

Common variation in the miR-659 binding-site of *GRN* is a major risk factor for TDP43-positive frontotemporal dementia

Rosa Rademakers, Jason L. Eriksen, Matt Baker, Todd Robinson, Zeshan Ahmed, Sarah J. Lincoln, NiCole Finch, Nicola J. Rutherford, Richard J. Crook, Keith A. Josephs, Bradley F. Boeve, David S. Knopman, Ronald C. Petersen, Joseph E. Parisi, Richard J. Caselli, Zbigniew K. Wszolek, Ryan J. Uitti, Howard Feldman, Michael L. Hutton, Ian R. Mackenzie, Neill R. Graff-Radford and Dennis W. Dickson

SUPPLEMENTARY RESULTS

Detailed genetic analyses of GRN genomic region

To study whether rs5848 is the likely functional variant underlying the association with FTL-D-U or whether another genetic variant in linkage disequilibrium (LD) with rs5848 could be responsible for the observed association we performed a detailed genetic analyses of the *GRN* genomic region. First, we determined the LD structure of *GRN* by sequencing 10.2 kb of *GRN* (UCSC genome browser, chr17:39775969-39785997), including 2kb upstream of the non-coding exon 0 and the complete 3'UTR, in 24 US individuals. In total, we identified 27 genetic variants of which 20 were observed more than once in the 48 chromosomes and were considered informative to determine the *GRN* LD haplotype structure (**Supplementary Figure S1**). Haplotype blocks were selected based on the algorithm defined by Gabriel et al. (2002)⁶⁸ using the Haploview software package v2.03. The first 14 SNPs were part of a single ~5kb haplotype block including the *GRN* promoter, 2kb of regulatory sequences, the non-coding exon 0 and intron 0 (defined as *GRN* promoter haplotype block) (**Supplementary Figure S1**). In contrast, little LD was observed in the *GRN* coding region, with the exception of two variants located in close proximity of *GRN* exon 4 (rs34424835 and rs850713). Importantly, none of the variants was in LD with rs5848 ($D' \leq 0.68$; $r^2 \leq 0.37$). In fact, using the genotype information available in the HapMap project we determined that rs5848 is located at the start of another more downstream haplotype block of ~24kb outside of *GRN* (**Supplementary Figure S2**).

We selected 13 SNPs to perform single SNP and haplotype association analyses in the FTL-D-U patient-control series: 8 tagging SNPs identified in the genomic

sequencing analyses that together capture 94% of the genetic diversity in the *GRN* region and 5 SNPs in considerable LD with rs5848 selected from the downstream haplotype block based on the HapMap data (**Figure 1**). Single SNP logistic regression analyses did not identify variants that were more strongly associated with FTLD-U than rs5848, although significant association with 4 of the 5 SNPs from the downstream haplotype block was observed ($p_{\text{adjusted}}=0.04$) (**Supplementary Table S1**). Haplotype association analyses adjusted for age and gender did not reveal significant association in the *GRN* promoter haplotype block, however borderline significance was obtained with Hap C (TCCGinsTGT), carrying the rs5848 risk T-allele, in the complete *GRN* genomic region ($p_{\text{adjusted}}=0.04$) (**Supplementary table S2**). However, when rs5848 was excluded from the analyses significant association was no longer observed ($p_{\text{adjusted}}=0.11$). Similarly, haplotype analyses in the downstream haplotype block showed a significant increase in Hap B (TACACC) from 30.3% in control individuals to 42.3% in FTLD-U patients ($p_{\text{adjusted}}=0.02$), which was no longer observed when rs5848 was excluded. Together these results strongly support the hypothesis that rs5848 is the functional variant in *GRN* associated with FTLD-U, rather than another variant in LD with rs5848 or a haplotype tagged by the T-allele of rs5848.

GRN haplotype diversity in homozygous rs5848 T-allele carriers

To provide additional evidence in favor of rs5848 being the functional variant we determined whether the risk T-allele in our patient population is found on different *GRN* haplotypic backgrounds. We genotyped 8 genetic variants covering the complete 8 kb *GRN* genomic region in a cohort of 35 FTLD patients from the Mayo Clinic FTLD series that were homozygous for the T-allele at rs5848. Within this population haplotype

estimations using Arlequin revealed 11 different haplotypes, including 8 haplotypes with frequencies > 5% (**Supplementary Table S3**). Based on the individual genotype data in our patient cohort, only 2 out of 35 FTLD patients (5.7%) were homozygous for the same haplotype (Hap B), while three FTLD patients were homozygous for other *GRN* haplotypes (one Hap C, one Hap H and one Hap I). When we included 2 simple tandem repeat (STR) markers flanking *GRN* (D17S1860 and GRN_GT15), spanning a total region of 65 kb, haplotype estimation revealed 40 possible haplotypes and none of the 35 patients were homozygous. These data are in agreement with the lack of LD in the *GRN* genomic region and support the hypothesis that the rs5848 T-allele is found on different *GRN* haplotypic backgrounds.

SUPPLEMENTARY TABLES

Supplementary table S1. Single SNP association analyses.

SNP	Genotypes	Controls (N=433)		Patients (N=59)		P _{adjusted}	OR	95% CI
		N	%	N	%			
rs4792937	TT	145	33.5	22	37.3	-	-	-
	CT	210	48.5	24	40.7	0.46	0.79	0.42-1.48
	CC	78	18.0	13	22.0	0.55	1.26	0.59-2.70
rs2879096	CC	244	56.4	34	57.6	-	-	-
	CT	150	34.6	18	30.5	0.72	0.90	0.48-1.66
	TT	39	9.0	7	11.9	0.38	1.50	0.61-3.68
c.-7-320C>G	CC	380	87.8	53	89.8	-	-	-
	CG	48	11.1	6	10.2	0.92	0.95	0.39-2.35
	GG	5	1.2	0	0	n/a	n/a	n/a
rs9897526	GG	337	77.8	47	79.7	-	-	-
	AG	90	20.8	11	18.6	0.60	0.83	0.41-1.68
	AA	6	1.4	1	1.7	0.92	1.12	0.13-9.75
rs34424835	wt/wt	245	56.6	28	47.5	-	-	-
	wt/ins	158	36.5	25	42.4	0.29	1.37	0.77-2.46
	ins/ins	30	6.9	6	10.2	0.25	1.77	0.67-4.64
rs25646	TT	408	94.2	55	93.2	-	-	-
	CT	25	5.8	4	6.8	0.71	1.23	0.41-3.71
	CC	0	0	0	0	n/a	n/a	n/a
c.835+7G>A	GG	376	86.8	51	86.4	-	-	-
	AG	54	12.5	8	13.6	0.69	1.18	0.53-2.64
	AA	3	0.7	0	0	n/a	n/a	n/a
rs5848	CC	199	46.0	21	35.6	-	-	-
	CT	191	44.1	23	39.0	0.74	1.12	0.59-2.10
	TT	43	9.9	15	25.4	0.003	3.18	1.50-6.73
rs708384	CC	157	36.3	17	28.8	-	-	-
	AC	210	48.5	25	42.4	0.75	1.11	0.59-2.14
	AA	66	15.2	17	28.8	0.04	2.20	1.04-4.63
rs850737	TT	148	34.2	18	30.5	-	-	-
	CT	206	47.6	22	37.3	0.72	0.89	0.46-1.72
	CC	79	18.2	19	32.2	0.08	1.90	0.94-3.86
rs5910	GG	151	34.9	18	30.5	-	-	-
	AG	210	48.5	22	37.3	0.73	0.89	0.46-1.73
	AA	72	16.6	19	32.2	0.04	2.13	1.05-4.34
rs5911	AA	152	35.1	18	30.5	-	-	-
	AC	209	48.3	22	37.3	0.75	0.90	0.46-1.75
	CC	72	16.6	19	32.2	0.04	2.14	1.05-4.36
rs850730	GG	152	35.1	18	30.5	-	-	-
	CG	209	48.3	22	37.3	0.75	0.90	0.46-1.75
	CC	72	16.6	19	32.2	0.04	2.14	1.05-4.36

^a Age and gender adjusted odds ratios (OR) are shown for carriers of one rare allele and for carriers of two rare alleles, each compared to homozygote carriers of the frequent allele (reference).

Supplementary table S2. Haplotype analyses in *GRN* genomic region and downstream haplotype block.

Haplotypes ^a		Control individuals N=433 (%)	Patients N=59 (%)	P _{adjusted} ^b
<i>GRN</i> promoter haplotype block: rs4792937 – rs2879096 – c.-7-320C>G				
Hap A	TCC	56.8	56.6	0.81 ^c
Hap B	CTC	26.3	26.1	0.70
Hap C	CCC	10.1	12.4	0.83
Hap D	CCG	5.8	3.1	0.54
Complete <i>GRN</i> genomic region: rs4792937 – rs2879096 – c.7.320C>G – rs9897526 – rs3442485 – rs25646 – c.835+7G>A – rs5848				
Hap A	TCCGwtTGC	26.3	27.0	0.27 ^c
Hap B	CTCGwtTGC	16.2	16.1	0.21
Hap C	TCCGinsTGT	7.1	13.5	0.74
Hap D	TCCGwtTGT	6.7	7.3	0.04
Hap E	TCCGwtTAC	6.4	2.6	0.24
Complete <i>GRN</i> genomic region excluding rs5848: rs4792937 – rs2879096 – c.7.320C>G – rs9897526 – rs3442485 – rs25646 – c.835+7G>A				
Hap A	TCCGwtTG	33.0	32.2	0.51 ^c
Hap B	CTCGwtTG	18.1	20.4	0.66
Hap C	TCCGinsTG	9.6	15.6	0.72
Hap D	TCCGwtTA	6.4	2.6	0.11
<i>GRN</i> downstream block: rs5848 – rs708384 – rs850737 – rs5910 – rs5911 – rs850730				
Hap A	CCTGAG	57.6	47.4	0.11 ^c
Hap B	TACACC	30.3	42.3	0.06
Hap C	CACACC	8.9	6.8	0.02
<i>GRN</i> downstream block excluding rs5848: rs708384 – rs850737 – rs5910 – rs5911 – rs850730				
Hap A	CTGAG	57.8	48.3	0.17 ^c
Hap B	ACACC	39.2	49.1	0.07

^a Haplotypes with an average frequency <5% were excluded from the analysis. The risk T-allele of rs5848 is shown in bold.

^b Simulated p-values corrected for age (age at death in patients, inclusion age for control individuals) and gender.

^c Global p-values.

Supplementary Table S3. Estimated frequencies of 8 kb *GRN* haplotypes containing the rs5848 T-allele.

ID	Haplotype	Frequency (%)
Hap A	T C G G wt T A G T	26.7
Hap B	T C G G ins T G G T	18.4
Hap C	C C G G wt T A G T	11.4
Hap D	C T G G wt T A G T	10.0
Hap E	C T G G ins T G G T	7.9
Hap F	C T G A wt T A G T	7.7
Hap G	T C G A ins T G G T	6.4
Hap H	C T G A ins T G G T	5.9
Hap I	T C G G ins T A G T	2.9
Hap J	C C G A wt C A G T	1.4
Hap K	C C C G wt T A G T	1.4

Note. Haplotypes are composed of rs4792937, rs2879096, c.7-320C>G, rs9897526, rs34424835, rs25646, rs850713, c.835+7G>A and rs5848 (cDNA numbering relative to NM_002087.2 starting at the translation initiation codon. The risk T-allele of rs5848 is shown in bold.

Supplementary Table S4. Predicted miRNA target sites in *GRN* 3'UTR.

miRNA	3'UTR Position		miRNA	3'UTR Position	
	Start	End		Start	End
hsa-miR-924	1	20	hsa-miR-631	216	237
hsa-miR-101	1	14	hsa-miR-502-3p	218	240
hsa-miR-604	7	27	hsa-miR-500*	218	239
hsa-miR-296-3p	10	29	hsa-miR-501-3p	218	240
hsa-miR-516a-5p	12	34	hsa-miR-519e	218	239
hsa-miR-526b	12	34	hsa-miR-518c	219	241
hsa-miR-525-5p	14	34	hsa-miR-339-3p	219	240
hsa-miR-23a*	15	37	hsa-miR-523	219	241
hsa-miR-23b*	15	37	hsa-miR-518e	220	240
hsa-miR-615-5p	16	38	hsa-miR-519d	220	241
hsa-miR-193b*	17	38	hsa-miR-518a-3p	220	241
rno-miR-336	30	50	hsa-miR-525-3p	220	241
hsa-miR-506	35	54	hsa-miR-518f	221	241
hsa-miR-342-5p	58	78	hsa-miR-330-3p	221	242
hsa-miR-659	68	89	hsa-miR-518d-3p	221	241
hsa-miR-198	82	102	hsa-miR-518b	221	241
hsa-miR-615-5p	84	105	hsa-miR-524-3p	221	241
hsa-miR-939	96	119	hsa-miR-135a*	228	248
hsa-miR-920	99	118	hsa-miR-299-3p	232	254
hsa-miR-617	113	135	hsa-miR-888*	243	264
hsa-miR-136	116	138	hsa-miR-506	243	263
hsa-miR-296-5p	121	141	hsa-miR-30e	245	264
hsa-miR-151-3p	128	148	hsa-miR-485-3p	248	270
hsa-miR-187*	157	178	hsa-miR-377	250	270
hsa-miR-96	158	180	hsa-miR-603	250	270
hsa-miR-18b*	159	178	hsa-miR-342-3p	250	271
hsa-miR-744*	161	182	hsa-miR-933	253	274
hsa-miR-644	170	189	hsa-miR-560	255	275
hsa-miR-142-3p	170	192	hsa-miR-574-3p	261	282
hsa-miR-542-5p	183	207	hsa-miR-523	261	283
hsa-miR-483-5p	195	215	hsa-miR-487b	262	283
hsa-miR-674	199	220	hsa-miR-139-3p	264	285
hsa-miR-181a-2*	200	221	hsa-miR-19a	276	299
hsa-miR-220c	207	229	hsa-miR-19b	276	299
hsa-miR-140-3p	211	231	hsa-miR-568	280	300
hsa-miR-588	215	233	hsa-let-7g*	280	300
hsa-miR-193a-3p	215	235	hsa-miR-648	284	302

Note. 3'UTR position starts from first base after TGA stop codon. Target sites predicted based on version 5 of the miRBASE registry (<http://microrna.sanger.ac.uk/>). A total of 70 different miRNAs are predicted to bind to the 304 bp 3'UTR of *GRN*, but only miR-659 is predicted to be affected by rs5848, located 78 nucleotides downstream of the ATG stop codon.

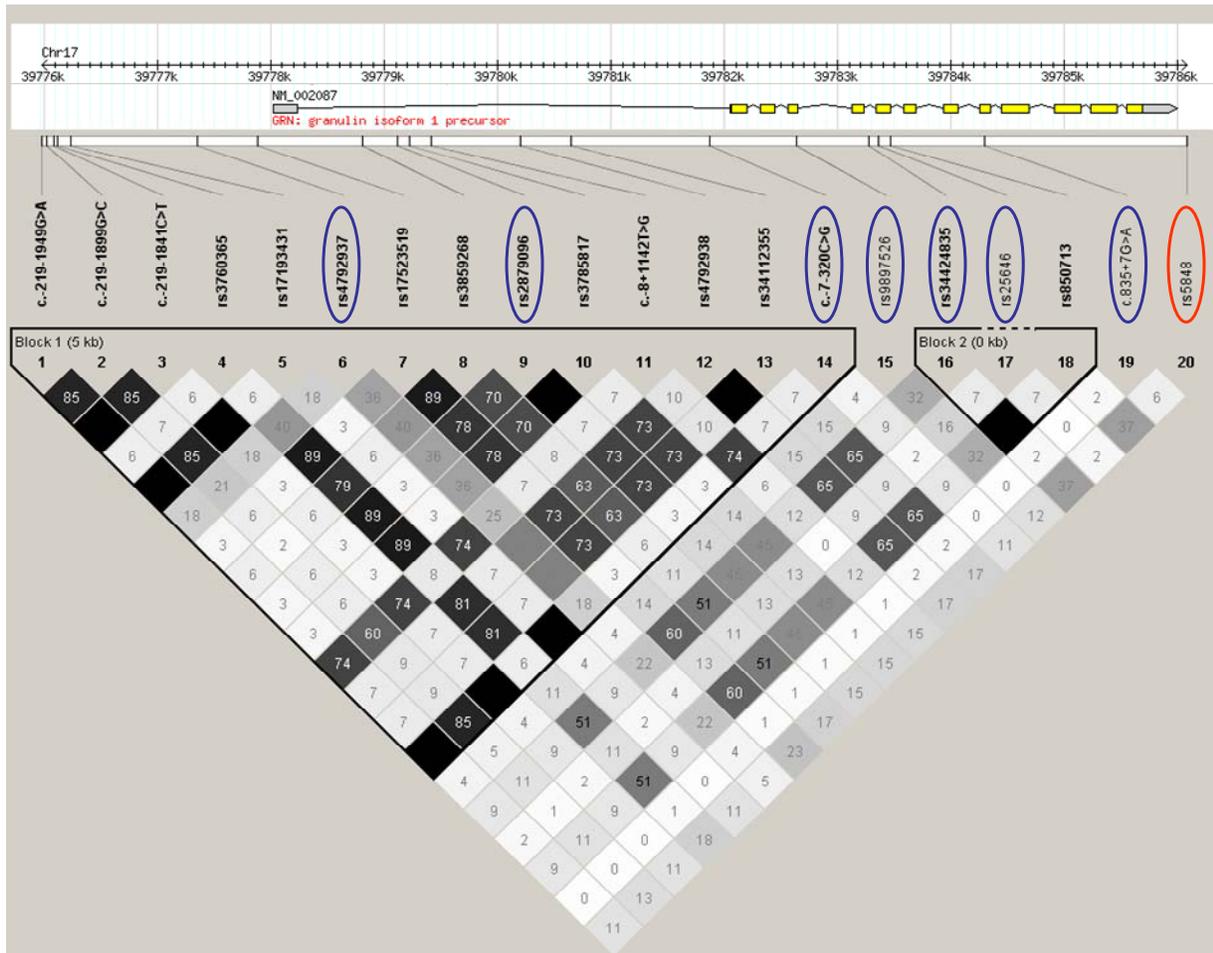
Supplementary Table S5. Information on Taqman genotyping assays.

Position relative to <i>GRN</i>	dbSNP ID	<i>ABI TaqMan Assay</i>	PCR Primers	Reporter primers (Dye)
GRN c.-4671C>T	rs4792937	Custom assay	F: GAGCGAGCCAGCTCAGTAG R: GCTCCCAAAGCGATTCTCCTA	CTGGGATTACAAATGTGAG (VIC) CTGGGATTACAAGTGTGAG (FAM)
GRN c.-2977C>T	rs2879096	C__15835934_10	N/A	N/A
GRN c.-327C>G	N/A	Custom assay	F: CTGCACAGATCAGACCCACAA R: CACTGGGCAGGCTTATGAGA	TCAGGAAGACGTGATTT (VIC) TCAGGAAGACCTGATTT (FAM)
GRN c.264+21G>A	rs9897526	C__2548248_10	N/A	N/A
GRN c.384T>C (D128D)	rs25646	C____14929_10	N/A	N/A
GRN c.835+7G>A	N/A	Custom assay	F: ACGGACCTCCTCACTAAGCT R: GCCCCACCCCTGTATCTG	ACAGGTACCAGAGGCA (VIC) AGGTACCAAAGGCA (FAM)
GRN c.*78C>T	rs5848	C__7452046_20	N/A	N/A
Downstream of <i>GRN</i>	rs708384	C__2259468_10	N/A	N/A
Downstream of <i>GRN</i>	rs850737	C__2548239_10	N/A	N/A
Downstream of <i>GRN</i>	rs5910	C__2548233_10	N/A	N/A
Downstream of <i>GRN</i>	rs5911	C__3017440_10	N/A	N/A
Downstream of <i>GRN</i>	rs850730	C__11880722_20	N/A	N/A

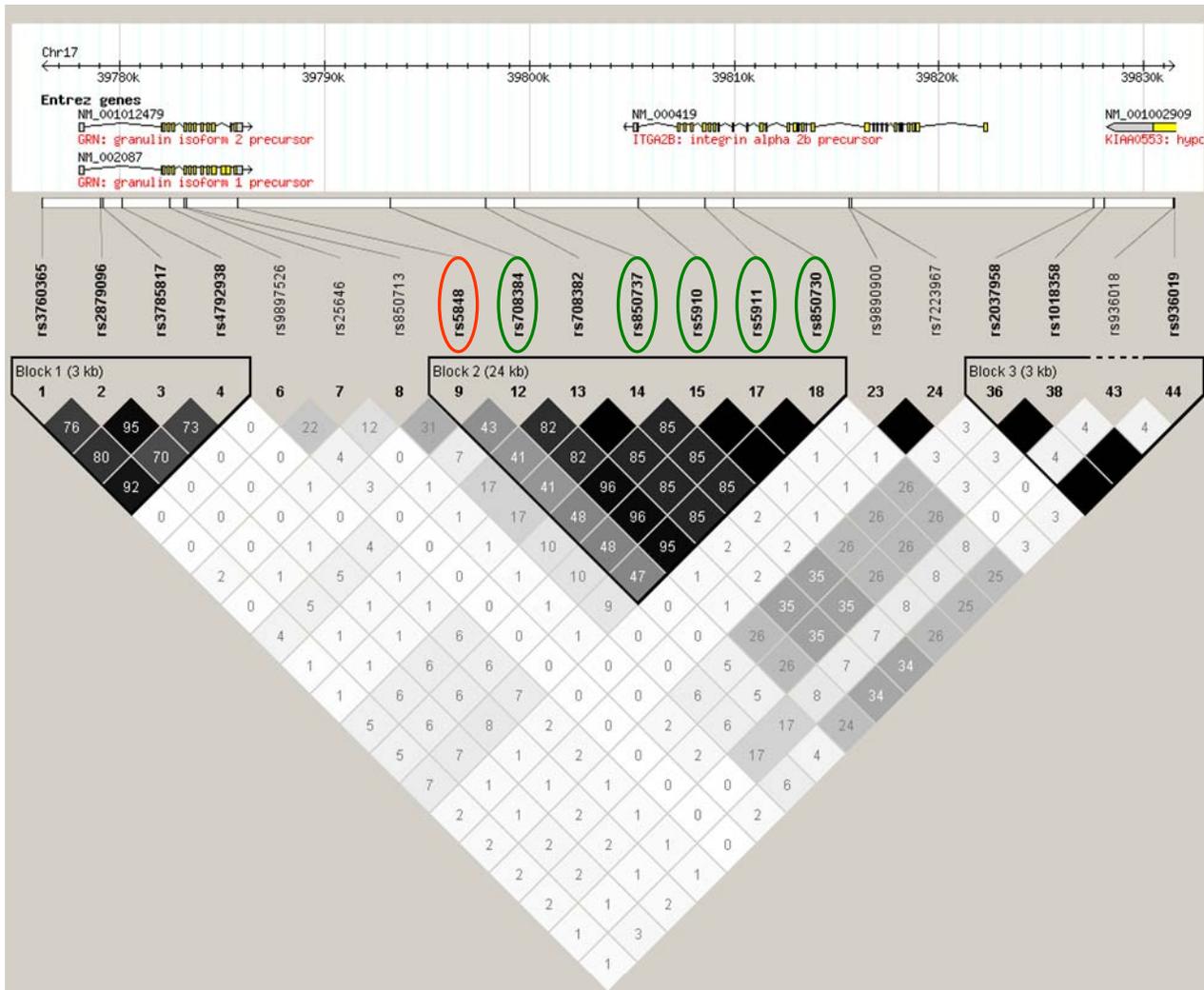
Supplementary Table S6. Information on marker genotyping assays.

Marker	UCSC genome browser position (Mar. 2006)	Primers
rs34424835	chr17:39783076-39783075	F: TAGTGTCACCCTCAAACCCAGT-FAM R: CCACGGAGTTGTTACCTGTGGAC
D17S1860	chr17:39725782-39726150	F: GCGCAATCTCAGGTCA-FAM R: ACCACCGTGTCTGGCTA
GRN_GT15	chr17:39790625-39790652	F: TCCATTTCTCCCTTCTAGTTG-FAM R: AAGTTGAGGCTGCAGGGTG

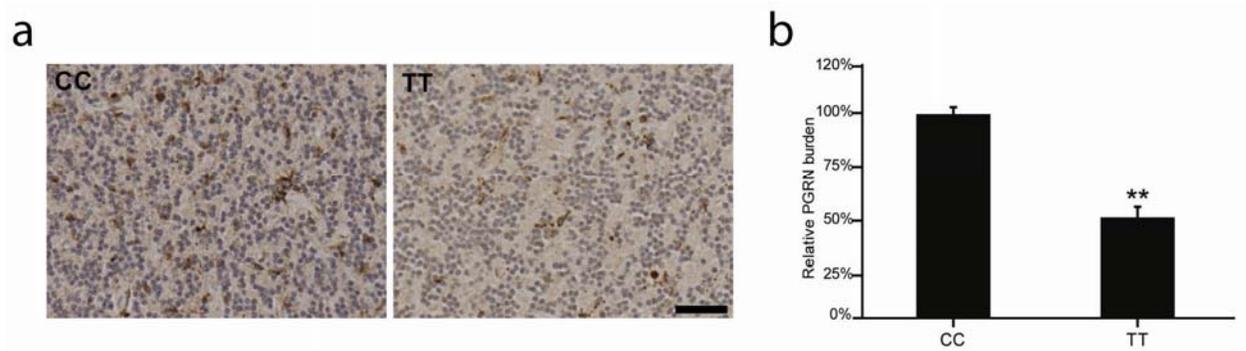
SUPPLEMENTARY FIGURES



Supplementary Figure S1. GRN LD structure analyses in US population. LD relationships between pairs of each of the 20 SNPs in the *GRN* genomic region as determined in 24 US individuals. Each square represents the pairwise strength and significance of LD (r^2). Black indicates no or minimal evidence of historic recombination ($r^2=1$), white indicates no LD ($r^2=0$), and shades of gray are uninformative LD. Tagging SNPs selected for genetic association studies are circled in blue, rs5848 located in the miR-659 binding site in the 3'UTR of *GRN* is circled in red.



Supplementary Figure S2. HapMap LD structure of *GRN* and downstream genomic region. HapMap data showing the LD relationship between pairs of each of the SNPs in *GRN* and its downstream genomic region. Each square represents the pairwise strength and significance of LD (r^2). Black indicates no or minimal evidence of historic recombination ($r^2=1$), white indicates no LD ($r^2=0$), and shades of gray are uninformative LD. SNPs located in the downstream haplotype block that were selected for genetic association studies are circled in green, rs5848 located in the miR-659 binding site in the 3'UTR of *GRN* is circled in red.

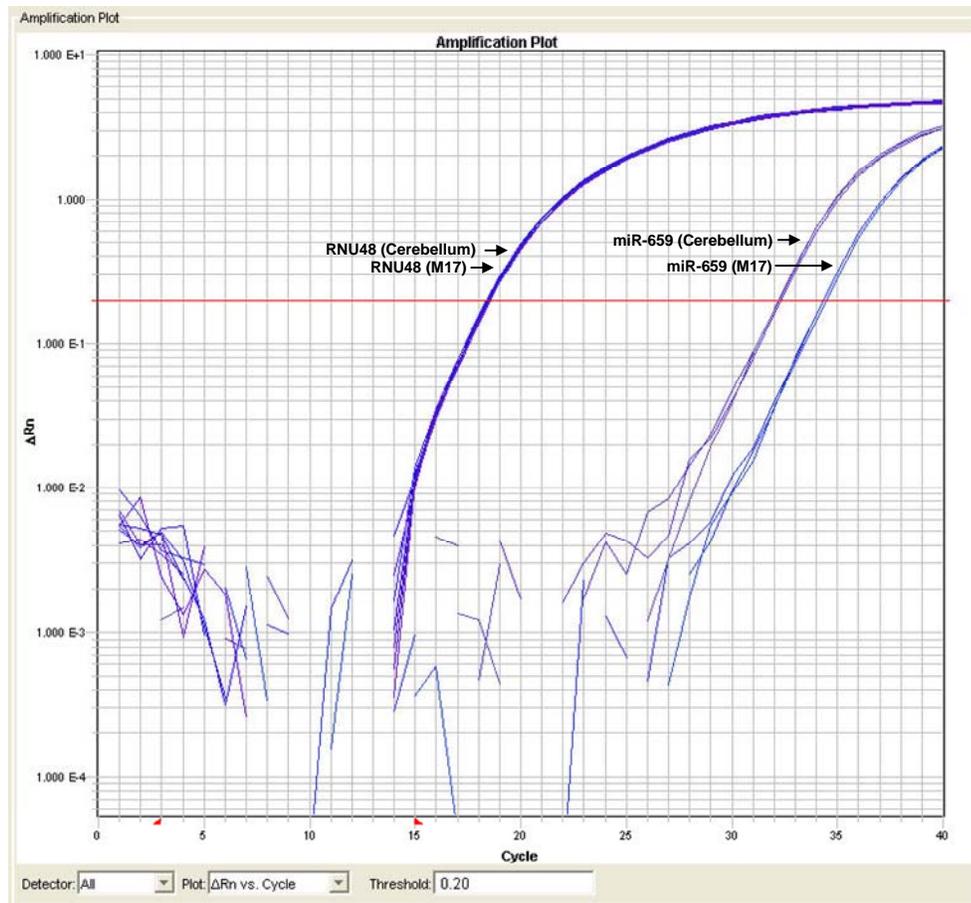


Supplementary Figure S3. Correlation of rs5848 genotypes with GRN

immunohistochemistry. (a) Representative *GRN* immunohistochemistry observed in granular cell layer of the cerebellum in FTLD-U patients homozygous for the wild-type C-allele of rs5848 (left panel) or homozygous for the risk T-allele (right panel).

Scale=50 μ M. (b) Quantification of the GRN burden in granular cell layer of the cerebellum in FTLD-U patients using image analysis (N=5 in each group). GRN

immunoreactivity was ~50% lower in FTLD-U TT carriers compared to CC carriers. (** indicates $p < 0.001$, two-tailed t-test).



Supplementary Figure S4. miR-659 is expressed in M17 cells and human cerebellar brain tissue. Real-time PCR on RNA isolated from M17 cells and human cerebellar tissue revealed positive expression of miR-659. Cycle threshold (C_t) values suggest slightly higher expression of miR-659 in cerebellum ($C_t=32$) compared to M17 cells ($C_t=34$). A probe specific to the ubiquitously expressed RNU48 (small nucleolar RNA, C/D box 48) was included as an endogenous control. Samples were run in triplicate and analyzed using SDSv2.2.2 software.