Supplementary appendix

SUPPLEMENTARY METHODS

Study subjects

Previously ascertained patients and controls were enrolled in the current study between October 2014 and October 2017. Details on the origin of patients and controls in each of the stages and their clinical characteristics are provided in **Table 1, Supplementary Table 1 and Supplementary Figure 1**. For the discovery stage, DNA was obtained from blood or brain samples of 493 FTLD patients with *GRN* mutations from 33 research centers. At time of recruitment, all research centers from Europe, North-America and Australia with publications reporting genetically characterized FTLD populations or *GRN* case reports were offered participation in the study without any restriction in terms of age, sex or race. Only patients with predicted loss of function mutations in *GRN* defined as either predicted to reduce GRN levels or predicted to create truncated GRN protein were included. These included nonsense variants, small insertion and deletions resulting in frameshift and premature stop codons, large deletions and splicing variants. Missense variants affecting the signal peptide sequence (p.A9D and p.L10S) as well as the p.C139R mutation previously shown to reduce plasma GRN levels were also included.[1] For a full overview of all *GRN* mutations included in the study see **Supplementary Table 2.** For the discovery stage, DNA was also obtained from 505 controls from Spain and Italy from 2 research centers which had also provided patients with *GRN* mutations for this stage. Moreover, genotype data from 1,986 healthy controls from the CIDR NGRC Parkinson's Disease Study, generated with Illumina Human Omni1-Quad arrays, was obtained from dbGaP (NCBI dbGaP phs000196.v3.p1 NINDS CIDR PD Environment), hereinafter referred to as the CIDR dataset and counted as a separate site. The CIDR dataset was combined with the controls from Spain and Italy for the selection of controls for the discovery stage. After quality control steps (described below) and matching of 3 controls to each patient, 382 patients and 1,146 controls were included in the discovery analysis.

The replication cohort consisted of 210 patients and 1,798 controls of European ancestry recruited at 26 research centers: 20 were overlapping with the discovery series and 6 centers were newly added. The total number of sites in our meta-analysis was thus 40 (**Supplementary Figure 1**).Inclusion criteria were the same as for the discovery cohort with the exception that only self-reported non-Hispanic white individuals were included and the fact that patients with pathologically-confirmed FTLD-TDP type A without *GRN* mutations were also allowed. To avoid confounding effects due to family structure, especially since GWAS data was not available in this cohort, only one patient per known family was included. Relatives of patients from the discovery cohort were also removed, as well as patients with the common Italian mutation p.Thr272Serfs*10. In addition, only one patient per pair of patients with the same mutation and recruited from the same country was included, even without known family relatedness. Among the 210 patients selected for replication, 67 patients carried *GRN* mutations whereas 143 patients had pathologically confirmed FTLD-TDP type A without mutations in *GRN*. Mutations in *C9ORF72* were also excluded in all GRN-negative FTLD-TDP type A patients.

Discovery stage

Genotyping, quality control, imputation

Genotyping was performed on all 493 patients and the 505 healthy controls from Italy and Spain using Illumina Human OmniExpress arrays at Mayo Clinic Rochester following manufacturer's protocol. Standard quality control (QC) procedures[2] were applied to this in-house generated dataset and to the CIDR dataset using PLINK v1.9[3]. Subjects with call rate <95% were excluded from analysis (n=13), as were subjects with sex discrepancies (n=5), or non-European ancestry (n=7). No samples showed unusually high heterozygosity. Duplicate samples (n=28) as well as one participant from each pair of related individuals as determined by identity by descent analysis (PI_HAT > 0·15) were also removed (n=85), which is equivalent to removing one individual from a pair of relatives below the first cousin level. A total of 382 patients, 478 controls from Italy and Spain and 1,986 controls from CIDR were retained after QC.

Variant level QC was performed as follows: variants with call rate <95%, variants with minor allele frequency (MAF) <0·01, or not in Hardy-Weinberg equilibrium (markers with $p<10^{-6}$ in the controls) were removed (n=194,890 of original 1,012,895 variants removed in the CIDR dataset; 83,019 of original 716,503 variants removed from the dataset generated by genotyping at Mayo Clinic). Only overlapping markers passing QC in both the in-house generated dataset and the CIDR dataset (601,202 variants) were further considered.

Genotypes of all patient and control samples included in the discovery stage were imputed together using Minimac3 on the Michigan Imputation Server, with the Haplotype reference Consortium r1.1 2016 data as the reference panel. For each imputed variant, imputation quality was estimated using the r^2 metric and only variants with r^2 above 0·8 were retained. A total 7,033,776 with MAF>0·01 variants fulfilled those criteria, and were retained for the association analyses.

Statistical analysis
Principal component analysis was performed using genome-wide SNP data to calculate the top principal components that capture the greatest degree of genetic variation in the study subjects[4]. Each of the 382 *GRN* mutation patients was then matched to three controls based on age, sex, country of origin and the first four principal components leading to a sample of 382 patients with *GRN* mutations and 1,146 controls for the discovery genome-wide association analyses. In this discovery set, the median age at onset of patients was 60·0 years (IQR 55·0 – 66·0) and the median age at death was 66·0 years (IQR 61·0 – 73·0) (**Table 1**). Females represented 55·2% (n=211) of the patients and 55·0% (n=630) of the controls.

Genotype associations with phenotypes (patient/control status and age of onset) were evaluated using variant allele dosage data, under the assumption of additive allele effects (or log-additive for disease risk). To test for variant association with disease risk, logistic regression analyses were performed using PLINK v1.9 adjusted for age, sex, and the first two PCs. Association of variants with age of onset was evaluated using linear regression adjusting for sex, and the first two PCs. QQ-plots are shown in **Supplementary Figure 3**.

In addition, a possible effect of rs5848 located in the 3'UTR of *GRN* on age at onset in patients with *GRN* mutations was assessed treating genotypes as categorical, allowing us to evaluate the differences in age at onset between the minor homozygotes and major homozygotes.

As an exploratory analysis, we also assessed the genome-wide association of genetic variants with memory, behavior, or language impairment as first clinical symptom, or presence/absence of parkinsonism during the disease course. Patients with the trait of interest were grouped and compared to patients without this trait using a logistic regression adjusting for age, sex, country of origin and first two PCs.

Haplotypes of 16 variants surrounding the *GRN* gene were estimated in patients using the haplo.stats R package v1.7.7

Replication stage

Genotyping and quality control
Forty-four suggestive loci ($p<10^{-5}$) were nominated for follow-up in the replication stage. For all suggestive loci, the lead variant and one proxy were included in the multiplex MassArray design (Agena Bioscience, San Diego, CA, USA). At the *TMEM106B* locus we also included the previously reported top GWAS variant rs1990622 and the coding *TMEM106B* variant rs3173615 in the assay design. No variants from the *GRN* locus at 17q21 were included. Five loci failed the design following manufacturer's recommendations (rs116550318, rs1989391, rs7044291, rs118011160 and rs78626473). MassArray iPLEX genotyping was performed in 210 patients and in 839 controls from the replication cohort and an additional 100 patients from the discovery cohort for QC purposes. Genotypes for an additional 959 neurologically normal controls that were part of the replication stage were extracted from whole genome sequence data previously generated by the Mayo Clinic Biobank (total controls in the replication stage is 839 + 959 = 1,798). 20 nanogram of DNA measured by spectrophotometer (Nanodrop; Wilmington, DE, USA) was used for genotyping on the MassArray iPLEX system (Agena Bioscience, San Diego, CA, USA) following the manufacturer's protocol. Variants with a call rate <95% or failing Hardy-Weinberg equilibrium in controls ($p<0.05$) were subsequently removed from the analysis (five variants, rs76035395, rs2702379, rs4430025, rs9421746, rs7926822). In addition, only one variant per locus was retained for analysis resulting in a total of 36 variants at 34 loci (three variants at *TMEM106B* locus).

Statistical analysis
Logistic and linear regression analyses were performed for disease risk and age-of-onset, adjusted for age and sex (logistic) and sex only (linear). Meta-analyses of the discovery and replication results were performed under a fixed effects model. We also calculated I^2 heterogeneity statistics to evaluate the degree of heterogeneity of the effects in

the discovery and replication stages, and for SNPs with I^2 suggesting moderate or high heterogeneity (I^2>0.3) we also performed a random effects meta-analysis, to verify that conclusions regarding association would not change under this model.

Power calculation
The discovery case/control GWAS with 382 cases and 1,146 controls, provided 80% power to detect ORs of 1·9 and 1·8 for SNPs with minor allele frequencies of 0·2 and 0·5, respectively, at a genome-wide significant level. With a less-stringent threshold of 10^{-5} to identify loci for replication, the discovery stage had 80% power to identify SNPs that had ORs exceeding 1·7 (MAF=0·2) or 1.6 (MAF=0·5). The replication stage with 210 cases and 1,798 controls had >91% power to detect those same effects.

The discovery GWAS of age-of-onset in 382 cases provided 80% power to detect SNPs that explain >9·8% of the variation in age of onset (i.e. R^2=0·098), at a genome-wide significant level. With a less-stringent threshold of 10^{-5} to identify loci for replication, the discovery stage had 80% power to identify SNPs that explain 7% of the variation in age of onset. The replication stage that included 210 cases had power >76% to detect the same effect on age of onset.

Annotation of variants
eQTL annotation of variants was performed using the Genotype-Tissue Expression (GTEx) Project which is supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the eQTL annotation were obtained from the GTEx Portal on 11/10/17. Detailed annotation of the variants in LD with rs36196656 was issued from the Haploreg database and included: central nervous system chromatin states and marks, expression quantitative trait loci from GTEx as well as potential transcription factor binding sites.

Expression analyses

Effect of rs36196656 on *GFRA2* expression
mRNA expression analysis of *GFRA2* was performed using 48 patients and healthy controls from the Mayo Clinic Brain Bank. We identified 24 individuals with the rarer 'AA' genotype at rs36196656 and then selected a similar number of 'CC' individuals matched to the best of our ability for disease, age and gender. In total our cohort consisted of nine patients with *GRN* mutations (four rs36196656-CC, five rs36196656-AA), 21 patients with FTLD-TDP type A pathology (12 rs36196656-CC, nine rs36196656-AA), and 18 healthy controls (eight rs36196656-CC, ten rs36196656-AA). RNA was extracted using RNeasy Qiagen Minikit (Qiagen, Hilden, Germany) and concentration was evaluated using Nanodrop. RNA quality was assessed using an Agilent 2100 bioanalyzer, RNA 6000 LabCHip kit (Agilent Technologies, Santa Clara, CA, USA). All RNA obtained had RIN >7·0 and was subsequently reverse transcribed using the Superscript III system (Life Technologies, Carlsbad, CA, USA). Quantitative real-time PCR was performed in quadruplicate for each sample on an ABI7900 PCR system (Applied Biosystems, Foster City, CA, USA), using TaqMan gene expression assays (Life Technologies, Carlsbad, CA, USA) and following the manufacturer's recommendations. *GFRA2* transcripts were measured using the probe Hs00176393_m1 for all *GFRA2* transcripts from Invitrogen. Specific probes for *GFRA2* variant A, B and C were designed (sequences available upon request). *MAP2* (probe Hs00258900_m1) and RPLPO (probe Hs00420895_gh) were used as reference genes. Results were analyzed using SDS software version 2.2 (Life Technologies, Carlsbad, CA, USA). *GFRA2* transcript levels were assessed using the ΔΔCt method normalized by the geometric mean of *MAP2* and *RPLPO* transcripts to account for both total cell number and the contribution of neuronal cells specifically. A Mann-Whitney test was performed assessing the difference of distribution of expression levels between genotypes.

Comparative analysis of *GFRA2* transcripts and relative expression to *GFRA1*
RNA from cerebellum, frontal cortex, motor cortex, hippocampus, medulla, amygdala, basal ganglia and thalamus from an FTLD-TDP type A patient was extracted as detailed above and reverse transcribed using the Superscript III system. PCR was performed on the cDNA samples using primers in *GFRA2* designed to amplify exons 1 to four using the following primers: forward (5'-3') CAAACGTCTTCTGCCTCTTC and reverse (5'-3') TGCAGTTGTCATTCAGGTTG and relative expression of each transcript variant among different brain regions was assessed. Amplicons for *GFRA2* variants had a size of 513bp for *GFRA2* variant A, 198bp for *GFRA2* variant B and 117bp for *GFRA2* variant C (**Supplementary Figure 4A)**. Quantification of the three main transcripts of

GFRA2 (variant A, B and C) and the combined expression of all *GFRA1* transcripts (probe Hs00237133_m1) was also performed by Taqman assays to allow semi-quantitative analysis of the relative expression of each *GFRA2* transcript across different brain regions and the relative expression of GFRA2 to GFRA1 in these brain regions (**Supplementary Figure 4B and 4D**).

Functional assays

Cell culture and transfections
HEK293T cells (purchased from American Type Culture Collection, Manassas, VA) were cultured and maintained in Eagle's Minimum Essential Medium (EMEM) supplemented with 10% FBS and 1% penicillin/streptavidin (pen/strep). The cells were maintained at 37 °C, five percent CO_2.

For the co-immunoprecipitation studies, HEK293T cells were transiently co-transfected with equal amounts of empty vector, PGRN, and GFRA2 constructs in different combinations using 2g/L polyethyleneimine (PEI) (Polysciences, Warrington, PA, USA) by mixing DNAs with PEI at the ratio 1:5 (µg:µl) in OptiMEM (Invitrogen, Carlsbad, CA, USA).

Co-immunoprecipitation
HEK293T cells were lysed in 50 mM Tris, pH8.0, 150 mM NaCl, 1% Triton X-100, and 0·1% deoxycholic acid with protease and phosphatase inhibitors (Thermo Fisher Scientific, Walthman, MA, USA) after transfection. After centrifugation at 14,000 rpm at 4°C for five minutes, the superrnatants were collected and pre-cleaned by incubating with protein G beads (Genscript, Piscataway, NJ, USA) at 4°C for two hours on a rotator. Five percent of the supernatants was saved as the input, and the remaining supernatant was subjected to either anti-PGRN antibody (R&D Systems, Minneapolis, MN, USA) or anti-GFRA2 antibody (R&D Systems, Minneapolis, MN, USA) followed by protein G Dynabeads (Thermo Fisher Scientific, Walthman, MA, USA). Beads were washed with 50 mM Tris, pH 8·0, 150 mM NaCl, and 1% Triton X-100 after four hours of incubation. The immunoprecipitation products were then eluted with x2 Tris-Glycine SDS sample buffer (Life Technologies, Carlsbad, CA, USA) supplemented with five percent β-mercaptoethanol

Western blot
The immunoprecipitated samples were denatured at 95°C for five minutes, separated using 10-20% Tris-glycine gels (Life Technologies, Carlsbad, CA, USA),and transferred to Immobilon membranes (Millipore, Burlington, MA, USA). The membranes were blocked in Odyssey blocking buffer (Licor Odyssey system, LI-COR Biosciences, Lincoln, NE, USA), and subsequently incubated with the primary antibodies diluted in Odyssey Blocking Buffer overnight at 4°C. The second day, membranes were washed three times with TBS-T and incubated with secondary antibodies conjugated to IRDye 680 or IRDye 800 (LI-COR Biosciences) for two hours at room temperature. Membranes were washed three more times with TBS-T, then imaged, and quantified using an Odyssey Infrared imaging system (LI-COR Biosciences).

Antibodies and constructs
The following primary antibodies were used for immunoblotting: goat anti-PGRN (1:1,000, R&D Systems, Minneapolis, MN, USA), and goat anti-GFRA2 (1:1000, R&D Systems, Minneapolis, MN, USA).

AAV-GFRA2 was generated by subcloning the *GFRA2* variant A cDNA from *GFRA2* v1 ORF (NM_001495, GenScript) into the pAAV using XhoI and HindIII restriction sites. Human *PGRN* in pCMV-Sport6 vector was obtained from Open Biosystems.

SUPPLEMENTARY RESULTS

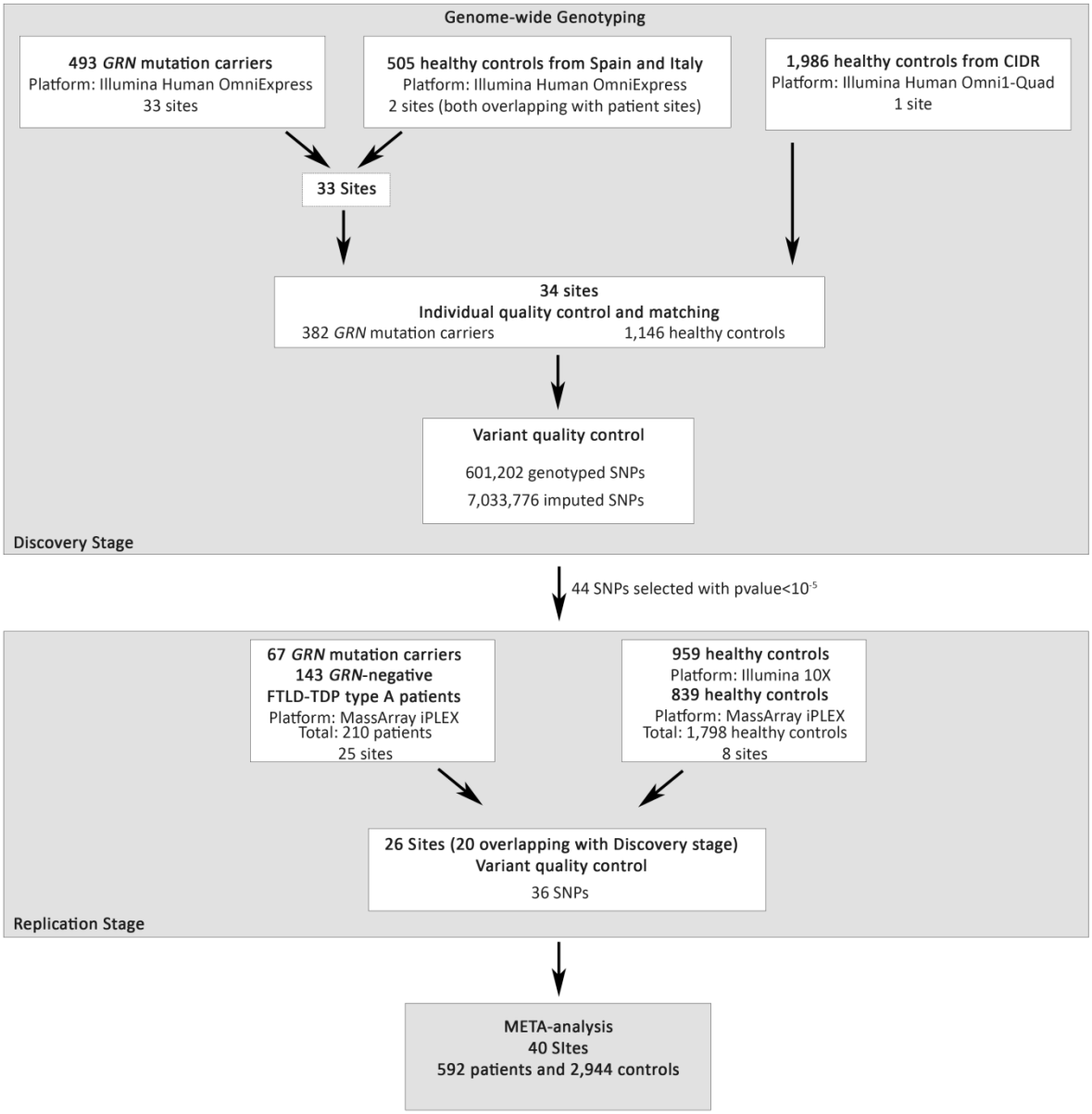
Associations with disease risk and age at onset including only *GRN* mutation carriers in the replication stage
To explore the contribution of *GRN* and non-*GRN* mutation carriers, we performed exploratory meta-analysis of the discovery and replication series in which only the subset of patients with *GRN* mutations was included. In the case/control analysis, six loci in addition to *GFRA2* and *TMEM106B* showed improved p-values as compared to the discovery analysis alone (lead variants within or in close proximity to *NDUFS1*, *BDH1*, *PJA2*, *STARD3NL*, *NEIL2* and *PALM2*; **Supplementary Table 5**). In the age at onset analysis, four loci showed improved p-values (*COL28A1*, *SHFM1*, *STMN2*, *RRBP1*; **Supplementary Table 8**).

Exploratory analysis of genome-wide associations with first clinical symptoms and parkinsonism
In our discovery cohort, variable first clinical symptoms were reported, including the presence of behavioral changes (n=262; 53·1%), language impairment (n=169; 34·3%) and memory dysfunction (n=139; 28·2%), with 109 (22·1%) patients reporting multiple clinical symptoms at disease onset. In addition, 37·3% (n=184) patients had parkinsonism during the course of their disease. While not the primary focus of our study, we tested within our discovery patients for association of genetic markers with the absence or presence of any of the 3 main first clinical symptoms (behavioral impairment, language dysfunction or memory impairment) or the presence or absence of parkinsonism during the disease course. No genome-wide significant associations were detected; however, 16 loci showed suggestive association ($p<10^{-5}$; **Supplementary Table 9**) and require study in future patient cohorts.

Genetic associations at previously reported modifier loci.
We also assessed the association of putative genetic modifier variants of disease presentation in known neurodegenerative disease genes, including apolipoprotein E (APOE) ε4 dosage (rs429358) for memory impairment as well as tau (MAPT) H1/H2 haplotype (rs1052553), glucosylceramidase beta (GBA, rs35749011) and alpha-synuclein (SNCA, rs356182) for parkinsonism. Only rs1052553 at the MAPT locus showed nominal significance with the presence of parkinsonism (p=0·02, OR=1·57, 95% CI: 1·20 – 1·94). Finally, we examined previously reported putative age at onset modifiers and did not identify significant association for rs9897526 at the *GRN* locus ($p=8·93\times10^{-2}$, Beta=-1·48, 95% CI: -3·18 – 0·22) and rs1052553 ($p=9·05\times10^{-2}$, Beta=1·07, 95% CI: -0·17 – 2·30).

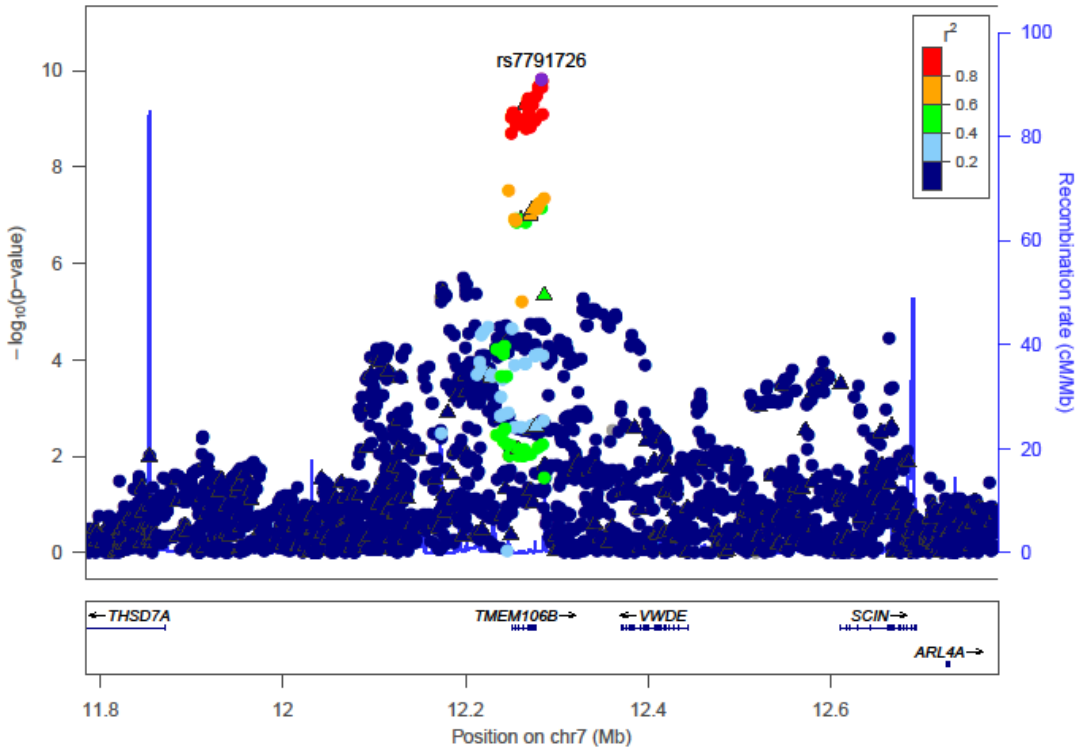
SUPPLEMENTARY FIGURES

Supplementary Figure 1: Analysis flow chart
The flow chart illustrates our strategy to identify new loci associated with disease risk (patients vs controls) and age at onset modifiers (patients only). SNP=single nucleotide polymorphism.

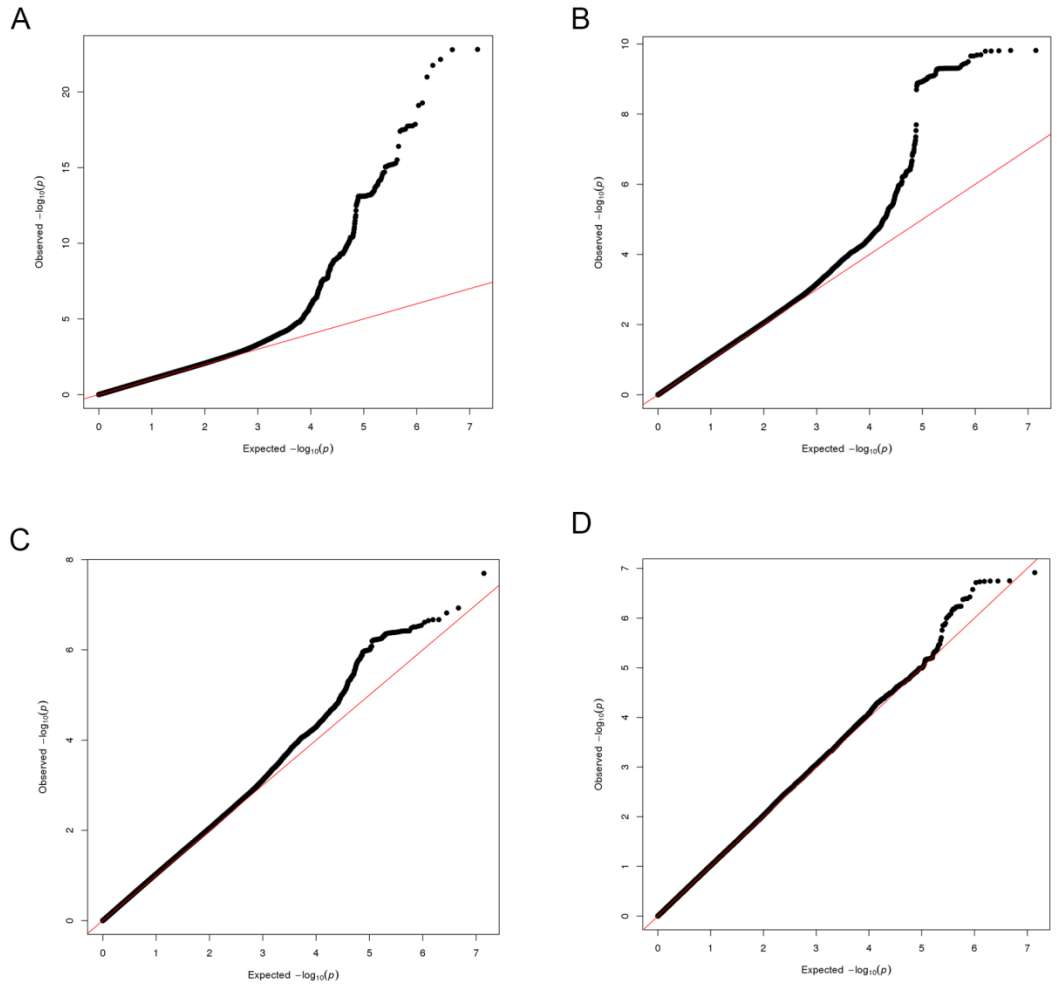


Supplementary Figure 2: Regional zoom in the *TMEM106B* locus

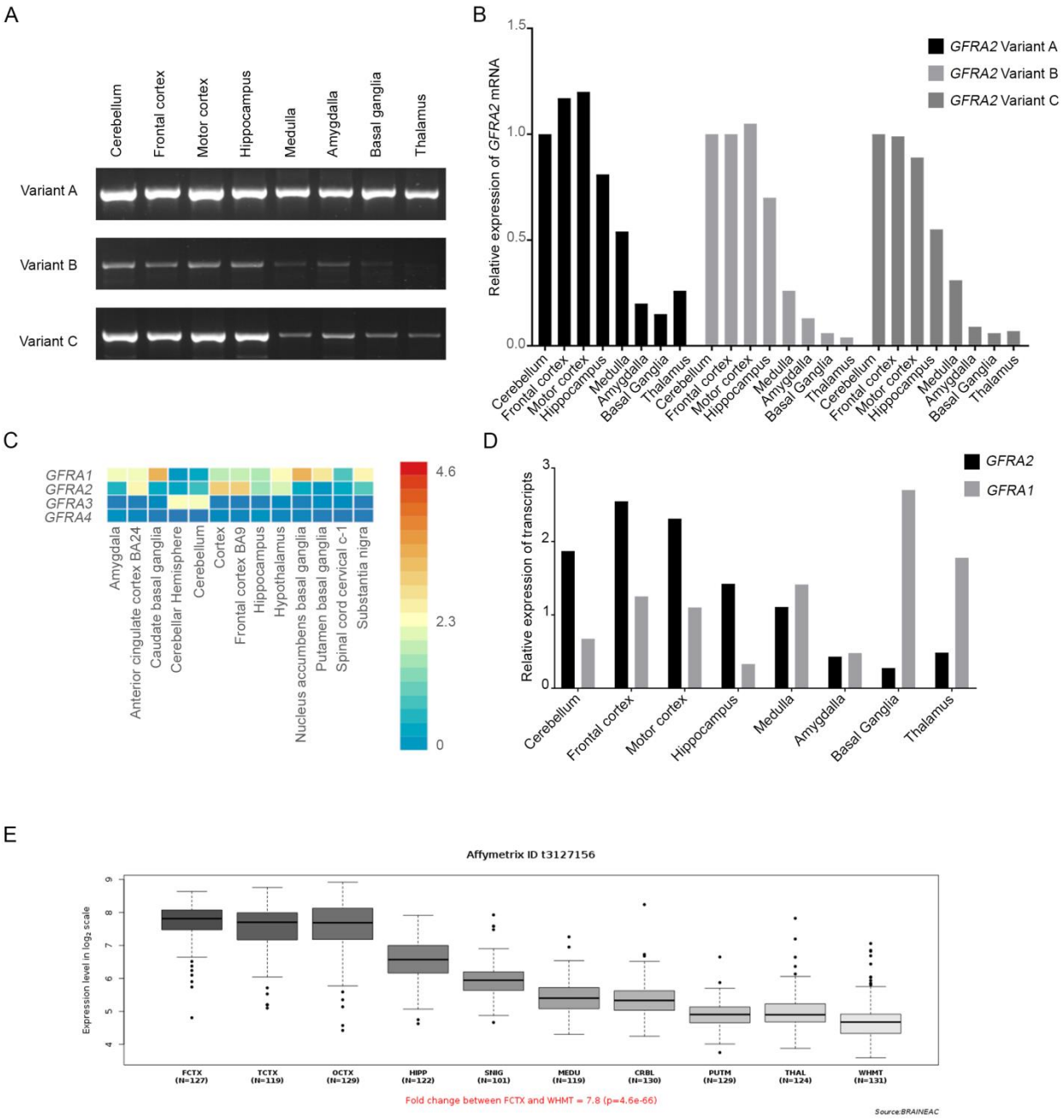
The *TMEM106B* locus zoom plot is presented with a zoom in the *TMEM106B* gene. The purple dot is the most significant variant (rs7791726) among variants in the region. Dots are colored according to the r^2 showing their degree of linkage disequilibrium with rs7791726. The blue line shows the estimated recombination rate.



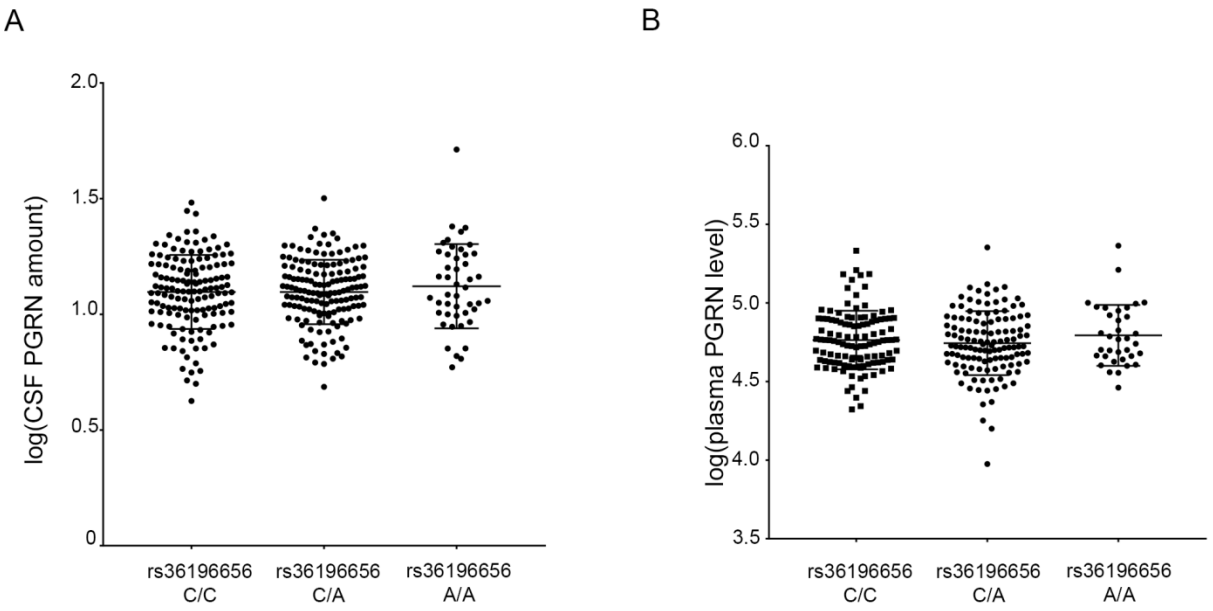
Supplementary Figure 3: QQ plot from the case/control (disease risk) and age at onset analysis.
(**A**) QQ plot of the logistic regression (disease risk) study (λ=1.07, λ_{1000}=1.12). (**B**) QQ plot of the logistic regression without the *GRN* locus (λ=1.05, λ_{1000}=1.08). (**C**) QQ plot of the logistic regression without the *GRN* locus and TMEM106B locus (λ=1.04, λ_{1000}=1.07). (**D**) QQ plot of the linear regression (age at onset) study (λ=1.02, λ_{1000}=1.052).



Supplementary Figure 4: *GFRA2* and *GFRA1* expression in multiple brain regions.
(**A**) Reverse transcriptase PCR showing the relative expression of *GFRA2* variant A, B and C across multiple brain regions in an FTLD-TDP type A patient. (**B**) Histograms representing the quantitative PCR-determined relative amounts of *GFRA2* transcripts A, B and C in multiple brain regions normalized to *MAP2* in an FTLD-TDP type A patient. For each variant (A, B and C) expression is shown normalized to the expression in the cerebellum for that particular variant. (**C**) Heat map issued from the GTex dataset (generated by RNA sequencing) showing the high expression of *GFRA2* in cortex and frontal cortex compared to other members of the GFRA family. (**D**) Histograms representing the quantitative PCR-determined relative amount of *GFRA2* (all variants) and *GFRA1* (all variants) transcripts across multiple brain regions normalized to *MAP2* in a FTLD-TDP type A patient. Note the highest expression of *GFRA2* in frontal and motor cortex, whereas *GFRA1* has its highest expression in basal ganglia, corroborating the GTex dataset. (**E**) mRNA expression of GFRA2 in multiple brain region based on the BRAINEAC dataset (http://www.braineac.org/) generated by microarrays. FCTX=frontal cortex; TCTX=temporal cortex; OCTX=occipital cortex; HIPP=hippocampus; SNIG=substantia nigra; MEDu=medulla; CRBL=cerebellum; PUTM=putamen; THAL=thalamus; WHMT=white matter


A
Cerebellum
Frontal cortex
Motor cortex
Hippocampus
Medulla
Amygdalla
Basal ganglia
Thalamus
Variant A
Variant B
Variant C
B
Relative expression of GFRA2 mRNA
1.5
1.0
0.5
0.0
GFRA2 Variant A
GFRA2 Variant B
GFRA2 Variant C
C
GFRA1
GFRA2
GFRA3
GFRA4
Amygdala
Anterior cingulate cortex BA24
Caudate basal ganglia
Cerebellar Hemisphere
Cerebellum
Cortex
Frontal cortex BA9
Hippocampus
Hypothalamus
Nucleus accumbens basal ganglia
Putamen basal ganglia
Spinal cord cervical c-1
Substantia nigra
4.6
2.3
0
D
Relative expression of transcripts
3
2
1
0
GFRA2
GFRA1
E
Affymetrix ID t3127156
Expression level in log2 scale
FCTX (N=127)
TCTX (N=119)
OCTX (N=129)
HIPP (N=122)
SNIG (N=101)
MEDU (N=119)
CRBL (N=130)
PUTM (N=129)
THAL (N=124)
WHMT (N=131)
Fold change between FCTX and WHMT = 7.8 (p=4.6e-66)
Source:BRAINEAC

Supplementary Figure 5: *GFRA2* rs36196656 does not affect CSF and plasma PGRN levels.
Log transformed cerebrospinal fluid (**A**) and plasma (**B**) PGRN levels in individuals from the Mayo Clinic Study of Aging are shown stratified by rs36196656 genotype (C/C, C/A, and A/A). CSF=cerebrospinal fluid



SUPPLEMENTARY TABLES

Supplementary Table 1: Overview of 40 sites participating in the discovery and replication stages.
Number of patients and controls included per site after QC are presented for the discovery stage and replication stage.

Site	Discovery patients	Replication patients	Discovery controls	Replication controls
University of Sydney, Sydney, Australia	9	5	0	0
University of British Columbia, Vancouver, BC, Canada	12	10	0	0
Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Canada	3	4	0	0
University of Western Ontario, London, ON, Canada	2	0	0	0
Copenhagen University Hospital, Copenhagen, Denmark	2	0	0	0
Brain and Spine Institute (ICM), Paris, France	55	0	0	0
University of Tübingen, Tübingen, Germany	1	5	0	0
Technische Universität München, Munich, Germany	0	9	0	0
Fatebenefratelli Research Institute, Brescia, Italy	42	0	336	0
Carlo Besta Neurological Institute, Milano, Italy	22	8	0	0
University of Brescia, Brescia, Italy	20	4	0	0
University of Milan, Milan, Italy	12	1	0	0
Regional Center of Neurogenetics, Catanzaro, Italy	7	0	0	0
University of Florence, Florence, Italy	7	0	0	0
St. Adalbert Hospital, Gdańsk, Poland	2	0	0	0
Hospital Universitario Donostia, San Sebastián, Spain	21	0	121	0
Hospital Clinic of Barcelona, Barcelona, Spain	9	0	0	0
Karolinska Institutet, Stockholm, Sweden	11	0	0	0
Erasmus Medical Center, Rotterdam, The Netherlands	11	6	0	0
University College London, London, UK	12	1	0	0
University of Manchester, Manchester, UK	12	0	0	0
King's College London, London, UK	0	14	0	0
University of Pennsylvania, Philadelphia, PA, USA	25	5	0	12
University of California, San Francisco, CA, USA	18	4	0	111
Mayo Clinic Brain Bank, Jacksonville, FL, USA	14	53	0	45
Mayo Clinic, Rochester, MN, USA	14	2	0	190
Mayo Clinic, Jacksonville, FL, USA	8	5	0	437
Washington University, St. Louis, MO, USA	8	11	0	0
Indiana University, Indianapolis, IN, USA	7	8	0	11
Northwestern University, Chicago, IL, USA	6	14	0	33
Banner Sun Health Research Institute, Sun City, AZ, USA	3	5	0	0
Columbia University, New York, NY, USA	2	4	0	0
Harvard Brain Bank, Boston, MA, USA	2	0	0	0
University Texas Southwestern, Dallas, TX, USA	2	16	0	0
Thomas Jefferson University, Philadelphia, PA, USA	1	0	0	0
Emory University, Atlanta, GA, USA	0	8	0	0
University of California, San Diego, CA, USA	0	3	0	0
University of Pittsburg, Pittsburgh, PA , USA	0	5	0	0
NINDS CIDR PD Environment	0	0	689	0
Mayo Clinic Biobank, MN, USA	0	0	0	959
TOTAL	**382**	**210**	**1146**	**1798**

Supplementary Table 2: *GRN* mutations included in the discovery and replication stages.
Counts of the different *GRN* mutations included in the study are presented. Each mutation is annotated with the HGNC nomenclature. The most frequent mutations are highlighted in bold. QC=Quality control.

cDNA change	Protein change	Count Discovery	Count Discovery after QC	Count Replication
c.-8+3A>T	p.?	1	0	0
c.-8+3A>G	p.?	0	0	3
c.1A>T	p.?	3	0	0
c.1A>G	p.?	2	1	0
c.2T>C	p.?	2	2	0
c.3G>A	p.?	1	1	0
c.26C>A	p.Ala9Asp	18	10	2
c.29T>C	p.Leu10Ser	2	1	0
c.57_58insC	p.Cys20Leufs*45	0	0	1
c.63_64insC	p.Asp22Argfs*43	4	2	0
c.70C>T	p.Gln24*	0	0	1
c.78C>A	p.Cys26*	2	1	0
c.87_90dup	p.Cys31Leufs*35	12	12	1
c.87dup	p.Cys30Leufs*35	2	2	0
c.88_89insC	p.Cys30Serfs*35	1	1	0
c.102del	p.Gly35Glufs*19	14	9	2
c.117dup	p.Ser40Glnfs*25	1	1	0
c.138+1G>A	p.?	2	2	0
c.139-1delG	p.Asp47fs	0	0	1
c.146G>A	p.Trp49*	0	0	1
c.154del	p.Thr52Hisfs*2	11	6	0
c.234_235del	p.Gly79Aspfs*39	2	2	2
c.243del	p.Ser82Valfs*174	13	6	0
c.255del	p.Phe86Serfs*170	1	1	0
c.264+2T>C	p.?	2	1	1
c.264G>A	p.()	1	1	1
c.295_308del	p.Cys99Profs*15	1	1	0
c.299del	p.Pro100Hisfs*156	0	0	1
c.314dup	p.Cys105Trpfs*14	1	1	0
c.328C>T	p.Arg110*	6	5	3
c.347C>A	p.Ser116*	2	1	0
c.348A>C	p.()	1	1	0
c.350-1G>T	p.?	2	1	0
c.361del	p.Val121Trpfs*135	4	3	0
c.373C>T	p.Gln125*	4	2	0
c.378C>A	p.Cys126*	1	1	0
c.380_381del	p.Pro127Argfs*2	5	4	0
c.384_387del	p.Gln130Serfs*125	1	1	1
c.388_391del	p.Gln130Serfs*125	7	5	1
c.415T>C	p.Cys139Arg	9	9	2
c.421_422del	p.Val141Tyrfs*18	2	2	0
c.443_444del	p.Gly148Valfs*11	2	1	0
c.445_446del	p.Cys149Lfs*10	4	4	1
c.462+1G>C	p.?	1	1	0
c.463-1G>A	p.?	1	1	0
c.468_474del	p.Cys157Lysfs*97	5	5	2

c.472_496dup	p.Pro166Leufs*2	1	1	0
c.481_482del	p.Arg161Glyfs*36	0	0	1
c.559del	p.Leu187Trpfs*69	1	1	0
c.560del	p.Leu187Argfs*69	0	0	1
c.592_593del	p.Arg198Glyfs*19	2	2	1
c.599_708del	p.Val200Glyfs*18	1	0	0
c.603dup	p.Ser203Valfs*15	1	1	0
c.607del	p.Ser203Profs*53	2	2	0
c.619dup	p.Met207Asnfs*11	1	1	0
c.675_676del	p.Ser226Trpfs*28	3	3	3
c.687T>G	p.Tyr229*	0	0	1
c.708+6_708+9del	p.?	8	8	4
c.708+1G>A	p.?	4	4	1
c.708+1G>C	p.?	1	0	0
c.709-1G>A	**p.?**	**31**	**22**	0
c.709-2A>G	p.?	9	8	3
c.709-4_713delTCAGGCCAC	p.?	0	0	1
c.745C>T	p.Gln249*	2	1	0
c.759_760del	p.Cys253*	3	3	2
c.761_762insTG	p.Leu255Alafs*2	1	1	0
c.768_769dup	p.Gln257Profs*27	3	2	0
c.813_816del	**p.Thr272Serfs*10**	**97**	**78**	0
c.833_834del	p.Thr278Serfs*7	2	2	0
c.834dup	p.Val279Serfs*7	1	1	0
c.835_835+1insCTGA	p.?	1	1	0
c.836-1G>C	p.?	6	4	0
c.882T>G	p.Tyr294*	3	2	1
c.898C>T	p.Gln300*	5	5	0
c.907del	p.Ala303Profs*58	1	1	0
c.909del	p.Trp304Glyfs*57	1	1	0
c.910_911insTG	p.Trp304Leufs*58	7	3	1
c.911G>A	p.Trp304*	1	0	0
c.912G>A	p.Trp304*	2	2	1
c.918C>A	p.Cys306*	2	2	0
c.933+1G>A	p.?	2	2	0
c.942C>A	p.Cys314*	3	3	0
c.975del	p.Phe326Leufs*35	1	1	0
c.988_989del	p.Thr330Alafs*6	2	1	0
c.988_989dup	p.Gln331Argfs*31	0	0	1
c.998del	p.Gly333Valfs*28	2	1	0
c.1009C>T	p.Gln337*	2	2	0
c.1014del	p.His340Thrfs*21	1	1	0
c.1021C>T	p.Gln341*	6	4	1
c.1072C>T	p.Gln358*	2	2	1
c.1095_1096del	p.Cys366*	5	4	0
c.1145del	p.Thr382Serfs*30	7	5	0
c.1145_1146insA	p.Ser383Valfs*31	7	5	0
c.1157G>A	p.Trp386*	4	3	0
c.1160dup	p.Cys388Leufs*26	0	0	1
c.1179+2T>C	p.?	1	1	0

c.1180-2A>G	p.?	1	0	0
c.1201C>T	p.Gln401*	3	2	0
c.1216C>T	p.Gln406*	1	1	0
c.1231_1232dup	p.Ala412*	2	1	0
c.1231_1232del	p.Val411Serfs*2	1	1	0
c.1243C>T	p.Gln415*	4	4	0
c.1246dup	p.Cys416Leufs*30	2	2	0
c.1252C>T	p.Arg418*	9	6	3
c.1256_1263dup	p.Ile422Glufs*72	3	2	0
c.1317_1318del	p.Asp441Hisfs*4	1	1	1
c.1354del	p.Val452Trpfs*39	5	4	0
c.1359dup	p.Gln454Alafs*58	2	2	0
c.1392G>A	p.Trp464*	0	0	1
c.1395dup	p.Cys466Leufs*46	1	1	0
c.1402C>T	p.Gln468*	2	2	1
c.1414-2A>G	p.?	1	0	0
c.1414-15_1590del194bp	p.Ala472_Glu548del	0	0	1
c.1420_1421del	p.Cys474Leufs*37	1	1	0
c.1428_1431del	p.Glu476Aspfs*14	1	1	0
c.1477C>T	**p.Arg493***	**35**	**30**	**6**
c.1494_1498del	p.Glu498Aspfs*12	6	5	0
c.1538_1539del	p.His513Argfs*16	1	1	0
c.1603C>T	p.Arg535*	0	0	1
Large gene deletions	p.?	7	7	1

Supplementary Table 3: Association results of the case/control analysis in discovery stage.
Summary of all suggestive variants from the discovery stage in *GRN* mutation carriers and controls are presented, irrespective of successful genotyping in the replication stage. Variant p-values and odds ratio were calculated using an additive genetic model. P-values for significant loci are shown in bold ($p<5\times10^{-8}$). Minor alleles were treated as effect alleles. The locus name is determined by the closest gene to the significant variant. The lead variant at the *TMEM106B* locus as well as the published GWAS hit rs1990622 and the non-synonymous variant rs3173615 are included. Variants with significant association on the *GRN* locus are not included. P=p-value; OR=odds ratio.

Variant	Position[a]	Major/minor allele	Locus name	Annotation	OR (95% CI)	P
rs76035395	1:39928482	C/T	*MYCL1*	intergenic	8·2 (3·70 – 18·19)	$2{\cdot}26\times10^{-7}$
rs13393316	2:206134615	A/G	*NDUFS1*	intronic	0·5 (0·38 – 0·67)	$2{\cdot}65\times10^{-6}$
rs4680382	3:157606472	A/G	*C3orf55*	intergenic	1·5 (1·26 – 1·78)	$4{\cdot}75\times10^{-6}$
rs13072484	3:197409951	G/A	*BDH1*	intergenic	1·54 (1·28 – 1·85)	$3{\cdot}79\times10^{-6}$
rs116550318	4:93514986	C/T	GRID2	intronic	3·43 (2·88 – 3·97)	$8{\cdot}78\times10^{-6}$
rs2702379	4:178450753	T/G	*LOC285501*	intergenic	1·49 (1·26 – 1·77)	$4{\cdot}05\times10^{-6}$
rs79095029	5:109519605	C/G	*PJA2*	intergenic	0·35 (0·23 – 0·55)	$5{\cdot}72\times10^{-6}$
rs146261599	5:124264446	T/G	*ZNF608*	intergenic	2·91 (1·82 – 4·64)	$7{\cdot}64\times10^{-6}$
rs181675566	5:169224908	T/C	*SLIT3*	intronic	3·86 (2·15 – 6·91)	$5{\cdot}72\times10^{-6}$
rs1989391	6:1784261	G/A	GMDS	intronic	1·83 (1·60 – 2·05)	$2{\cdot}15\times10^{-7}$
rs6904835	6:17809964	T/C	*KIF13A*	intronic	1·5 (1·25 – 1·8)	$9{\cdot}67\times10^{-6}$
rs3173615	7:12229791	C/G	*TMEM106B*	non-synonymous	0·55 (0·45 – 0·66)	$\mathbf{7{\cdot}81\times10^{-10}}$
rs7791726	7:12243703	G/C	*TMEM106B*	intergenic	0·53 (0·44 – 0·64)	$\mathbf{1{\cdot}53\times10^{-10}}$
rs1990622	7:12244161	A/G	*TMEM106B*	intergenic	0·53 (0·44 – 0·64)	$\mathbf{1{\cdot}61\times10^{-10}}$
rs62443267	7:38113711	C/T	*STARD3NL*	intergenic	0·62 (0·5 – 0·76)	$6{\cdot}83\times10^{-6}$
rs4430025	7:98445147	G/A	*BAIAP2L1*	intergenic	1·91 (1·45 – 2·52)	$4{\cdot}11\times10^{-6}$
rs141226303	7:104610766	A/G	*LHFPL3*	intronic	3·73 (2·11 – 6·59)	$5{\cdot}61\times10^{-6}$
rs3110811	7:135717900	A/G	*SLC13A4*	intronic	1·55 (1·29 – 1·87)	$3{\cdot}50\times10^{-6}$
rs10101195	8:11765703	C/A	*NEIL2*	intergenic	0·62 (0·51 – 0·77)	$7{\cdot}50\times10^{-6}$
rs36196656	8:21763735	C/A	*GFRA2*	intronic	1·51	$9{\cdot}44\times10^{-6}$

					(1·26 – 1·82)	
rs7044291	9:2202182	G/A	SMARCA2	intergenic	0·65 (0·46 – 0·83)	$5·00\times10^{-6}$
rs10816848	9:109659155	T/A	*PALM2*	intronic	0·68 (0·57 – 0·8)	$8·74\times10^{-6}$
rs9421746	10:47245881	G/A	*GDF10*	intergenic	1·6 (1·31 – 1·95)	$4·33\times10^{-6}$
rs78781776	11:36444983	A/G	*PRR5L*	intronic	2 (1·47 – 2·71)	$8·88\times10^{-6}$
rs10791882	11:66551842	A/G	*ACTN3*	intronic	1·49 (1·26 – 1·77)	$5·01\times10^{-6}$
rs7926822	11:78686972	G/T	*ODZ4*	intronic	2·32 (1·65 – 3·26)	$1·35\times10^{-6}$
rs10860097	12:96805878	A/T	*NEDD1*	intronic	3·43 (2·14 – 5·5)	$2·88\times10^{-7}$
rs61965655	13:74138778	T/A	*KLF12*	intergenic	2·33 (1·61 – 3·39)	$8·52\times10^{-6}$
rs847358	14:72313813	G/A	*RGS6*	intronic	1·46 (1·24 – 1·73)	$7·43\times10^{-6}$
rs12605286	18:43570202	G/A	*SYT4*	intergenic	0·6 (0·48 – 0·74)	$2·37\times10^{-6}$
rs7240419	18:79168989	G/A	*ATP9B*	intronic	1·62 (1·34 – 1·94)	$3·80\times10^{-7}$
rs6076187	20:24101942	G/A	*FLJ33581*	intergenic	2·47 (1·71 – 3·57)	$1·53\times10^{-6}$

[a]Positions are based on the Human Genome version 38 (hg38).

Supplementary Table 4: Association results of the age at onset analysis in discovery stage.
Summary of all suggestive variants from the discovery stage ($p<10^{-5}$) in age at onset analysis are presented, irrespective of successful genotyping in the replication stage. Variant p-values and odds ratio were calculated using an additive genetic model. Minor alleles were treated as effect alleles. The locus name is determined by the closest gene to the significant variant. P, p-value; MAF=minor allele frequency.

Variant	Position[a]	Major/minor allele	Locus name	Annotation	Discovery		
					MAF patients	Association	
						Beta (95% CI)	P
rs116316277	2:185834886	C/T	*ZNF804A*	intergenic	0·03	8·09 (4·72 – 11·46)	$3·58\times10^{-6}$
rs6809184	3:170888198	C/T	*TNIK*	intronic	0·05	-6·78 (-9·24 – -4·32)	$1·22\times10^{-7}$
rs12189587	6:165332257	C/T	*C6orf11*	intergenic	0·11	-4·05 (-5·76 – -2·34)	$4·83\times10^{-6}$
rs6962939	7:7524226	T/A	*COL28A1*	intronic	0·04	-6·02 (-8·61 – -3·43)	$7·00\times10^{-6}$
rs2922921	7:96398079	G/A	*SHFM1*	intergenic	0·02	9·65 (5·58 – 13·72)	$4·65\times10^{-6}$
rs77466830	7:151529171	C/A	*PRKAG2*	intronic	0·32	2·91 (1·64 – 4·18)	$9·49\times10^{-6}$
rs9792144	8:53081551	C/G	*ST18*	intronic	0·12	3·99 (2·3 – 5·68)	$4·88\times10^{-6}$
rs3922636	8:80383502	G/A	*STMN2*	intergenic	0·19	3·28 (1·89 – 4·67)	$5·08\times10^{-6}$
rs12943707	17:73317510	C/G	*GRB2*	intronic	0·29	-2·8 (-4·00 – -1·6)	$6·40\times10^{-6}$
rs1561819	18:2712629	G/A	*SMCHD1*	intronic	0·49	-2·41 (-3·46 – -1·36)	$8·96\times10^{-6}$
rs6108746	20:10902771	T/C	*JAG1*	intergenic	0·19	3·54 (2·19 – 4·89)	$4·23\times10^{-7}$
rs6111609	20:17664546	C/A	*RRBP1*	intergenic	0·22	2·86 (1·61 – 4·11)	$9·83\times10^{-6}$
rs118011160	11:66782152	G/A	*SYT12*	intergenic	0·04	-6·86 (-9·53 – -4·19)	$7·10\times10^{-7}$
rs78626473	17:7378856	C/T	*ZBTB4*	intronic	0·17	-3·25 (-4·66 – -1·84)	$8·22\times10^{-6}$

[a]Positions are based on the Human Genome version 38 (hg38).

Supplementary Table 5: Loci identified in case/control analysis aimed at identifying modifiers of disease risk when only including *GRN* patients in replication stage.

Suggestive variants identified in the discovery stage ($p<10^{-5}$) and followed-up in the replication stage, as well the meta-analyses using a fixed-effect model are presented. These results were obtained when only *GRN* mutation carriers were included in the replication stage. Variant p-values and odds ratio were calculated using an additive genetic model. Minor alleles were treated as effect alleles. P-values for significant loci are shown in bold. The locus name is determined by the closest gene to the significant variant. MAF=minor allele frequency; OR=odds ratio; P=p-value.

Variant	Position[a]	Major/ minor allele	Locus name	Discovery			Replication *GRN+*			Meta-analysis[b]		
				MAF patients/ controls	Association OR (95%CI)	Association P	MAF patients/ controls	Association OR (95% CI)	Association P	OR (95% CI)	P	I^2
rs13393316	2:206999339	A/G	*NDUFS1*	0·10/0·16	0·50 (0·38 – 0·67)	$2·65\times10^{-6}$	0·12/0·14	0·65 (0·36 – 1·19)	$1·61\times10^{-1}$	0·53 (0·41 – 0·68)	$1·28\times10^{-6}$	0
rs4680382	3:157324261	G/A	*C3orf55*	0·59/0·32	1·50 (1·26 – 1·78)	$4·75\times10^{-6}$	0·36/0·35	1·03 (0·72 – 1·49)	$8·66\times10^{-1}$	1·40 (1·20 – 1·64)	$2·56\times10^{-5}$	69·2
rs13072484	3:197136822	G/A	*BDH1*	0·29/0·21	1·54 (1·28 – 1·85)	$3·79\times10^{-6}$	0·28/0·22	1·37 (0·93 – 2·30)	$1·14\times10^{-1}$	1·51 (1·28 – 1·78)	$1·19\times10^{-6}$	0
rs79095029	5:108855306	C/G	*PJA2*	0·03/0·08	0·35 (0·23 – 0·55)	$5·72\times10^{-6}$	0·05/0·08	0·57 (2·03 – 1·33)	$1·94\times10^{-1}$	0·39 (0·27 – 0·59)	$3·99\times10^{-6}$	0·5
rs146261599	5:123600139	T/G	*ZNF608*	0·05/0·02	2·91 (1·82 – 4·64)	$7·64\times10^{-6}$	0·02/0·03	0·59 (0·14 – 2·44)	$4·62\times10^{-1}$	2·49 (1·60 – 3·88)	$5·75\times10^{-5}$	77·2
rs181675566	5:168651912	T/C	*SLIT3*	0·04/0·01	3·86 (2·15 – 6·90)	$5·72\times10^{-6}$	0·01/0·02	0·37 (0·05 – 2·69)	$3·24\times10^{-1}$	3·20 (1·83 – 5·60)	$4·59\times10^{-5}$	79·8
rs6904835[c]	6:17810195	T/C	*KIF13A*	0·32/0·24	1·50 (1·25 – 1·80)	$9·67\times10^{-6}$	0·20/0·27	0·72 (0·47 – 1·10)	$1·32\times10^{-1}$	1·34 (1·14 – 1·59)	$4·68\times10^{-4}$	89·6
rs3173615[cd]	7:12269417	C/G	*TMEM106B*	0·27/0·39	0·55 (0·45 – 0·66)	**$7·81\times10^{-10}$**	0·21/0·42	0·37 (0·24 – 0·57)	$6·32\times10^{-6}$	0·51 (0·43 – 0·61)	**$9·05\times10^{-14}$**	62·0
rs7791726[cd]	7:12283329	G/C	*TMEM106B*	0·26/0·39	0·53 (0·44 – 0·64)	**$1·53\times10^{-10}$**	0·22/0·42	0·38 (0·25 – 0·58)	$9·87\times10^{-6}$	0·50 (0·42 – 0·60)	**$1·95\times10^{-14}$**	49·0
rs1990622[cd]	7:12283787	A/G	*TMEM106B*	0·26/0·39	0·53 (0·44 – 0·65)	**$1·61\times10^{-10}$**	0·22/0·42	0·40 (0·26 – 0·61)	$1·67\times10^{-5}$	0·50 (0·42 – 0·60)	**$2·68\times10^{-14}$**	32·4
rs62443267	7:38153313	C/T	*STARD3NL*	0·19/0·26	0·62 (0·50 – 0·76)	$6·83\times10^{-6}$	0·22/0·25	0·78 (0·50 – 1·2)	$2·58\times10^{-1}$	0·65 (0·23 – 0·78)	$5·51\times10^{-6}$	0
rs141226303	7:104251213	A/G	*LHFPL3*	0·04/0·01	3·73 (2·11 – 6·59)	$5·61\times10^{-6}$	0·01/0·01	0·51 (0·07 – 3·68)	$5·06\times10^{-1}$	3·21 (1·86 – 5·54)	$2·93\times10^{-5}$	72·2
rs3110811[c]	7:135402648	A/G	*SLC13A4*	0·29/0·2	1·55 (1·29 – 1·87)	$3·50\times10^{-6}$	0·21/0·23	0·89 (0·58 – 1·36)	$5·83\times10^{-1}$	1·42 (1·20 – 1·69)	$5·28\times10^{-5}$	81·8
rs10101195[c]	8:11623212	C/A	*NEIL2*	0·18/0·26	0·62 (0·51 – 0·77)	$7·50\times10^{-6}$	0·15/0·23	0·62 (0·38 – 1·01)	$5·55\times10^{-2}$	0·62 (0·51 – 0·75)	$1·11\times10^{-6}$	0
rs36196656[c]	8:21621247	C/A	*GFRA2*	0·46/0·37	1·51 (1·26 – 1·82)	$9·44\times10^{-6}$	0·47/0·35	1·69 (1·19 – 2·40)	$3·11\times10^{-3}$	1·55 (1·32 – 1·82)	$1·17\times10^{-7}$	0
rs10816848	9:112421435	T/A	*PALM2*	0·42/0·49	0·68 (0·57 – 0·80)	$8·74\times10^{-6}$	0·43/0·49	0·80 (0·56 – 1·13)	$2·08\times10^{-1}$	0·70 (0·60 – 0·82)	$5·39\times10^{-6}$	79·5

rs78781776	11:36466533	A/G	*PRR5L*	0·10/0·05	2·00 (1·47 – 2·72)	$8{\cdot}88\times10^{-6}$	0·07/0·07	1·01 (0·50 – 2·02)	$9{\cdot}86\times10^{-1}$	1·79 (1·35 – 2·37)	$4{\cdot}63\times10^{-5}$	68·1
rs10791882[c]	11:66319313	G/A	*ACTN3*	0·46/0·37	1·49 (1·26 – 1·77)	$5{\cdot}01\times10^{-6}$	0·44/0·39	1·20 (0·85 – 1·71)	$3{\cdot}04\times10^{-1}$	1·43 (1·23 – 1·67)	$5{\cdot}33\times10^{-6}$	14·9
rs10860097	12:97199656	A/T	*NEDD1*	0·05/0·02	3·43 (2·14 – 5·50)	$2{\cdot}88\times10^{-7}$	0·01/0·02	0·35 (0·05 – 2·57)	$3{\cdot}04\times10^{-1}$	3·04 (1·92 – 4·81)	$1{\cdot}98\times10^{-6}$	79·0
rs61965655	13:74712915	T/A	*KLF12*	0·07/0·04	2·33 (1·61 – 3·39)	$8{\cdot}52\times10^{-6}$	0·02/0·05	0·47 (0·15 – 1·50)	$2{\cdot}02\times10^{-1}$	2·01 (1·41 – 2·86)	$1{\cdot}21\times10^{-4}$	85·4
rs847358	14:72780521	G/A	*RGS6*	0·53/0·44	1·46 (1·24 – 1·73)	$7{\cdot}43\times10^{-6}$	0·42/0·46	0·86 (0·60 – 1·22)	$3{\cdot}89\times10^{-1}$	1·33 (1·14 – 1·54)	$2{\cdot}22\times10^{-4}$	86·1
rs12605286	18:41150167	G/A	*SYT4*	0·23/0·31	0·60 (0·48 – 0·74)	$2{\cdot}37\times10^{-6}$	0·30/0·26	1·23 (0·84 – 1·82)	$2{\cdot}91\times10^{-1}$	0·71 (0·59 – 0·85)	$2{\cdot}79\times10^{-4}$	90·2
rs7240419[c]	18:76928989	G/A	*ATP9B*	0·31/0·22	1·62 (1·34 – 1·94)	$3{\cdot}80\times10^{-7}$	0·24/0·23	1·11 (0·74 – 1·67)	$6{\cdot}13\times10^{-1}$	1·52 (1·28 – 1·79)	$1{\cdot}35\times10^{-6}$	62·5
rs6076187	20:24082578	G/A	*FLJ33581*	0·07/0·03	2·47 (1·71 – 3·57)	$1{\cdot}53\times10^{-6}$	0·01/0·04	0·20 (0·03 – 1·47)	$1{\cdot}14\times10^{-1}$	2·27 (1·58 – 3·26)	$9{\cdot}13\times10^{-6}$	83·1

[a]Positions are based on the Human Genome version 38 (hg38). [b]For SNPs with $I^2>0.3$, a random effects meta-analysis was also performed. The p-values were generally larger in the random effects meta-analysis, and the results were consistent with the fixed effects, showing that no SNPs other than those in *TMEM106B* were significantly associated with the outcome. [c]Variants annotated as eQTL in the GTex database. [d]Variants that are study-wide significant at the replication stage after Bonferroni correction.

Supplementary Table 6: Loci identified in case/control analysis aimed at identifying modifiers of disease risk when only including *GRN*-negative FTLD-TDP type A patients in replication stage.

Summary of suggestive variants ($p<10^{-5}$) identified in the discovery stage (*GRN* patients versus controls) and followed-up in the replication stage when only patients with *GRN*-negative FTLD-TDP type A and controls were included. Variant p-values and odds ratio were calculated using an additive genetic model. Minor alleles were treated as effect alleles. P-values for significant loci are shown in bold. The locus name is determined by the closest gene to the significant variant. MAF=minor allele frequency; OR=odds ratio; P=p-value. Variants genome-wide significant are highlighted in bold.

Variant	Position[a]	Major/ minor allele	Locus name	Discovery			Replication (*GRN*-negative FTLD-TDP Type A only)		
					Association			Association	
				MAF patients/controls	OR (95%CI)	P	MAF patients/controls	OR (95%CI)	P
rs13393316	2:206999339	A/G	*NDUFS1*	0·10/0·16	0·50 (0·38 – 0·67)	$2·65\times10^{-6}$	0·19/0·23	0·83 (0·55 – 1·23)	$3·50\times10^{-1}$
rs4680382	3:157324261	G/A	*C3orf55*	0·59/0·32	1·50 (1·26 – 1·78)	$4·75\times10^{-6}$	0·35/0·45	1·00 (0·77 – 1·31)	$9·84\times10^{-1}$
rs13072484	3:197136822	G/A	*BDH1*	0·29/0·21	1·54 (1·28 – 1·85)	$3·79\times10^{-6}$	0·34/0·33	0·80 (0·65 – 1·22)	$4·69\times10^{-1}$
rs79095029	5:108855306	C/G	*PJA2*	0·03/0·08	0·35 (0·23 – 0·55)	$5·72\times10^{-6}$	0·15/0·15	1·18 (0·73 – 1·10)	$4·90\times10^{-1}$
rs146261599	5:123600139	T/G	*ZNF608*	0·05/0·02	2·91 (1·82 – 4·64)	$7·64\times10^{-6}$	0·06/0·05	1·44 (0·65 – 3·10)	$3·73\times10^{-1}$
rs181675566	5:168651912	T/C	*SLIT3*	0·04/0·01	3·86 (2·15 – 6·90)	$5·72\times10^{-6}$	0·03/0·04	0·78 (0·27 – 2·25)	$6·51\times10^{-1}$
rs6904835[b]	6:17810195	T/C	*KIF13A*	0·32/0·24	1·50 (1·25 – 1·80)	$9·67\times10^{-6}$	0·45/0·38	1·34 (1·02 – 1·76)	$3·50\times10^{-2}$
rs3173615[bc]	7:12269417	C/G	*TMEM106B*	0·27/0·39	0·55 (0·45 – 0·66)	**$7·81\times10^{-10}$**	0·38/0·48	0·63 (0·48 – 0·82)	$8·36\times10^{-4}$
rs7791726[bc]	7:12283329	G/C	*TMEM106B*	0·26/0·39	0·53 (0·44 – 0·64)	**$1·53\times10^{-10}$**	0·39/0·48	0·66 (0·50 – 0·86)	$2·56\times10^{-3}$
rs1990622[bc]	7:12283787	A/G	*TMEM106B*	0·26/0·39	0·53 (0·44 – 0·65)	**$1·61\times10^{-10}$**	0·38/0·48	0·65 (0·40 – 0·85)	$1·81\times10^{-3}$
rs62443267	7:38153313	C/T	*STARD3NL*	0·19/0·26	0·62 (0·50 – 0·76)	$6·83\times10^{-6}$	0·34/0·39	1·01 (0·75 – 1·35)	$9·55\times10^{-1}$
rs141226303	7:104251213	A/G	*LHFPL3*	0·04/0·01	3·73 (2·11 – 6·59)	$5·61\times10^{-6}$	0·04/0·03	1·36 (0·55 – 3·37)	$5·06\times10^{-1}$
rs3110811[b]	7:135402648	A/G	*SLC13A4*	0·29/0·20	1·55 (1·29 – 1·87)	$3·50\times10^{-6}$	0·29/0·37	0·80 (0·50 – 1·10)	$1·74\times10^{-1}$
rs10101195	8:11623212	C/A	*NEIL2*	0·18/0·26	0·62 (0·51 – 0·77)	$7·50\times10^{-6}$	0·33/0·36	0·88 (0·65 – 1·10)	$3·98\times10^{-1}$
rs36196656[b]	8:21621247	C/A	*GFRA2*	0·46/0·37	1·51 (1·26 – 1·82)	$9·44\times10^{-6}$	0·51/0·45	1·40 (1·08 – 1·82)	$1·08\times10^{-2}$

rs10816848	9:112421435	T/A	*PALM2*	0·42/0·49	0·68 (0·57 – 0·80)	$8{\cdot}74\times10^{-6}$	0·46/0·48	0·20 (0·71 – 1·1)	$5{\cdot}37\times10^{-1}$
rs78781776	11:36466533	A/G	*PRR5L*	0·10/0·05	2·00 (1·47 – 2·72)	$8{\cdot}88\times10^{-6}$	0·18/0·13	1·46 (0·30 – 2·31)	$1{\cdot}03\times10^{-1}$
rs10791882[b]	11:66319313	G/A	*ACTN3*	0·46/0·37	1·49 (1·26 – 1·77)	$5{\cdot}01\times10^{-6}$	0·54/0·47	0·50 (0·73 – 1·25)	$7{\cdot}33\times10^{-1}$
rs10860097	12:97199656	A/T	*NEDD1*	0·05/0·02	3·43 (2·14 – 5·50)	$2{\cdot}88\times10^{-7}$	0·06/0·04	1·75 (0·90 – 3·38)	$9{\cdot}81\times10^{-2}$
rs61965655	13:74712915	T/A	*KLF12*	0·07/0·04	2·33 (1·61 – 3·39)	$8{\cdot}52\times10^{-6}$	0·09/0·09	1·17 (0·66 – 2·08)	$5{\cdot}88\times10^{-1}$
rs847358	14:72780521	G/A	*RGS6*	0·53/0·44	1·46 (1·24 – 1·73)	$7{\cdot}43\times10^{-6}$	0·47/0·49	1·07 (0·82 – 1·38)	$6{\cdot}17\times10^{-1}$
rs12605286	18:41150167	G/A	*SYT4*	0·23/0·31	0·60 (0·48 – 0·74)	$2{\cdot}37\times10^{-6}$	0·37/0·38	1·17 (0·88 – 1·55)	$2{\cdot}93\times10^{-1}$
rs7240419[b]	18:76928989	G/A	*ATP9B*	0·31/0·22	1·62 (1·34 – 1·94)	$3{\cdot}80\times10^{-7}$	0·37/0·35	1·12 (0·83 – 1·51)	$4{\cdot}70\times10^{-1}$
rs6076187	20:24082578	G/A	*FLJ33581*	0·07/0·03	2·47 (1·71 – 3·57)	$1{\cdot}53\times10^{-6}$	0·11/0·07	1·17 (0·63 – 2·15)	$6{\cdot}24\times10^{-1}$

[a]Positions are based on the Human Genome version 38 (hg38), build 38. [b]Variants annotated as eQTL in the GTex database. [c]Variants that are study-wide significant at the replication stage after Bonferroni correction.

Supplementary Table 7: Variants in linkage disequilibrium with rs36196656.

Variants in linkage disequilibrium (r^2>0·8) with the lead variant rs36196656 at the *GFRA2* locus are presented along with the central nervous system chromatin states and marks and eQTL characteristic for *GFRA2* expression levels from GTex and any transcription factor binding sites that are disrupted. All data is extracted from the Haploreg database. GTex p-value for rs36144451 was obtained using rs372215503 as rsID. eQTL=expression quantitative trait loci; E069=cingulate Gyrus; E067=angular gyrus; E073=dorsolateral prefrontal cortex; NA=not available.

Variant	Position[a]	Major/ Minor allele	r2	D'	Annotation	Chromatin States	Chromatin Marks	eQTL pvalue	Transcription Factor binding sites
rs144692383	8:21762856	C/A	1	1	intronic	NA	NA	$3{\cdot}12\times10^{-9}$	DMRT7;Hoxa13;Hoxb13; Hoxc10;Hoxd10;Pou5f1;Sox4
rs36196656	8:21763735	C/A	1	-1	intronic	E073,H3K4me1_Enh; E067,H3K27ac_Enh	NA	$2{\cdot}06\times10^{-9}$	AP-1_disc7;AP-2rep; HNF4;LF-A1;NF-E2_disc4; RORalpha1_3;TR4_disc2
rs36144451	8:21765791	TCTC/T	0·99	1	intronic	NA	E067,H3K27ac_Enh; E073,H3K4me1_Enh	$1{\cdot}80\times10^{-14}$	Evi-1_2;p300_disc5
rs150047054	8:21766061	C/T	0·98	-0·99	intronic	NA	E069,H3K4me1_Enh; E073,H3K4me1_Enh	$1{\cdot}68\times10^{-10}$	Myb_2

Supplementary Table 8: Loci identified in the age at onset association when only including *GRN* patients in replication stage.

Suggestive variants identified in the discovery stage ($p<10^{-5}$) and followed-up in the replication stage, as well the meta-analyses are presented. These results were obtained when only *GRN* mutation carriers were included in the replication stage. Variant p-values and beta values were calculated using an additive genetic model. Minor alleles were treated as effect alleles. The locus name is determined by the closest gene to the significant variant. MAF=minor allele frequency; P=p-value.

Variant	Position[a]	Major/minor allele	Locus name	Discovery			Replication *GRN+*			Meta-analysis[b]		
				MAF patients	Association		MAF patients	Association				
					Beta (95%CI)	P		Beta (95%CI)	P	Beta (95%CI)	P	I^2
rs116316277	2:185834886	C/T	*ZNF804A*	0·03	8·09 (4·72 – 11·46)	$3·58\times10^{-6}$	0·14	1·13 (-3·78 – 6·05)	$6·53\times10^{-1}$	5·86 (3·08 – 8·65)	$3·56\times10^{-5}$	80·9
rs6809184	3:170888198	C/T	*TNIK*	0·05	-6·78 (-9·24 – -4·32)	$1·22\times10^{-7}$	0·11	1·45 (-2·92 – 5·82)	$5·19\times10^{-1}$	-4·80 (-6·94 – -2·65)	$1·18\times10^{-5}$	90·3
rs12189587	6:165332257	C/T	*C6orf11*	0·11	-4·05 (-5·76 – -2·34)	$4·83\times10^{-6}$	0·16	-0·64 (` – 3·35)	$7·54\times10^{-1}$	-3·52 (-5·1 – -1·95)	$1·15\times10^{-5}$	57·9
rs6962939	7:7524226	T/A	*COL28A1*	0·04	-6·02 (-8·61 – -3·43)	$7·00\times10^{-6}$	0·05	-3·25 (` – 4·82)	$4·33\times10^{-1}$	-5·76 (-8·22 – -3·29)	$4·63\times10^{-6}$	0
rs2922921	7:96398079	G/A	*SHFM1*	0·02	9·65 (5·58 – 13·72)	$4·65\times10^{-6}$	0·03	7·80 (-5·56 – 21·16)	$2·61\times10^{-1}$	9·50 (5·6 – 13·39)	$1·76\times10^{-6}$	0
rs77466830	7:151529171	C/A	*PRKAG2*	0·32	2·91 (1·64 – 4·18)	$9·49\times10^{-6}$	0·44	-2·47 (-5·10 – 0·17)	$7·16\times10^{-2}$	1·90 (0·75 – 3·04)	$1·15\times10^{-3}$	92·3
rs9792144	8:53081551	C/G	*ST18*	0·12	3·99 (2·30 – 5·68)	$4·88\times10^{-6}$	0·14	0·48 (-4·45 – 5·40)	$8·50\times10^{-1}$	3·62 (2·03 – 5·22)	$8·67\times10^{-6}$	43·0
rs3922636	8:80383502	G/A	*STMN2*	0·19	3·28 (1·89 – 4·67)	$5·08\times10^{-6}$	0·37	1·43 (-1·77 – 4·62)	$3·85\times10^{-1}$	2·98 (1·71 – 4·25)	$4·36\times10^{-6}$	8·0
rs12943707	17:73317510	C/G	*GRB2*	0·29	-2·80 (-4·00 – -1·60)	$6·40\times10^{-6}$	0·43	0·71 (-2·02 – 3·44)	$6·12\times10^{-1}$	-2·23 (-3·33 – -1·13)	$6·72\times10^{-5}$	81·2
rs1561819	18:2712629	G/A	*SMCHD1*	0·49	-2·41 (-3·46 – -1·36)	$8·96\times10^{-6}$	0·54	1·00 (-1·53 – 3·53)	$4·43\times10^{-1}$	-1·91 (-2·88 – -0·94)	$1·12\times10^{-4}$	83·2
rs6108746	20:10902771	T/C	*JAG1*	0·19	3·54 (2·19 – 4·89)	$4·23\times10^{-7}$	0·22	0·36 (-3·79 – 4·51)	$8·64\times10^{-1}$	3·24 (1·96 – 4·52)	$7·44\times10^{-7}$	51·0
rs6111609	20:17664546	C/A	*RRBP1*	0·22	2·86 (1·61 – 4·11)	$9·83\times10^{-6}$	0·28	1·97 (-1·93 – 5·87)	$3·27\times10^{-1}$	2·78 (1·59 – 3·97)	$4·89\times10^{-6}$	0

[a]Positions are based on the Human Genome version 38 (hg38). [b]For SNPs with $I^2>0.3$, a random effects meta-analysis was also performed. The p-values were generally larger in the random effects meta-analysis, and the results were consistent with the fixed effects, showing that none of the SNPs were significantly associated with the outcome.

Supplementary Table 9: Summary of loci associated with presentation of specific first clinical symptoms and presence of parkinsonism.
Summary of the suggestive variants identified in the discovery stage ($p<10^{-5}$) when four exploratory genome-wide analyses were performed focused on the presence of specific first clinical symptoms (behavior changes, language impairment or memory dysfunction) and presence of parkinsonism at any time during the disease course. Variant p-values and odds ratio were calculated using an additive genetic model. Minor alleles were treated as effect alleles. The locus name is determined by the closest gene to the significant variant. OR=odds ratio; P=p-value.

Variant	Position[a]	Symptom associated	Major/minor allele	Locus name	Annotation	Frequency in patients without associated symptom	Frequency in patients with associated symptom	Association	
								OR (95% CI)	P
rs73191662	7:107837310	Behavior	C/T	*SLC26A3*	intergenic	0·28	0·15	0·37 (-0·04 – 0·78)	$2{\cdot}46\times10^{-6}$
rs1455740	6:126184256	Behavior	G/C	*TRMT11*	intergenic	0·32	0·18	0·41 (0·02 – 0·80)	$6{\cdot}55\times10^{-6}$
rs4888280	16:73431477	Behavior	G/A	*ZFHX3*	intergenic	0·33	0·50	2·00 (1·71 – 2·29)	$7{\cdot}22\times10^{-6}$
rs7673707	4:144284561	Behavior	C/A	*GYPA*	intergenic	0·55	0·38	0·49 (0·17 – 0·81)	$9{\cdot}75\times10^{-6}$
rs12671090	7:157088226	Language	G/C	*UBE3C*	intergenic	0·40	0·23	0·44 (0·09 – 0·79)	$4{\cdot}46\times10^{-6}$
rs10959548	9:1121085	Language	C/T	*DMRT2*	intergenic	0·30	0·47	2·12 (1·79 – 2·45)	$5{\cdot}93\times10^{-6}$
rs10058086	5:72119802	Language	C/T	*MAP1B*	intronic	0·15	0·27	2·56 (2·15 – 2·97)	$9{\cdot}75\times10^{-6}$
rs2732260	8:33431749	Memory	G/A	*FUT10*	intronic	0·03	0·12	5·64 (4·93 – 6·35)	$1{\cdot}37\times10^{-6}$
rs1034797	7:79474577	Memory	T/G	*MAGI2*	intergenic	0·08	0·17	3·84 (3·27 – 4·41)	$4{\cdot}93\times10^{-6}$
rs28649357	17:32040171	Memory	A/G	*LRRC37B*	intronic	0·14	0·28	2·61 (2·20 – 3·02)	$7{\cdot}71\times10^{-6}$
rs13012521	2:152703130	Memory	A/T	*PRPF40A*	intronic	0·09	0·21	2·99 (2·50 – 3·48)	$7{\cdot}94\times10^{-6}$
rs6010061	22:50713296	Memory	T/C	*SHANK3*	intronic	0·64	0·47	0·46 (0·11 – 0·80)	$8{\cdot}27\times10^{-6}$
rs10084892	4:125446831	Memory	T/C	*FAT4*	intronic	0·10	0·22	2·94 (2·47 – 3·41)	$8{\cdot}87\times10^{-6}$
rs3773384	3:374758	parkinsonism	C/T	*CHL1*	intronic	0·42	0·24	0·43 (0·08 – 0·78)	$3{\cdot}51\times10^{-6}$
rs34656641	12:53403452	parkinsonism	A/C	*SP1*	intronic	0·08	0·20	3·26 (2·75 – 3·77)	$7{\cdot}61\times10^{-6}$
rs17833740	14:59154251	parkinsonism	G/T	*DAAM1*	intergenic	0·29	0·46	2·17 (1·84 – 2·50)	$8{\cdot}81\times10^{-6}$

[a]Positions are based on the Human Genome version 38 (hg38).

Supplementary Table 10: rs36196656 minor allele frequency per main country in the discovery stage.
Frequencies of the minor allele of rs36196656 are presented in function of the country of patient origin. MAF=minor allele frequency; OR=odds ratio.

	MAF patients	MAF controls	OR
Italian	0·45	0·35	1·55
USA	0·48	0·35	1·67
Spain	0·43	0·38	1·31
All	0·46	0·36	1·51

REFERENCES

1. Finch N, Baker M, Crook R, et al. Plasma progranulin levels predict progranulin mutation status in frontotemporal dementia patients and asymptomatic family members. *Brain* 2009; **132**(Pt 3): 583-91.
2. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc* 2010; **5**(9): 1564-73.
3. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; **4**: 7.
4. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**(8): 904-9.