Published in final edited form as:
Gene. 2012 June 10; 501(1): 63–78. doi:10.1016/j.gene.2012.01.049.

Structural and functional characterization of H2 haplotype *MAPT* promoter: Unique neurospecific domains and a hypoxia–inducible element would enhance rationally targeted tauopathy research for Alzheimer's disease

Bryan Maloney[a] and **Debomoy K. Lahiri**[a,b]

[a]Laboratory of Molecular Neurogenetics, Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN 46202, USA

[b]Laboratory of Molecular Neurogenetics, Department of Medical and Molecular Genetics, Institute of Psychiatric Research, Indiana University School of Medicine, Indianapolis, IN 46202, USA

Abstract

Alzheimer's disease (AD) is the leading cause of dementia in the elderly. Extraneuronal plaque consisting primarily of amyloid β peptide and intraneuronal tangles of hyperphosphorylated microtubule–associated τ protein (τ, gene *MAPT*) are typical of AD. Misfolded τ is also implicated in Parkinson's disease and frontotemporal dementia. We aim to understand the regulation of the human *MAPT* promoter by mapping its functional domains. We subcloned a 4868 base pair fragment from human BAC RPCI–11 100C5. Sequence analysis revealed an H2 haplotype *MAPT* promoter, 5′–UTR, and intronal fragment. Database analysis of the fragment showed 50%–75% homology with mouse and >90% with rhesus monkey. Comparison with human H1 sequences revealed differences that crossed predicted transcription factor sites. DNA–protein interaction studies by electrophoretic mobility shift assay suggested an active specificity protein 1 (SP1) site in the 5′–untranslated region and hypoxia response. Transfection of a series of *MAPT* promoter deletions revealed unique functional domains. The distal–most had different response in neuronal vs. non–neuronal cells. We have cloned, sequenced, and functionally characterized a 4868 bp fragment of the human *MAPT* 5′–flanking region, including the core promoter region (−302/+4), neurospecific domains (−4364/−1992 and +293/+504, relative to +1 TSS), and a hypoxia–inducible element (+60/+84). Our work extends functional analysis of the *MAPT* sequence further upstream, and explores cell–type specificity of *MAPT* promoter activity. Finally, we provided direct comparison of likely transcription factor binding sites, which are useful to understand differences between H1/H2 pathogenic associations.

Address correspondence to: Debomoy K. Lahiri, PhD, 791 Union Drive, Indianapolis, IN 46202, USA, 317–274–2706., FAX: 317–274–1365, dlahiri@iupui.edu.

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

Authors' Contributions
BM substantially contributed to cloning and subcloning of JF412354. BM carried out DNA sequencing, EMSA, phylogeny, and data/statistical analysis. DKL and BM shared writing of the manuscript and planning all experiments. DKM had final approval authority.

Keywords

cell type specificity; gene regulation; H2 haplotype; neurofibrillary tangles; promoter; tau

1. Introduction

Alzheimer's disease (AD) is the leading cause of dementia in the elderly (Thies and Bleiler, 2011). AD is characterized by neuropathological traits that include senile plaques consisting primarily of amyloid–β peptide (Aβ) and neurofibrillary tangles consisting primarily of hyperphosphorylated microtubule–associated protein τ. It has been predicted that over 113 million people will suffer from dementia by 2050 (Brodaty et al., 2011). We report the structural and functional characteristics of H2 haplotype τ protein gene (*MAPT*) promoter, and the identification of a distal neurospecific domain and a proximal hypoxia–inducible element within the 5′–flanking region of the *MAPT* gene. Intracellular tangles of microtubule–associated protein (MAP) τ also play an important role in the disorder (De Strooper, 2010). Tauopathies (τ–related disorders) are not limited to AD, but also include frontotemporal dementia (FTD) (Adamec et al., 2001) and over 20 other disorders (Delacourte, 2008), particularly including Parkinson's disease (PD) (Williams-Gray et al., 2009). In AD, τ is hyperphosphorylated, and this hyperphosphorylated τ is the primary component of the τ tangle (Iqbal et al., 2010). Generally, hyperphosphorylated τ dysfunction is relegated to "secondary" status behind Aβ peptide in AD progression, but it has been found that elevation of cerebrospinal τ levels and τ:Aβ ratio correlated with more rapid cognitive decline in AD with mild dementia (Snider et al., 2009). Total neurofibrillary tangles in the CA1 region of the hippocampus correlated better with cognitive decline than did Aβ plaque deposition (Giannakopoulos et al., 2009). Furthermore, loss of medial septum cholinergic neurons in a mouse AD model has been linked to τ pathology, and such loss may mean a critical role for τ pathology in loss of cholinergic neurons typical of the disease (Belarbi et al., 2011). It has been generally presumed that aberrant aggregation is the primary etiological contribution of τ to the tauopathies. However, the possibility may exist that the underlying pathologies could begin due to changes in τ function, with aggregation following (Delacourte, 2008). Significant progress has been made in *pre–mortem* detection and modulation of τ aggregates (Kim et al., 2010), and induction of τ disaggregation may prove a promising therapy (Duff et al., 2010).

While it is a modified version of τ protein, i. e., hyperphosphorylated τ, that is usually associated with neurological disorders, studies have shown that variations in the τ gene (*MAPT*) and its expression levels may be associated with neuropathology. The C↔G polymorphism at +347, within the intron between the 5′–UTR and exon 1, was associated with upregulation of *MAPT* expression and greater AD risk (Sun and Jia, 2009). *MAPT* has two known haplotypes, specifically H1 and H2 (Kwok et al., 2004). The haplotype polymorphism has been variously claimed to only exist within *MAPT* (Kwok et al., 2004) or to extend as much as 1.3 megabases (Oliveira et al., 2004). It has also been found to be part of a 900 kb chromosomal H1 vs. H2 inversion (Stefansson et al., 2005), which also covers the gene for corticotrophin releasing hormone receptor 1 (*CRHR1*). H1 consists of several subhaplotypes, all of which are associated with some level of greater risk for PD (Kwok et al., 2004; Seto-Salvia et al., 2011). H1 drives higher levels of *MAPT* gene transcription in cell culture (Kwok et al., 2004). *MAPT* haplotype influence on PD risk is cumulative with genetic and environmental factors such as *APOE* genotype, tobacco smoking, and coffee intake (McCulloch et al., 2008). While specific association between overall onset risk for AD and H1 has been rejected, (Seto-Salvia et al., 2011), progression from mild cognitive impairment (MCI) to dementia is faster in patients with both *MAPT* H1 and the *APOE* ε4 allele (Samaranch et al., 2010). On the other hand, H2 is associated with 17q21.31

microdeletion syndrome, a more common form of mental retardation (Rao et al., 2010). Likewise, while H2 does not appear to alter absolute risk for FTD, it does reduce age at onset and shows a more severe decline of frontal lobe glucose utilization (Laws et al., 2007) and overall worse FTD prognosis once disease is diagnosed (Borroni et al., 2011). H2 has been claimed to be under positive selective pressure in the Icelandic population (Stefansson et al., 2005).

Early studies of AD molecular biology focused on protein sequence variations, such as found in the now classic APP_{SWE} (Mullan et al., 1992) and $PSEN_{delE9}$ (Thinakaran et al., 1996) familial AD (FAD) mutations. Unfortunately, FAD only represents a small minority of AD cases (Thies and Bleiler, 2011), and age of onset, severity, and rate of progression have all been linked to levels of apolipoprotein E (*APOE*), *MAPT*, and other AD–associated genes (Yamagata et al., 2001; Sun and Jia, 2009). Specific promoter polymorphisms in the *APOE*, *APP*, β–amyloid site cleaving enzyme 1 (*BACE1*), and *MAPT* genes have been linked to AD risk (Lahiri et al., 2005b; Lv et al., 2008; Zuo and Jia, 2009; Maloney et al., 2010; Wang and Jia, 2010). Therefore, both molecular biology and epidemiology point to the vital importance of investigating promoter regulatory effects in AD etiology.

In our present work, we have cloned and sequenced a 4868bp fragment of a human chromosome 17 BAC, containing 4364 bp of the 5′–flanking (promoter) sequence of the *MAPT* gene, the sequence of the 5′–UTR exon, conventionally referred to as “exon 0” (Sundar et al., 2007), and a portion of the intron between exons 0 and 1 (GenBank ID: JF412354). Our structural analysis determined that it was an H2 lineage clone. We compared JF412354 with other human *MAPT* 5′–flanking sequences and located both proximal and distal H1 vs. H2 polymorphisms, several of which corresponded to predicted transcription factor (TF) binding sites. A previous direct comparison of H1 and H2 covered 1039bp, of which 808 were considered to be the “promoter” sequence (Kwok et al., 2004), with a single full–length expression clone constructed. No sequence from that analysis has been deposited in available databases.

Our functional study of JF412354 included DNA protein interaction studies using the electrophoretic mobility shift assay (EMSA). Double–stranded oligomer probes derived from short sequences within the 5′–UTR were radiolabeled to probe mammalian cell nuclear extracts. One of these probes, corresponding to +60/+84 from the +1 transcription start site (TSS) had specific DNA–protein interaction activity in human neuroblastoma (NB) nuclear extracts. Competition EMSA independently against unlabeled probes for SP1, estrogen receptor (ER), progesterone receptor (PR), and retinoic acid receptor (RAR) eliminated EMSA signal for this probe. Of these TF sites, a GC box/SP site was predicted to exist within +60/+84.

Functional expression studies were carried out by DNA transfection experiments in different cell types. For this, a library of 8 forward and 4 reverse promoter deletion fusion clones was constructed in a promoterless chloramphenicol acetyltransferase (CAT) expression vector. Different recombinant *MAPT* promoter–reporter clones were transiently transfected into human and rat cells. All cell transfection results indicated several active sites of significant basal promoter activity. For assays with neuronal–type cultures, the core promoter lies within −302/+4, and all extensions of the JF412354 sequence upstream of this increased promoter activity. In addition, the inclusion of “exon 0”, which is exclusively in the 5′–UTR, significantly increased apparent promoter activity in NB and PC12 cell cultures. We used this data to determine functional domains within the JF412354 sequence, looking at both k ratio (Bayesian) and pairwise *g* for each comparison. In short, when k ratio was 100 or greater and *g* was ≥ 1.0 $\sqrt{MSE}$ (square root of mean square error for the ANOVA)

between adjusted signals for compared clones, the clones were deemed to belong to different "domains".

Finally, we aligned JF412354 vs. H1 and H2 human *MAPT* sequences in GenBank, and against corresponding sequences for chimpanzee, gorilla, orangutan, marmoset, mouse, and rat to estimate a phylogenic tree. Our proposed phylogenic tree agreed with several already published estimations that had set the divergence of human *MAPT* from chimpanzee before any split between H1 and H2 haplotypes within humans.

2. Methods

2.1. Reagents

Unless otherwise specified, reagents were purchased from Sigma (St. Louis, MO) and were of "molecular biology" or "analytic" quality. Restriction enzymes were purchased from New England Biological (Ipswich, MA). Other enzymes were purchased from Roche (Indianapolis, IN). Cell culture reagents were purchased from Invitrogen (Carlsbad, CA).

Statistical analysis

Data was analyzed with the R environment for statistical computing (R Development Core Team, 2011), using the "nortest" (Gross, 2006) and "agricolæ" (de Mendiburu, 2010) supplemental packages. In addition to statistical significance, omnibus ω^2 (Olejnik and Algina, 2003) effect sizes were calculated by $\frac{SST-(dfT\times MSE)}{SST+SSE+MSE}$ and unbiased Ψ (Harlow et al., 1997) were calculated by $\sqrt{\frac{1}{k-1}\times\frac{\sum\left(\bar{x}_j-\bar{X}\right)^2}{MSE}}\times\frac{\Gamma(a/2)}{\sqrt{a/2\Gamma\left(\frac{a-1}{2}\right)}}$. Within each experiment, modified *g* (Hedges and Olkin, 1985) for pairwise effect sizes were calculated for clone comparisons by the formula $\left|\frac{\bar{x}_1-\bar{x}_2}{\sqrt{MSE}}\times\frac{\Gamma(a/2)}{\sqrt{a/2\Gamma\left(\frac{a-1}{2}\right)}}\right|$. This is modified from a two–sample *g* by using $\sqrt{MSE}$ in place of pooled standard deviation, to reflect the overall ANOVA context of a given pairwise comparison. In these equations, *SST* is the sums of squares for treatments; *dfT* is the degrees of freedom for treatments; *MSE* is the mean square error (aka MSwithin or MSresiduals), *SSE* is the sums of squares for the error, *k* is the number of treatments; $\bar{x}_j$ is the mean value for a specific clone's adjusted CAT level within an ANOVA; $\bar{X}$ is the overall "grand mean" adjusted CAT level for an ANOVA; $\bar{x}_1$ is the mean adjusted CAT level of one clone in a given pairwise comparison; $\bar{x}_2$ is the mean adjusted CAT level of the other clone; Γ is the gamma function; and *a* is either $(n_j + k - 2)$ for ANOVA or $(n_1 + n_2 - 2)$ for pairwise comparisons, with n_j as the number of replicates within each clone for the ANOVA or n_1 and n_2 as the number of replicates for each clone, respectively in a pairwise comparison. Results for Ψ and *g* were rounded to one decimal place. The units of Ψ and *g* are $\sqrt{MSE}$ and these effect size estimates are analogous to Cohen's *d*, while ω^2 estimates "variance explained" and is analogous to r^2. The Γ function factor adjusts for sample size induced bias in effect size estimation (Nakagawa and Cuthill, 2007); ω^2 is considered to be a "less–biased" estimator of variance explained compared to similar effect size measures (Olejnik and Algina, 2003).

2.2 Cloning of ~5kb MAPT 5′–flanking sequence from human BAC RPCI–11 100C5

The BAC RPCI–11 100C5 was acquired from Roswell Park Cancer Institute (RPCI) (Buffalo, NY, USA). Two pairs of PCR primers (τD–F: 5′–GATGACTTCTCTCTCATGCA–3′, τD–R: 5′–TAGGACGGGATTTCACCA–'3, τP–F: 5′–TGAGCAAGAGAGTGACA–3′, τP–R: 5′–GTGCATCCCTTGCTCA–3′) were

synthesized (Invitrogen) based on the human genome reference chromosome 17 sequence (Sayers et al., 2010). Each pair was designed to cover approximately 2.5kb DNA sequence, with the distal and proximal fragments partially overlapping (Fig. 1A). PCR was carried out similarly to methods described elsewhere (Maloney et al., 2009). Briefly, the "FailSafe" PCR buffer kit (Epicentre) was used with an initial T_{anneal} of 58.6 °C, reduced by 0.5 °C per cycle for 15 cycles, followed by 20 cycles at T_{anneal} 51.6 °C. Reactions were then run on 1% agarose–TAE gel. For the distal fragment reaction, FailSafe buffer "H" produced a strong product at 2.5 kb, while FailSafe buffer "D" had a 2.5 kb product for the proximal fragment reaction. Each respective band was gel purified and cloned into the pCR2.1–Topo (Invitrogen) vector (Fig 1B, D). Clones were screened by *Eco*RI enzyme digest, and candidates with inserts were subjected to end sequencing with M13–F and M13–R primers (Macrogen USA). Four clones (two distal, pτTA–DF & pτTA–DR, and two proximal, pτTA–PF & pτTA–PR, in each orientation) were then chosen for further subcloning and sequencing. Sequencing was carried out using primers based on the human genome reference chromosome 17 sequence (Table 1).

2.3. Analysis of MAPT BAC RPCI–11 100C5 sequence

DNA sequence contigs were assembled with CAP3 (Huang, 1996). The resulting 4868bp fragment was submitted to GenBank and given the accession number JF412354. JF412354 was then used to BLAST or BLAT probe the genomic databases for human, chimpanzee, gorilla, orangutan, rhesus monkey, marmoset, mouse, and rat. In addition, the "nr" database of GenBank was probed with JF412354. Homologous sequences are summarized in Table 2. JF412354 was analyzed for G+C content, CpG dinucleotides, and GG dinucleotides in a 100bp window. JF412354 was also probed with RepeatMasker (Chen, 2004) and Censor (Jurka et al., 1996) to locate common repeat sequences, such as Alu. In addition, JF412354 was aligned separately to the human chromosome 17 genetic contig (GenBank ID: NT_010738), the rhesus monkey chromosome 16 genetic scaffold (GenBank ID: NW_001102974), and the mouse chromosome 11 genetic contig (GenBank ID: NT_165773). Homologies in a 100bp window were calculated and graphed. Finally, JF412354 human *MAPT* sequences were used to probe the TransFac database via MatInspector (Quandt et al., 1995) and H1 vs. H2 predicted transcription factor sites were compared.

2.3.1 Preparation of nuclear extracts for electrophoretic mobility shift assay (EMSA)—Nuclear extracts from different cell lines, specifically human cervical epithelial (HeLa), SK–N–BE human neuroblastoma (NB), rat adrenal pheochromocytoma (PC12), and human glioblastoma–astrocytoma (U373), as well as nuclear extracts from mouse brain, were obtained from commercial sources (ActiveMotif, Carlsbad, CA) or prepared as described previously (Lahiri and Ge, 2000).

2.3.2 Electrophoretic mobility shift assay (EMSA) of 5′–UTR fragments of JF412354—Double–stranded oligomer pairs with sequences from JF412354 (Table 3) were synthesized (Invitrogen) and labeled with γ–^{32}P–ATP (Amersham) with polynucleotide kinase (Roche, Indianapolis). Reactions were purified by passing through the G–25 minicolumn and activity was determined by β particle count. An activity of 10,000cpm/μl (approx. 60pg/μl) was used to react with human SK–N–BE (NB) neuroblastoma cell nuclear extract and mouse brain nuclear extract. Reactions were carried out at 8 °C in EMSA buffer (Tris HCl pH7.5, 10mM; NaCl, 50mM; Triton X–100, 0.05%; DTT, 0.5mM; glycerol, 5%; poly dIdC, 100μg/ml). For some reactions, self–competition against unlabeled pair was carried out. In that case, reactions were pre–incubated for 30 minutes with 200× molar excess unlabeled homologous oligomer pair. All reactions were incubated for 30 minutes with labeled probe. Reactions were then analyzed in 5%

polyacrylamide–TGE (50 mM Tris at pH 8.5, 0.384 mM glycine, and 2 mM EDTA) gel. The gel was dried and exposed to X–ray film for autoradiography at −70 °C. Free unbound oligonucleotides ran at the bottom of the gel, and different protein–DNA complexes were detected as mobility–retarded bands as described previously (Ge et al., 2004a).

2.3.3 Competition EMSA of specific transcription factor–binding double–stranded oligomers against the UTR1B probe—Double–stranded oligomer pairs for the binding sites of transcription factors AP1, AP2, SP1, GATA, estrogen receptor (ER), progesterone receptor (PR), and retinoic acid receptor (RAR) were used at 200× molar excess to compete against radioactive UTR1B oligomer pair probe. Reactions were separated electrophoretically on 5% polyacrylamide–TGE gel, dried, and exposed to X–ray film as per 4.4.2.

2.4 Subcloning of MAPT 5′–flanking sequence into pBLCAT3 chloramphenicol acetyltransferase (CAT) expression backbone and creation of promoter deletion series library

Clone pτTA–PF was digested with *Xba*I and *BamH*I. A 2.5kb fragment was purified and cloned into the *Xba*I/*BamH*I sites of expression vector pBLCAT3 to produce pτ2 (Fig. 1C). The distal *MAPT* promoter fragment from pτTA–D was cut out from its backbone with *Xba*I and *BstE*II and cloned into the *Xba*I and *BstE*II sites of pτ2 to produce pτ1 (Fig. 1E). The promoter–reporter fusion clone pτ2 was digested with *Hin*dIII. A 738bp fragment was gel purified and cloned into the *Hin*dIII site of pBLCAT3 to produce both pτ3 and pτ3R. Orientation was checked by end sequencing. The same *Hin*dIII digestion of pτ2 was self–ligated to produce pτ4. Clone pτ2 was also digested with *Psp*XI and the backbone self–ligated to produce pτ5. Clone pτ5 was digested with *Sal*I and the 595bp fragment was cloned into the *Sal*I site of pBLCAT3 to produce both pτ6 and pτ6R. The *Sal*I digested backbone of pτ5 was self–ligated to produce pτ8. Finally, pτ6 was digested with *Eag*I and self ligated to produce pτ7.

2.5 Expression of transient transfections by a reporter chloramphenicol acetyltransferase (CAT)–JF412354 deletion series library

Rat pheochromocytoma (PC12) and glioblastoma (C6) and human SK–N–SH neuroblastoma (NB) cell cultures were acquired from ATCC and routinely cultured in our laboratory according to ATCC instructions (Ghosh et al., 2000). Tissue culture reagents were obtained from Invitrogen. PC12 cultures were grown in duplicate, one set treated with 30 ng/ml nerve growth factor (NGF) for 7–10 days. The NB and PC12 cell lines were chosen as commonly–used representative neuronal cell cultures. PC12 was further treated with NGF as PC12 + NGF cultures resemble differentiated neurons (Kim et al., 1994). C6 cells were chosen to represent non–neuronal cells that would otherwise be found in close proximity to neuronal cells in the brain. C6, NB, PC12, and PC12 + NGF cell cultures were transfected with empty pBLCAT3 vector or one of either 12 JF412354–*CAT* reporter constructs (C6, NB, PC12) or 7 such constructs (PC12 + NGF) by Lipofectamine and Plus reagent (Invitrogen), as described previously (Ghosh et al., 2000). DNA transfection was carried out in 2×10^6 cells per 60mm plate in triplicate, with 2.7 μg of promoter–reporter clone or pBLCAT3 DNA. To monitor transfection efficiency, cells were cotransfected with 0.3μg pSVβGAL (Promega) under the same conditions. Cells were harvested, extracts prepared, protein concentration determined by modified Bradford assay (BioRad), CAT reporter protein levels were measured by a sensitive enzyme–labeled immunosorbent assay (ELISA), and β–galactosidase (β–gal) activity levels were measured colorimetrically. The reporter CAT protein levels were adjusted by β–gal and total protein and adjusted data further analyzed. This information was used to characterize 6 distinct regulatory domains (domains I–V upstream of the +1 TSS, domain 0 downstream of the +1 TSS) within the

JF412354 *MAPT* promoter sequence (Table 3, Fig. 10). All calculations and estimations were performed on transformed data. Data was back–transformed for presentation.

2.6 Multiple alignment and phylogeny of JF412354 to selected mammalian and human MAPT sequences

JF412354 was subject to multiple alignment with all non–human sequences extracted from GenBank and ENSEMBL and with non–redundant human sequences from GenBank (Table 2) via PRANK (Fletcher and Yang, 2010). Gaps were then encoded with FastGap (Borchsenius, 2009), which implements the algorithm of Simmons & Ochoterena (Simmons and Ochoterena, 2000). Encoded gaps were appended to the end of the sequence alignment. Alignment was then used as input to MrBayes (Ronquist and Huelsenbeck, 2003) for maximum likelihood and to PARS (Fink, 1986) for parsimony phylogeny estimation. For branch lengths, the PARS consensus tree was used as input to FITCH (Fink, 1986), with a distance matrix generated by Tree–Puzzle (Schmidt et al., 2002), using the HKY nuclear substitution model.

3. Results

3.1.1 Subcloning and Sequencing of a 4868 base–pair (bp) fragment of the MAPT 5′–flanking region

BAC RPCI–11 100C5 was subject to parallel PCR as described herein. Two 2.5 kb DNA fragments were isolated and cloned into pCR2.1–Topo PCR cloning vector (Fig. 1). These clones were used as templates for simultaneous parallel DNA sequencing with primers (Table 1) based on the human genome *MAPT* sequence. DNA sequencing revealed that the RPCI–11 100C5 clone was 4868 bp between the τD–F and τP–R primers (Fig. 2A). This sequence was deposited with GenBank and given the accession number JF412354, which will be used for the remainder of this paper to indicate the 4868 bp RPCI–11 100C5–derived sequence.

3.2 Structure of the JF412354 MAPT gene 5′–flanking sequence

Fragment JF412354 included the +1 TSS, exon zero (5′–UTR), and a portion of the first intron, in addition to > 4kb of upstream promoter sequence. The sequence was screened for known repetitive elements with RepeatMasker (Chen, 2004) and Censor (Jurka et al., 1996). Fifteen separate repeats were found (Fig. 2B, Table 4). These included three Alu and two L2b elements. Structurally, JF412354 is typical of a “housekeeping” promoter (Fig. 3), with moderate (30%–50%) G + C content upstream of approximately −250 from the +1 TSS. Total G + C proportion increased to about 75% for the remainder of the sequence. As would be expected, CpG density was fairly low for most of the sequence. From −4364/−357, CpG density was 2.5% or less in a 100bp window. CpG density then rose to as much as 10%, peaking around the TSS. Density of GG dimers has three peaks that exceed 10% frequency at −2112/−2051, −363/−292, and +434/>+454. GGG trimers rise above 5% frequency at −2114/−2031, −364/−314, and +433/>+454. Dimer and trimer frequencies were calculated in 100bp windows. The C in CpG dimers is the most common site of DNA methylation in mammals (Grayson et al., 2005), while GGG and GG sequences are the most and second–most likely sites of DNA oxidation (Lee et al., 2008). DNA methylation and oxidation are both fundamental to gene–environment interaction models of idiopathic disease, such as the “Latent Early–life Associated Regulation (LEARn) model for AD (Bales et al., 2000; Lahiri and Maloney, 2010).

3.2.1. Structure of the JF412354 MAPT gene 5′–flanking sequence: homology of JF412354 with selected mammalian 5′–flanking MAPT sequences—

Alignment with available human upstream *MAPT* sequences in GenBank (Table 2) revealed

that JF412354 is in the H2 haplotype, a minority inversion polymorphism that covers MB on chromosome 17. Homology comparison with BX544879 vs. the H1 sequences (AC010792, NT_010783, AC091628, NG_007398) established several specific dimorphisms between H1 and H2 haplotype sequence groups (Fig. 4). Nevertheless, homology between JF412354 and the reference human genome sequence (coincidentally H1 haplotype) was still high, above 95% throughout (Fig. 5A). Homology with rhesus monkey sequence was around 90% from approximately −3750 to +500, but it fell off to 50% upstream of −3750 (Fig. 5B). Homology with mouse sequence was not strong, generally between 50% and 75%, with several gaps.

3.2.2 Structure of the JF412354 MAPT gene 5′–flanking sequence: predicted transcription factor (TF) sites in JF412354 that cross H1/H2 dimorphic regions

—JF412354 was used to probe the TransFac database with Genomatix/MatExplorer. A large number of putative TF sites were found. These included but were not limited to activator protein (AP)1, AP2, and related factors, CCAAT/enhancer binding protein (C/EBP), cAMP responsive element binding proteins (CREB), Heat shock factors (HSF), Hypoxia–inducible factors (HIF, HRE), Interferon regulatory factors, nuclear factor κ–B (NF–κ–B), and GC box/SP1 family factors. A recently discovered Aβ interacting domain (AβID) decamer (Bailey et al., 2011; Maloney and Lahiri, 2011) was predicted to occur four times in JF412354, although none of the predicted AβID sites crossed an H1/H2 dimorphism. The four H1 sequences were also used to probe TransFac, and a majority presence of a given site was taken to mean that the site was “typical” of the H1 haplotype. Several putative TF binding sites did differ between H1/H2 dimorphisms (Table 5). Of particular note is that some of these sites differed at H1/JF412354 dimorphisms for all instances of a specific TF, specifically the AP1–related factor neural retina leucine zipper (NRL), hypoxia–inducible factor (HIF)1, and retinoic acid receptor–related orphan receptor α (RORα). In addition, individual sites were present exclusively in H1 or JF412354, including cAMP–responsive element binding proteins nuclear factor, interleukin 3 regulated (NFIL3) and x–box binding protein (XBP)1; CCAAT/enhancer binding protein (C/EBP)α and C/EBPβ; E–box binding factors activating transcription factor (ATF)6 and upstream stimulatory factor (USF); and the NF–κ–B site binding protein zinc finger protein (ZFP)40. Other instances of these TF sites also appeared in both JF412354 and H1. In addition, potential binding sites for TFs of interest to AD and tauopathies in general were found in both H1 and JF412354 (Table 6). These included AP1 (at −3853, −464), AP2 (at −1140, −350, −251, −69), GC Box (−4205, 77, 429), LXR (at −2267), Aβ/AβID (at −2352, −2069), SP1 (at −2654, −323, −318, −256, +119, +217, +357, +452), STAT1 (at −2357, −975), and STAT3 (−4139, −1472, −686, 467).

3.3 Proximal EMSA of JF412354 reveals a potential active GC box/SP site

A set of six oligomer pairs (Table 3) that correspond to two regions within the *MAPT* 5′–UTR were synthesized, radiolabeled, and used to probe nuclear extracts from neuronal and glial cell lines and from mouse brain. Two regions, UTR1 (+35/+114) and UTR2 (+195/+306) (Fig. 6A) were chosen based on density of predicted TF sites (data not shown). An initial screening vs. neuronal (SK–N–BE) nuclear extracts (Fig. 6B) suggested that a specific DNA–protein interaction was occurring when UTR1B and UTR2B were used to probe extracts. Some weaker specificity may have occurred with UTR1C. Probing of mouse brain nuclear extracts (Fig. 6C) produced a single sharp band for UTR1B (lane 2) and strong interaction for all UTR2 probes (lanes 4–6). It is possible that the specific nuclear proteins differ between the NB cell culture and mouse brains. Likewise, the more complex environment of whole brain may lead to protein degradation, resulting in smaller available fragments for *in vitro* DNA–protein interaction.

When UTR1B was used as a probe in competition EMSA, its self specificity in PC12 (Fig. 7, lane 5), hypoxic PC12 (lane 12), and HeLa (lane 14) nuclear extracts was fairly apparent

(Fig. 7). All interaction was eliminated by self competition in PC12 extract (lanes 1 vs. 2). Band III interaction was reduced, if not eliminated with hypoxic PC12 (lanes 12 vs. 13) and HeLa (lanes 14 vs. 15) nuclear extracts. Unexpectedly, competition with UTR1C also eliminated UTR1B–PC12 interaction signal (lane 1 vs., 4) unlike UTR1A (lane 3). However, this can be explained by the very strong competition offered by unlabeled SP1 binding oligomer pair vs. UTR1B (lanes 7 and 1). Both UTR1B and UTR1C are predicted to have a GC box/SP family TF binding site. Thus, either of these two sites has the potential to be active, synergistically or antagonistically, *in vitro*. In addition to SP1 competition, unlabeled oligomer pairs for PR (lane 10) and RAR (lane 11) competed out the UTR1B signal. Genomatix did not predict for either of these particular TFs to bind this sequence. However, this does not exclude the possibility that a marginal PR or RAR site may be able to bind DNA in an *in vitro* EMSA.

3.4 Chloramphenicol acetyltransferase (CAT) reporter gene expressing fusion clones reveal different regulatory domains within JF412354 MAPT gene promoter

A library of JF412354–*CAT* gene reporter fusion clones was constructed as described herein. Eight specific forward–orientation and four reverse–orientation clones were constructed (Fig. 8). These clones were transiently transfected into NB, PC12, PC12 treated with NGF, and C6 cell cultures as described herein. Cells were harvested and total protein was measured by Bradford assay. CAT protein and β–gal levels were measured by ELISA. CAT protein levels were adjusted for clone insert length, β–gal levels, and total protein. Examination of residuals revealed significant non–normality by Anderson–Darling test for each of the four cell line/treatment combinations. Data was adjusted by Box–Cox (NIST/SEMATCH, 2005) transformation (Supplemental Figure). Transformed data was analyzed by Waller–Duncan multiple range test (Steel et al., 1997) to separate values for each clone. In each case, transformed data revealed acceptable differences in promoter activity depending upon specific JF412354 fragment (Fig. 9, Table 7). We required both the generally accepted minimum k ratio of 100 in our analysis for the Waller–Duncan test and a minimum *g* of 1.0 to accept a difference as valid.

In NB and PC12 cells, highest activity occurred when only domain V (the distal–most 2.5kb) was removed (Fig. 9 τ2 construct). The presence of this distal fragment significantly reduced promoter activity in both of the untreated neuronal cell lines (Fig. 9, τ1 vs. τ2 construct). By contrast, in the glial C6 line, highest promoter activity was detected only when domain V was part of a reporter clone (Figs 9 and 10). For the three untreated cell lines, inclusion of domain 0 (partial intron–containing fragment +293/+504) significantly reduced apparent promoter activity. On the other hand, when PC12 cells were treated with NGF, inclusion of domain 0 resulted in increased promoter activity. The “intron–only” clone τ8 had no apparent promoter activity on its own except in PC12–NGF cell culture, in which it had small but significant promoter activity vs. promoterless pBLCAT3. The core promoter region appeared to be between −302/+4 for all three cell lines, although inclusion of the 5′–UTR (+4/+293) significantly increased CAT protein levels in all three cell lines.

The reverse–orientation clones (τ1R, τ2R, τ3R, τ6R) drove CAT levels that were the same as or lower than pBLCAT3 levels in C6, PC12, and PC12 + NGF cells. In none of these cases did reverse–orientation JF12354 insert clones exceed the CAT levels produced by the promoterless vector clone pBLCAT3. In NB cells, on the other hand, two of the reverse clones, τ1R and τ2R differed from pBLCAT3, driving relative adjusted CAT levels of 2.04 and 1.56, respectively. Both of these clones contain the +293/+504 domain 0 intronic fragment. No functional gene has been verified that would correspond to such promoter–like activity. However, a hypothetical LOC100128977 has been reported immediately upstream of and in opposite orientation to the *MAPT* gene (Strausberg et al., 2002).

3.5 Phylogeny of the H2 haplotype agrees with an intermediate pre −H1/H2 sequence for MAPT promoter regions

Rodent and primate *MAPT* sequences were aligned with non–redundant human *MAPT* sequences, including JF412354 and gap–coded as described herein (Supplemental Alignment). The alignment was then analyzed by maximum likelihood (Bayesian) and parsimony phylogeny estimation (Fig. 11, Supplemental Trees). Both methods predicted that the H1 and H2 sequences are more likely to share some common ancestor that diverged from the ape lineage rather than either H1 descending from or being the ancestor of H2. When aligned sequences were compared (chimpanzee vs. H1 and chimpanzee vs. H2) there were 65 points of sequence difference that both H1 and H2 shared with chimpanzee. H2 had a further 28 differences with chimpanzee while H1 had 18 (Supplemental Table, summarized in Table 8). Fisher's exact test reported $p = 0.231$ (odds ratio 1.552), indicating that neither human haplotype has significantly more differences in primary DNA sequence from chimpanzee than does the other in comparison to the number of shared differences. To extend this analysis to a somewhat more functional model, we predicted transcription factor binding sites on the chimpanzee sequence with MatInspector. Both H1 and H2 had 266 shared different predicted TF sites vs. chimpanzee. H2 had 85 unique predicted differences, while H1 had 65. Fisher's exact test reported $p = 0.075$ (odds ratio 1.416). As expected, the result is similar, i.e., non–significant, to direct primary DNA sequence comparison.

4. Discussion

Molecular mechanisms that underlie AD pathological findings and the clinical symptoms are still poorly understood (Holtzman et al., 2011). The amyloid cascade hypothesis posits that accumulation of Aβ initiates a cascade of molecular events leading to neurodegeneration and dementia (Hardy and Selkoe, 2002); however, Aβ–based drugs have recently failed in clinical trials (Sambamurti et al., 2011). Inflammatory responses, mitochondrial dysfunction, oxidative stress, calcium dysregulation, impaired axonal transport, membrane damage and τ based neurotoxicity, among others mechanisms, are likely to participants in this cascade (Sambamurti et al., 2006; Querfurth and LaFerla, 2010). A great deal of work investigating amyloid plaque and τ tangle hypotheses emphasizes the role of protein products. Our premise is to understand the regulation of several genes important to AD, such as *APOE* (Maloney et al., 2010), *APP* (Lahiri and Robakis, 1991; Lahiri et al., 2005a; Lahiri et al., 2005b), *BACE1* (Ge et al., 2004b; Sambamurti et al., 2004; Lahiri et al., 2006a), and *BACE2* (Lahiri et al., 2006b). This present work further reinforces the role of functional domains present in these gene promoters.

Microtubule associated protein τ (gene *MAPT*) is indispensible for normal neuronal function. Diseases as diverse as Parkinson's disease (PD), frontotemporal dementia, progressive supranuclear palsy, Pick disease, and several other tauopathies also have well–known association with aberrant τ deposition or activity (Delacourte, 2008; Borroni et al., 2011; Seto-Salvia et al., 2011), although some forms of frontotemporal dementia are associated with mutations in the gene for progranulin with no associated *MAPT* genetic abnormality (Baker et al., 2006). Many tauopathies have been traced to aberrant post–transcriptional processing, especially selection of specific τ isoforms (Goedert et al., 1989). However, the possibility exists that even moderate alteration of overall τ expression levels regardless of isoform may also contribute tauopathies and even to disorders not specifically characterized as tauopathies simply because τ's normal function includes regulation of intraneuronal transport by dynein and kinesin (Dixit et al., 2008). It is prone to be found in axons, and has been long known to be the primary component of intraneuronal fibrillary tangles (Goedert et al., 1989) in AD. Normal τ function is to interact with tubulin in the cytoskeleton, and disruption of this function contributes to cellular degeneration in AD (Alonso et al., 1996).

The τ protein has a fundamental role in AD progression. Intraneuronal tangles of hyperphosphorylated τ are one of the two archetypal molecular markers of the disorder. Of particular note, τ load is pronounced in the inferior colliculus of AD brains when compared to control brains (Dugger et al., 2011). Antioxidant capacity (a measure of oxidative stress), directly correlates with Braak τ tangle stage in AD (Tayler et al., 2010). While it has been shown that aberrations in τ mRNA splicing contribute to AD and other tauopathies (Chen et al., 2010), several studies indicate *MAPT* gene regulatory contributions to pathogenesis. For example, the C allele of a *MAPT* SNP at +347 (G→C) was significantly associated with AD, as was the C/C genotype. When stratified for ApoE status, these associations persisted at the same significance level for non–ApoEε4 subjects. In addition, this SNP corresponded to greater levels of *MAPT* promoter activity in cell culture (Sun and Jia, 2009). Greater understanding of the *MAPT* promoter would, therefore, contribute to progress against AD and other tauopathies.

Structure and function of the *MAPT* promoter has been previously investigated to some extent (Andreadis et al., 1996). A six clone deletion series derived from a *Hin*dIII/*Sal*I fragment of an H1 *MAPT* promoter, corresponding to −1255/+293 on JF412354, was subcloned and transfected into human cervical epithelial (HeLa) and neuroblastoma SK–N–SH (NB) cells. Andreadis *et al* concluded that the promoter region upstream of "exon 0" was not neuron specific, given that NB and HeLa expression levels only differed twofold (Andreadis et al., 1996). However, they did allow the possibility that a more distal sequence may confer some level of neuronal specificity. On the other hand, a deletion series based on *MAPT* fragments from −2869/+66 suggested that the entirety of the *MAPT* promoter is cell type specific, exercising strong neuronal preference throughout its length (Sadot et al., 1996). We reconcile this apparent contradiction as follows:

Comparison of Andreadis *et al* with our own promoter deletion series leads us to agree that the sequence from *Hin*dIII to *Sal*I (−1255/+293) does not appear to have neuronal preference when NB cells are the neuronal cell line. However, when we compared our transfected PC12 cell culture results to our C6 results over a longer promoter region that roughly corresponds to Sadot *et al*'s deletion series (Sadot et al., 1996), we find that the *MAPT* promoter does have "neuronal preference" throughout the length that corresponds to Sadot *et al*'s experiments. Reporter protein levels from *MAPT*–transfected cell extracts were higher vs. promoterless vector in neuronal (PC12) vs. glial (C6) cells from the same species. This occurs both on the basis of relative adjusted CAT levels and pairwise effect size (*g*). This trend is for PC12 cells that were untreated and those treated with NGF. Interestingly enough, when we tested sequences more distal form the +1 TSS than those surveyed by the other groups, such sequences actually suppressed promoter activity in both NB and PC12 cells, while increasing activity in C6 cells. The presence of a distal regulatory element within the *APP* gene (Lahiri et al., 1999) and a distal *APP* single–nucleotide polymorphism associated with AD (Lahiri et al., 2005b) both raise the possibility of distal sequence participation in a broader regulatory environment in other AD–associated genes. Of interest to questions of neuropathology is that induced differentiation of PC12 cells by NGF reversed the apparent activity of a proximal (+293/+504) intronic fragment. Bringing this all together, we conclude that neuronal specificity for *MAPT* is most likely to lie in the "more global context for correct expression" (Andreadis et al., 1996), which relies on the activities of cells within tissues, tissues within organs, organs within an organism, all within an environment.

Also of note, our work suggests an additional possibility. Specifically, comparing clones τ5 with τ6 shows that the presence of short intronic fragment 0 (+293/+504) functions to significantly reduce expression levels in unstimulated cell lines. This reduction in occurred in all three untreated cell lines we tested, but in PC12 cells, adjusted CAT protein levels

were still significantly higher than promoterless vector. When we treated PC12 cells with NGF, the 0 region acted to upregulate promoter activity. Regulation of gene transcription through introns is known (Jones et al., 2011), and in some cases, this process is mediated through transcription factors such as NF–κ–B and SP1 (Deshane et al., 2010; Ying et al., 2010), both of which have association with neurodegenerative disorders (Du et al., 2005; Dosunmu et al., 2009). Therefore, our examination of both *MAPT* distal promoter sequences and a fragment of its first intron, along with differing results that depend upon what one accepts as a "neuronal" cell model, leads us to suggest that both distal promoter and elements downstream of the +1 TSS may work to confer neuronal specificity on *MAPT* expression, and that this specificity may be particularly sensitive to the overall context of expression.

One variable that has not been taken into account is that our *MAPT* sequence was H2, and the earlier studies used H1 sequences. One direct comparison of H1 vs. H2 promoter activity has been done. This consisted of a comparison of ~1kb *MAPT* promoter/5′–UTR fragments fused to a luciferase reporter vector. The H2 clone in that study had 20% to 25% lower luciferase activity than did the H1 clone (Kwok et al., 2004). However, no deletion series of the promoter was constructed to establish if any differences in specific domains existed. Our current experiments were to provide a broad analysis of the specific H2 clone. While a side–by–side H1/H2 cloning series would have been most ideal, it would have also been beyond the scope of a single manuscript. Therefore, any functional conclusions drawn between H1 and H2 *MAPT* promoters are tentative and useful for generating hypotheses for further investigation.

In addition to detailed molecular and functional anatomy of the *MAPT* promoter, we have undertaken some exploration of potential specifics of its proximal regulation. The presence of putative SP1 sites in the proximal *MAPT* promoter has been previously reported (Andreadis et al., 1992; Andreadis et al., 1996; Sadot et al., 1996), and we also predict GC box/SP family sites in the proximal *MAPT* region, specifically the 5′–UTR coded by "exon 0". We began our investigation of potential proximal transcription factor binding by running EMSA with selected oligomer pairs synthesized from short sequences near the JF412354 3′–end. Of the six probes tested, one sequence (+60/+84) showed self–specificity in DNA–protein interaction. We further tested this specificity by competition EMSA against several unlabeled (cold) oligomer pairs that contain specific transcription factor binding sites. Interestingly, we observed that SP1, ER, GR, and RAR oligomer pairs all successfully competed against the +60/+84 probe, as did an unlabeled oligomer pair that corresponded to +85/+114. This competitor has a fragmentary GC box site at its 5′–most end, which may explain the competition. TransFac prediction of TF sites within JF412354 did not predict active ER, GR, or RAR sites within +60/+84. The presence of a potentially active SP1 site in JF412354 is worth interest in light of our findings that closely associate SP1 levels with mRNA levels of the AD–associated genes *APP* and β–secretase enzyme (*BACE1*) over the lifespan of rodents and monkeys (Dosunmu et al., 2009).

Characterization of the *MAPT* promoter indicates that our sequence is of the H2 haplotype, which spans over 1 megabase (Oliveira et al., 2004) and includes a 900kb inversion (Stefansson et al., 2005). The H1 haplotype of *MAPT* is associated with several tauopathies, including PD and to faster progression to dementia in AD (Wider et al., 2010; Seto-Salvia et al., 2011). The H2 allele has been associated with 17q21.31 microdeletion mental retardation and more severe decline in FTD (Rao et al., 2010; Borroni et al., 2011). H2 is found primarily in European–originating populations (Stefansson et al., 2005; Donnelly et al., 2010). Hypotheses based on recent mutation of H2 from H1 have been generally discounted, and both H1 and H2 are accepted as ancient lineages (Stefansson et al., 2005; Zody et al., 2008). Likewise, the suggestion that H2 was somehow infused from Neanderthal

contact with archaic *H. sapiens* has been rejected in light of available Neanderthal sequence information (Pennisi, 2009). A case has been made that H2 is the ancestral lineage (Zody et al., 2008). Alternatively, others grant neither lineage ancestral status, based on roughly equal distance each haplotype has vs. chimpanzee (Cruts et al., 2005; Stefansson et al., 2005). Our own data tends to agree with the latter hypothesis, that H1 and H2 are each descendents of a third, ancestral sequence.

However, the ancestral status of H1 vs. H2 is not as immediately important as functional differences between the two. Our present analysis has concentrated on bringing the state of detailed knowledge for the H2 *MAPT* promoter region up at least to the level of H1. Now that specific regulatory domains have been mapped out for the H2 *MAPT* promoter, it can be compared “head to head” with the same or at least very similar analysis of H1 *MAPT* sequences to elucidate detailed, specific functional differences that could be amenable to direct pharmaceutical intervention. The presence of two well–defined *MAPT* lineages with distinct effects on neurodegenerative disorders ought to be an excellent workbench for elucidating the specific mechanistic traits of each that contribute to or protect from respective diseases. In addition to our promoter functional analysis and EMSA data highly suggestive of a functioning proximal SP1 site, we have come across several predicted TF recognition sequences that cross H1/H2 points of dimorphism. Most notably, these include C/EBPβ, CMYC, HIF1, and NF–κ–B, all of which are associated with inflammatory response, often a key element in etiology of neurodegenerative disorders. In addition, C/EBP and CREB signaling are intimately involved in memory persistence, AD progression, and neuronal plasticity (Saura and Valero, 2011). These particularly putative sites exclusively cross H1/H2 dimorphic sites. These sites offer particularly inviting specific targets for further research. Several GC box/SP factor sites also cross H1/H2 dimorphism locations. SP1 is strongly implicated in models of gene–environment interaction for AD etiology, particularly models that differentiate among familial and sporadic AD (Bales et al., 2000; Lahiri and Maloney, 2010). NF–κ–B activity generally functions in neuroinflammation (Bales et al., 2000) and specifically acts to enable Aβ–mediated upregulation of *APOE* gene expression (Du et al., 2005). The SP1 TF also has been important in development of the “latent early–life associated regulation” (LEARn) model of sporadic disease etiology (Lahiri and Maloney, 2010). In short, the LEARn model proposes that transcription factor sites can be “hot spots” for disease etiology due to containing CpG (DNA methylation) or GG (DNA oxidation) sites. Early–life exposure or stress induces epigenetic DNA changes in the form of DNA methylation and/or oxidation. These epigenetic markers have latent effect until an additional trigger event occurs, at which time effects exceed a clinical threshold, resulting in disease.

We have extended specific knowledge of *MAPT* promoter activity and structure further upstream than previous publications have. Our functional promoter deletion series has explicitly laid out challenges inherent in model cell cultures for *MAPT* promoter analyses and suggests that *MAPT* regulation is as much a matter of *in vivo* context as local DNA sequence/TF interaction. We also have opened up the possibility of intronic regulation of *MAPT* expression, and that this regulation may be cell type specific. Our extension of the understanding of *MAPT* has been both physical and conceptual and points to mechanistic–oriented studies with definable and useful results. We have determined that a potentially active SP1 site exists in the 5′–UTR of the sequence. Finally, we have identified several potential TF binding sites that cross H1/H2 haplotype dimorphisms. Given the well–established differences in disease risk posed by H1 vs. H2, these specific TF sites are good candidates for rationally targeted tauopathy research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Alzheimer's Association and National Institutes of Health AG18379 and AG18884. We thank Junrong Ma and Yuan–Wen Ge, for vital technical assistance, including assisting with cloning and subcloning of JF412354 and performing cell culture transfections and CAT protein assays.

List of Abbreviations

AD	Alzheimer's disease
AP	activator protein
APOE	apolipoprotein E gene
APP	amyloid β peptide precursor gene
Aβ	amyloid β peptide
bp	base–pair
C/EBP	CCAAT/enhancer binding protein
C6	rat glioblastoma cells
CAT	chloramphenicol acetyltransferase protein
CAT	chloramphenicol acetyltransferase gene
CMYC	myelocytomatosis viral oncogene homolog
CREB	cyclic AMP responsive element binding proteins
CRHR1	corticotrophin releasing hormone receptor 1
E2F4	E2F transcription factor 4
EMSA	electrophoretic mobility shift assay or gel shift assay
ER	estrogen receptor
FTD	frontotemporal dementia
HIF	hypoxia–inducible factor
HRE	hypoxia responsive element
MAPT	microtubule–associated τ gene
MARE	musculoaponeurotic fibrosarcoma oncogene homolog recognition element
NB	human neuroblastoma SK–N–BE extract or human neuroblastoma SK–N–SH cells
NF–κ–B	nuclear factor κ B
NRL	neural retina leucine zipper
PC12	rat pheochromocytoma/neuronal cells
PD	Parkinson's disease
PDEF	prostate derived Ets transcription factor
PR	progesterone receptor

RAR	retinoic acid receptor
SP	specificity protein
TF	transcription factor
TSS	hnRNA transcription start site
U373	human glioblastoma–astrocytoma
UTR	mRNA untranslated region
τ	microtubule–associated τ protein

References

Adamec E, Chang HT, Stopa EG, Hedreen JC, Vonsattel JP. Tau protein expression in frontotemporal dementias. Neuroscience Letters. 2001; 315:21–4. [PubMed: 11711205]

Alonso AC, Grundke-Iqbal I, Iqbal K. Alzheimer's disease hyperphosphorylated tau sequesters normal tau into tangles of filaments and disassembles microtubules. Nature Medicine. 1996; 2:783–7.

Andreadis A, Brown WM, Kosik KS. Structure and novel exons of the human tau gene. Biochemistry. 1992; 31:10626–33. [PubMed: 1420178]

Andreadis A, Wagner BK, Broderick JA, Kosik KS. A tau promoter region without neuronal specificity. J Neurochem. 1996; 66:2257–63. [PubMed: 8632146]

Bailey JA, Maloney B, Ge YW, Lahiri DK. Functional activity of the novel Alzheimer's amyloid beta-peptide interacting domain (AbetaID) in the APP and BACE1 promoter sequences and implications in activating apoptotic genes and in amyloidogenesis. Gene. 2011; 488:13–22. [PubMed: 21708232]

Baker M, Mackenzie IR, Pickering-Brown SM, Gass J, Rademakers R, Lindholm C, Snowden J, Adamson J, Sadovnick AD, Rollinson S, Cannon A, Dwosh E, Neary D, Melquist S, Richardson A, Dickson D, Berger Z, Eriksen J, Robinson T, Zehr C, Dickey CA, Crook R, McGowan E, Mann D, Boeve B, Feldman H, Hutton M. Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. Nature. 2006; 442:916–9. [PubMed: 16862116]

Bales KR, Du Y, Holtzman D, Cordell B, Paul SM. Neuroinflammation and Alzheimer's disease: critical roles for cytokine/Abeta-induced glial activation, NF-kappaB, and apolipoprotein E. Neurobiol Aging. 2000; 21:427–432. [PubMed: 10858588]

Belarbi K, Burnouf S, Fernandez-Gomez FJ, Desmercieres J, Troquier L, Brouillette J, Tsambou L, Grosjean ME, Caillierez R, Demeyer D, Hamdane M, Schindowski K, Blum D, Buee L. Loss of Medial Septum Cholinergic Neurons in THY-Tau22 Mouse Model: What Links with tau Pathology? Curr Alzheimer Res. 2011; 8:633–8. [PubMed: 21605043]

Borchsenius, F. FastGap 1.2. 2009. http://192.38.46.42/aubot/fb/FastGap_home.htm

Borroni B, Grassi M, Archetti S, Papetti A, Del Bo R, Bonvicini C, Comi GP, Gennarelli M, Bellelli G, Di Luca M, Padovani A. Genetic Background Predicts Poor Prognosis in Frontotemporal Lobar Degeneration. Neurodegener Dis. 2011

Brodaty H, Breteler MM, Dekosky ST, Dorenlot P, Fratiglioni L, Hock C, Kenigsberg PA, Scheltens P, De Strooper B. The World of Dementia Beyond 2020. J Am Geriatr Soc. 2011; 59:923–927. [PubMed: 21488846]

Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2004; Chapter 4(Unit 4):10. [PubMed: 18428725]

Chen S, Townsend K, Goldberg TE, Davies P, Conejero-Goldberg C. MAPT isoforms: differential transcriptional profiles related to 3R and 4R splice variants. J Alzheimers Dis. 2010; 22:1313–29. [PubMed: 20930284]

Cruts M, Rademakers R, Gijselinck I, van der Zee J, Dermaut B, de Pooter T, de Rijk P, Del-Favero J, van Broeckhoven C. Genomic architecture of human 17q21 linked to frontotemporal dementia uncovers a highly homologous family of low-copy repeats in the tau region. Human Molecular Genetics. 2005; 14:1753–62. [PubMed: 15888485]

de Mendiburu, F. Agricolae: Practical Manual. 2010. http://tarwi.lamolina.edu.pe/~fmendiburu/index-filer/download/PracticalManualAgricolae.pdf

De Strooper B. Proteases and proteolysis in Alzheimer disease: a multifactorial view on the disease process. Physiol Rev. 2010; 90:465–94. [PubMed: 20393191]

Delacourte A. Tau pathology and neurodegeneration: an obvious but misunderstood link. J Alzheimers Dis. 2008; 14:437–40. [PubMed: 18688095]

Deshane J, Kim J, Bolisetty S, Hock TD, Hill-Kapturczak N, Agarwal A. Sp1 regulates chromatin looping between an intronic enhancer and distal promoter of the human heme oxygenase-1 gene in renal cells. J Biol Chem. 2010; 285:16476–86. [PubMed: 20351094]

Dixit R, Ross JL, Goldman YE, Holzbaur EL. Differential regulation of dynein and kinesin motor proteins by tau. Science. 2008; 319:1086–9. [PubMed: 18202255]

Donnelly MP, Paschou P, Grigorenko E, Gurwitz D, Mehdi SQ, Kajuna SL, Barta C, Kungulilo S, Karoma NJ, Lu RB, Zhukova OV, Kim JJ, Comas D, Siniscalco M, New M, Li P, Li H, Manolopoulos VG, Speed WC, Rajeevan H, Pakstis AJ, Kidd JR, Kidd KK. The distribution and most recent common ancestor of the 17q21 inversion in humans. American Journal of Human Genetics. 2010; 86:161–71. [PubMed: 20116045]

Dosunmu R, Wu J, Adwan L, Maloney B, Basha MR, McPherson CA, Harry GJ, Rice DC, Zawia NH, Lahiri DK. Lifespan profiles of Alzheimer's disease-associated genes and products in monkeys and mice. Journal of Alzheimers Disease. 2009; 18:211–30.

Du Y, Chen X, Wei X, Bales KR, Berg DT, Paul SM, Farlow MR, Maloney B, Ge YW, Lahiri DK. NF-(kappa)B mediates amyloid beta peptide-stimulated activity of the human apolipoprotein E gene promoter in human astroglial cells. Brain Res Mol Brain Res. 2005; 136:177–88. [PubMed: 15893602]

Duff K, Kuret J, Congdon EE. Disaggregation of tau as a therapeutic approach to tauopathies. Curr Alzheimer Res. 2010; 7:235–40. [PubMed: 20088807]

Dugger BN, Tu M, Murray ME, Dickson DW. Disease specificity and pathologic progression of tau pathology in brainstem nuclei of Alzheimer's disease and progressive supranuclear palsy. Neurosci Lett. 2011; 491:122–6. [PubMed: 21236314]

Fink WL. Microcomputers and phylogenetic analysis. Science. 1986; 234:1135–9. [PubMed: 17778959]

Fletcher W, Yang Z. The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. Mol Biol Evol. 2010; 27:2257–67. [PubMed: 20447933]

Ge Y-W, Ghosh M, Song W, Maloney B, Lahiri D. Mechanism of promoter activity of the beta-amyloid precursor protein gene in different cell types. Identification of a specific 30 bp fragment in the proximal promoter region. Journal of Neurochemistry. 2004a; 90:1432–44. [PubMed: 15341527]

Ge Y-W, Maloney B, Sambamurti K, Lahiri DK. Functional characterization of the 5′ flanking region of the BACE gene: identification of a 91 bp fragment involved in basal level of BACE promoter expression. FASEB Journal. 2004b; 18:1037–1039. [PubMed: 15059977]

Ghosh C, Song W, Lahiri DK. Efficient DNA transfection in neuronal and astrocytic cell lines. Molecular Biology Reports. 2000; 27:113–121. [PubMed: 11092558]

Giannakopoulos P, Kovari E, Gold G, von Gunten A, Hof PR, Bouras C. Pathological substrates of cognitive decline in Alzheimer's disease. Front Neurol Neurosci. 2009; 24:20–9. [PubMed: 19182459]

Goedert M, Spillantini MG, Jakes R, Rutherford D, Crowther RA. Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease. Neuron. 1989; 3:519–26. [PubMed: 2484340]

Grayson DR, Jia X, Chen Y, Sharma RP, Mitchell CP, Guidotti A, Costa E. Reelin promoter hypermethylation in schizophrenia. Proc Natl Acad Sci U S A. 2005; 102:9341–6. [PubMed: 15961543]

Gross, J. Five omnibus tests for the composite hypothesis of normality. 2006. http://cran.r-project.org/web/packages/nortest/

Hardy J, Selkoe DJ. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. Science. 2002; 297:353–356. [PubMed: 12130773]

Harlow, LL.; Mulaik, SA.; Steiger, JH. What if there were no significance tests?. Lawrence Erlbaum Associates Publishers; Mahwah, N.J: 1997.

Hedges, LV.; Olkin, I. Statistical Method for Meta-Analysis. Academic Press; Waltham, Massachussetts: 1985.

Holtzman DM, Morris JC, Goate AM. Alzheimer's disease: the challenge of the second century. Sci Transl Med. 2011; 3:77sr1. [PubMed: 21471435]

Huang X. An improved sequence assembly program. Genomics. 1996; 33:21–31. [PubMed: 8617506]

Iqbal K, Liu F, Gong CX, Grundke-Iqbal I. Tau in Alzheimer disease and related tauopathies. Curr Alzheimer Res. 2010; 7:656–64. [PubMed: 20678074]

Jones EA, Lopez-Anido C, Srinivasan R, Krueger C, Chang LW, Nagarajan R, Svaren J. Regulation of the PMP22 Gene through an Intronic Enhancer. J Neurosci. 2011; 31:4242–50. [PubMed: 21411665]

Jurka J, Klonowski P, Dagman V, Pelton P. CENSOR--a program for identification and elimination of repetitive elements from DNA sequences. Comput Chem. 1996; 20:119–21. [PubMed: 8867843]

Kim S, Jensen JR, Cisek K, Funk KE, Naphade S, Schafer K, Kuret J. Imaging as a strategy for premortem diagnosis and staging of tauopathies. Curr Alzheimer Res. 2010; 7:230–4. [PubMed: 20088805]

Kim SJ, Park K, Rudkin BB, Dey BR, Sporn MB, Roberts AB. Nerve growth factor induces transcription of transforming growth factor-beta 1 through a specific promoter element in PC12 cells. Journal of Biological Chemistry. 1994; 269:3739–44. [PubMed: 8106419]

Kwok JB, Teber ET, Loy C, Hallupp M, Nicholson G, Mellick GD, Buchanan DD, Silburn PA, Schofield PR. Tau haplotypes regulate transcription and are associated with Parkinson's disease. Annals of Neurology. 2004; 55:329–34. [PubMed: 14991810]

Lahiri DK, Ge Y-W, Maloney B. Characterization of the APP proximal promoter and 5′-untranslated regions: identification of cell-type specific domains and implications in APP gene expression and Alzheimer's disease. FASEB Journal. 2005a; 19:653–655. [PubMed: 15703276]

Lahiri DK, Ge Y. Electrophoretic mobility shift assay for the detection of specific DNA-protein complex in nuclear extracts from the cultured cells and frozen autopsy human brain tissue. Brain Research. Brain Research Protocols. 2000; 5:257–265. [PubMed: 10906491]

Lahiri DK, Maloney B. The "LEARn" (Latent Early-life Associated Regulation) model integrates environmental risk factors and the developmental basis of Alzheimer's disease, and proposes remedial steps. Experimental Gerontology. 2010; 45:291–6. [PubMed: 20064601]

Lahiri DK, Maloney B, Ge Y-W. BACE1 gene promoter is differentially regulated: Detection of a novel promoter region for its cell type–specific regulation. Journal of Molecular Neuroscience. 2006a; 28:193–210. [PubMed: 16679558]

Lahiri DK, Maloney B, Ge Y-W. Functional Domains of the BACE1 and BACE2 promoters and mechanisms of transcriptional suppression of the BACE2 promoter in normal neuronal cells. Journal of Molecular Neuroscience. 2006b; 29:65–80. [PubMed: 16757811]

Lahiri DK, Maloney B, Zawia NH. The LEARn model: an epigenetic explanation for idiopathic neurobiological diseases. Molecular Psychiatry. 2009; 14:992–1003. [PubMed: 19851280]

Lahiri DK, Nall C, Ge YW. Promoter activity of the beta-amyloid precursor protein gene is negatively modulated by an upstream regulatory element. Brain Research. Molecular Brain Research. 1999; 71:32–41. [PubMed: 10407184]

Lahiri DK, Robakis NK. The promoter activity of the gene encoding Alzheimer beta-amyloid precursor protein (APP) is regulated by two blocks of upstream sequences. Brain Research. Molecular Brain Research. 1991; 9:253–257. [PubMed: 1851527]

Lahiri DK, Wavrant De-Vrieze F, Ge Y-W, Maloney B, Hardy J. Characterization of two APP gene promoter polymorphisms that appear to influence risk of late-onset Alzheimer's disease. Neurobiology of Aging. 2005b; 26:1329–1341. [PubMed: 16243604]

Laws SM, Perneczky R, Drzezga A, Diehl-Schmid J, Ibach B, Bauml J, Eisele T, Forstl H, Kurz A, Riemenschneider M. Association of the tau haplotype H2 with age at onset and functional alterations of glucose utilization in frontotemporal dementia. Am J Psychiatry. 2007; 164:1577–84. [PubMed: 17898350]

Lee YA, Durandin A, Dedon PC, Geacintov NE, Shafirovich V. Oxidation of guanine in G, GG, and GGG sequence contexts by aromatic pyrenyl radical cations and carbonate radical anions: relationship between kinetics and distribution of alkali-labile lesions. J Phys Chem B. 2008; 112:1834–44. [PubMed: 18211057]

Lv H, Jia L, Jia J. Promoter polymorphisms which modulate APP expression may increase susceptibility to Alzheimer's disease. Neurobiol Aging. 2008; 29:194–202. [PubMed: 17112637]

Maloney B, Ge YW, Petersen RC, Hardy J, Rogers JT, Perez-Tur J, Lahiri DK. Functional characterization of three single-nucleotide polymorphisms present in the human APOE promoter sequence: Differential effects in neuronal cells and on DNA-protein interactions. American Journal of Medical Genetics part B: Neuropsychiatric Genetics. 2010; 153B:185–201.

Maloney B, Lahiri DK. The Alzheimer's amyloid beta-peptide (Abeta) binds a specific DNA Abeta-interacting domain (AbetaID) in the APP, BACE1, and APOE promoters in a sequence-specific manner: Characterizing a new regulatory motif. Gene. 2011; 488:1–12. [PubMed: 21699964]

Maloney B, Ray B, Hayden EP, Nurnberger JI Jr, Lahiri DK. Development and validation of the high-quality 'rapid method for swab' to genotype the HTTLPR serotonin transporter (SLC6A4) promoter polymorphism. Psychiatr Genet. 2009; 19:72–82. [PubMed: 19668112]

McCulloch CC, Kay DM, Factor SA, Samii A, Nutt JG, Higgins DS, Griffith A, Roberts JW, Leis BC, Montimurro JS, Zabetian CP, Payami H. Exploring gene-environment interactions in Parkinson's disease. Hum Genet. 2008; 123:257–65. [PubMed: 18210157]

Mullan M, Crawford F, Axelman K, Houlden H, Lilius L, Winblad B, Lannfelt L. A pathogenic mutation for probable Alzheimer's disease in the APP gene at the N-terminus of beta-amyloid. Nat Genet. 1992; 1:345–7. [PubMed: 1302033]

Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev Camb Philos Soc. 2007; 82:591–605. [PubMed: 17944619]

NIST/SEMATCH e-Handbook of Statistical Methods. Box-Cox normality plot. 2005. http://www.itl.nist.gov/div898/handbook/eda/section3/boxcoxno.htm

Olejnik S, Algina J. Generalized eta and omega squared statistics: measures of effect size for some common research designs. Psychological Methods. 2003; 8:434–47. [PubMed: 14664681]

Oliveira SA, Scott WK, Zhang F, Stajich JM, Fujiwara K, Hauser M, Scott BL, Pericak-Vance MA, Vance JM, Martin ER. Linkage disequilibrium and haplotype tagging polymorphisms in the Tau H1 haplotype. Neurogenetics. 2004; 5:147–55. [PubMed: 15459824]

Pennisi E. Neandertal genomics. Tales of a prehistoric human genome. Science. 2009; 323:866–71. [PubMed: 19213888]

Quandt K, Frech K, Karas H, Wingender E, Werner T. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Research. 1995; 23:4878–4884. [PubMed: 8532532]

Querfurth HW, LaFerla FM. Alzheimer's disease. New England Journal of Medicine. 2010; 362:329–44. [PubMed: 20107219]

R Development Core Team. R: A language and environment for statistical computing. 2011. http://www.R-project.org/

Rao PN, Li W, Vissers LE, Veltman JA, Ophoff RA. Recurrent inversion events at 17q21.31 microdeletion locus are linked to the MAPT H2 haplotype. Cytogenet Genome Res. 2010; 129:275–9. [PubMed: 20606400]

Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003; 19:1572–4. [PubMed: 12912839]

Sadot E, Heicklen-Klein A, Barg J, Lazarovici P, Ginzburg I. Identification of a tau promoter region mediating tissue-specific-regulated expression in PC12 cells. J Mol Biol. 1996; 256:805–12. [PubMed: 8601831]

Samaranch L, Cervantes S, Barabash A, Alonso A, Cabranes JA, Lamet I, Ancin I, Lorenzo E, Martinez-Lage P, Marcos A, Clarimon J, Alcolea D, Lleo A, Blesa R, Gomez-Isla T, Pastor P. The effect of MAPT H1 and APOE epsilon4 on transition from mild cognitive impairment to dementia. J Alzheimers Dis. 2010; 22:1065–71. [PubMed: 20930301]

Sambamurti K, Greig NH, Utsuki T, Barnwell EL, Sharma E, Mazell C, Bhat NR, Kindy MS, Lahiri DK, Pappolla MA. Targets for AD treatment: conflicting messages from gamma-secretase inhibitors. J Neurochem. 2011; 117:359–74. [PubMed: 21320126]

Sambamurti K, Kinsey R, Maloney B, Ge YW, Lahiri DK. Gene structure and organization of the human beta-secretase (BACE) promoter. FASEB Journal. 2004; 18:1034–1036. [PubMed: 15059975]

Sambamurti K, Suram A, Venugopal C, Prakasam A, Zhou Y, Lahiri DK, Greig NH. A partial failure of membrane protein turnover may cause Alzheimer's disease: a new hypothesis. Curr Alzheimer Res. 2006; 3:81–90. [PubMed: 16472208]

Saura CA, Valero J. The role of CREB signaling in Alzheimer's disease and other cognitive disorders. Rev Neurosci. 2011; 22:153–69. [PubMed: 21476939]

Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Mizrachi I, Ostell J, Panchenko A, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, John Wilbur W, Yaschenko E, Ye J. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2010; 38:D5–16. [PubMed: 19910364]

Schmidt HA, Strimmer K, Vingron M, von Haeseler A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. 2002; 18:502–4. [PubMed: 11934758]

Seto-Salvia N, Clarimon J, Pagonabarraga J, Pascual-Sedano B, Campolongo A, Combarros O, Mateo JI, Regana D, Martinez-Corral M, Marquie M, Alcolea D, Suarez-Calvet M, Molina-Porcel L, Dols O, Gomez-Isla T, Blesa R, Lleo A, Kulisevsky J. Dementia Risk in Parkinson Disease: Disentangling the Role of MAPT Haplotypes. Arch Neurol. 2011; 68:359–64. [PubMed: 21403021]

Simmons MP, Ochoterena H. Gaps as characters in sequence-based phylogenetic analyses. Syst Biol. 2000; 49:369–81. [PubMed: 12118412]

Snider BJ, Fagan AM, Roe C, Shah AR, Grant EA, Xiong C, Morris JC, Holtzman DM. Cerebrospinal fluid biomarkers and rate of cognitive decline in very mild dementia of the Alzheimer type. Arch Neurol. 2009; 66:638–45. [PubMed: 19433664]

Steel, RGD.; Torrie, JH.; Dickey, DA. Principles and procedures of statistics: a biometrical approach. McGraw-Hill; New York: 1997.

Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG, Desnica N, Hicks A, Gylfason A, Gudbjartsson DF, Jonsdottir GM, Sainz J, Agnarsson K, Birgisdottir B, Ghosh S, Olafsdottir A, Cazier JB, Kristjansson K, Frigge ML, Thorgeirsson TE, Gulcher JR, Kong A, Stefansson K. A common inversion under selection in Europeans. Nature Genetics. 2005; 37:129–37. [PubMed: 15654335]

Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD, Collins FS, Wagner L, Shenmen CM, Schuler GD, Altschul SF, Zeeberg B, Buetow KH, Schaefer CF, Bhat NK, Hopkins RF, Jordan H, Moore T, Max SI, Wang J, Hsieh F, Diatchenko L, Marusina K, Farmer AA, Rubin GM, Hong L, Stapleton M, Soares MB, Bonaldo MF, Casavant TL, Scheetz TE, Brownstein MJ, Usdin TB, Toshiyuki S, Carninci P, Prange C, Raha SS, Loquellano NA, Peters GJ, Abramson RD, Mullahy SJ, Bosak SA, McEwan PJ, McKernan KJ, Malek JA, Gunaratne PH, Richards S, Worley KC, Hale S, Garcia AM, Gay LJ, Hulyk SW, Villalon DK, Muzny DM, Sodergren EJ, Lu X, Gibbs RA, Fahey J, Helton E, Ketteman M, Madan A, Rodrigues S, Sanchez A, Whiting M, Madan A, Young AC, Shevchenko Y, Bouffard GG, Blakesley RW, Touchman JW, Green ED, Dickson MC, Rodriguez AC, Grimwood J, Schmutz J, Myers RM, Butterfield YS, Krzywinski MI, Skalska U, Smailus DE, Schnerch A, Schein JE, Jones SJ, Marra MA. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proceedings of the National Academy of Sciences of the United States of America. 2002; 99:16899–16903. [PubMed: 12477932]

Sun W, Jia J. The +347 C promoter allele up-regulates MAPT expression and is associated with Alzheimer's disease among the Chinese Han. Neurosci Lett. 2009; 450:340–3. [PubMed: 19070646]

Sundar PD, Yu CE, Sieh W, Steinbart E, Garruto RM, Oyanagi K, Craig UK, Bird TD, Wijsman EM, Galasko DR, Schellenberg GD. Two sites in the MAPT region confer genetic risk for Guam ALS/PDC and dementia. Hum Mol Genet. 2007; 16:295–306. [PubMed: 17185385]

Tayler H, Fraser T, Miners JS, Kehoe PG, Love S. Oxidative balance in Alzheimer's disease: relationship to APOE, Braak tangle stage, and the concentrations of soluble and insoluble amyloid-beta. J Alzheimers Dis. 2010; 22:1363–73. [PubMed: 20930272]

Thies W, Bleiler L. 2011 Alzheimer's disease facts and figures. Alzheimers & Dementia. 2011; 7:208–44.

Thinakaran G, Borchelt DR, Lee MK, Slunt HH, Spitzer L, Kim G, Ratovitsky T, Davenport F, Nordstedt C, Seeger M, Hardy J, Levey AI, Gandy SE, Jenkins NA, Copeland NG, Price DL, Sisodia SS. Endoproteolysis of presenilin 1 and accumulation of processed derivatives in vivo. Neuron. 1996; 17:181–90. [PubMed: 8755489]

Wang S, Jia J. Promoter polymorphisms which modulate BACE1 expression are associated with sporadic Alzheimer's disease. Am J Med Genet B Neuropsychiatr Genet. 2010; 153B:159–66. [PubMed: 19441127]

Wider C, Vilarino-Guell C, Jasinska-Myga B, Heckman MG, Soto-Ortolaza AI, Cobb SA, Aasly JO, Gibson JM, Lynch T, Uitti RJ, Wszolek ZK, Farrer MJ, Ross OA. Association of the MAPT locus with Parkinson's disease. Eur J Neurol. 2010; 17:483–6. [PubMed: 19912324]

Williams-Gray CH, Evans JR, Goris A, Foltynie T, Ban M, Robbins TW, Brayne C, Kolachana BS, Weinberger DR, Sawcer SJ, Barker RA. The distinct cognitive syndromes of Parkinson's disease: 5 year follow-up of the CamPaIGN cohort. Brain. 2009; 132:2958–69. [PubMed: 19812213]

Yamagata K, Urakami K, Ikeda K, Ji Y, Adachi Y, Arai H, Sasaki H, Sato K, Nakashima K. High expression of apolipoprotein E mRNA in the brains with sporadic Alzheimer's disease. Dementia and Geriatric Cognitive Disorders. 2001; 12:57–62. [PubMed: 11173875]

Ying S, Kojima T, Kawada A, Nachat R, Serre G, Simon M, Takahara H. An intronic enhancer driven by NF-kappaB contributes to transcriptional regulation of peptidylarginine deiminase type I gene in human keratinocytes. J Invest Dermatol. 2010; 130:2543–52. [PubMed: 20596086]

Zody MC, Jiang Z, Fung HC, Antonacci F, Hillier LW, Cardone MF, Graves TA, Kidd JM, Cheng Z, Abouelleil A, Chen L, Wallis J, Glasscock J, Wilson RK, Reily AD, Duckworth J, Ventura M, Hardy J, Warren WC, Eichler EE. Evolutionary toggling of the MAPT 17q21.31 inversion region. Nature Genetics. 2008; 40:1076–83. [PubMed: 19165922]

Zuo X, Jia J. Promoter polymorphisms which modulate insulin degrading enzyme expression may increase susceptibility to Alzheimer's disease. Brain Res. 2009; 1249:1–8. [PubMed: 18996360]

Highlights

We cloned a previously unpublished 4kb tau (*MAPT*) gene promoter region.

We further analyzed the 4kb sequence via EMSA and *CAT* reporter gene clone series.

The *MAPT* sequence was for the less common H2 haplotype.

EMSA showed potentially important transcription factor interactions.

Reporter clone series revealed promoter regulatory domains.

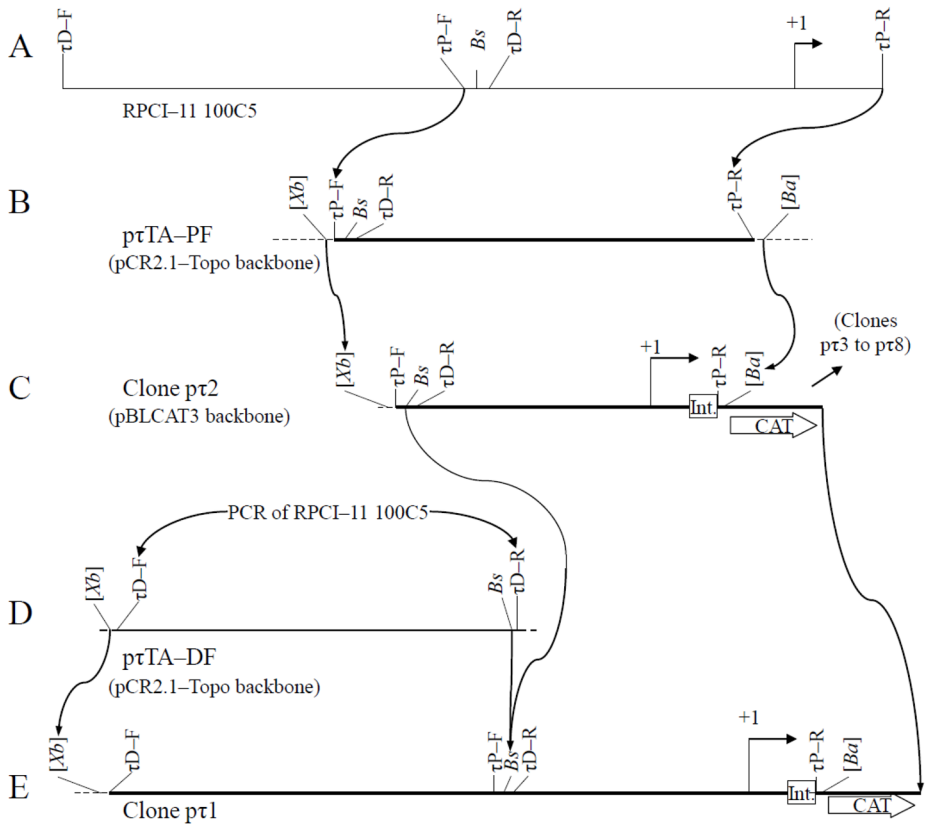


Fig. 1. Generation of *MAPT* 5′–flanking DNA sequence clones from clone RPCI–11 100C5

A) BAC RPCI–11 100C5 was subjected to PCR with two primer pairs, τD–F/τD–R (distal) and τP–F/τP–R (proximal), as described in the text. PCR produced partially overlapping ~2.5kb fragments. PCR products were cloned into the pCR2.1–Topo vector (Invitrogen) in random orientation. Inserts and orientations were confirmed by digestion with appropriate restriction enzymes, and pCR2.1–Topo backbone clones were used for sequencing and subcloning into expression vectors. The proximal PCR fragment was used to produce B) pτTA–P in both orientations (the "pτTA–PF" forward orientation clone is shown). C) Clone pτTA–PF was digested with *Xba*I and *Bam*HI. A 2.5kb fragment was purified and cloned into the *Xba*I/*Bam*HI sites of expression vector pBLCAT3 (Promega) to produce pτ2. D) The distal *MAPT* promoter PCR fragment from RPCI–11 100C5 was cloned into pCR2.1–Topo to produce pτTA–D in both orientations (the "pτTA–DF" forward orientation clone is shown). E) Clone pτTA–DF was digested with *Xba*I and *Bst*EII and the 2.5kb *MAPT* fragment was cloned into the *Xba*I and *Bst*EII sites of pτ2 to produce pτ1. The *Bst*EII site was native to the RPCI–11 100C5 BAC, making the pτ1 clone is a reconstruction of the original BAC sequence fragment. In addition, pτ2 was used in further digests, as described in the text, to produce clones pτ3 to pτ8. The "R" (reverse) orientation clones were produced in a similar fashion, starting with opposite–oriented pCR2.1–Topo backbone clones. Restriction enzyme names in the figure are abbreviated as follows: Ba, *Bam*HI; Bs, *Bst*EII; Xb, *Xba*I.

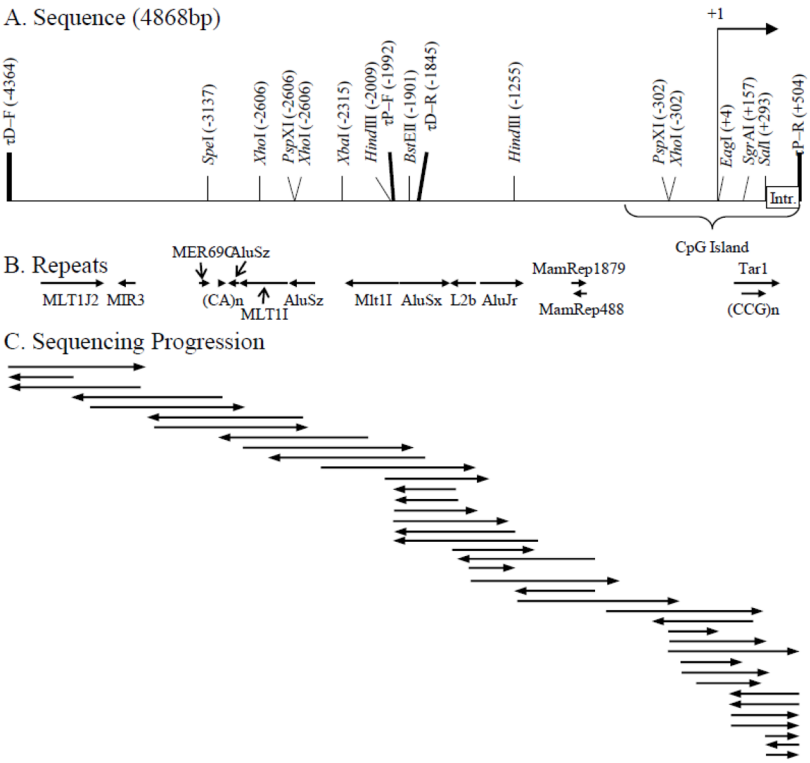


Fig. 2. Sequence diagram of the 4868bp *MAPT* 5′–flanking region fragment, JF412354

A) The 4868 bp sequence, showing +1 transcription start site, locations of PCR primers, and restrictions enzymes used in subsequent subcloning of *CAT* gene reporter expression clones. A partial intron at the 3′–end of the sequence is also shown. A CpG island is indicated. B) Positions of repeats found by RepeatMasker (Chen, 2004) and Censor (Jurka et al., 1996). Specific repeat positions by bp are in Table 4. C) Sequencing “walk”, showing all overlapping progress of sequencing progression for JF412354.

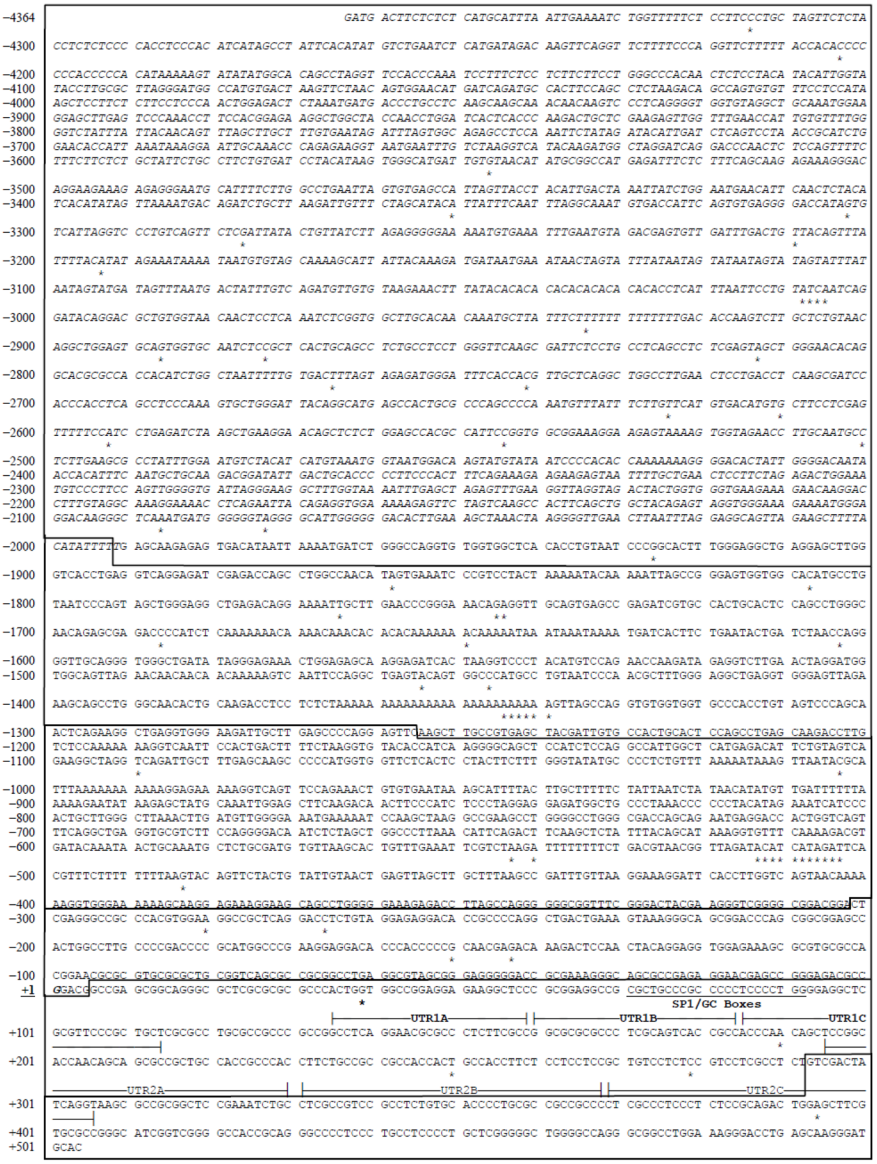


Fig. 3. Sequence of the 4868bp *MAPT* 5′–flanking region fragment, JF412354

Clones of a DNA fragment containing 4868 bp were sequenced in both directions, as described in the text. DNA sequence was aligned with *MAPT* upstream sequences from both H1 and H2 haplotypes. Bases where H2 sequences (including JF412354) disagreed with the majority of H1 sequences have "*" beneath them. The locations of six oligonucleotide pairs (UTR1A–C, UTR2A–C) used for EMSA analysis of nuclear protein–proximal promoter interactions (Table 3, Fig. 6–7) are indicated. Two predicted GC box/SP1 sites in the 5′–UTR region with possible activity as identified by EMSA (Fig. 7) are indicated. The locations of oligomer pairs used for EMSA analysis are indicated. A potential distal neurospecific promoter region at −4364/−1992 (Figs 9–10) is italicized. The +1 TSS base is boldface and italic. Functional domains (Fig. 10) are indicated.

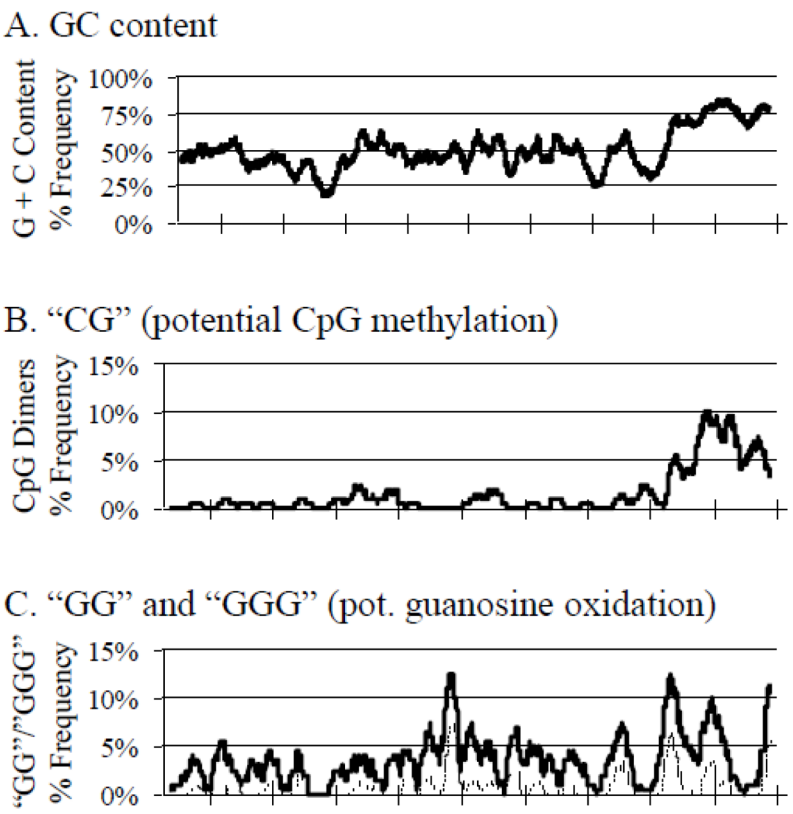


Fig. 4. GC content and structural attributes JF412354

A) Percent GC content of JF412354 (4868 bp) in a 100bp window. B) Percent of sequence dimers in 100bp window that are "CpG" (potential DNA methylation) sites. C) Percent of sequence dimers in 100bp window that are "GG" (DNA "oxidation hotspots"). CpG and GG dimers are relevant in explaining the biological basis of AD etiology as per the previously proposed "LEARn" model of gene–environment interactions behind sporadic disorders (Lahiri et al., 2009).

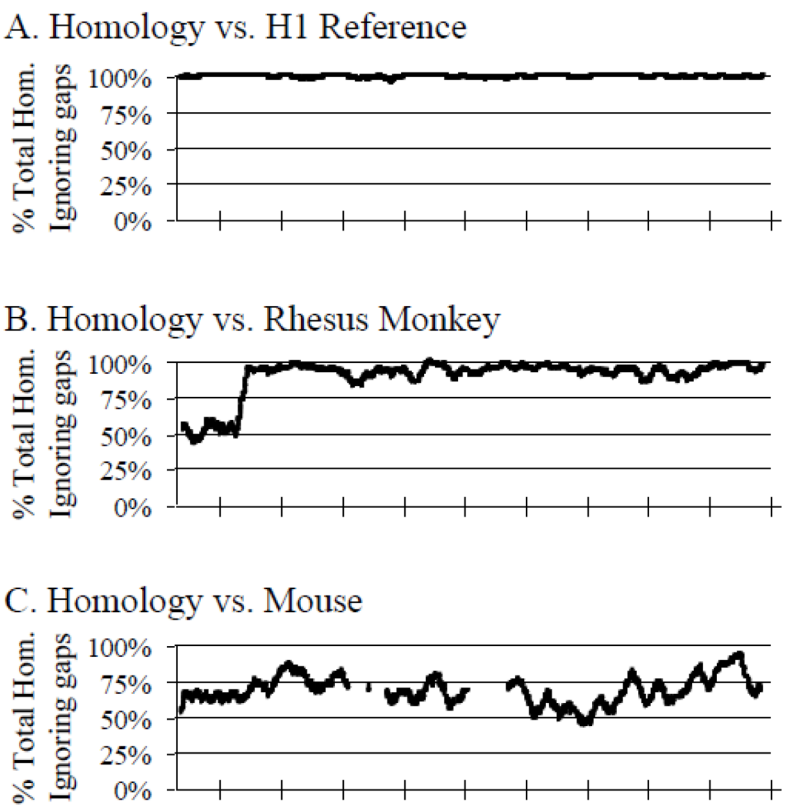


Fig. 5. Sequence homology of JF412354 with selected *MAPT* 5′–flanking regions in different species

DNA sequences from human, mouse, and rhesus monkey genomic reference databases were aligned with JF412354 as described in the text. Homology was calculated, ignoring gaps, in a 100bp window, for each alignment. A) Homology with the human reference *MAPT* genomic sequence (H1 haplotype). B) Homology with the rhesus monkey *MAPT* reference genomic sequence. C) Homology with the mouse *MAPT* reference sequence.

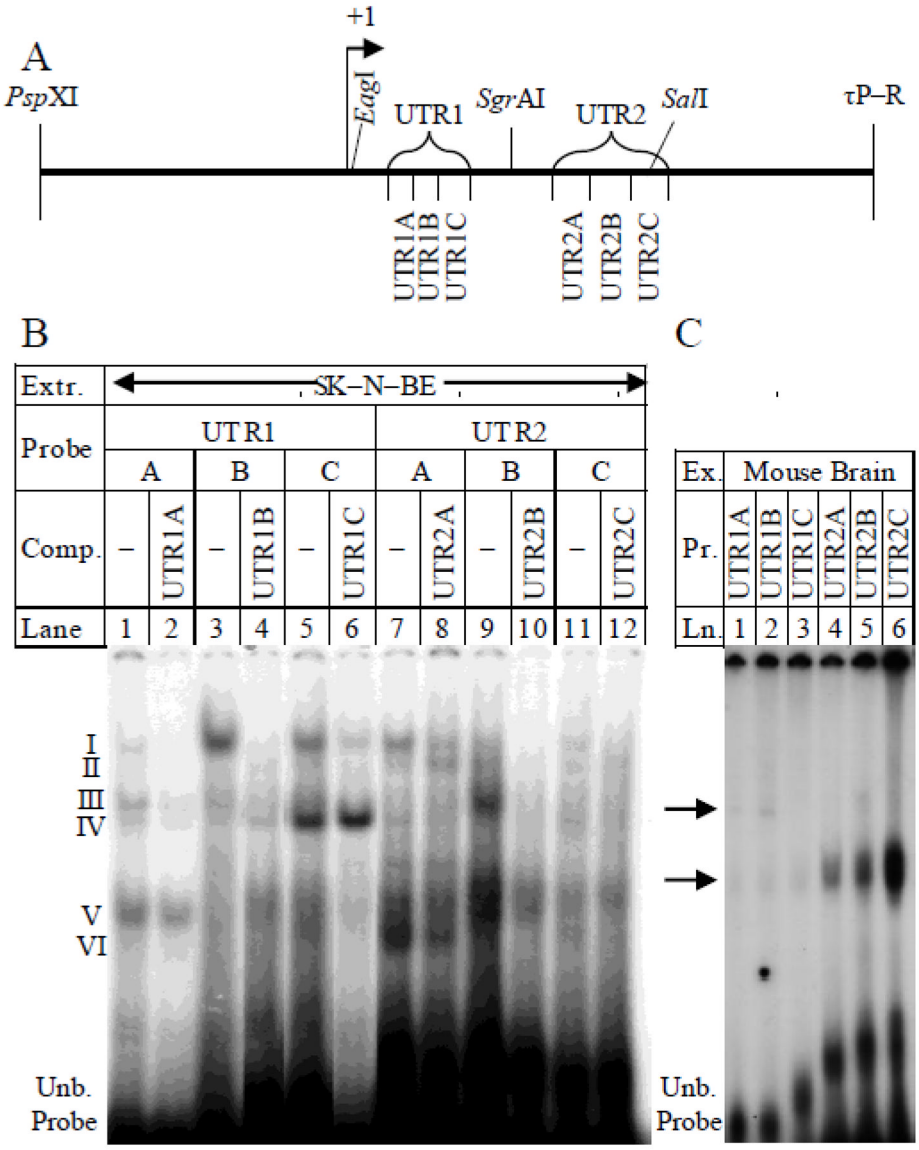


Fig. 6. Electrophoretic mobility shift (EMSA/gel shift) assay of 5′–UTR fragments of JF412354 in human neuroblastoma cell nuclear extracts

A) Oligomer pairs corresponding to positions +35/+59 (UTR1 A), +60/+84 (UTR1B), +85/+114 (UTR1C), +195/+230 (UTR2A), +231/+268 (UTR2B), and +269/+306 (UTR2C) were synthesized according to the JF412354 sequence and independently used as radioactive probes in SK–N–BE (NB) and mouse brain (MB) cell nuclear extract EMSA. B) Autoradiogram of EMSA results in NB cell nuclear extracts. Probes used were UTR1A (lanes 1–2), UTR1B (lanes 3–4), UTR1C (lanes 5–6), UTR2A (lanes 7–8), UTR2B (lanes 9–10), and UTR2C (lanes 11–12). Lanes 4–6 were loaded out of order and have been re–ordered in the figure. Original loading order was 6, 5, 4. All lanes are from a single experiment and gel. Each reaction was incubated without competition (lanes 1,3,5,7,9,11) and with autocompetition against non–radioactive probe (lanes 2,4,6,8,10,12). Major bands are indicated by roman numerals. Autocompetition with UTR1B or with UTR2B resulted in elimination of strong EMSA bands at “I” and “III”/”IV”, respectively. C) EMSA in mouse brain nuclear extracts. Probes used were UTR1A (lane 1), UTR1B (lane 2), UTR1C (lane 3),

UTR2A (lane 4), UTR2B (lane 5), and UTR2C (lane 6). All lanes are from a single experiment and gel. Major bands are indicated by arrows. UTR1B produced a faint but distinct band in MB nuclear extract, while all UTR2 probes produced a, presumably nonspecific, band in MB nuclear extract.

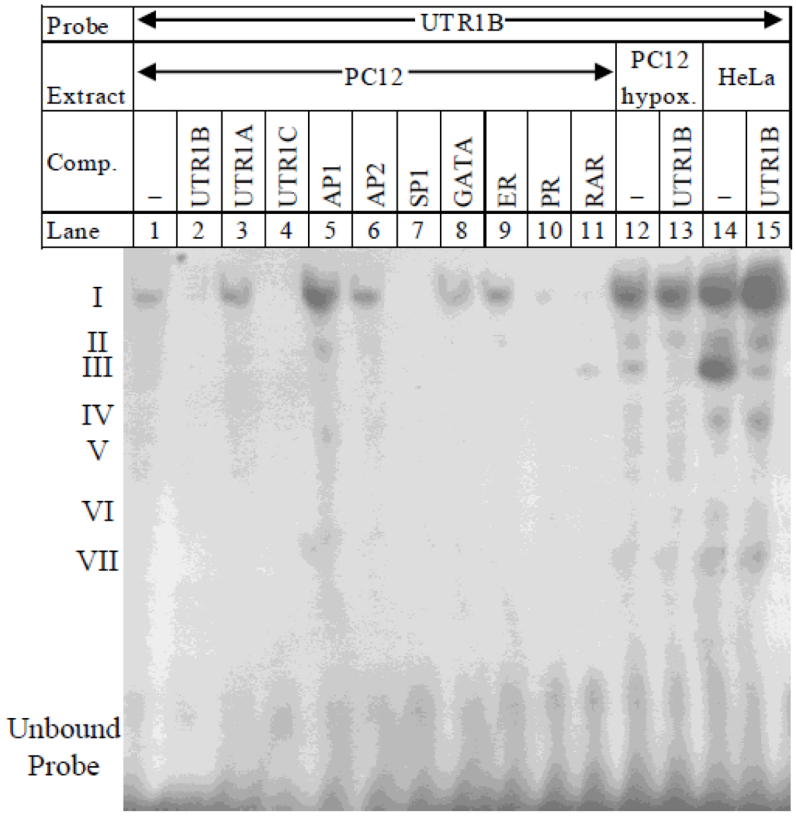


Fig. 7. Competition EMSA of UTR1B oligomer pair vs. specific transcription factor–binding oligomer pairs or in multiple cell nuclear extracts

EMSA was performed with the UTR1B probe in nuclear extracts from PC12 (lanes 1–11), PC12 under hypoxia (lanes 12–13), or HeLa (lanes 14–15). All lanes are from a single experiment and gel. All reactions were run uncompeted (lanes 1,12,14) and autocompeted against unlabeled UTR1B (lanes 2,13,15). In addition, PC12 reactions were run in competition against UTR1A (lane 3), UTR1C (lane 4), and commercial unlabeled oligomer pairs that bound known transcription factors AP1 (lane 5), AP2 (lane 6), SP1 (lane 7), GATA (lane 8), ER (lane 9), PR (lane 10), and RAR (lane 11). In addition to self–competition in PC12 (band I), PC12/hypoxia and HeLa (band III), and U373 (bands I, II, and III), UTR1 and oligomers to SP1, PR, and RAR competed out signal in the PC12–UTR1B reaction.

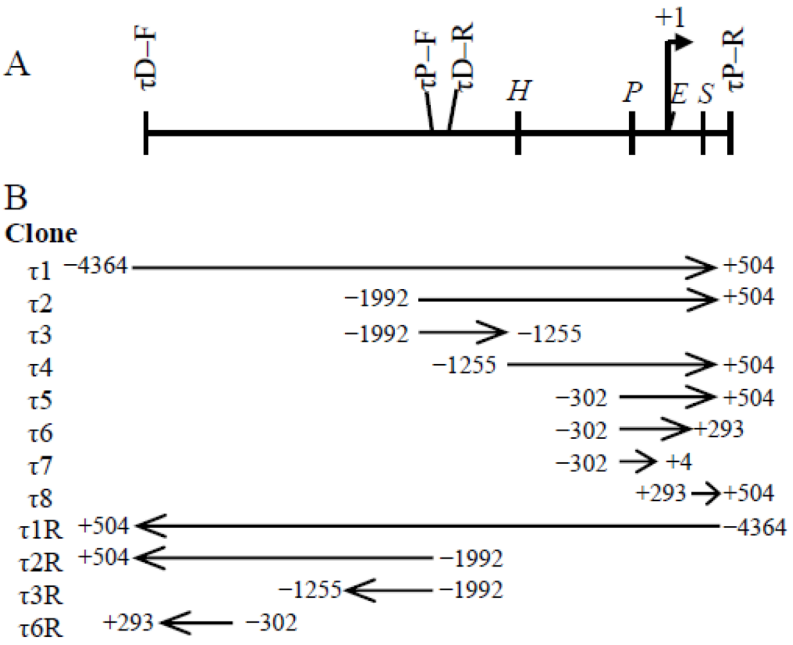


Fig. 8. *MAPT–CAT* reporter vector deletion series cloning

A) Scale diagram of *MAPT* promoter sequence JF412354. B) The *CAT* reporter gene fusion vector pBLCAT3 was cloned to *MAPT* 5′–fragments as described in the text, producing successive forward deletion clones from pτ1 through pτ8. In addition, reverse cloning produced pτ1R, pτ2R, pτ3R, and pτ6R.

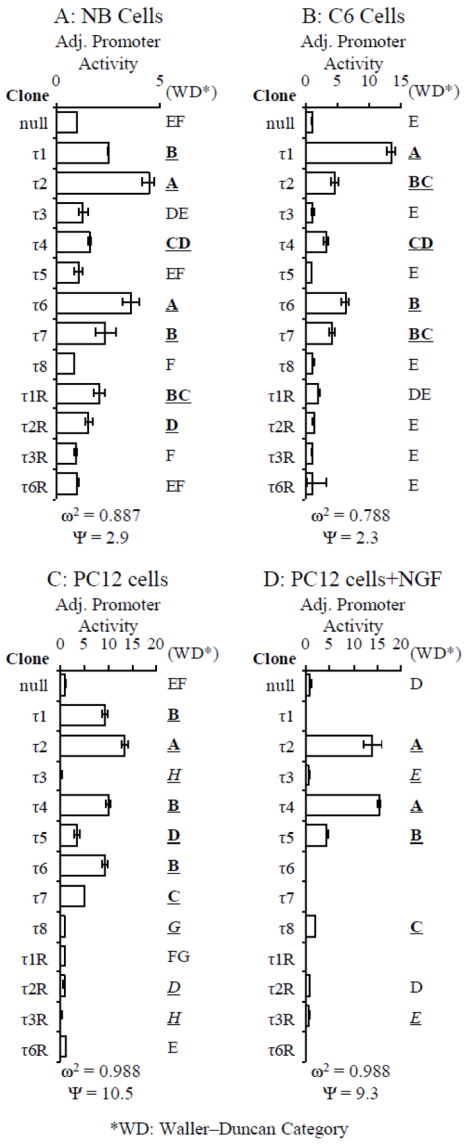
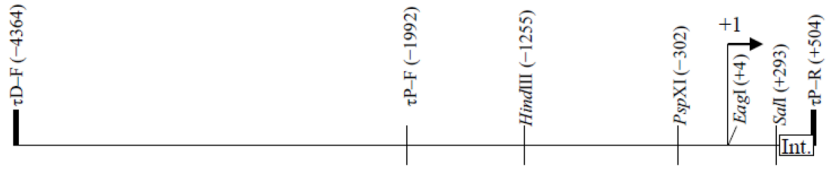


Fig. 9. *MAPT–CAT* reporter functional assay by transient expression of clones in three cell lines

MAPT–CAT fusion reporter clones were transfected into A) NB human neuroblastoma, B) PC12 rat pheochromocytoma, C) PC12 +NGF, and D) rat C6 glioma, along with pSVβGAL as a transfection efficiency control vector. Cells were harvested as described in the text. The CAT reporter protein levels were measured by ELISA and then adjusted for β gal levels and total protein content. Residuals were normalized with Box–Cox transformation and Waller–Duncan multiple range test was used to separate means. Letters to right of each graph indicate Waller–Duncan categories. Clones that share a Waller–Duncan category are not significantly different. Transformed data were back–transformed for presentation purposes. Relative clone expression levels are shown. Effect size estimates (ω^2 for variance and Ψ for standard magnitude) are indicated for each experiment.



Domain	V	IV	III	II	I	0
NB	2.6−	4.6+	1.9+	3.8+	2.0+	5.5−
PC12	3.3−	2.5+	8.1+	9.4+	4.8+	7.4−
PC12 + NGF	not tested	0.4 no effect	5.8+	5.8+		5.0+
C6	3.0+	0.7 no effect	2.0+	2.3+	1.0+	3.5−

Fig. 10. Functional domains of *MAPT* 5′–flanking 4.9kb region

Data from reporter clone transfection studies was used to determine how various segments of the JF412354 sequence influenced CAT reporter protein levels. Assays were assessed separately for A) NB, B) PC12, C) PC12 treated with NGF, and D) C6 cells. Effect size was calculated for selected pairwise comparisons (Table 7). The result was rounded to one decimal place and is represented as "+" or "−" in the figure. NB and PC12 cells were approximately in agreement. Treatment of PC12 cells with NGF reversed the effect of domain 0 (intronic +293/+504) from negative to positive. C6 cells had a similar expression profile to NB and untreated PC12 cells, except for the distal–most 2.5kb (−4364/−1992), which had positive effects upon CAT protein levels in C6 cells, in contrast to unstimulated neuronal cell cultures, in which it had a negative effect. Distal regions of this *MAPT* promoter may have neurospecific activity and the proximal region of the first intron may have further specificity within different neuronal tissues or brain regions.

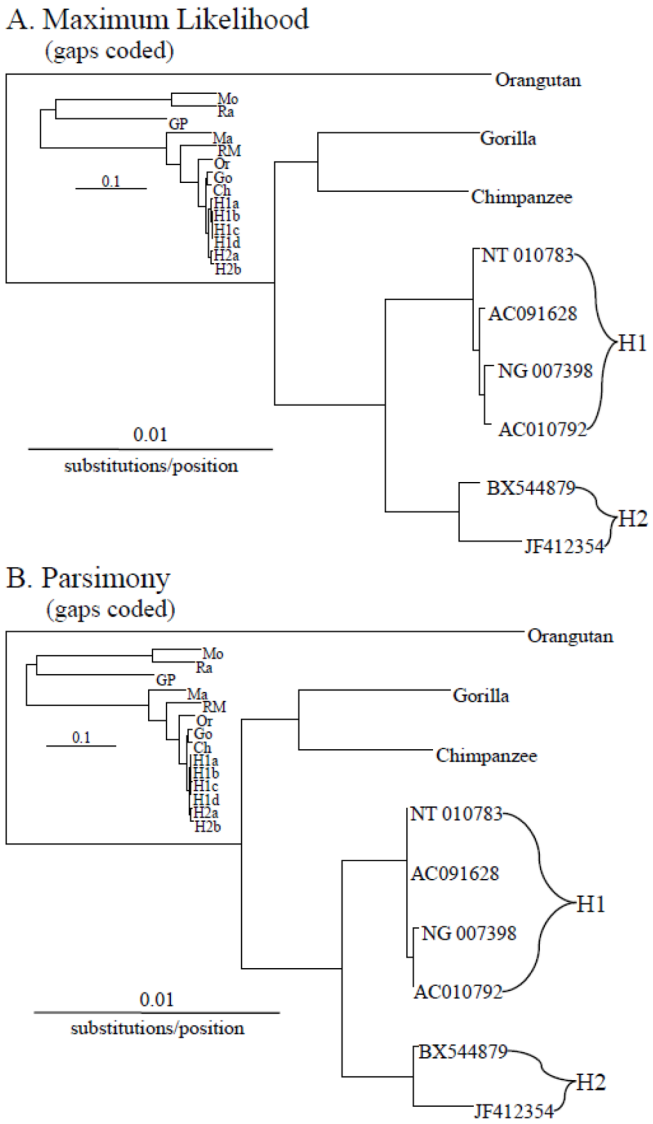


Fig. 11. Phylogeny of H2 vs. H1 *MAPT* haplotype sequences in context of other primates and rodents

Sequences of mouse, rat, marmoset, rhesus monkey, orangutan, gorilla, chimpanzee, our own JF412354 *MAPT* sequence, an additional H2 haplotype *MAPT* promoter sequence, and promoter sequences from three H1 haplotype sources were aligned by MUSCLE and, gaps coded, and phylogenies inferred by A) maximum likelihood (MrBayes) and B) discrete character parsimony (PHYLIP PARS). Non–ape branches are trimmed from larger trees for clarity of illustrating H1/H2 vs. chimpanzee relationship. Of particular note, both phylogenies indicate that the human H2 and H1 sequences, regardless of chromosomal orientation, descend from an intermediate sequence, with neither ancestral to the other.

Table 1

Sequencing Primers

Primer	Sequence	Position[a]	Primer	Sequence	Position[a]
τseq1F	5′– GTTCTCTACCTCTCTCCCCACCT–3′	−4308	τseq1R	5′–ATCCAGGTTGGTAGCCAGCCTTC–3′	−3850
τseq2F	5′– GCTACCAACCTGGATCACTCACC–3′	−3864	τseq2R	5′– CCTCACACTGAATGGTCACATTTGC–3′	−3312
τseq3F	5′–CAAATGTGACCATTCAGTGTGAGG–3′	−3335	τseq3R	5′–TGAACCCAGGAGGCAGAGGCT–3′	−2844
τseq4F	5′–CTCTGCCTCCTGGGTTCAAGCG–3′	−2861	τseq4R	5′–CTGAAAGTGGGAAGGGGGTGC–3′	−2346
τseq5F	5′–GTCCCTTCCAGTTGGGGTGATTAG–3′	−2299	τseq5R	5′–CACCATGTTGGCCAGGCTGGTC–3′	−1860
τseq6F	5′–GGGACACTTGAAAGCTAAACTAAGGG–3′	−2062	τseq6R	5′– ATCTCCTTGCTCTCCAGT–3′	−1554
τseq7F	5′–ACTGGAGAGCAAGGAGAT–3′	−1571	τseq7R	5′– GTCCTCATTCTGCTGGTC–3′	−712
τseq8F	5′–GACCAGCAGAATGAGGAC–3′	−729	τseq8R	5′– CTCTCCTACAGAAGTCCT–3′	−255
τseq9F	5′–CTCAGGACTTCTGTAGGA–3′	−275	τseq9R	5′–GGTGGCTGTGGTGGCG–3′	256
τseq10F	5′–CGCCACCACAGCCACC–3′	241	M13F	5′–GTAAAACGACGGCCAGT–3′	*b*
			M13R	5′–GCGGATAACAATTTCACACAGG–3′	*b*

[a]Position is 5′–most base, reading downstream on *MAPT* sequence for “F” primers and 3′–most base, reading upstream on *MAPT* sequence for “R” primers. Positions are numbered relative to the +1 transcription start site.

[b]These primers are on plasmid vectors flanking the *MAPT* inserts.

Table 2

Database *MAPT* Sequences from GenBank and Ensembl.

Accession[a]	Clone/Name	Haplotype[b]	Notes
AC010792	2524N8	H1	
NT_010783	Chr. 17 genomic contig, GRCh37.p2	H1	
AC217771	RP11–105N13	H1	incorp. into NT_010783
CR936218	RP11–669E14	H1	incorp. into NT_010783
CR932417	XX–D_24I13	H1	incorp. into NT_010783
AC091628	chromosome 17 map 17q21.1	H1	
AC003662	CTB–104N19	H1	incorp. into NG_007398
NG_007398	*MAPT*, RefSeqGene	H1	
BX544879	RP11–769P22	H2	
NT_165773	mouse chr 11 genomic contig, MGSCv37 C57BL/6J		
NW_047340	rat chr 10 genomic scaffold, RGSC_v3.4		
NW_003183508	marmoset chr 5 genomic scaffold, Callithrix jacchus–3.2		
NW_001102974	rhesus chr 16 genomic scaffold, Mmul_051212		
NW_002888732	orangutan chr 17 genomic scaffold, P_pygmaeus_2.0.2		
gorGor3:5[c]	gorilla chromosome 5		
NW_001226917	chimpanzee chr 17 genomic contig, reference assembly (based on Pan_troglodytes–2.1)		

[a] Italic indicates human sequence.

[b] Haplotype only applies to human sequences.

[c] Gorilla sequence is from Ensembl.

Table 3

Oligomers for EMSA of JF412354 UTR sub–regions.

Oligo	Position[a]	Sequence	Oligo	Position	Sequence
UTR1 A–F	+34/+58	5′–ACTGGTGGCCGGAGGAGAAGGCTCC–3′	UTR2 A–F	+195/+230	5′–TCCGGCACCAACAGCAGCGCCGCTGCCACCGCCCAC–3′
UTR1 A–R		5′–GGAGCCTTCTCCTCCGGCCACCAGT–3′	UTR2 A–R		5′–GTGGGCGGTGGCAGCGGCGCTGCTGTTGGTGCCGGA–3′
UTR1 B–F	+59/+84	5′–CGCGGAGGCCGCGCTGCCCGCCCCC–3′	UTR2 B–F	+231/+268	5′–CTTCTGCCGCCGCCACCACTGCCACCTTCTCCTCCTCC–3′
UTR1 B–R		5′–GGGGGCGGGCAGCGCGGCCTCCGCG–3′	UTR2 B–R		5′–GGAGGAGGAGAAGGTGGCAGTGGTGGCGGCGGCAGAAG–3′
UTR1 C–F	+85/+114	5′–TCCCCTGGGGAGGCTCGCGTTCCCGCTGCT–3′	UTR2 C–F	+299/+336	5′–GCTGTCCTCTCCGTCCTCGCCTCTGTCGACTATCAGGT–3′
UTR1 C–R		5′–AGCAGCGGGAACGCGAGCCTCCCCAGGGGA–3′	UTR2 C–R		5′–ACCTGATAGTCGACAGAGGCGAGGACGGAGAGGACAGC–3′

[a]Position is for double–stranded oligomer pair, relative to +1 TSS of *MAPT*.

Table 4

Repeat Sequences in JF412354.

Start[a]	End	Repeat	Orientation[b]
−4166	−3780	MLT1J2	forward
−3690	−3580	MIR3	reverse
−3195	−3126	MER59C	forward
−3047	−3026	(CA)n	forward
−3012	−2944	MLT1I	reverse
−2943	−2644	AluSz	reverse
−2643	−2480	MLT1I	reverse
−2292	−1960	L2b	reverse
−1959	−1650	AluSx	forward
−1649	−1494	L2b	reverse
−1463	−1189	AluJr	forward
−902	−807	MamRep1879	reverse
−890	−807	MamRep488	forward
+101	+386	TAR1	forward
+147	+296	(CCG)n	forward

[a] Start and end positions are relative to the +1 TSS.

[b] Orientation is relative to the *MAPT* reading frame.

Table 5

Transcription Factors that Differ between JF412354 and Majority of H1 Sequence.

TF Family	Factor	Position[a]
cAMP–responsive element binding proteins	NFIL3	*−3699*, −3698, **−3557,**[b] *−3205*, −3071, −2477, *−2010*[c] 295
	XBP1	−612, **−296**
CCAAT/Enhancer Binding Protein	C/EBPα	**−3555,** −3069, *−2961*, −1997, −1084, *−616*
	C/EBPβ	−2890, *−2753*, [−1864]
Cellular and viral myb–like transcriptional regulators	c–myb	−3715, *−2022*, *−1503*, −595, *−529*, **185**
	v–myb	[−3359], *−3276*, *−3219*, −529, −468
E–box binding factors	ATF6	*33*, [180]
	USF	*−4082*, *−2626*, −1902, [−292]
Estrogen response elements	ER	**−1098, 169,** 458
Hypoxia inducible factor, bHLH/PAS protein family	HIF1	**−295**
MAF and AP1 related factors	NRL	**−2838**
Nuclear factor κ B/c–rel	HIVEP1	*−2307*, [−1775]
Peroxisome proliferator–activated receptor	PPAR/RXR	*−3421*, [−1761]
Retinoic acid receptor/RXR heterodimer binding sites	RAR/RXR	−3661, [−2889], *−2728*, −2077, *−202*
v–ERB and RAR–related orphan receptor α	REv–ERBα	*−4277*, *−4084*, **−3311**
	RORα	**−1101**
	RORγ	*−3445*, [−1622]

[a] Position is relative to +1 TSS of JF412354 sequence, H1 *MAPT* sequences aligned to this for comparison.

[b] Boldface indicates a specific site appears in JF412354 (H2) but not the majority of H1 sequences.

[c] Italic indicates a reverse–orientation site.

[d] Square brackets indicate a site that appears in the majority of H1 sequences but not in JF412354.

Table 6

Transcription Factor Sites Shared between JF412354 and Majority H1 Sequences.

TF Family	Factor	Position[a]
Activator protein 1	AP1	*−3853*, −464
Activator protein 2	AP2	*−1140*,[b] −350, −251, *−69*
	AP2α	−3934, *−3934*, −371, *−371*
	AP2β	*84*
	AP2ε	−747, *−747*, *−251*, −69
Aβ interacting domain	AβID	*−2352*, −2069
Calsenilin/presenilin binding protein	CSEN	*−1181*
cAMP–responsive element binding proteins	CREB	−540
	CREB2	−1986, −537
CCAAT/Enhancer Binding Protein	C/EBP	*−4215*, *−3793*, −2392
E–box binding factors	c–myc	−293, *−292*
	MAX	−609
	MYCMAX	−4083, −2803, *−2619*, *−184*
EGR/nerve growth factor induced protein C & related factors	CKROX	*−4293*, −58, *78*, *369*
	EGR1	*−4210*, *−4202*, −348, *−162*, *74*, *162*, *237*
	EGR3	*−297*, −96
	WT1	*−4208*, *−4206*, *−4200*, *−166*, −56, *76*
Estrogen response elements	ERR	*−2742*, *−210*
	ESRRα	*−1220*
	ESRRβ	−3651, −1194, −983
GATA binding factors	GATA1	*−3432*, *−3271*, −3153, −1528
	GATA3	−3156, *−2591*, −2590, −1616
GC–Box related factors	GC Box	*−4205*, *77*, *429*
	SP1	*−2654*, −323, −318, *−256*, *119*, *217*, *357*, 452
	SP2	−2071
	SP4	*−2560*, −345, *−300*, −66, *71*
	TIEG	*−4213*, *−4211*, *−3037*, *−3035*, *−165*
Glucocorticoid responsive and related elements	ARE	*−3507*
	GRE	*−1494*
Heat shock factors	HSF1	*−3291*
	HSF2	*−3534*, −3430, *−1179*, 38
Hypoxia inducible factor, bHLH/PAS protein family	ARNT	*−2749*, *−507*

TF Family	Factor	Position[a]
	HRE	*−294*
NeuroD, β2, HLH domain	NEUROD1	*−3640*, *−843*
	NEUROG	−4083
	NGN/NEUROD	*−3074*, −2791
Nuclear factor κ–B/c rel	NFκB	−2756, *−1861*
p53 tumor suppressor	p53	−4256, −3951, *−3560*, −1554, *−1543*, −209
Peroxisome proliferator–activated receptor	PPARα	*−2013*, −1533
	PPARγ	*−2207*, −2202, −1203, −992, −240, −49
Retinoic acid receptor/RXR heterodimer binding sites	CAR/RXR	−3806, −3651, *−3486*, −2226, *−1913*, *−1190*
	LXRE	−2595, *−2267*
	PXR/RXR	*−3471*
	THR	−1907
	VDR/RXR	*−3587*, −1893, −993, *−838*, *267*, *367*, *427*
v–ERB and RAR–related orphan receptor α	RORα1	−1197, −986
	RORα2	*−3455*, *−1989*, −423
	v–ERBα	−3654, *−2725*, −1900
Serum response element binding factor	SRF	−3713, −3675, *−3549*, *−2795*, −2794, *−2496*, −2495, −1589
Signal transducer and activator of transcription	STAT1	*−2357*, −975
	STAT3	−4139, −1472, −686, *467*

[a]Position is relative to +1 TSS of JF412354 sequence.

[b]Italic indicates a reverse–orientation site.

Table 7

Standardized Effect Sizes for Selected Pairwise Clone Comparisons in Deletion Series of JF412354

Segment	Position[a]	Clone Comparison	Unbiased d^b			
			NB[d]	PC12[c]		C6[f]
				−NGF[e]	+NGF	
0	+293/+504 (intron)	τ5 vs. τ6	5.5	7.4	5.0	3.5
I	+4/+293 (5′–UTR)	τ6 vs. τ7	2.0	4.8	5.8[g]	1.0
II	−302/+4	pBLCAT vs. τ7	3.8	9.4		2.3
III	−1255/−302	τ4 vs. τ5	1.9	8.1	5.8	2.0
IV	−1992/−1255	τ2 vs. τ4	4.6	2.5	0.4[h]	0.7[h]
V	−4364/−1992	τ1 vs. τ2	2.6	3.3	nc	3.0

[a] Relative to +1 transcription start site.

[b] Lack of statistically significant promoter/repressor activity is indicated by “nsa”. No clone corresponded to domain V for PC12 treated with NGF, indicated by “nc”.

[c] Rat neuronal cells.

[d] Human neuroblastoma cells.

[e] Neurite growth factor.

[f] Rat glioblastoma cells.

[g] A single clone combining domains I and II was tested for PC12 + NGF. The *d* estimated is for the combined −302/+293 fragment.

[h] These domains were also not significant at the k = 100 Waller–Duncan cutoff.

Table 8

Count of Shared vs. Unique Differences between Chimpanzee and Human H1 or H2 *MAPT* Promoter Sequences.

Category	H1	H2
Shared Single Differences	65[a]	65
Differences Unique by Haplotype	18	28
Total within Each Haplotype	83	93

[a]Counts are of individual differences, not percentages, with a substitution counting as 1 and a single indel of any length counting as 1. Counts were analyzed by Fisher's exact test, with no significant difference (p = 0.231).