Supplementary materials

Supplemental methodology

Genotyping

University of California, San Francisco academic cohort

C9orf72 GGGGCC pathogenic expansion was detected using a two-step protocol. First, in all samples, the hexanucleotide repeat was PCR amplified using one fluorescently labeled primer followed by fragment length analysis on an automated ABI3730 DNA analyzer as previously described.[1] All patients who appeared homozygous in this assay were further analyzed using a repeat-primed PCR method. A characteristic stutter amplification pattern on the electropherogram was considered evidence of a pathogenic repeat expansion. A standard comparable protocol was used to sequence the commercial laboratory samples as well.

Variant nomenclature

Sequence variants identified in both cohorts were named and numbered according to the guidelines of Human Genome Variation Society (HGVS). Variants identified in *GRN* were numbered using reference transcript NM_002087.2, and variants identified in *MAPT* were numbered using NM_005910.4. For both genes, the "A" of the initiation codon (ATG) is labeled as nucleotide #1.

Table S1 Transcript IDs

	GRN	*MAPT*
UCSF	NM_002087.2	NM_005910.4
Clinical laboratory	NM_002087.2	NM_005910.4

Notes: Variants are mapped to human genome build GRCh37 (hg19). The "A" of the ATG start codon is numbered as nucleotide position one (c.1). The first exon in both *GRN* and *MAPT* is noncoding. It is described as exon 1 here, but other literature may describe it as Exon 0. cDNA numbering may also differ subtly from other literature, which may describe the first nucleotide of the first, non-coding exon as c.1.
Abbreviations: UCSF, University of California, San Francisco; IVS, intervening sequence.

Table S2 *GRN* variants included in this study; count in the commercial clinical laboratory/UCSF cohort and overall ExAC frequencies[a]

cDNA (NM_002087.2)	Protein (NP_002078.1)	Clinical laboratory count (n=24)	UCSF count (n=20)	Instances in ExAC, (total frequency)	Major ethnic group in ExAC[c]
c.1A>T	p.Met1Leu	0	1	1, (8.21×10^{-5})	SAS
c.2T>C	p.Met1Thr	1	0	1, (8.24×10^{-6})	NFE
c.26C>A	p.Ala9Asp	2	0	0	
c.154delA	p.Thr52Hisfs*2	0	1	0	
c.295_308delTGCCCACGGGGCTT	p.Cys99Profs*15	0	1	0	
c.264+1 G>A	Splice site	1	0	0	
c.264+2 T>C	Splice site	1	0	0	
c.328C>T	p.Arg110*	2	2	1, (8.51×10^{-6})	NFE
c.347 C>A	p.Ser116*	0	1	0	
c.349+1 G>A	Splice site	1	0	0	
c.388_391delCAGT	p.Gln130Serfs*125	2	0	1, (8.24×10^{-6})	NFE
c.415 T>C[b]	p.Cys139Arg	0	1	22, (1.81×10^{-4})	14X NFE, 7X SAS,1X AMR
c.472_496dup	p.Pro166Leufs*2	0	1	0	
c.592_593 delAG	p.Arg198Glyfs*19	2	1	0	
c.708+1G>A	Splice site	1	0	1, (8.28×10^{-6})	FIN
c.745C>T	p.Gln249*	1	0	0	
c.709-2A>G	Splice site	3	0	0	
c.836-1 G>C	Splice site	0	1	0	
c.898C>T	p.Gln300*	1	2	0	
c.910_911 dupTG	p.Trp304Cysfs*58	1	0	1, (8.26×10^{-6})	NFE
c.1145 del C	p.Thr382Serfs*30	0	1	0	
c.1157G>A	p.Trp386*	1	0	0	
c.1216C>T	p.Gln406*	0	1	0	
c.1263_1264insGAAGCGAG	p.Ile422Glufs*72	0	1	0	
c.1336C>T	p.Gln446*	1	0	0	
c.1477C>T	p.Arg493*	1	5	0	
c.1562G>A	p.Cys521Tyr	1	0	5, (4.12×10^{-5})	3X in AMR, 2X in NFE
c.1414-15_1591del	Splice site	1	0	0	

Notes: [a]A total of 28 different *GRN* variants were found in the 44 unrelated subjects included in the final analysis. [b]c.415C>T was classified as a risk allele. [c]The ExAC contains exome sequence data for >60,000 individuals of various ethnic backgrounds. SAS indicates South Asian, NFE indicates Non-Finnish European, AMR indicates Latino, and FIN indicates Finnish.
Abbreviations: UCSF, University of California, San Francisco; ExAC, Exome Aggregation Consortium.

Table S3 Pathogenic/likely pathogenic *MAPT* variants included in this study; count in commercial clinical laboratory/UCSF cohort and overall ExAC frequencies[a]

cDNA (NM_005910.4)	Protein	Clinical laboratory count	UCSF count	Instances in ExAC, (total frequency)	Major ethnic group in ExAC[b]
c.742 G>A	p.Val248Met	0	1	3, (2.52×10^{-5})	2X NFE, 1X EAS
c.837 T>A	p.Asn279Lys	1	0	0	
c.902 C>T	p.Pro301Leu	11	3	0	
c.1066T>A	p.Ser356Thr	0	1	0	
c.915+14C>T	Splice site	1	0	0	
c.915+16 C>T	Splice site	3	2	2, (8.08×10^{-5})	1X AFR, 1X NFE
c.915+19 C>G	Splice site	1	0	1, (8.08×10^{-5})	1X AFR, 1X NFE
c.1165 G>A	p.Gly389Arg	2	0	0	
c.1216C>T	p.Arg406Trp	0	2	1, (8.27×10^{-6})	NFE

Notes: [a]A total of nine different *MAPT* variants were found in 28 unrelated subjects included in the final analysis. [b]The ExAC contains exome sequence data for >60,000 individuals of various ethnic backgrounds. NFE indicates Non-Finnish European, EAS indicates East Asian, AFR indicates African/African American.
Abbreviations: UCSF, University of California, San Francisco; ExAC, Exome Aggregation Consortium.

Table S4 Type and quantity of overlap in *MAPT* and *GRN* variants between cohorts from UCSF and a commercial clinical laboratory

Gene	Protein	UCSF, n (%)	Clinical laboratory, n (%)	Total occurrences
***GRN* (NM_002087.2)**				
	p.Arg110*	2 (10)	2 (8)	4
	p.Arg198Glyfs*19	1 (5)	2 (8)	3
	p.Gln300*	2 (10)	1 (4)	3
	p.Arg493*	5 (25)	1 (4)	6
***MAPT* (NM_005910.4)**				
	c.915+16 C>T[a]	2 (22)	3 (15)	5
	p.Pro301Leu	3 (33)	11 (58)	14

Note: [a]*MAPT* c.915+16 C>T is located in the IVS region and has no protein product.
Abbreviations: UCSF, University of California, San Francisco; IVS, intervening sequence; Arg, arginine; Pro, proline; Leu, leucine; Glyfs, glycine frame shift; Gln, glutamine.

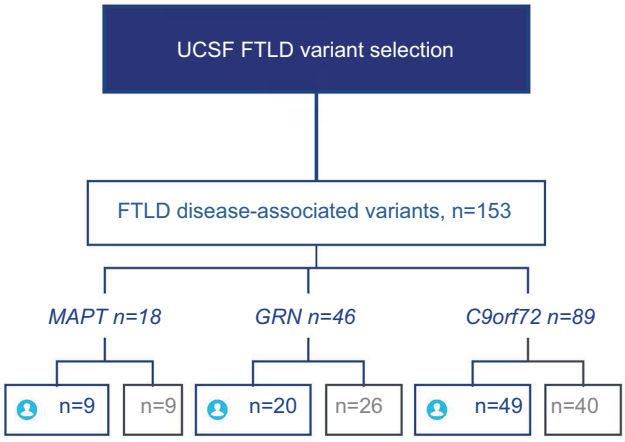

Figure S1 Selection of participants from UCSF.
Notes: The UCSF dataset included a total of 153 pathogenic/likely pathogenic FTLD variants, 89 of which were *C9orf72* pathogenic expansions, 46 *GRN*, and 18 *MAPT*. We obtained the subset of unrelated patients (blue icon) by removing variants found in family members of probands.
Abbreviations: UCSF, University of California, San Francisco; FTLD, frontotemporal lobar degeneration.

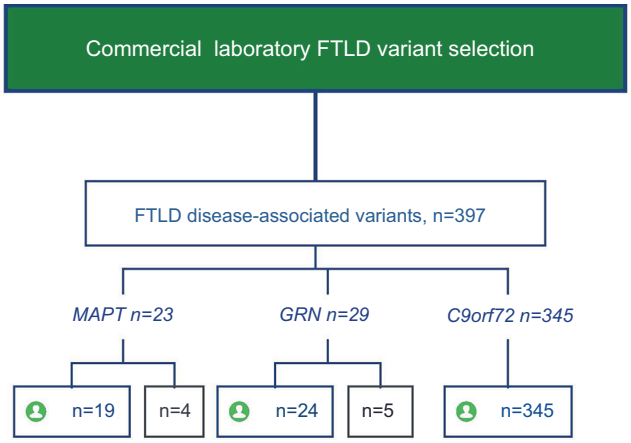

Figure S2 Selection of participants from a commercial clinical laboratory.
Notes: The commercial clinical laboratory dataset included a total of 397 pathogenic/likely pathogenic FTLD variants, 345 of which were *C9orf72* pathogenic expansions, 29 *GRN*, and 23 *MAPT*. We obtained the subset of unrelated patients (green icon) by removing variants ordered through family test codes (gray). There was no family test code for *C9orf72*; thus, all pathogenic expansions were included in the final analysis.
Abbreviation: FTLD, frontotemporal lobar degeneration.

Reference

1. Dejesus-Hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*. 2011;72(2):245–256.