this alternative promoter has a complex regulatory mechanism and/or structure.

5′RACE-PCR and characterization of the shorter transcripts. In order to assess promoter B’s functional status in human brain tissue, we performed a Rapid amplification of cDNA 5′end (5′RACE) and cloned the shortest MAPT mRNA isoforms. 5′RACE amplification was performed with a probe located at the junction between MAPT exons 11 and 12 (Fig. 3A), and enabled us to clone 64 different cDNA transcripts. Fifty-four
of these were in-frame and could potentially be translated into tau proteins. More specifically, we cloned 31 cDNAs starting from different sites in exon 0 (between −205 and −11 bp upstream of the exon’s 5′ end; Fig. 3B). Interestingly we found 14 shorter clones beginning at exon 1 in all three study pools (i.e. non-clinical controls, AD patients and PSP patients). We also cloned 6 cDNAs starting from exon 2 (Supplemental Table 1). The existence of transcripts with different 5′ termini reflected the usage of the alternative promoter B of MAPT gene in human brain tissue.

A bioinformatics analysis of the MAPT sequence revealed the presence of three ATG codons respecting the consensus Kozak sequence. The first translation initiation site is located at the 18th nucleotide in exon 1 (c.1, Nm_005910.5; GRCh37/Hg19); this is the conventional initiation start site for tau proteins. The second translation initiation site is located at the 48th nucleotide in exon 1 (c.31), and the third and last is located at the 6th nucleotide in exon 5 (c.379). With reference to the 2N4R tau isoform containing 441 amino acids (Np_005901; GRCh37/Hg19), these three translation initiation sites correspond to methionines (Met) 1, 11 and 127, respectively. These new isoforms might subsequently be translated into normal or N-terminal-truncated tau proteins.

Figure 1. A bioinformatics analysis of the human MAPT gene’s alternative promoter. (A) The various transcripts of the MAPT gene reported in Ensembl: the six conventional transcripts of MAPT and five shorter transcripts starting in exon 1. (B) The exon 1 of MAPT and its upstream region are conserved in mammals, as shown by alignment with the human Hg19 DNA sequences of MAPT orthologues in various organisms, using the ECR browser (https://www.ecrbrowser.dcode.org). Exon 1’s coding sequences are shown in blue and the non-coding sequences are shown in yellow. (C) Sequence of the putative alternative promoter of MAPT. Exon 1’s sequence is depicted in red capital letters. Putative transcription factor binding sites (obtained with RSAT software and the TRANSFAC database) are marked with a white box, a gray box and in blue. The GC box is underlined in pink. Two ATG codons (corresponding to methionines 1 and 11) are framed in red.
Short isoforms of MAPT transcripts are overexpressed in AD and PSP brains. Elevated levels of MAPT transcripts not starting at exon 0. To measure the expression level of short MAPT transcripts, we used TaqMan® quantitative reverse transcriptase-PCRs (qRT-PCR) with a different set of probes. Firstly, we measured the expression level of total MAPT transcripts (i.e. those initiated by all putative promoters) by using a probe shared by all known coding transcripts (the exon 12/13 junction) (Fig. 4A). Secondly, the expression of transcripts initiated by the constitutive promoter A was measured with a specific probe for the exon 0/1 junction (Fig. 4B). In AD patients, MAPT mRNA levels were not significantly higher in either area (Fig. 4A), but were associated with low levels of transcripts initiated by promoter A in both regions ($p = 0.0002$ and $p = 0.003$, after Dunnett’s adjustment for multiple comparisons) (Fig. 4B). In samples from PSP patients, we observed significantly higher levels of total MAPT mRNA in tissue from the frontal cortex ($p = 0.006$) but not in the occipital cortex (which is free of neurofibrillary degeneration) (Fig. 4A). However, in frontal cortex, this elevated levels of total MAPT transcripts was associated with normal levels of transcripts initiated by promoter A ($p = 0.70$) (Fig. 4B). The discrepancy between the overall MAPT mRNA expression and expression of mRNA from promoter A strongly suggests the presence of high levels of MAPT transcripts not starting at exon 0 in AD and PSP brains.

Semi-quantitative assessment of MAPT transcripts starting at exon 1. Based on the results of the 5‘RACE cloning and qPCR experiments, we next sought to directly measure the expression levels of transcripts beginning at exon 1 and thus assess their putative involvement in AD and PSP. Hence, we performed two independent 5‘RACE-PCRs with mRNA from 34 brain samples (13 controls, 11 AD brains and 10 PSP brains). In line with the results of the 5‘RACE-PCR using gene specific primer 11-12 (GSP 11-12), we observed a large number of exon-0-containing MAPT transcripts of different lengths. These appeared to be the major transcripts in the brain tissue. In all patients (but not in all controls), shorter transcripts starting at exon 1 were revealed (i) by the 5‘RACE-PCR using GSP1 as a faint band of around 190 bp and (ii) by the 5‘RACE-PCR using GSP4 as bands of between 240-300 bp. The electrophoresis gel and its interpretation are described in detail in Supplemental Figures S1 and S2. Hence, two independent 5‘RACE-PCRs confirmed that a small proportion of MAPT transcripts in the brain start at exon 1.

Semi-quantitative assessment using Fiji ImageJ software showed significant greater levels of shorter transcripts in the frontal cortex from AD and PSP brains than in controls. This difference was observed in 5‘RACE assays.
Figure 3. *MAPT* transcripts with different transcription start sites in human brain. (A) Schematic representation of the *MAPT* gene. Exons and introns are shown as boxes and horizontal lines, respectively. Untranslated regions are indicated as black boxes, constitutive exons are marked in white, alternative exons in the brain are marked in light gray, and exons presenting alternative splicing in other tissues are marked in dark gray. The different transcription initiation sites are indicated by black perpendicular arrows. The locations of the start codon (M1), the two putative alternative start codons (M11 and M127) and the GSP11-12 (arrow) are given below the *MAPT* gene (Nm_005910.5/Np_005901; GRCh37/Hg19). (B) Schematic representations of the different in-frame mRNAs cloned by 5′RACE, using the GSP11-12. Exons are shown as boxes and spliced introns are depicted as lines. For new transcripts with TSSs in exon 1, the predicted start codons are depicted to the left of the transcripts.

Figure 4. Quantitative qPCRs of the transcripts initiated from promoter B. (A) Quantification of all the coding *MAPT* transcripts in postmortem brain tissues, using the TaqMan® probe Hs00902194_m1. (B) Quantification of *MAPT* transcripts initiated from promoter A, using a custom TaqMan® probe overlapping with exon 0 and exon 1. The median [Interquartile range] level of *MAPT* expression is reported for each study group. *p < 0.05 calculated using ANOVA with posthoc Dunnet's test AD: Alzheimer's disease; PSP: progressive supranuclear palsy.
using GSP1 (Fig. 5A; \( p = 0.008 \) in the AD brain and \( p = 0.085 \) in the PSP brain) and GSP4 (Fig. 5B; \( p = 0.0006 \) in the AD brain and \( p = 0.005 \) in the PSP brain).

Overall, the results of the qRT-PCRs and the two independent 5′RACE-PCRs showed that although transcripts initiated by promoter B (upstream exon 1) of MAPT account for a small proportion of total MAPT transcripts, their levels are higher in AD and PSP brains than in control brains.

**Discussion**

Our present results showed that like many genes expressed in the brain, MAPT has an alternative promoter. This promoter is located upstream of exon 1 and is active in brain tissue. Moreover, our analysis of brain tissue samples from control individuals, AD patients and PSP patients revealed (i) new, shorter transcripts issued from this alternative promoter, and (ii) additional transcripts starting from exon 0 with different transcription start sites. Lastly, we showed that MAPT’s alternative promoter is active in brain tissues affected by the tauopathies AD and PSP.

Although we queried three promoter prediction programs, only one predicted a promoter in exon 1. Hence, promoter prediction is a valuable tool but the outputs must be interpreted with a degree of caution. In fact, these prediction techniques are based on the accumulation of indirect markers of the presence of a promoter (such as GC, TATA and CAAT boxes). In the case of MAPT, it has already been reported that the constitutive promoter in exon 0 is a TATA- and CAAT-less promoter. Accordingly, we did not expect the alternative promoter to contain a TATA box. We therefore focused our attention on transcription factor binding sites and elements indicating the presence of a promoter in the exon 1 region. Interestingly, we observed an accumulation of transcription factor binding sites at the start of the exon 1; these included Sp1 (a transcription factor that binds to GC-rich motifs in many promoters), E2F and YY1 (transcription factors that recruit histone-deacetylases and acetylases (for a review, see16)). This type of element might therefore exert epigenetic control over a promoter located near to exon 1.

Although luciferase expression assays showed weak activity for promoter B, the results are consistent with the weak expression observed for the short transcripts initiated by the alternative promoter in our 5′RACE-PCR. Moreover, a weak activity has already been observed for alternative promoters, and is often associated with tissue-specific expression17,18. For example, the APOE gene (the main risk factor described for AD) has many promoters and is reportedly down-regulated by a factor of 2.13 in the AD temporal lobe - a region in which three transcriptional isoforms are expressed19. One of these isoforms is generated by an alternative promoter upstream of the second exon, and two transcription start sites (TSSs) are identified. These isoforms are specifically and differentially regulated in the AD temporal lobe - constituting a switch in promoter usage in healthy vs. AD brains. The PSEN2 gene is also associated with an inherited predisposition to AD. The gene’s two promoters (P1 and P2) drive two alternative transcripts that differ with regard to their 5′ UTR20,21. Although most tissues express
both transcripts, some tissues only express one - suggesting that TSS usage may be regulated in a tissue-specific manner. Moreover, P1 and P2 are regulated by distinct transcription factors (Sp1 and Egr-1, respectively). Egr-1 can induce a partial shift in promoter usage by downregulating P1's activity two-fold in neuronal cells. One can imagine that a similar mechanism applies to APU for MAPT, i.e. the activation of a secondary promoter in a particular cell type under pathological conditions. Alternatively, an enhancer sequence upstream of the promoter B (i.e. a sequence missing from our construct) might potentiate MAPT's activity in vivo. Indeed, we recently characterized a regulatory region in intron 0 (upstream of a short CpG island and containing a CpG site that is hypomethylated in patients with PSP [13]), located 13kb upstream of exon 1. We suspect that this site influences MAPT expression, as has already been described for other elements (such as SNPs and long, non-coding RNAs) in intron 0 [19–22]. Our work described new shorter MAPT transcripts; short tau isoforms have already been attracting increasing attention in the field of neurodegenerative disorders. Indeed, Western blotting with antibodies against the C- or N-terminal end of the tau proteins revealed products that were initially considered to be degradation products or artifacts. Truncated tau proteins have been detected in patient brains and in murine disease models [23–25], and induced a neurofibrillary pathology in the brain of transgenic rats [26–27]. It has been suggested that truncated tau may seed the aggregation of the full-length protein [28]. Interestingly, it has been shown that a 35 kDa C-terminal tau fragment lacking the N-terminus but containing four microtubule binding-repeats is over-represented in PSP and in neurodegenerative disorders in which 4R tau isoforms predominate but not in AD [29]. These observations confirm that (i) smaller tau isoforms can be expressed in various neurodegenerative diseases, and (ii) specific aggregative mechanisms may be linked to the isoform's structural features. To date, it has been assumed that these small tau isoforms resulted from post-translational truncation at several sites within the full-length protein; however, the proteases responsible for this fragmentation have not been fully identified and characterized (for a review, see [30]). Our present results suggest that short isoforms are rare proteins resulting from an alternative transcription pattern. Indeed, we detected in-frame transcripts for which the first ATG was located in the 48th position on the cDNA (corresponding to Met11 in the protein) or in exon 5 (corresponding to Met127). These isoforms have already been described [31]. By using a proteomics approach to study human brain tissue, Derisbourg et al. identified N-terminally truncated tau isoforms starting from Met11 and Met127. Lastly, we identified several transcripts with an atypical exon start or splicing sites that are not in frame. We cannot assume that these unusual transcripts are necessarily expressed in brain tissue. Indeed, they might be targeted by the nonsense-mediated decay pathway. It should be noted that some in-frame transcripts lacked exon 4, which is not supposedly an alternative exon in the literature.

These observations raise the question of the functional consequences of APU and alternative transcription start sites for the MAPT gene. In fact, APU diversifies the transcriptional regulation of a gene by modulating alternative splicing [32–37], tissue specificity [2], regional specificity [3,4], subcellular specificity [5], and/or the expression level. As suggested in the review by Davuluri et al. [38], APU probably enables the mammalian genome to produce the required mRNA isoform in the right place at the right time. Our hypothesis is that under pathological conditions associated to epigenetic changes at the MAPT locus, a promoter switching occurs and could lead to a harmful process. Moreover, this APU could drive a specific splicing of the shorter transcripts. Indeed, it is now well established that splicing is physically and functionally coupled to the transcription and can be modified by the transcriptional rate (for a review, see [39]). The promoter B usage could promote a preferential exclusion or inclusion of exon 10 for instance, thereby altering the 3 R:4 R tau proteins ratio.

Our preliminary study is the first to show that the MAPT gene has two promoters: a more active constitutive promoter, plus an alternative promoter associated with elevated expression under pathological conditions. The function of the shorter transcripts initiated by this second promoter has yet to be established. However, we propose a new model in which the previously observed truncated tau isoforms are generated by APU and the use of alternative transcription start sites, rather than by post-translational modification. The present study's main finding is the identification of an alternative promoter that could become a specific therapeutic target. A detailed understanding of the alternative promoter's regulatory sites and factors might therefore enable specific blockade of the APU and (potentially) the neurodegenerative process.

**Patients, Materials and Methods**

**Human brain tissue samples.** Post-mortem brains were obtained from brain banks at university medical centers in Lille (France), Paris (France) and Geneva (Switzerland), following approval by the local institutional review board and the provision of written, informed consent by the donor's family. All methods were performed in accordance with the relevant approved guidelines and regulations. Frontal and occipital brain regions were dissected by trained neuropathologists and then frozen at −80 °C prior to sample preparation and genetic analysis. The present study included brain tissue samples from 35 patients (22 with AD and 13 with PSP) and 18 non-clinical controls. Most of these cases have been described previously [34].

**Extraction of DNA and RNA.** DNA and RNA were extracted from 50 mg samples from the frontal lobe (Brodmann area 10) and the occipital lobe (Brodmann area 18). Genomic DNA was obtained by phenol-chloroform extraction. Total RNA was isolated using the RNeasy® Lipid Tissue Mini Kit (Qiagen, Courtaboeuf, France), according to the manufacturer’s instructions. Assessment of the RNA quality was based on the RNA integrity number (RIN) given by the Agilent Bioanalyzer 2100 and a RNA 6000 Nano Kit (Agilent, Courtaboeuf, France). All samples had a RNA concentration ≥ 100 ng/μl and only samples with a RIN ≥ 7 were used for subsequent qRT-PCRs. Only frontal lobe samples with a RIN ≥ 7 (n = 34) were selected for the various 5′ rapid amplification of cDNA ends (RACE)-PCRs.
**MAPT alternative promoter constructs.** To assess the functionality of the alternative promoter upstream of exon 1 (arbitrarily referred to here as promoter B), we first cloned a 2 kb region encompassing this putative sequence (1879 bp upstream of exon 1 and the first 121 bp of exon 1; forward primer: GTCCCTTTCTGGAATTTCCACTACAC; reverse primer: TGGTCTTGGTGCACTGTTGAC). As a positive control, we cloned a 1.8 kb region encompassing the constitutive promoter upstream of exon 0 (referred to here as promoter A); 1509 bp upstream of exon 0 and the first 302 bp of exon 0; forward primer: AGGTCTTGAACTAGATGTTGGC; reverse primer: GATAGTGCAAGAGGCGAGGAC. Both sequences were obtained from a control individual by using long-range PCR with Takara LA Taq DNA Polymerase (Clontech, Mountain View, CA, USA). To define the minimal essential region for promoter B, we prepared four sequential 5′-truncated constructs (727 bp, 414 bp, 227 bp and 91 bp in length) from the full-length sequence. After purification with the Nucleospin® Gel and PCR Clean-up kit (Macherey Nagel, Düren, Germany), PCR products were cloned into pGL3-Basic vector (Promega, Madison, WI, USA) using the In-Fusion® HD cloning kit (Clontech). Constructs were purified using the Nucleobond® Plasmid No Lid kit (Macherey Nagel) and sequenced by GATC Biotech (Zurich, Switzerland).

**Cell culture, transfection and luciferase assays.** HeLa and SH-SY5Y cells were cultured as monolayers in six-well plates using Dulbecco's Modified Essential Medium (Invitrogen, California, USA) supplemented with 10% fetal bovine serum and 0.5% penicillin/streptomycin, at 37 °C in a humidified 5% CO2 incubator. Cells grown to a confluence of ~60% were transiently cotransfected with 2 μg of plasmid DNA (pGL3-promoter A, pGL3-promoter B, or pGL3-basic lacking an insert) and the pRL-TK Renilla Luciferase Control Reporter Vector as a control (test vector/co-reporter vector ratio: 10:1), using the TransporTer™ 5 Transfection Reagent (Polysciences, Warrington, PA, USA). Luciferase activity was analyzed 72 h post-transfection with the Dual-Luciferase® Reporter Assay System (Promega) on a Mithras LB 940 Multimode Microplate Reader device (Berthold Technologies, Bad Wildbad, Germany). Each experiment was performed in triplicate on at least three independent culture passages. The pGL3-promoters’ activities were normalized against the pGL3-basic vector lacking an insert.

**Quantitative RT-PCR.** One μg of total RNA was used to generate cDNA by using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Villeurbanne, France) with Multiscribe Reverse Transcriptase and random primers. MAPT expression levels were determined by qRT-PCR, using TaqMan® Gene Expression Assays, TaqMan® Universal PCR Master Mix II, and uracil-N glycosylase (Life Technologies, Carlsbad, CA, USA). To quantify total MAPT mRNA levels, we used the Hs00902194_m1 probe located at the junction between exons 12 and 13 (i.e. within a region present in all coding transcripts). To assess the cDNA expression of transcripts from promoter A, we designed a custom qPCR TaqMan® assay with a probe overlapping exons 0 and 1. UBC (Hs00824723_m1) was used as a control housekeeping gene. For each sample, all qRT-PCRs were performed in triplicate on an ABI PRISM® 7900 HT instrument (Applied Biosystems), according to the manufacturer's protocol. The comparative CT method (2−ΔΔCT) was used to calculate relative mRNA expression levels.

**5′ RACE and cloning.** To avoid the selection of patient-specific isoforms, we pooled the patients’ RNAs in equimolar concentrations. Only patients with a RIN≥7 were selected. Hence, the three pools corresponded respectively to 15 controls, 10 AD patients and 11 PSP patients. Rapid amplification of cDNA ends was carried out using a SMARTer™ RACE 5′ kit (Clontech), according to the manufacturer's instructions. For cDNA synthesis, 500 ng of total RNA was converted into RACE-ready first-strand cDNA. A 50 μl PCR reaction mixture was then prepared. It was composed of 2.5 μl 5′ RACE ready cDNA, 5 μl of 10X Universal Primer Mix, 1 μl of 10 μM gene-specific primer (GSP), and 41.5 μl of Master Mix (Clontech). The MAPT-specific primer (GSP11-12: GATTACGCCAACCTTTGCCCACCTCCTGTTATATGAGAT) designed for use in a 5′ RACE reaction was complementary to a specific mRNA sequence for the overlap of exons 11 and 12. Positive and negative controls were prepared according to the manufacturer's protocols. The 5′ RACE products were separated by electrophoresis on a 2% agarose gel (18 h at 50 mV). The smaller bands (corresponding to transcripts lacking exon 0) were cut out, and the DNA contained therein was extracted using the Nucloet Spin® Gel and PCR Clean-up kit (Macherey Nagel). Transcripts were then cloned into a pRACE cloning vector (Clontech) with the In-Fusion® HD Cloning kit (Clontech). Plasmid DNAs were purified using the Nucleobond® Plasmid No Lid kit (Macherey Nagel) and sequenced by GATC Biotech.

**Semi-quantitative assessment of short MAPT transcripts using 5′ RACE-PCRs.** In order to select short MAPT mRNA isoforms, we performed two 5′ RACE-PCRs using the protocol described above. The primers were GSP1 (GATTACGCCAACCTTTGCCCACCTCCTGTTATATGAGAT) and GSP4 (GATTACGCCAACCTTTGCCCACCTCCTGTTATATGAGAT) located in exons 1 and 4, respectively. The PCR products were separated on an 8% polyacrylamide gel. For each individual and each of the two 5′ RACE-PCRs, we used Fiji ImageJ software to measure the intensity of the bands for the different MAPT transcripts. For GSP1, the transcripts (starting at exon 1) were expected to be 194 bp or less in length. For GSP4, the expected lengths were 270, 357 and 444 bp, depending on whether or not the alternative exons 2 and 3 were included. The values corresponded to the ratio between the area under the curve (AUC) for transcripts starting at exon 1 and the AUC for total MAPT transcripts (expressed as a percentage).

**Statistical analysis.** The luciferase activities of the promoter B constructs were compared to the empty pGL3-basic vector (i.e. reference value of 1) using one sample Student’s t test. Comparisons of MAPT expression levels (qPCR and 5′ RACE semi-quantitative analysis) between controls and each disease groups were performed using analysis of variance (ANOVA) with post-hoc Dunnett’s test; normality of residuals were checked using
histogram and Shapiro-Wilk test. Statistical testing was performed at the 2-tailed $\alpha$ level of 0.05. Data were analyzed using the software package SAS, release 9.3 (SAS Institute, Cary, North Carolina, USA).

References

1. Landry, J.-R., Mager, D. L. & Wilhelm, B. T. Complex controls: the role of alternative promoters in mammalian genomes. Trends Genet. TIG 19, 640–648 (2003).
2. Golovine, K., Schwerin, M. & Vanselow, J. Three different promoters control expression of the aromatase cytochrome p450 gene (cyp19) in mouse gonads and brain. Biol. Reprod. 68, 978–984 (2003).
3. Pan, Y. X. et al. Generation of the mu opioid receptor (MOR-1) protein by three new splice variants of the Oprm gene. Proc. Natl. Acad. Sci. USA 98, 14084–14089 (2001).
4. Pardo, L. M. et al. Regional differences in gene expression and promoter usage in aged human brains. Neurobiol. Aging 34, 1825–1836 (2013).
5. Chen, T., Ueda, Y., Xie, S. & Li, E. A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. J. Biol. Chem. 277, 38746–38754 (2002).
6. Saleh, A., Makrigiannis, A. P., Hodge, D. L. & Anderson, S. K. Identification of a novel lifelong promoter that is active in bone marrow and fetal thymus. J. Immunol. Baltim. Md 1950 168, 5163–5169 (2002).
7. McMahon, A. C. et al. SynGAP isoforms exert opposing effects on synaptic strength. Nat. Commun. 3, 900 (2012).
8. Twine, N. A., Janitz, K., Wilkins, M. R. & Janitz, M. Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. PLoS One 6, e16266 (2011).
9. Renbaum, P. et al. Ege-1 upregulates the Alzheimer’s disease presenilin-2 gene in neuronal cells. Gene 318, 113–124 (2003).
10. Ounallah-Saad, H. et al. Transcriptional regulation of the murine Presenilin-2 gene reveals similarities and differences to its human orthologue. Gene 446, 81–89 (2009).
11. Andreadis, A. Tau gene alternative splicing: expression patterns, regulation and modulation of function in normal brain and neurodegenerative diseases. Biochim. Biophys. Acta 1739, 91–103 (2005).
12. Sergeant, N., Delacourte, A. & Buée, L. Tau protein as a differential biomarker of tauopathies. Biochim. Biophys. Acta 1739, 179–197 (2005).
13. Huin, V. et al. The MAPT gene is differentially methylated in the progressive supranuclear palsy brain. Mov. Disord. Off. J. Mov. Disord. Soc. 31, 1883–1890 (2016).
14. Cailliet-Boudin, M.-L., Buée, L., Sergeant, N. & Lefebvre, B. Regulation of human MAPT gene expression. Mol. Neurodegener. 10, 28 (2015).
15. Andreadis, A., Wagner, B. K., Broderick, J. A. & Kosik, K. S. A tau promoter region without neuronal specificity. J. Neurochem. 66, 2257–2263 (1996).
16. Deng, Z., Cao, P., Wan, M. M. & Sui, G. Yin Yang 1: a multifaceted protein beyond a transcription factor. Transcription 1, 81–84 (2010).
17. Borowsky, R. & Hoffman, B. J. Analysis of a gene encoding two glycine transporter variants reveals alternative promoter usage and a novel gene structure. J. Biol. Chem. 272, 29077–29085 (1998).
18. Udono, T. et al. Structural organization of the human microphthalmia-associated transcription factor gene containing four alternative promoters. Biochem. Biophys. Acta 1491, 205–219 (2000).
19. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature 533, 539–542 (2016).
20. Coupland, K. G. et al. DNA methylation of the MAPT gene in Parkinson’s disease cohorts and modulation by vitamin E. J. Neurochem. Off. J. Neurochem. Soc. 29, 1660–1664 (2014).
21. Coupland, K. G. et al. Role of the Long Non-Coding RNA MAPT-AS1 in Regulation of Microtubule Associated Protein Tau (MAPT) Expression in Parkinson’s Disease. PLoS One 11, e0157924 (2016).
22. Li, Y. et al. An epigenetic signature in peripheral blood associated with the haplotype on 17q21.31, a risk factor for neurodegenerative tauopathy. PLoS Genet. 10, e1004211 (2014).
23. Rissman, R. A. et al. Caspase-cleavage of tau is an early event in Alzheimer disease tangle pathology. J. Clin. Invest. 114, 121–130 (2004).
24. de Calignon, A. et al. Caspase activation precedes and leads to tangles. Nature 464, 1201–1204 (2010).
25. Matsumoto, S.-E. et al. The twenty-four kDa C-terminal tau fragment increases with aging in tauopathy mice: implications of prion-like properties. Hum. Mol. Genet. 24, 6403–6416 (2015).
26. Zilk, N. et al. Truncated tau from sporadic Alzheimer’s disease suffices to drive neurofibrillary degeneration in vivo. FEBS Lett. 580, 3582–3588 (2006).
27. Filippic, P. et al. First transgenic rat model developing progressive cortical neurofibrillary tangles. Neurobiol. Aging 33, 1448–1456 (2012).
28. Wang, Y. P., Bierrat, J., Pickhardt, M., Mandelkow, E. & Mandelkow, E.-M. Stepwise proteolysis liberates tau fragments that nucleate the Alzheimer-like aggregation of full-length tau in a neuronal cell model. Proc. Natl. Acad. Sci. USA 104, 10252–10257 (2007).
29. Cray, S., Saxton, M., Anderton, B. H. & Hanger, D. P. Direct analysis of tau from PSP brain identifies new phosphorylation sites and a major fragment of N-terminally cleaved tau containing four microtubule-binding repeats. J. Neurochem. 105, 2343–2352 (2008).
30. Hanger, D. P. & Wray, S. Tau cleavage and tau aggregation in neurodegenerative disease. Biochem. Soc. Trans. 38, 1016–1020 (2010).
31. Berdyshev, M. et al. Role of the Tau N-terminal region in microtubule stabilization revealed by new endogenous truncated forms. Sci. Rep. 5, 9659 (2015).
32. Pecci, A., Viegas, L. R., Baranao, J. L. & Beato, M. Promoter choice influences alternative splicing and determines the balance of isoforms expressed from the mouse bcl-x gene. J. Biol. Chem. 276, 21062–21069 (2001).
33. Loette, E. et al. The human caspase-2 gene: alternative promoters, pre-mRNA splicing and AUG usage direct isoform-specific expression. Oncogene 22, 935–946 (2003).
34. Nogues, G., Kadener, S., Cramer, P., Bentley, D. & Kornblith, A. R. Transcriptional activators differ in their abilities to control alternative splicing. J. Biol. Chem. 277, 43110–43114 (2002).
35. Auboeuf, D. et al. Differential recruitment of nuclear receptor coactivators may determine alternative RNA splice site choice in target genes. Proc. Natl. Acad. Sci. USA 101, 2270–2274 (2004).
36. Kadener, S., Fededa, J. P., Rosbash, M. & Kornblith, A. R. Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. Proc. Natl. Acad. Sci. USA 99, 8185–8190 (2002).
37. Cramer, P. et al. Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. Mol. Cell 4, 251–258 (1999).
38. Davuluri, R. V., Suzuki, Y., Sugano, S., Plass, C. & Huang, T. H.-M. The functional consequences of alternative promoter use in mammalian genomes. Trends Genet. TIG 24, 167–177 (2008).
39. Montes, M., Becerra, S., Sánchez-Álvarez, M. & Suñé, C. Functional coupling of transcription and splicing. Gene 501, 104–117 (2012).
40. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682 (2012).
Acknowledgements

We thank the patients and their families for participating in the present study. We thank Claude-Alain Maurage and Vincent Deramecourt (Lille Neurobank), Charles Duyckaerts (NeuroCEB) and Eniko Kovari (HUG Brain bank) for providing and analyzing the brain tissue samples. This study was funded by the French government's LabEx program ("Development of Innovative Strategies for a Transdisciplinary approach to Alzheimer’s disease" - DISTALZ) and the Paralysie Supranucléaire Progressive France charity.

Author Contributions

V.H. was involved in the research project (conception, design of the work, acquisition, analysis and interpretation of data for the work), statistical analysis and manuscript preparation (drafting the work and revision of the manuscript). L.B. was involved in the study follow-up and manuscript preparation (revising the draft for critical intellectual content). H.B. was involved in the statistical analysis and manuscript preparation (revision of the manuscript). J.L. was involved in the statistical analysis and manuscript preparation (revision of the manuscript). B.S. was involved in the research project (interpretation of data), and manuscript preparation (revising the draft for critical intellectual content). C.M.D. was involved was involved in the research project (conception, interpretation of data), and manuscript preparation (drafting the manuscript, and revising it for critical intellectual content). All authors revised and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-12955-7.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017