Supplementary Tables

Table S1. Sample sizes and age ranges in each Late-onset Alzheimer' disease (LOAD) case-control study of Alzheimer's Disease Sequencing Project (ADSP)

Consortium	Study	Cases		Controls	
		n	Age range	n	Age range
ADGC		4,966	40-99+	3,209	42-99+
	ACT	273	69-89	996	68-89
	ADC	2,417	60-90+	839	64-90+
	CHAP	27	68-90+	204	78-90+
	EFIGA	160	59-90+	171	42-90+
	GDF	111	59-90+	96	77-90+
	NIA-LOAD	364	37-90+	111	78-90+
	MAP	132	71-90+	283	72-90+
	MAYO	250	60-87	99	78-90+
	MAYO PD	181	59-89	14	79-90+
	MIA	316	56-88	15	78-89
	MIRAGE	0	-	20	74-90+
	NCRAD	160	58-90+	0	-
	RAS	46	56-88	0	-
	ROS	144	63-90+	207	67-90+
	TARCC	132	60-90+	12	80-89
	TOR	9	40-84	0	-
	VAN	210	60-90+	26	79-90+
	WHICAP	34	73-90+	116	78-90+
CHARGE		805	60-99+	1,927	61-99+
	ARIC	39	67-89	18	77-85
	ASPS	121	60-89	5	78-86
	CHS	251	68-90+	583	76-90+
	ERF	45	60-88	0	-
	FHS	126	65-90+	455	61-90+
	RS	223	61-90+	866	76-90+

ADGC = Alzheimer's Disease Genetics Consortium; CHARGE = Cohorts for Heart and Aging Research in Genomic Epidemiology; ACT =Adult Changes in Thought; ADC = NIA Alzheimer Disease Centers; CHAP = Chicago Health and Aging Project; EFIGA = Estudio Familiar de la Influencia Genetica en Alzheimer; GDF = Genetic Differences; NIA-LOAD = National Institute on Aging (NIA) Late Onset Alzheimer's Disease Family Study; MAP = Memory and Aging Project; MAYO = Mayo Clinic; MAYO PD = Mayo PD; MIA = University of Miami; MIRAGE = Multi-Institutional Research in Alzheimer's Genetic Epidemiology; NCRAD = National Cell Repository for Alzheimer's Disease; RAS = University of Washington Families; ROS = Religious Orders Study; TARCC = Texas Alzheimer's Research and Care Consortium; TOR = University of Toronto; VAN = Vanderbilt University; WHICAP = Washington Heights-Inwood Columbia Aging Project; ARIC = Atherosclerosis Risk in Communities Study; ASPS = Austrian Stroke Prevention Study; CHS = Cardiovascular Health Study; ERF = Erasmus Rucphen Family; FHS = Framingham Heart Study; RS = Rotterdam Study

Table S2. Variant level-filtering results for both the Broad Institute (BROAD) and the Baylor College of Medicine (BAYLOR) pipelines

Filter value	# of variants
Single nucleotide variants	1,586,687
BOTH_passed	1,431,108
BROAD_passed_BAYLOR_uncalled	69,945
BROAD_passed_BAYLOR_failed	62,422
BROAD_failed_BAYLOR_passed	12,807
BAYLOR_passed_BROAD_uncalled	10,405
Insertions/deletions	49,244

Table S3. Plasmid concentrations used for transfection and co-immunoprecipitation experiments

Transfection combination	pCMV-Empty vector	pCMV-AP2A2	pCMV-Tau
1	3.0 µg		
2	1.5 µg	1.5 µg	
3	1.5 µg		1.5 µg
4		1.5 µg	1.5 µg

pCMV6-XL5 empty vector; pCMV6-XL5-MYC/DDK[FLAG]-AP2A2, and pCMV6-XL5-Tau (no tag) were all sourced from Origene. Transfections were performed using Lipofectamine 3000 (Thermo Fisher) according to manufacturer's protocol. Shown in the table are the amount of each plasmid used for the 4 sets of transfections.

Supplementary Figures

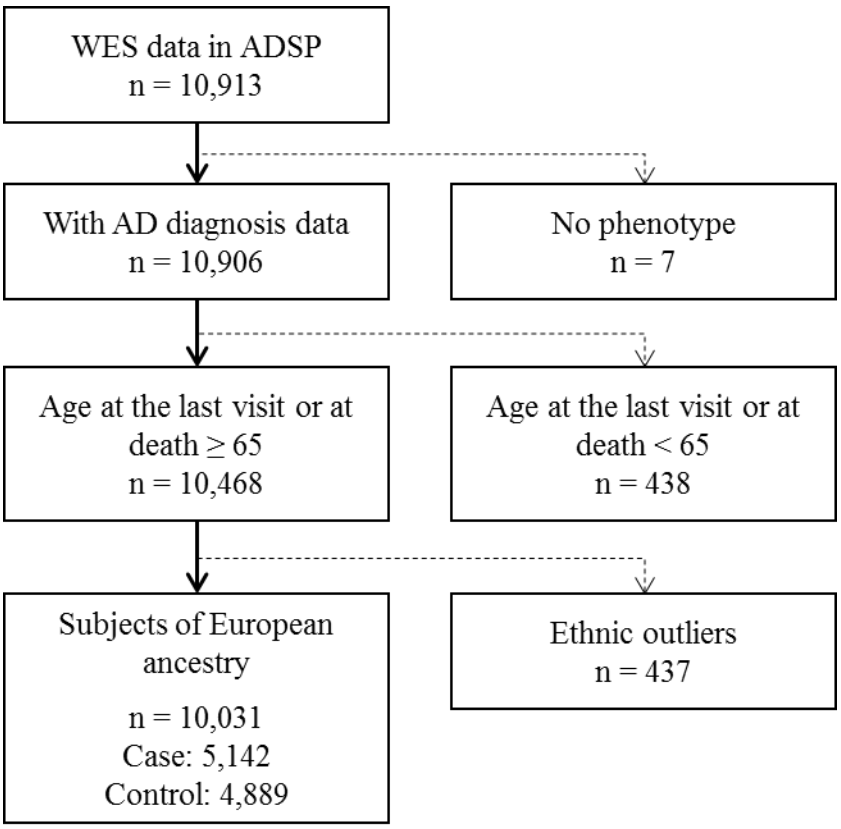


Figure. S1. Flow diagram of the subjects in the ADSP WES analyses, with sample sizes and exclusion criteria. WES = whole-exome sequencing; ADSP = Alzheimer's Disease Sequencing Project

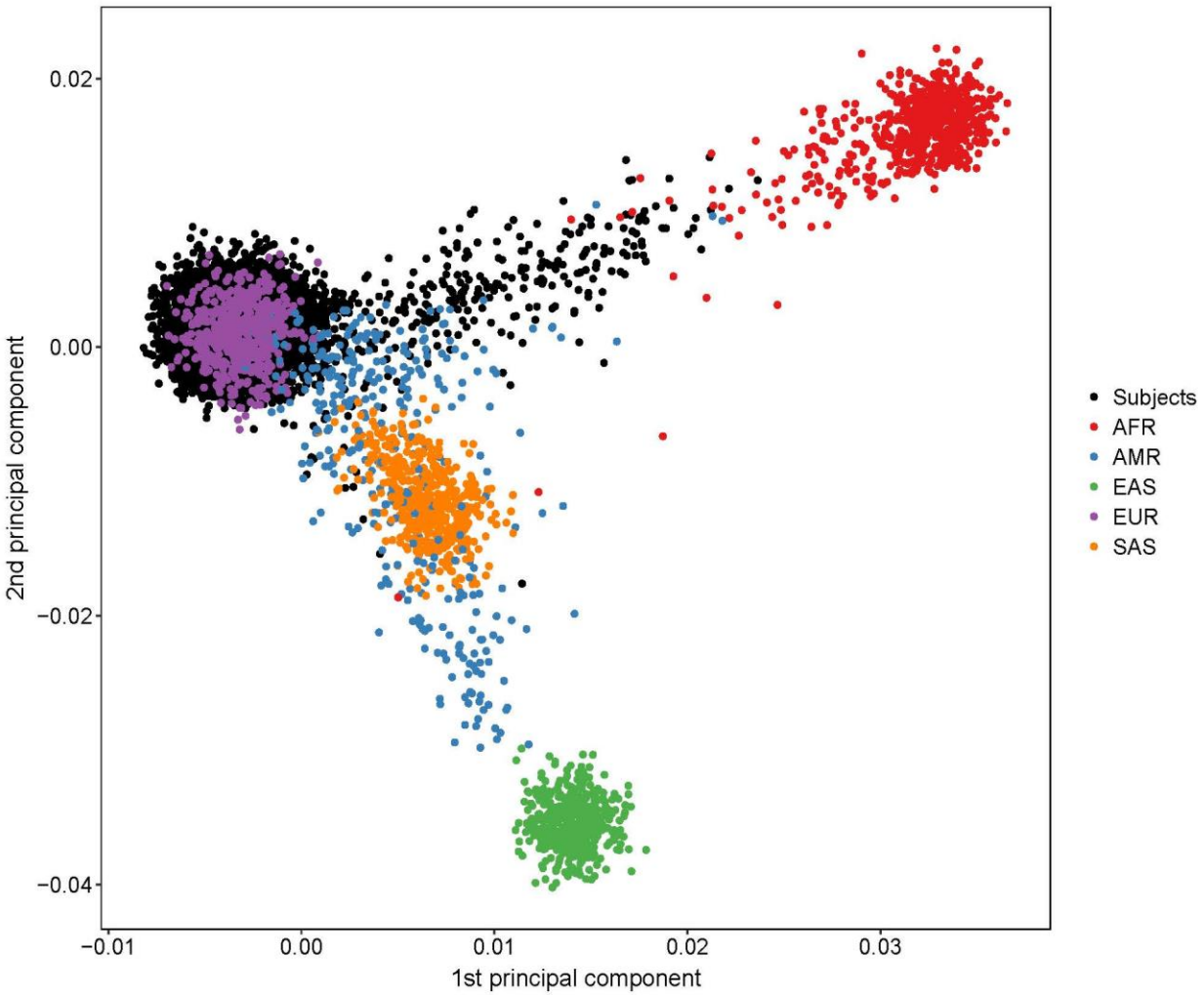


Figure. S2. First and second principal components plots along with 1000 genome reference samples. Block dots indicate individuals in this study. We analyzed data among individuals within the red dashed circle based on Euclidean distance. AFR = African; AMR = Admixed American; EAS = East Asian; EUR = European; SAS = South Asian

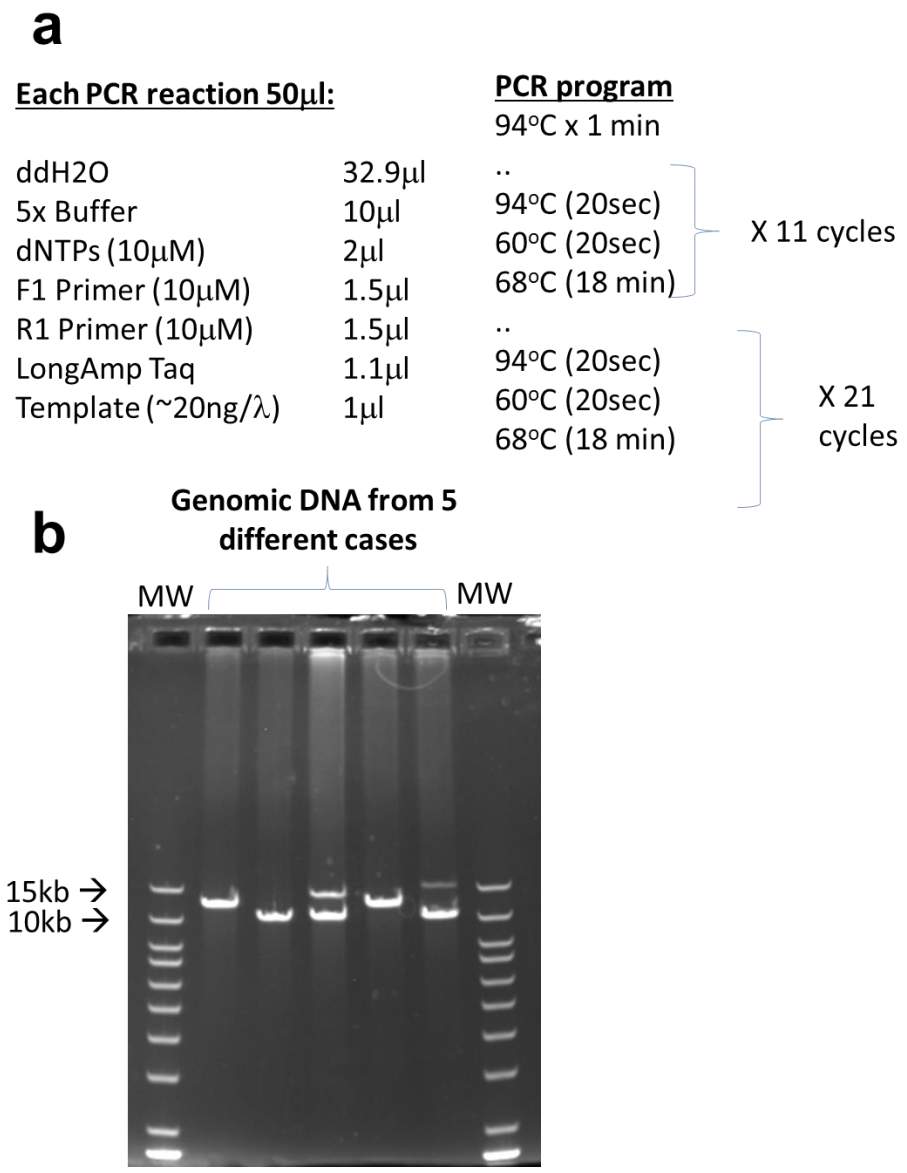


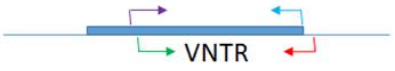
Figure. S3. The Replication cohort PCRs were performed using 0.8% agarose gels and 1kb Plus molecular weight (MW) markers. Since some of the DNA (particularly that from blood samples) was of lower quality (in comparison with the Discovery cohort, where all DNA samples were isolated from brain), we experimented with other long-range DNA polymerases, including the LongAmp Taq Polymerase (NEB). The protocol that we used (including the PCR program) for this polymerase is shown (**a**), along with representative results (**b**). Note that this assay, with lower % agarose gel and larger MW (up to 15kb) marker, conveyed somewhat larger amplicons than the prior (**Fig. 2**) assay. All of the Replication cohort samples were run using this larger MW marker and lower % agarose gel. Some of the Replication cohort samples never generated an amplicon. These DNA samples were not included and were not among the n=167 sample size. All the results were scored and/or reviewed by a worker blinded to the clinical and pathologic information.

Long range PCR using different primer pairs that recognize the *MUC6* VNTR

a

Non-overlapping primers

name	seq	size	Tm
m612-F1	CAGGTGAGATGGAGACAATGG	21	60.5
m612-F5	ATCGACCAATCAGGAACTGC	20	60.1
m612-R1	TGGTTGCAGAACTCAAGTGG	20	59.9
m612-R4	CCAAGACGAGGAGGATATGAAG	22	60.1



b

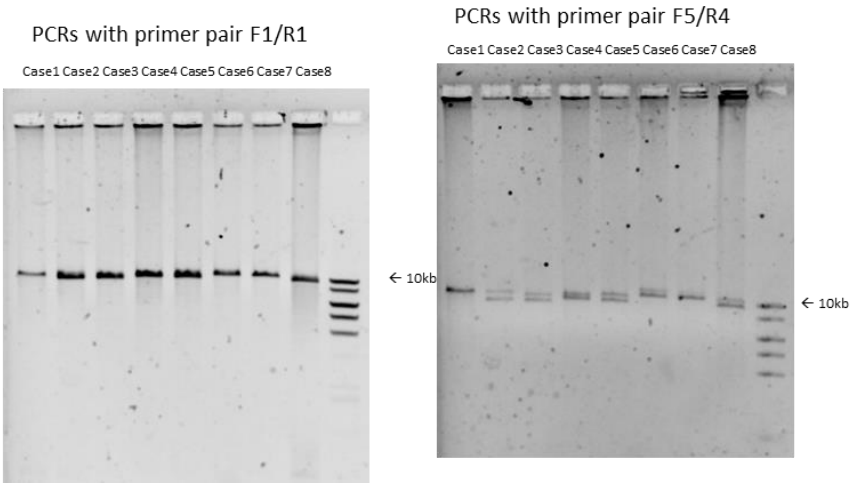
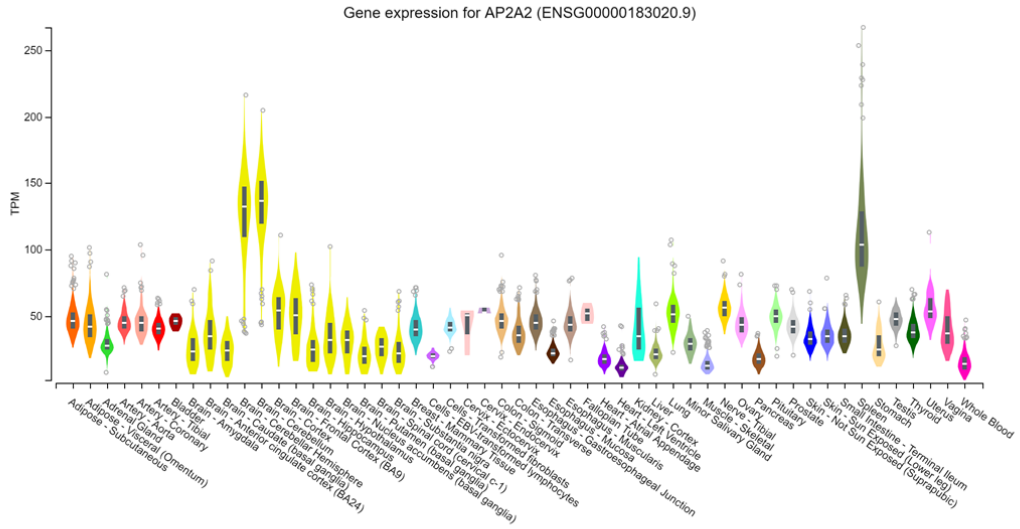


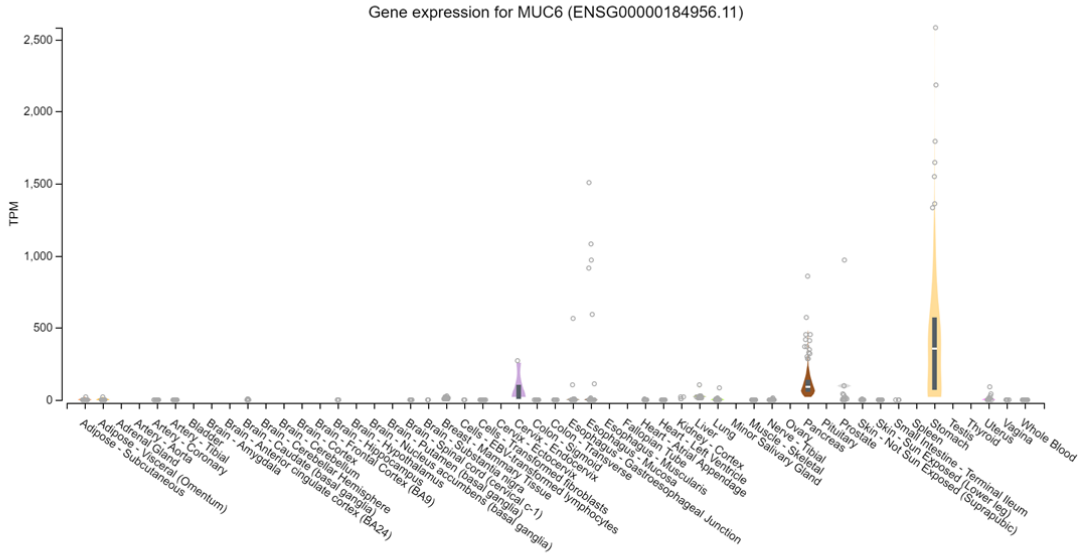
Figure. S4. To help verify that the PCR amplicons were targeting the correct genomic region (the *MUC6* VNTR region), an additional set of separate nested primers was used. The sequences of the 2 sets of primers, and their location and orientation as indicated by color coding, are depicted in panel (**a**). Panel (**b**) shows the results of PCRs performed using the separate primer pairs (F1/R1 on the left, F5/R4 on the right) on a subsample of individuals' DNA. As expected, PCR with these different primers showed the same sized amplicons.

a Human tissue gene expression: *AP2A2*



Source https://gtexportal.org/home/gene/AP2A2

b Human tissue gene expression: *MUC6*



Source https://gtexportal.org/home/gene/MUC6

Figure. S5. Screenshots from the Genotype-Tissue Expression (GTEx) data set depicting tissue-specific gene expression patterns for *AP2A2* (a) and *MUC6* (b). Note that *AP2A2* is expressed at high levels in the human cerebellum, but also in other brain regions. By contrast, *MUC6* is expressed in epithelial tissues but not appreciably the central nervous system.

Figure. S6. Manhattan plot of gene-based analysis of Alzheimer's Disease Sequencing Project (ADSP) whole exome sequencing (WES) data. Optimal unified sequence kernel association test (SKAT-O) association results are depicted for n = 10,031 (5,142 AD cases and 4,889 controls) subjects. The gene with lowest p-value was Mucin 6 (*MUC6*) and the second lowest p-value was Triggering receptor expressed on myeloid cells 2 (*TREM2*).

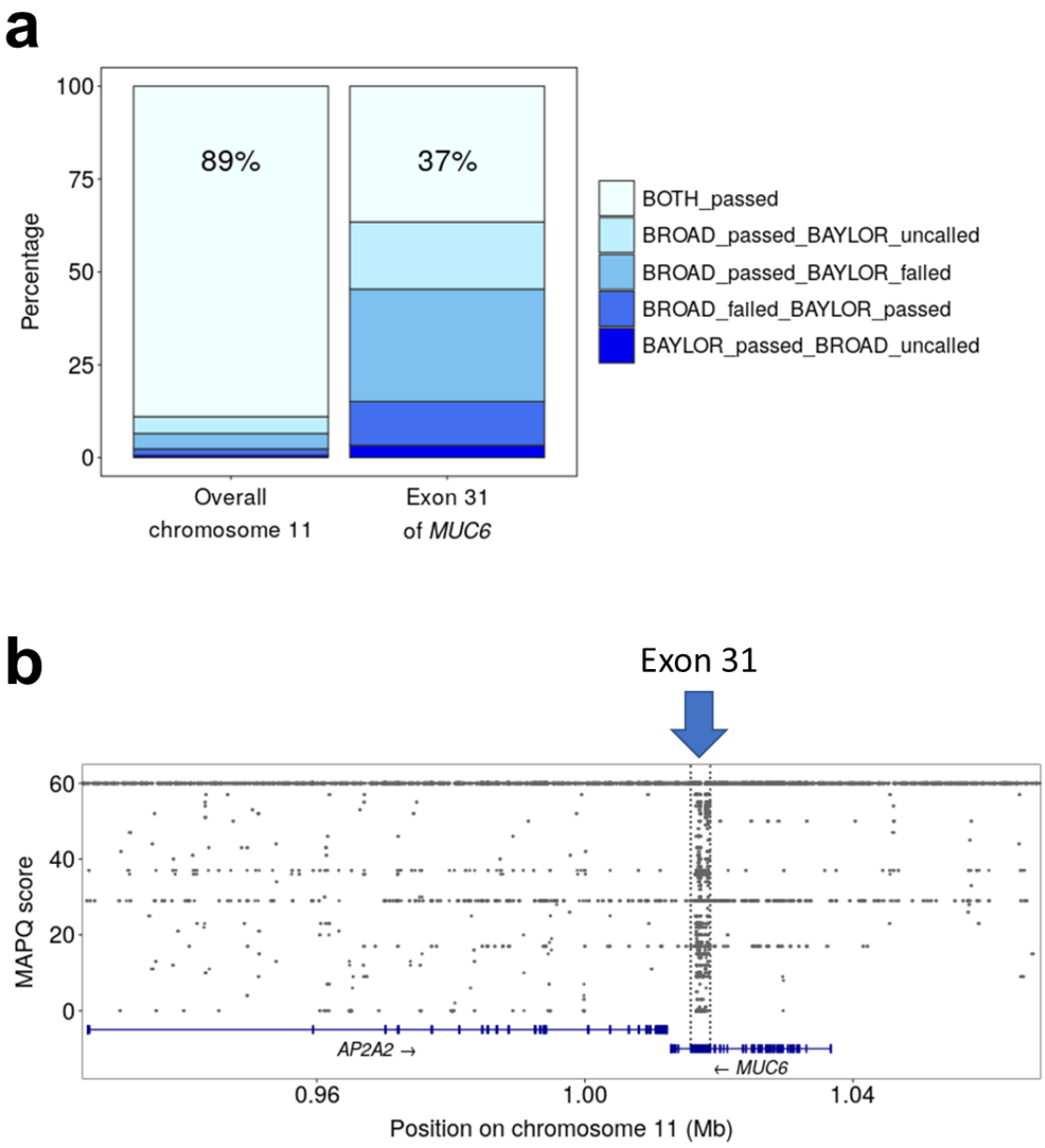


Figure. S7. Percentage of single nucleotide variants (SNVs) that passed quality filters in Baylor University (BAYLOR) and Broad Institute (BROAD) analyses (a) and mapping quality (MAPQ) score in the *AP2A2*/*MUC6* genomic region (b). The percentages were calculated from variant level-filtering results for the BROAD and BAYLOR pipelines contained in consensus VCF file of Alzheimer's Disease Sequencing Project (ADSP) whole exome sequencing (WES) data (**a**). The MAPQ scores were obtained from exome alignment BAM file for subject ID = NA06984 (Utah Residents with Northern and Western European Ancestry) in 1000 Genomes Project Phase 3 (**b**).

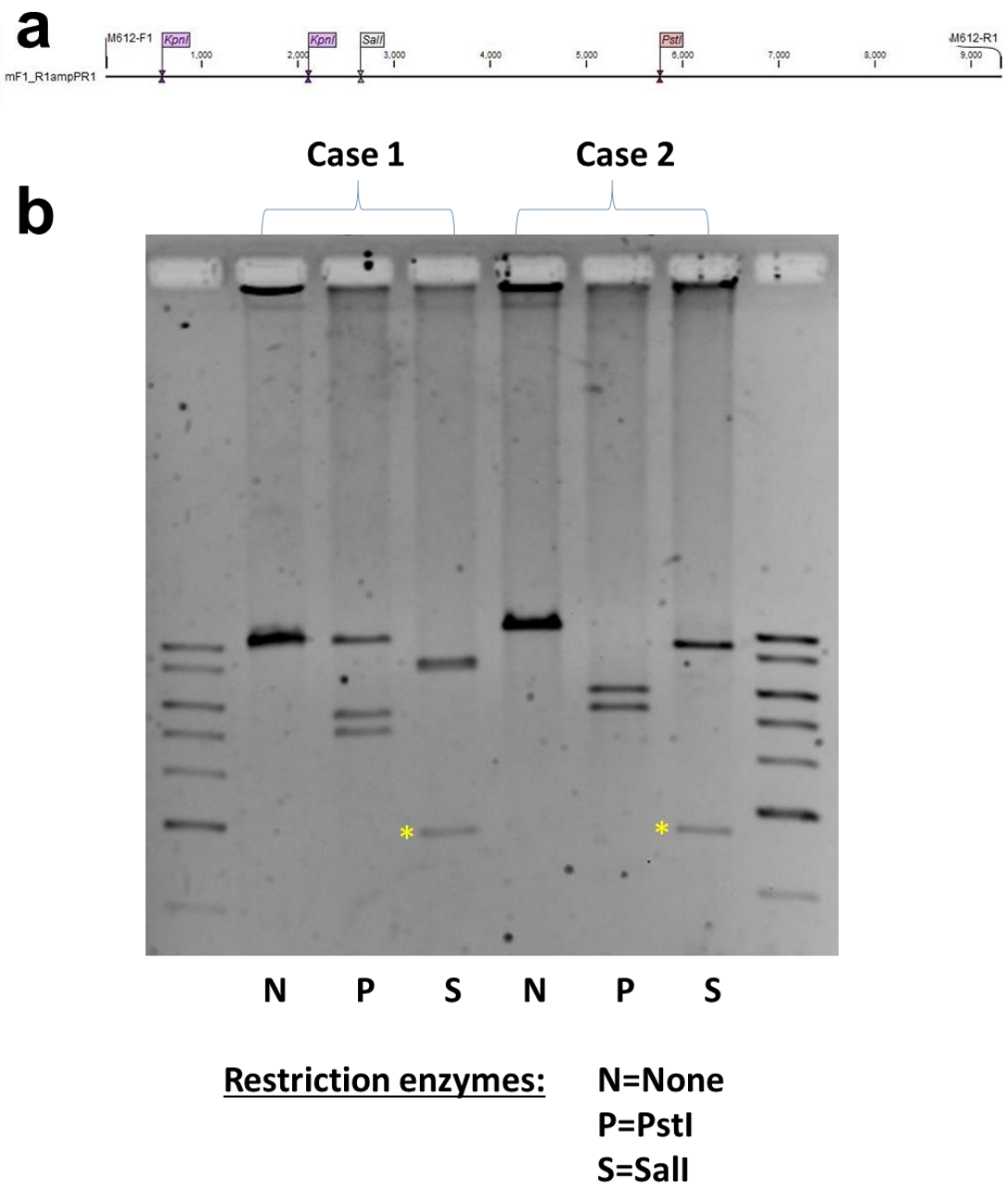


Figure. S8. DNA resolving agarose gels preparations from two representative cases showing the results of digesting the PCR amplicons using restriction enzymes PstI and SalI. For both of the DNA samples, the DNA was digested with either no enzyme (N), PstI (P) and SalI (S). The predicted sites of the enzyme-mediated cleavage are shown in panel (**a**). In panel (**b**) are the gels. Note the smaller fragment generated by the SalI enzyme (yellow asterisks), which were cloned into plasmids and sequenced.

Direct sequence results of digested amplicons (cloned for sequencing) from two cases

>1278-5'. Small Sal1 fragment of Large MUC6 case amplicon
CAGGTGAGATGGANACAATGGGGCAGGCTGGAGTGCCCAGCAGGGGCCATGTCACAGGAA
CAGAGGCACAGACAGGCAAGAAAAAGGTCACATAGACAAAAGGACGGGCCAGCGGAGGTC
AGGGTAGAGAAACAAAAACAATAACGATGACAACTTCACCAATTCCCACAGCAAAATCGA
CCAATCAGGAACTGCCAGGAACAACGGCCACCCAGACGACAGGCCCACGTCCAACCCCAG
CAAGCACCACAGGCCCAACCACCCCACAGCCAGGACAACCCACGAGGCCCACAGCCACAG
AGACCACTCAAACAAGAACGACTACTGAATACACAACGCCCCAAACCCCACACACCACAC
ACTCCCCGCCTACGGCGGGGAGTCCCGTCCCTTCCACAGGTCCTGTCACTGCAACATCTT
TCCATGCCACCACTACCTATCCAACCCCATCACACCCTGAGACCACACTTCCCACTCACG
TTCCACCTTTCTCCACCTCCTTGGTGACTCCAAGTACTCACACAGTCATCACCCCTACCC
ACGCACAGATGGCCACATCTGCCTCCATCCACTCAGCGCCAACAGGTACCATTCCTCCAC
CAACAACGCTCAAGGCCACAGGGTCCACCCACACAGCCCCACCAATAACGCCGACCACCA
GTGGGACCAGCCAAGCCCACAGCTCATTCAGCACAAACAAAACACCTACCTCGCTACATT
CACACACTTCCTCCACACACCATCCTGAAGTCACCCCAACTTCTACTACCACGATTACTC
CCAACCCCACTAGTACACGCACCAGAACCCCTGTGGCCCACACCAACTCAGCCACCAGCA
GCAGGCCACCACCACCCTTCACCACACACTCCCCACCTACAGGGAGCAGTCCCTTCTCTT
CCACAGGTCCCATGACGGCAACATCCTTCAAGACCACCACTACCTATCCAACCCCATC
>1278-3' Small Sal1 fragment of Large MUC6 case amplicon
GTCGACCCTGTGGGCATGCGCGTTGTCAGTGGAGGAACGGTGCCTGTTGGCGTTGAGTGG
ATCGAAGCAGAAGTGGACATTTGTGCGTGGGTAGGGGTGATGACTGTGTGAGTACTTGGA
GTGACTGATGAGGTGGAGAAAGGTGGAACATGAGTGGTAAGTGTGGTCTGAGGGTGTGAT
GGGGTTGGATAGGTCGTGGTGGTCTTGATGGATG
>1282-5' Small Sal1 fragment of Small MUC6 case amplicon
CAGGTGAGATGGANACAATGGGGCAGGCTGGAGTGCCCAGCAGGGGCCATGTCACAGGAA
CAGAGGCACAGACAGGCAAGAAAAAGGTCACATAGACAAAAGGACGGGCCAGCGGAGGTC
AGGGTAGAGAAACAAAAACAATAACGATGACAACTTCACCAATTCCCACAGCAAAATCGA
CCAATCAGGAACTGCCAGGAACAACGGCCACCCAGACGACAGGCCCACGTCCAACCCCAG
CAAGCACCACAGGCCCAACCACCCCACAGCCAGGACAACCCACGAGGCCCACAGCCACAG
AGACCACTCAAACAAGAACGACTACTGAATACACAACGCCCCAAACCCCACACACCACAC
ACTCCCCGCCTACGGCGGGGAGTCCCGTCCCTTCCACAGGTCCTGTCACTGCAACATCTT
TCCATGCCACCACTACCTATCCAACCCCATCACACCCTGAGACCACACTTCCCACTCACG
TTCCACCTTTCTCCACCTCCTTGGTGACTCCAAGTACTCACACAGTCATCACCCCTACCC
ACGCACAGATGGCCACATCTGCCTCCAACCACTCAGCGCCAACAGGTACCATTCCTCCAC
CAACAACGCTCAAGGCCACAGGGTCCACCCACACAGCCCCACCAATAACGCCGACCACCA
GTGGGACCAGCCAAGCCCACAGCTCATTCANCACAAACAAAACACCTACC
>1282-3' Small Sal1 fragment of Small MUC6 case amplicon
GTCGACCCTGTGGGCATGCGCGTTGTCAGTGGAGGAACGGTGCCTGTTGGCGTTGAGTGG
ATCGAAGCAGAAGTGGACATTTGTGCGTGGGTAGGGGTGATGACTGTGTGAGTACTTGGA
GTGACTGATGAGGTGGAGAAAGGTGGAACATGAGTGGTAAGTGTGGTCTGAGGGTGTGAT
GGGGTTG

Figure. S9. Sequences of DNA amplicon portions (from 5' and 3' ends) that were digested and subcloned into sequencing plasmids. The sequences from two different amplicons (from cases with different banding patterns, as shown in **Supplementary Fig. 8**), showed sequences that exactly match the 5' portion of the *MUC6* variable number of tandem repeat (VNTR).

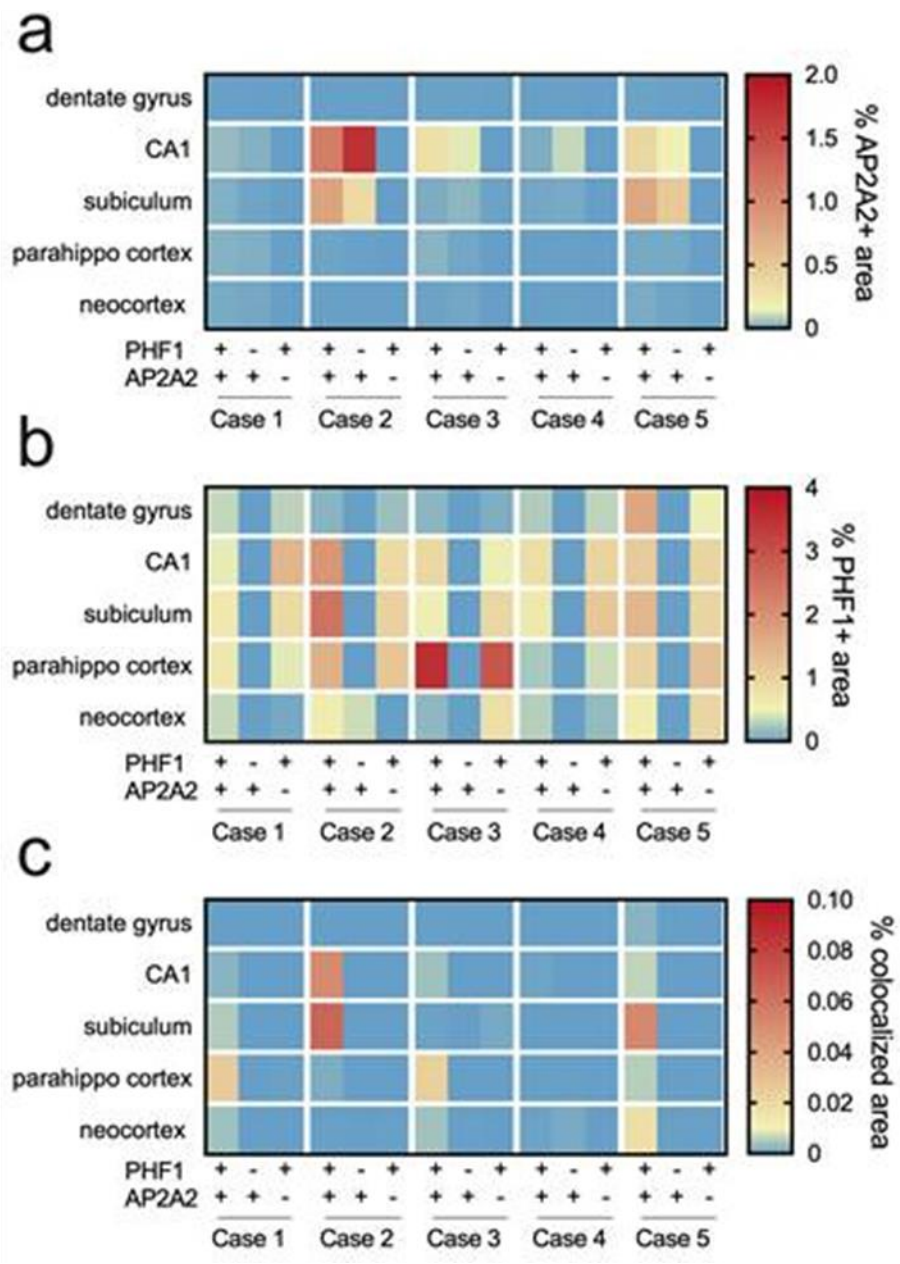


Figure. S10. AP2A2 and pTau colocalization and quantification in LOAD cases. The antibody PHF1 was used to immunostain pTau. Quantification of the single antibody positive and $AP2A2^{+}PHF^{+}$ double positive stained serial sections is shown for the percent area that is $AP2A2^{+}$ **(a)**, $pTau^{+}$ **(b)** and $AP2A2^{+}PHF^{+}$ **(c)** double positive.

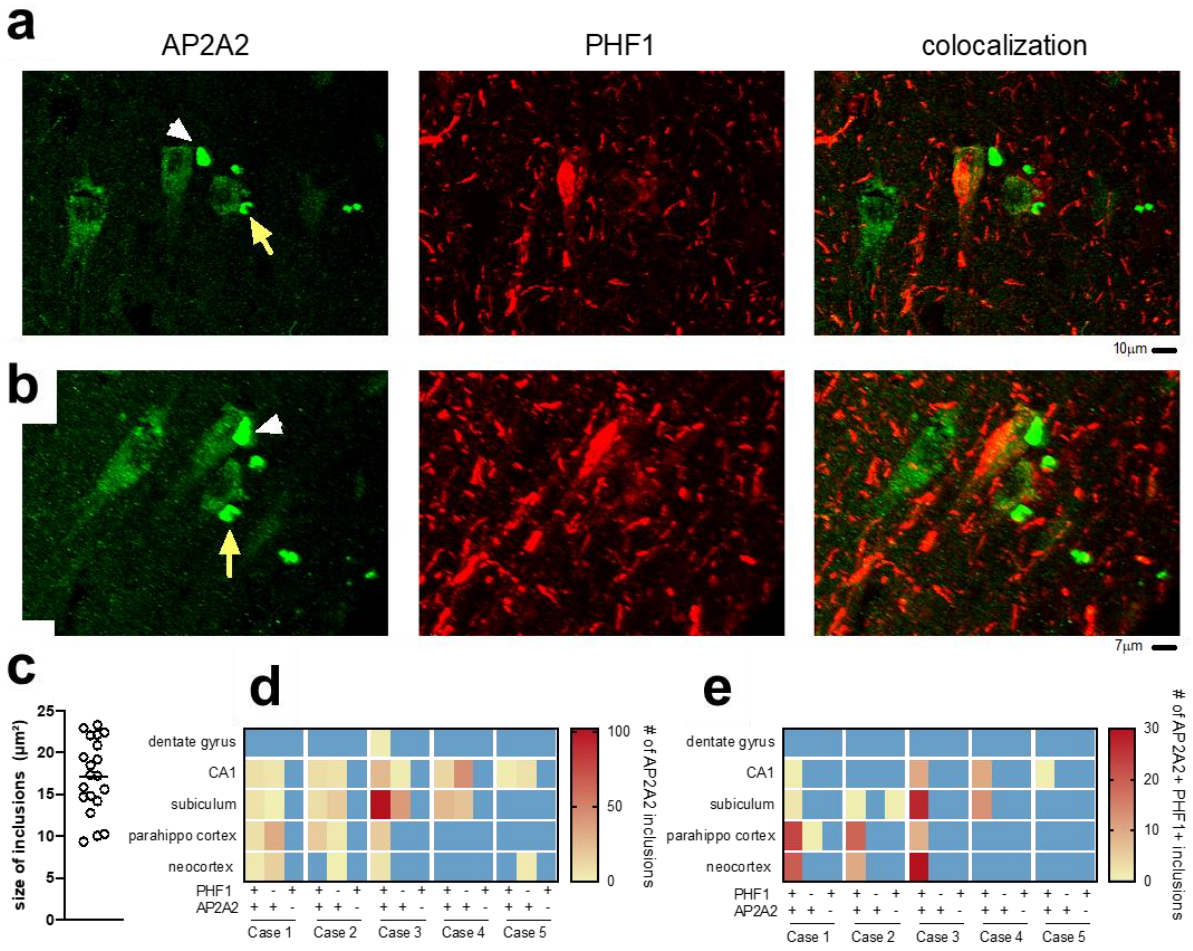


Figure. S11. In addition to AP2A2^{+}PHF^{+} NFTs, other intense AP2A2-immunoreactive structures were identified in the LOAD cases. (a) Representative maximum projection image from the confocal z-stack shows the AP2A2 inclusions. The arrow and arrowhead indicate the same inclusions **(a)**, shown in the orthogonal projection image in (**b**). The size of the inclusions was quantified using the HALO image analysis software **(c)**. The number of the small-bright AP2A2^{+} inclusions were quantified in the five regions of interest, and the number of inclusions per unit area is shown in **(d)**. The antibody PHF1 was used to immunostain pTau. The number of AP2A2^{+} inclusions that contained some amount of pTau^{+} staining is shown in **(e)**.

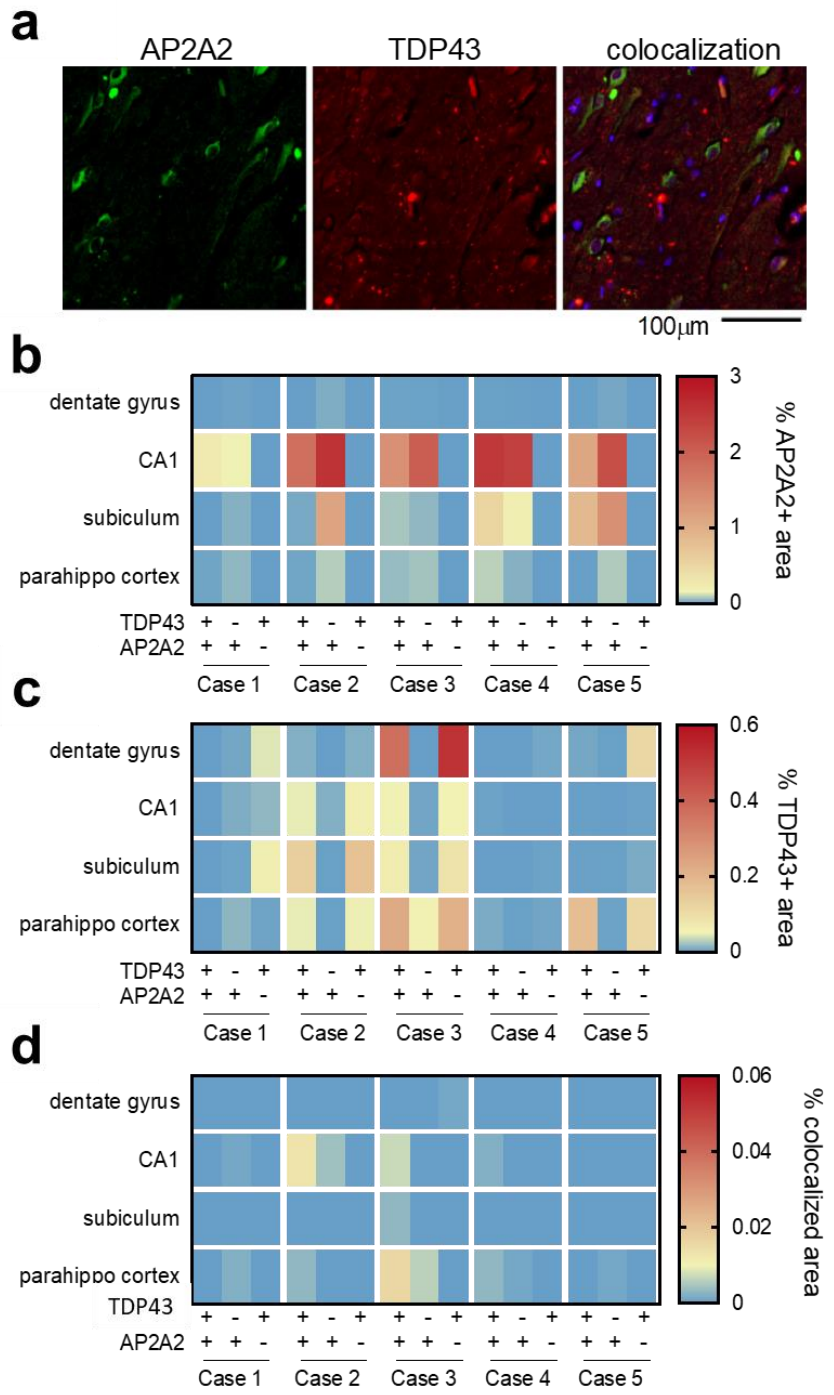


Figure. S12. Digital quantification of AP2A2 and phospho-TDP-43 immunoreactivity in brains with comorbid LOAD and limbic predominant age-related TDP-43 encephalopathy neuropathologic changes (LATE-NC). (a) Representative photomicrograph of AP2A2 and phospho-TDP-43 staining. Digital imaging quantification of the area fraction of tissue that was positive for AP2A2 **(b),** TDP-43 **(c)**, and $AP2A2^{+}TDP\text{-}43^{+}$ **(d)**.

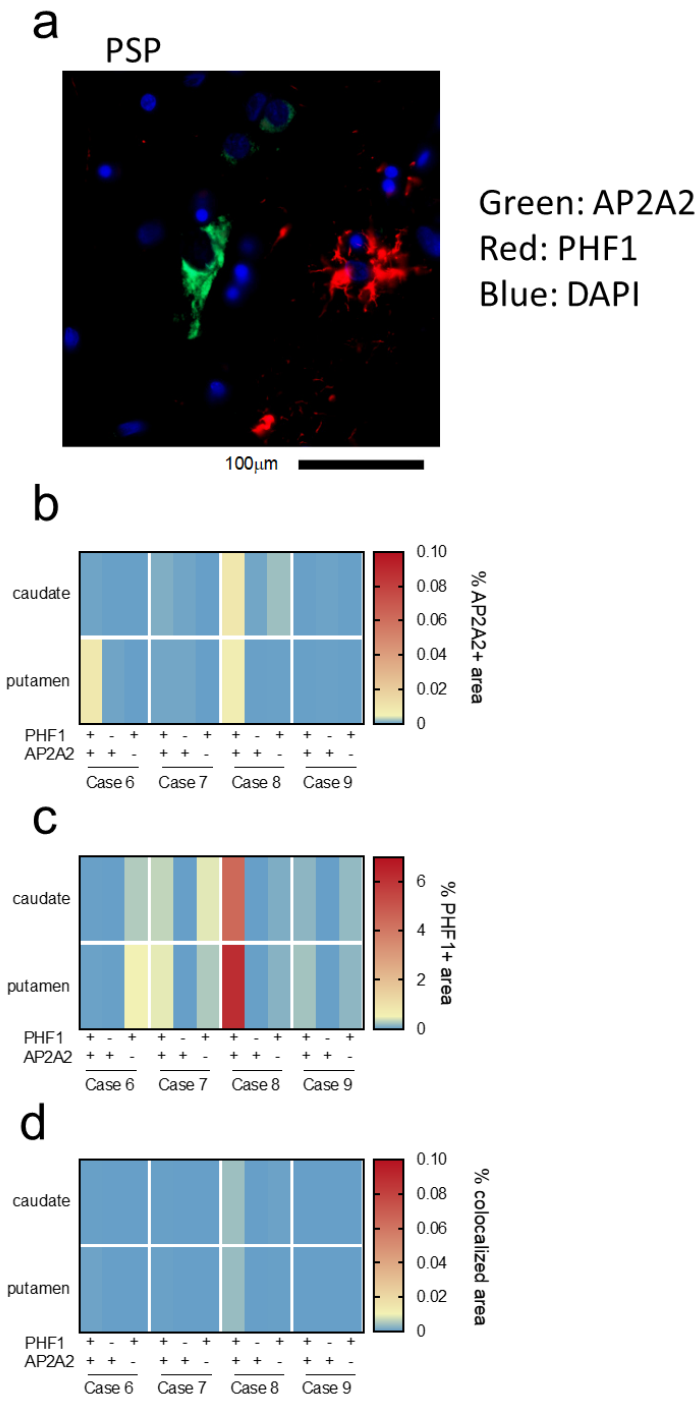


Figure. S13. Digital quantification of AP2A2 and pTau in progressive supranuclear palsy (PSP) brains. The antibody PHF1 was used to immunostain pTau. **(a)** Representative photomicrograph of AP2A2 and pTau staining. Digital imaging quantification of the area fraction of tissue that was positive for AP2A2 **(b)**, pTau **(c)**, and $AP2A2^{+}pTau^{+}$ **(d)**.

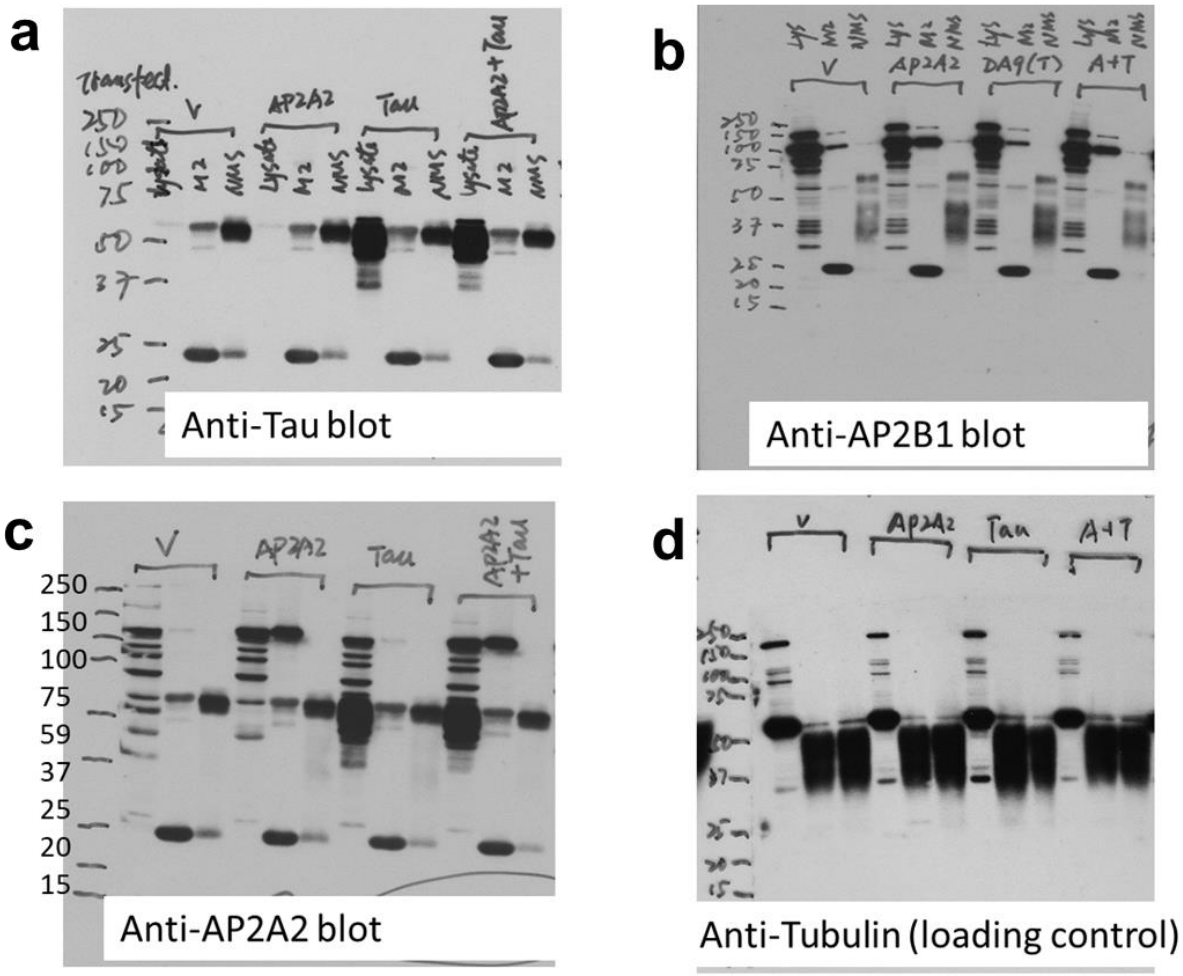


Figure. S14. Western blots depicting results of the co-immunoprecipitation (co-IP) experiments. These blots were shown partially in **Fig. 8**, for anti-Tau (**a**), anti-AP2B2 (**b**), anti-AP2A2 (**c**), and anti-tubulin (**d**). Labelled atop the blots are the plasmids transfected: empty vector (v), AP2A2, Tau, and AP2A2+Tau. For each plasmid transfected, three lanes' results are shown: lysate, M2 (Flag co-IP), and NMS (non-immunized mouse serum co-IP).