An AluYb8 mobile element further characterises a risk haplotype of TMEM106B associated in neurodegeneration

Alex Salazar[1], Niccolò Tesi[1,2], Marc Hulsman[1,2], Lydian Knoop[1], Sven van der Lee[1,2,4], Sanduni Wijesekera[1], Jana Krizova[1], and Henne Holstege[1,2,3,4]

1. Section Genomics of Neurodegenerative Diseases and Aging, Department of Clinical Genetics, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands
2. Delft Bioinformatics Lab, Delft University of Technology, Delft, The Netherlands
3. Alzheimer Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The Netherlands.
4. Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands

Abstract

TMEM106B may play a central role in neurodegenerative diseases. Genome-wide association studies have pointed to a risk and protective haplotype in TMEM106B across different study cohorts, including frontotemporal lobar degeneration with TAR DNA binding protein inclusions (FTD-LD) and Alzheimer's disease (AD). However, the exact mechanism in which the risk haplotype of TMEM106B (genetically) exerts its pathological consequence remains elusive. Based on long-read whole-genome sequencing data of 256 individuals, we identified an AluYb8 mobile element that is inserted in the 3'UTR of TMEM106B of the risk haplotype. AluYb8s are known to be active in the human population, and can consequently (dys)regulate transcription and translation of nearby genes. Here, we propose a mechanism for how TMEM106B, PGRN, TDP-43, and age interact in neurodegeneration via a negative-feedback loop that leads to a progressive decline of the endolysosomal system.

Background

TMEM106B likely plays a central role in neurodegenerative diseases. The gene encodes for a 274 amino-acid transmembrane protein highly expressed in the central nervous system, especially in neurons and oligodendrocytes[1,2]. Although its functions are not entirely clear, TMEM106B localizes in the endolysosomal system, where it is involved in maintaining lysosome integrity and function, such as morphology, movement, acidification, and internal enzyme activity[1,2]. Genome-wide association studies have highlighted the presence of two common haplotypes in *TMEM106B*, one associated with increased risk of neurodegenerative diseases, and hence the other with a protective effect[2–10] (see Figure 1 A). The haplotypes are characterised by several common variants (allele-frequency ~50%) that are in strong-to-complete linkage disequilibrium (LD) with each other, and associated with various neurodegenerative diseases[3–8] (see Figure 1 A and B). This includes the rs1990622 variant, first associated with protection against frontotemporal lobar degeneration with TAR DNA binding protein inclusions (FTDLD-TDP) (OR=0.61; $p = 1.08 \times 10^{-11}$)[3]; rs13237518, associated with protection against Alzheimer's disease (OR=0.96, $p = 4.90 \times 10^{-11}$)[11]; and rs3173615, a non-synonymous coding variant (Thr185Ser) in exon 6 of *TMEM106B* . All three variants are in strong linkage disequilibrium (LD) with both each other, and characterize the protective haplotype[9] (see Figure 1 A and B). Notably, the association of the protective haplotype in *TMEM106B* with FTDLD-TDP is more pronounced in carriers of pathogenic *GRN* variants (OR=0.49, $p = 1.34 \times 10^{-9}$)[2,3], a gene encoding for progranulin (PGRN)[12]. Since both genes are essential for lysosome function[12], especially in the degradation of misfolded proteins[13], their effects may converge in the impairment of lysosomal mechanisms.

Previous studies have investigated the functional consequences of the protective haplotype of *TMEM106B*, but there is currently no consensus about its biological mechanism(s). Several reports focused on the possible effect of rs3173615, which substitutes a threonine residue (risk) for a serine residue (protective), often referred to as the T185S substitution[9]. The rs3173615 variant is predicted to be 'benign' by Polyphen, SIFT and REVEL (score of 0.0267), and 'moderate' by CADD (score of 21.4), suggesting that it is unlikely that T185S alters TMEM106B function. However, the T185S substitution juxtaposes an *N*-glycosylation site at position 183, which might affect post-translational protein maturation[14]. Recent studies reported that TMEM106B protein aggregates in the human brain[15–19]. More specifically, neurodegenerative and aged human brains were reported to have aggregate amyloid fibrils composed of the C-terminal fragment (CTF) of TMEM106B[15–19], with the risk haplotype correlating with higher amounts of CTF[20]. Interestingly, cryogenic-electron microscopy showed that the aggregated fibrils have three distinct structural polymorphisms, with their abundance correlating with the risk and protective haplotypes[15,16,19]. More specifically, folds I and III were the most common folds, where fold I predominantly exists in individuals homozygous for the risk allele threonine (TT) and heterozygous carriers (TS), while fold III exists primarily in individuals homozygous for the protective allele serine (SS)[15–19], suggesting that fold III may have lower aggregation potential than folds I and II. Hence, several studies hypothesized that rs3173615 may alter *N*-glycosylation such that the risk haplotype induces a post-translational modification of the TMEM106B protein leading to differential protein folding with higher aggregation potential[14–20]. However, *N*-glycosylation sites commonly occur at a N-X-S/T motif, suggesting that substituting the N_{183}-I_{184}-T_{185} motif to N_{183}-I_{184}-S_{185} motif is unlikely to have any effect[14]. Notably, the protective T185S substitution did not lead

to an observable functional change in a CRISPR-based knockin-mouse model representing SS carriers[21].

Alternatively, the protective haplotype may involve changes in *TMEM016B* expression, but there is currently no consensus that either haplotype leads to increased or decreased gene expression levels or TMEM106B protein abundance[2,5,8,10,21–25]. For example, the rs1990620 variant reportedly influences the recruitment of CCCTC-binding factors (CTCF), affecting long-range chromatin interactions that consequently perturb gene expression of *TMEM106B*[10]. The protective variant rs1990620—and hence, the protective haplotype—reportedly decreases recruitment of CTCF leading to lower gene expression of *TMEM106B*[10]. However, analysis of multi-tissue eQTL databases such UKBEC and GTEx of the rs1990622 and rs3173615 do not support that the protective haplotype decreases *TMEM106B* gene expression[5]. Instead, GTEx indicates an association between the protective haplotype and with *increased TMEM106B* expression in the cerebellum, cortex, and cerebellar hemisphere[5].

Together, the protective haplotype associates with protection against neurodegenerative disease phenotypes[2,6–8,22]. In fact, the protective rs13237518 variant was strongly enriched in genomes from cognitively healthy centenarians compared to middle-aged population matched-individuals and AD patients ($p = 1.16x10^{-2}$, protective odds-ratio of 0.81)[4]. Notably, of all AD GWAS loci, the protective alleles in the *TMEM106B*, *SORT1*, and *GRN*—all involved in lysosome activity—are among the variants most strongly enriched in centenarians (REF), suggesting that lysosome-related pathways are highly important to maintain cognitive health at extreme old age[4].

It is important to note that, thus far, studies have exclusively focused on single nucleotide variants from GWAS and/or short-read whole-genome sequencing data. Hence, larger uncharacterised structural variants in linkage with either the risk or the protective haplotype of *TMEM106B* have been understudied. However, such variants may include the potential drivers of the genetic association with altered disease risk.

Results and discussion

Here, we report the existence of a large, common structural variant (SV) in the 3' untranslated region (UTR) of the TMEM106B risk haplotype. In a PacBio whole-genome long-read sequencing dataset of 92 AD and 117 centenarian genomes, we observed a ~330 bp insertion in the 3' UTR of *TMEM106B* in the risk haplotype (see Table 1). Similarly, analysis of PacBio long-read sequencing data of 43 genomes from the Human Pangenome Reference Consortium (HPRC) representing genomes from global human populations[26] showed that this SV is also common in these genomes (see Table 1).

A closer Investigation of the long-read sequencing data showed that this SV is in LD with the reported variants characterising the risk haplotype of *TMEM106B* (r^2 = 0.98 – 1.0, $p < 2.2x10^{-16}$; see Figure 1C). The SV is ~12.7 Kbp downstream of rs3173615. Interestingly, LD is lower in the HPRC dataset (r^2=0.5-1.0, $p < 2.2x10^{-16}$, see Figure 1C), driven primarily by genomes of African ancestry harbouring alternate genotype combinations, suggesting the presence of additional *TMEM106B* haplotypes outside of European-ancestral populations.

Annotations from the GRCh38 reference genome, along with sequence alignment to the DFAM repeat database[30], showed that the SV is an Alu-mobile element of the Yb8 lineage

sub-family. Alu-elements are retrotransposons, and extensively abundant in the human genome comprising of ~11% of all human genomic sequences[31–33]. They propagate through a "copy-and-paste" mechanism by using reverse-transcription of their mRNA transcripts to then insert themselves at different regions in the genome[31–33]. Alu-elements are primate-specific, and the Alu Yb8-lineage sub-family is not only among the youngest lineage of Alu-elements (integrated ~2.39 million years ago), but also dominantly expanded in humans[31,34,35].

Alu-based SV has been associated in various human diseases, as they can have regulatory effects on nearby genes by influencing alternative splicing, gene expression, and translation[32,36,37]. Additionally, Alu-elements can be transcriptionally active, and expression is particularly induced by cellular stress[38–41], leading to expression and translation dysregulation of nearby genes[32,42]. Specifically, known Alu-elements embedded in UTRs of other genes have been shown to decrease their expression[32]. Together, the mobile element in the *TMEM106B* UTR may similarly decrease its expression and/or translation. This supports observations that the risk haplotype of *TMEM106B*, which harbours the Alu-element, has a relatively decreased gene expression compared to the protective haplotype, in multi-tissue (brain) eQTL and bulk RNA-sequence analysis[5,8].

To support the functional effects of the AluYb8-mobile element on *TMEM106B*, we highlight its intricate synergy with GRN and TDP-43—two central factors of various neurodegenerative diseases (see Figure 2). First, the associated risk in the *TMEM106B* haplotype is more pronounced in pathogenic GRN carriers[2,3], possibly because both genes are essential for lysosome activity, such that impairment of both leads to profound additive effects [12,13,43]. Second, PGRN appears to (indirectly) interact with TDP-43, a protein encoded by the *TARDBP*

gene, which is commonly aggregated in various neurodegenerative diseases[44], and negatively correlates with cognitive function[45]. TDP-43 itself is involved in RNA regulation, influencing transcription, mRNA stabilisation, and splicing[44,46]. TDP-43-depleted iPSC-derived neurons showed increased levels of mis-splicing, leading to intron retention, and the synthesis of alternate, non-functional, protein sequences[46]. Cellular uptake of TDP-43 by endocytosis is required for the regulation TDP-43 toxicity and turnover[49], and deficiency in PGRN has been observed to lead to higher levels of TDP-43 aggregation[47,48]. Additionally, age also leads to higher levels of TDP-43 aggregation[45,47,48], likely due to age-related demethylation of the 3' UTR of *TARDBP*, which reportedly dysregulates its splicing and expression[47].

Interestingly, TDP-43 suppresses Alu-element transcription[42,50]. In fact, dysfunction of TDP-43 leads to higher Alu-transcription[42,50], consequently disrupting the expression and/or translation of nearby genes, potentially due to competition between RNA polymerase II and III[42]. Hence, TDP-43 may actively suppress transcription of the Alu-element in the risk haplotype of *TMEM106B*, and when TDP-43 function is compromised, it can lead to a dysregulated transcription and/or translation of *TMEM106B* relative to the protective haplotype (absence of the Alu-element, see Figure 2). In fact, a recent study using iPSC-derived neurons with non-functional TDP-43 showed that the protein abundance of TMEM106B is reduced relative to iPSC-derived neurons with functional TDP-43[46], although the exact genotype (i.e. Alu-element presence) is unclear. Consequently, dysregulation of TMEM106B via a transcriptionally active Alu-element further compromises lysosome activity, leading to increased neuronal toxicity from protein aggregation via a negative-feedback loop (see Figure 2).

Ultimately, we acknowledge that functional studies are needed to better understand how the Alu-element in the risk haplotype influences gene expression and/or translation of *TMEM106B*. We particularly highlight that the Alu-element alone is likely not enough to lead to the pathological consequences of the risk haplotype of *TMEM106B*, but instead requires specific states of PGRN, TDP-43, and most importantly, advanced age. As such, conflicting reports on transcriptional and translational effects of the risk and protective haplotype may be due, in part, to differences in genotypes that include or exclude the Alu-element[2,5,8,10,21–25]. Furthermore, the reduced LD of the Alu-element along with other known neurodegenerative-associated variants in genomes with African ancestry implies that different haplotype-architectures exist in *TMEM106B*, further highlighting the importance of identifying AD-risk variants outside of individuals form European-ancestry[51]. Additionally, if the Alu-element does indeed (partly) drive molecular risk of the *TMEM106B* risk haplotype, exploiting existing evolutionary-selected defense mechanisms against retrotransposon transcription may provide strategies for development of therapeutics[52]. Nevertheless, further characterisations of the TMEM106B haplotypes should account for the regulatory consequences from the presence/absence of the Alu-element.

Conclusion

An AluYb8 mobile element residing in the 3'UTR of *TMEM106B* is commonly present in the human population, and is in complete linkage with the risk haplotype of *TMEM106B* in genomes from Alzheimer's disease patients and cognitively healthy centenarians from The Netherlands. This linkage is reduced in genomes with African-based ancestry. Future functional studies aiming to characterise the risk and protective haplotypes should therefore

account for the effect of the AluYb8 mobile element. Notably, a negative feedback loop mediated by the Alu-element provides a mechanism for how TMEM106B, PGRN, TDP-43, and age may interact: pathogenic *GRN* variation leads to compromised endolysosomal activity, in turn, increasing aggregation of TDP-43 due to incomplete degradation. Increased age leads to hypomethylation of *TARDP* (gene encoding TDP-43), which dysregulates TDP-43 protein synthesis. As TDP-43 supresses Alu-element transcription, reduced functionality of TDP-43 leads to increased Alu-element transcription, consequently dysregulation transcription and/or translation of TMEM106B, further compromising endolysosome activity.

Methods

Sample populations

Alzheimer disease (AD) individuals in this study are patients from the VU University Medical Center, Amsterdam, and were diagnosed as previously described Tesi *et al.* 2023[4]. Similarly, cognitive healthy centenarians in this study are individuals from the 100-plus Study[29], who are at least 100 years old. All individuals from both study-cohorts provided informed consent for this study.

PacBio long-read whole-genome sequencing

Genomic DNA of AD and centenarian individuals were extracted from whole blood, and gDNA integrity was assessed with Qubit system 2 (Qubit dsDNA HS Assay Ref. No. Q32854) and Nanodrop (NanoDrop 2000). gDNA was sheared with the Megaruptor from Diagenode with a long hydropore cartridge (Cat. No. E07010003) and a specialized shearing protocol. Qubit

dsDNA BR assay was used to quantify DNA concentration. All libraries were prepared using SMRTbell Template Prep Kit 2.0 (Pacific Biosciences Ref. No. 100-938-900) following manufacturer's instructions. Briefly, DNA was sheared using the Megaruptor, after which size selection was performed using the BluePippin high-pass DNA size selection (sage Science BLU0001; 0.75% Agarose Cassettes, with Marker S1). The range selection mode was set from 8,000-50,000 Kbp. DNA damage repair after size selection was performed with the reaction volumes described in the protocol. Qubit dsDNA HS assay was used to quantify DNA concentration. Sequencing primers v3 were annealed to the SMRTbell library. Polymerase was bound using the Sequel Binding Kit 2.0. SMRTbell complexes were purified using MicroSpin columns S-400 for diffusion loading. Sequencing reaction was performed using the Sequel sequencing plate 2.0 on a SMRT-cell 1M chip. All samples were sequenced at the Department of Clinical Genetics at the Amsterdam University Medical Center (AUMC) using a Pacific Biosciences (PacBio) Sequel II System. Overall, 92 AD genomes were sequenced at a median coverage of 18.2x, with median read-lengths of 14.8 Kbp; 117 centenarian genomes were sequenced at a median coverage of 20.1x, with median read-lengths of 14.7 Kbp.

PacBio HiFi data processing

After sequencing, raw reads were collected and analyzed through an in-house pipeline (freely available at https://github.com/holstegelab/snakemake_pipeline). Briefly, raw reads first underwent PacBio's *ccs* algorithm (v6.0.0, https://github.com/PacificBiosciences/ccs) to generate high-fidelity (HiFi) reads with custom parameters (min-passes 0, min-rq 0, keeping kinetics information). Using these parameters, we retained both *HiFi reads* (read-quality >99%, number of passes >3), and lower-quality *non-HiFi reads*, which are typically excluded when running the *ccs* algorithm with default settings. HiFi and non-HiFi reads were then

aligned to GRCh38 (patch release 14) reference genomes using *pbmm2* (v1.3.0, https://github.com/PacificBiosciences/pbmm2) with suggested alignment presets for HiFi and non-HiFi reads, generating one BAM-file per SMRT-cell. When available, individual BAM-files from the same sample were merged into a single BAM-file using *samtools*[53] (v1.13).

We used HiFi-based publicly available whole-genomes to contextualise our findings to global human genetic diversity. Specifically, we downloaded 47 GRCh38-aligned BAM-files from the Human Pangenome Research Consortium (HPRC)[26] from their respective AWS S3 buckets (https://github.com/human-pangenomics/HPP_Year1_Data_Freeze_v1.0). Long-reads were extracted from the BAM-files using *samtools* and re-aligned to GRCh38 using *minimap2*[54] (version 2.21-r1071).

Variant calling and linkage analysis

Structural variants across AD, centenarians, and HPRC genomes were identified using *sniffles2*[55] (version 2.0.6) with default parameters. Manual inspection of the structural variants in the TMEM106B gene using the UCSC genome browser (https://genome.ucsc.edu/) showed a large deletion overlapping the 3' UTR nearby known GWAS variants. The deletion exactly overlapped an AluYb8 mobile element based on annotations from the UCSC genome browser. This AluYb8-element (chr7:12242091-12242407) along with GWAS variants rs1020004, rs3173615, rs13237518, rs1990622, rs1990621, and rs1990620 were then locally assembled across all AD, centenarian, and HPRC genomes using the *"assemble"* command from *otter* (version 0.1.0) with the "*-F 0.2*" parameter. The locally assembled sequence of the deletion were aligned to DFAM[30] (release 3.7), confirming the structural variant is indeed an AluYb8 mobile element. The locally assembled sequences were then used to jointly genotype

all variants using the "*genotype"* command from *otter* with the *"-e 0.5 -m 10*", ensuring that only the genotypes of variants with at least 10x of total coverage per region per sample were reported. Pairwise linkage-disequilibrium (LD) was then performed using the "genetics" package (version 1.3.8.1.3) in RStudio (version 2023.06.0+421). GWAS-variants and transcripts in TMEM106B were visualized with *ggbio*[56] (version 1.42.0) using the *EnsDB.Hsapiens.v86* package[28], as well as LocusZoom[27] using the "*Alzheimer_NatGenet2022*" study.

Acknowledgements

The authors are grateful to all study participants, their family members, the participating medical staff, general practitioners, pharmacists and all laboratory personnel involved in patient diagnosis, blood collection, blood biobanking, DNA preparation and sequencing. Part of the work in this manuscript was carried out on the Cartesius supercomputer, which is embedded in the Dutch national e-infrastructure with the support of SURF Cooperative. Computing hours were granted to H. H. by the Dutch Research Council ('100plus': project# vuh15226, 15318, 17232, and 2020.030; 'Role of VNTRs in AD'; project# 2022.31, 'Alzheimer's Genetics Hub' project# 2022.38). This work is supported by a VIDI grant from the Dutch Scientific Counsel (#NWO 09150172010083) and a public-private partnership with TU Delft and PacBIo, receiving funding from ZonMW and Health~Holland, Topsector Life Sciences & Health (PPP-allowance), and by Alzheimer Nederland WE.03-2018-07. H.H., S.L., are recipients of ABOARD, a public-private partnership receiving funding from ZonMW (#73305095007) and Health~Holland, Topsector Life Sciences & Health (PPP-allowance; #LSHM20106). S.L. is recipient of ZonMW funding (#733050512). H.H. was supported by the Hans und Ilse Breuer

Stiftung (2020), Dioraphte 16020404 (2014) and the HorstingStuit Foundation (2018). Acquisition of the PacBio Sequel II long read sequencing machine was supported by the ADORE Foundation (2022).

HH has a collaboration contract with Muna Therapeutics, PacBio, Neurimmune and Alchemab. She serves in the scientific advisory boards of Muna Therapeutics and is an external advisor for Retromer Therapeutics. Research of Alzheimer center Amsterdam is part of the neurodegeneration research program of Amsterdam Neuroscience. Alzheimer Center Amsterdam is supported by Stichting Alzheimer Nederland and Stichting Steun Alzheimercentrum Amsterdam. The clinical database structure was developed with funding from Stichting Dioraphte.

Author contributions

Conceived the study: HH; Wrote the manuscript: AS, HH; Patient selection: NT, HH, Patient blood collection and sequencing, MG, MH, DD, SW, JK LK; Data management: NT, MH, SvdL, Bioinformatic analysis: AS.

References

1. Feng, T., Lacrampe, A. & Hu, F. Physiological and pathological functions of TMEM106B: a gene associated with brain aging and multiple brain disorders. *Acta Neuropathol. (Berl.)* **141**, 327–339 (2021).
2. Nicholson, A. M. & Rademakers, R. What we know about TMEM106B in neurodegeneration. *Acta Neuropathol. (Berl.)* **132**, 639–651 (2016).
3. Van Deerlin, V. M. *et al.* Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. *Nat. Genet.* **42**, 234–239 (2010).
4. Tesi, N. *et al. Cognitively Healthy Centenarians are genetically protected against Alzheimer's disease specifically in immune and endo-lysosomal systems*. http://medrxiv.org/lookup/doi/10.1101/2023.05.16.23290049 (2023) doi:10.1101/2023.05.16.23290049.
5. Hu, Y. *et al.* rs1990622 variant associates with Alzheimer's disease and regulates TMEM106B expression in human brain tissues. *BMC Med.* **19**, 11 (2021).
6. Tropea, T. F. *et al.* TMEM106B Effect on cognition in Parkinson disease and frontotemporal dementia. *Ann. Neurol.* **85**, 801–811 (2019).
7. Mao, F. *et al.* TMEM106B modifies TDP-43 pathology in human ALS brain and cell-based models of TDP-43 proteinopathy. *Acta Neuropathol. (Berl.)* **142**, 629–642 (2021).
8. Li, Z. *et al.* The TMEM106B FTLD-protective variant, rs1990621, is also associated with increased neuronal proportion. *Acta Neuropathol. (Berl.)* **139**, 45–61 (2020).
9. Van Der Zee, J. *et al.* TMEM106B is associated with frontotemporal lobar degeneration in a clinically diagnosed patient cohort. *Brain* **134**, 808–815 (2011).

10. Gallagher, M. D. *et al.* A Dementia-Associated Risk Variant near TMEM106B Alters Chromatin Architecture and Gene Expression. *Am. J. Hum. Genet.* **101**, 643–663 (2017).

11. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).

12. Rhinn, H., Tatton, N., McCaughey, S., Kurnellas, M. & Rosenthal, A. Progranulin as a therapeutic target in neurodegenerative diseases. *Trends Pharmacol. Sci.* **43**, 641–652 (2022).

13. Butler, V. J. *et al.* Progranulin Stimulates the In Vitro Maturation of Pro-Cathepsin D at Acidic pH. *J. Mol. Biol.* **431**, 1038–1047 (2019).

14. Lang, C. M. *et al.* Membrane Orientation and Subcellular Localization of Transmembrane Protein 106B (TMEM106B), a Major Risk Factor for Frontotemporal Lobar Degeneration. *J. Biol. Chem.* **287**, 19355–19365 (2012).

15. Schweighauser, M. *et al.* Age-dependent formation of TMEM106B amyloid filaments in human brains. *Nature* **605**, 310–314 (2022).

16. Jiang, Y. X. *et al.* Amyloid fibrils in FTLD-TDP are composed of TMEM106B and not TDP-43. *Nature* **605**, 304–309 (2022).

17. Chang, A. *et al.* Homotypic fibrillization of TMEM106B across diverse neurodegenerative diseases. *Cell* **185**, 1346-1355.e15 (2022).

18. Fan, Y. *et al.* Generic amyloid fibrillation of TMEM106B in patient with Parkinson's disease dementia and normal elders. *Cell Res.* **32**, 585–588 (2022).

19. Fan, Y. *et al.* Newly identified transmembrane protein 106B amyloid fibrils in the human brain: pathogens or by-products? *Ageing Neurodegener. Dis.* **3**, 4 (2023).

20. T Vicente, C. *et al.* C-terminal TMEM106B fragments in human brain correlate with disease-associated *TMEM106B* haplotypes. *Brain* awad133 (2023) doi:10.1093/brain/awad133.

21. Cabron, A.-S. *et al.* Lack of a protective effect of the Tmem106b "protective SNP" in the Grn knockout mouse model for frontotemporal lobar degeneration. *Acta Neuropathol. Commun.* **11**, 21 (2023).

22. Harding, S. R. *et al.* The TMEM106B risk allele is associated with lower cortical volumes in a clinically diagnosed frontotemporal dementia cohort. *J. Neurol. Neurosurg. Psychiatry* **88**, 997–998 (2017).

23. Nicholson, A. M. *et al.* TMEM106B p.T185S regulates TMEM106B protein levels: implications for frontotemporal dementia. *J. Neurochem.* **126**, 781–791 (2013).

24. Ren, Y. *et al.* TMEM106B haplotypes have distinct gene expression patterns in aged brain. *Mol. Neurodegener.* **13**, 35 (2018).

25. Yang, H.-S. *et al.* Genetics of Gene Expression in the Aging Human Brain Reveal TDP-43 Proteinopathy Pathophysiology. *Neuron* **107**, 496-508.e6 (2020).

26. Wang, T. *et al.* The Human Pangenome Project: a global resource to map genomic diversity. *Nature* **604**, 437–446 (2022).

27. Boughton, A. P. *et al.* LocusZoom.js: interactive and embeddable visualization of genetic association study results. *Bioinformatics* **37**, 3017–3018 (2021).

28. Rainer, J. EnsDb.Hsapiens.v86. (2017) doi:10.18129/B9.BIOC.ENSDB.HSAPIENS.V86.

29. Holstege, H. *et al.* The 100-plus Study of cognitively healthy centenarians: rationale, design and cohort description. *Eur. J. Epidemiol.* **33**, 1229–1249 (2018).

30. Storer, J., Hubley, R., Rosen, J., Wheeler, T. J. & Smit, A. F. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob. DNA* **12**, 2 (2021).

31. Deininger, P. Alu elements: know the SINEs. *Genome Biol.* **12**, 236 (2011).

32. Chen, L.-L. & Yang, L. ALU ternative Regulation for Gene Expression. *Trends Cell Biol.* **27**, 480–490 (2017).

33. Solyom, S. & Kazazian, H. H. Mobile elements in the human genome: implications for disease. *Genome Med.* **4**, 12 (2012).

34. Feusier, J. *et al.* Discovery of rare, diagnostic AluYb8/9 elements in diverse human populations. *Mob. DNA* **8**, 9 (2017).

35. Carter, A. B. *et al.* Genome-wide analysis of the human Alu Yb-lineage. *Hum. Genomics* **1**, 167 (2004).

36. Deininger, P. L. & Batzer, M. A. Alu Repeats and Human Disease. *Mol. Genet. Metab.* **67**, 183–193 (1999).

37. Gussakovsky, D. & McKenna, S. A. Alu RNA and their roles in human disease states. *RNA Biol.* **18**, 574–585 (2021).

38. Hwang, Y. E., Baek, Y. M., Baek, A. & Kim, D.-E. Oxidative stress causes Alu RNA accumulation via PIWIL4 sequestration into stress granules. *BMB Rep.* **52**, 196–201 (2019).

39. Li, T.-H. & Schmid, C. W. Differential stress induction of individual Alu loci: implications for transcription and retrotransposition. *Gene* **276**, 135–141 (2001).

40. Pandey, R., Mandal, A. K., Jha, V. & Mukerji, M. Heat shock factor binding in Alu repeats expands its involvement in stress through an antisense mechanism. *Genome Biol.* **12**, R117 (2011).

41. Larsen, P. A., Hunnicutt, K. E., Larsen, R. J., Yoder, A. D. & Saunders, A. M. Warning SINEs: Alu elements, evolution of the human brain, and the spectrum of neurological disease. *Chromosome Res.* **26**, 93–111 (2018).

42. Morera, A. A., Ahmed, N. S. & Schwartz, J. C. TDP-43 regulates transcription at protein-coding genes and Alu retrotransposons. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1862**, 194434 (2019).

43. Clayton, E. L. & Isaacs, A. M. Progranulin and TMEM106B: when two become wan. *EMBO Rep.* **21**, e51668 (2020).

44. Jo, M. *et al.* The role of TDP-43 propagation in neurodegenerative diseases: integrating insights from clinical and experimental studies. *Exp. Mol. Med.* **52**, 1652–1662 (2020).

45. Zhang, M. *et al.* The correlation between neuropathology levels and cognitive performance in centenarians. *Alzheimers Dement.* alz.13087 (2023) doi:10.1002/alz.13087.

46. Seddighi, S. *et al. Mis-spliced transcripts generate* de novo *proteins in TDP-43-related ALS/FTD*. http://biorxiv.org/lookup/doi/10.1101/2023.01.23.525149 (2023) doi:10.1101/2023.01.23.525149.

47. Liu, G. *et al.* Endocytosis regulates TDP-43 toxicity and turnover. *Nat. Commun.* **8**, 2092 (2017).

48. Zhang, J. *et al.* Neurotoxic microglia promote TDP-43 proteinopathy in progranulin deficiency. *Nature* **588**, 459–465 (2020).

49. Beel, S. *et al.* Progranulin reduces insoluble TDP-43 levels, slows down axonal degeneration and prolongs survival in mutant TDP-43 mice. *Mol. Neurodegener.* **13**, 55 (2018).

50. Li, W., Jin, Y., Prazak, L., Hammell, M. & Dubnau, J. Transposable Elements in TDP-43-Mediated Neurodegenerative Disorders. *PLoS ONE* **7**, e44099 (2012).

51. Sherva, R. *et al.* African ancestry GWAS of dementia in a large military cohort identifies significant risk loci. *Mol. Psychiatry* **28**, 1293–1302 (2023).

52. Goodier, J. L. Restricting retrotransposons: a review. *Mob. DNA* **7**, 16 (2016).

53. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

54. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

55. Smolka, M. *et al. Comprehensive Structural Variant Detection: From Mosaic to Population-Level*. http://biorxiv.org/lookup/doi/10.1101/2022.04.04.487055 (2022) doi:10.1101/2022.04.04.487055.

56. Yin, T., Cook, D. & Lawrence, M. ggbio: an R package for extending the grammar of graphics for genomic data. *Genome Biol.* **13**, R77 (2012).

Table 1. An AluYb8 mobile element is commonly present in the 3' UTR of TMEM106B. Observed relative frequencies of homozygous non-carriers (0/0), heterozygous (0/1), and homozygous carriers (1/1) of the AluYb8 mobile elements in long-read sequencing datasets of Alzheimer's disease (AD) and cognitively healthy centenarian genomes from The Netherlands (100-plus study), as well as genetically diverse genomes from the Human Pangenome Research Consortium (HPRC).

Study	Cohort	Genomes	0/0	0/1	1/1
100-plus					
	AD	88	17% (15)	44% (39)	39% (34)
	centenarians	112	27.5% (31)	37.5% (42)	35% (39)
HPRC		47	17% (8)	43% (20)	40% (19)

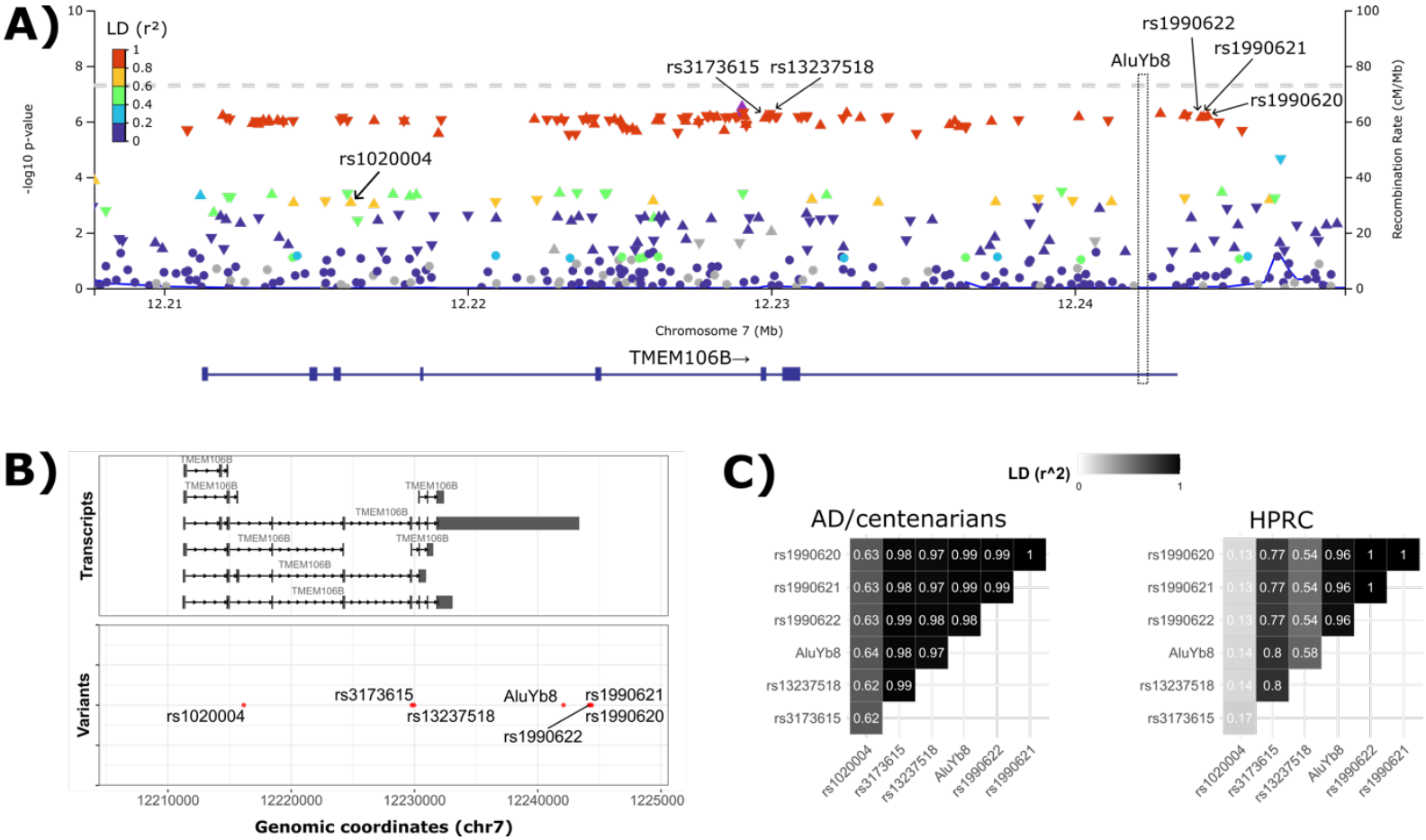

Figure 1. An AluYb8 mobile element further characterizes risk and protective haplotypes of *TMEM106B* associated in neurodegeneration. (A) Modified overview of the TMEM106B gene from LocousZoom[27] displaying genome-wide associated (GWAS) variants (triangles) from the Bellenguez *et al.* 2022 study[11]; the colour of each variant represents LD (r^2) values based on genomes with European ancestry. **(B)** Annotated transcripts of TMEM106B from Ensembl[28] (top), along with the corresponding location of the AluYb8-element and known GWAS-variants (bottom). **(C)** LD (r^2) values based on 92 AD genomes and 117 centenarian genomes from The Netherlands[4,29] (left), and 43 genomes from Human Pangenome Reference Consortium (HPRC)[26] (right).

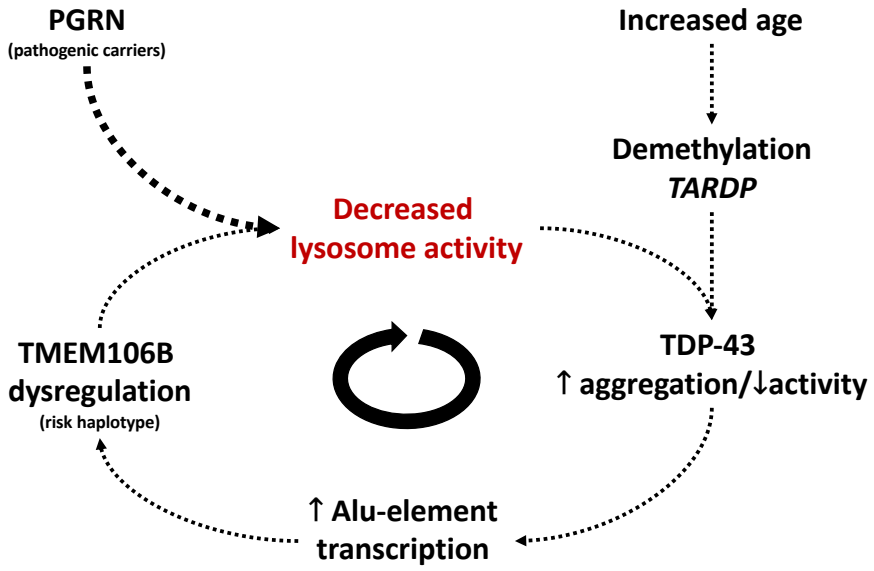

Figure 2. Endolysosomal processes depend on GRN, TMEM106B, TDP-43 and advanced age interactions. Transcription/translation of *TMEM106B* risk haplotype requires functional TDP-43. Therefore, the risk haplotype—when TDP-43 becomes compromised/aggregated with age—may contribute to a progressive dysregulation of endolysosomal function via a negative-feedback loop. Since pathogenic PGRN variants further affect TDP-43 integrity, carriers may be especially vulnerable for impaired endolysosomal function.