Article

Amyloid fibrils in FTLD-TDP are composed of TMEM106B and not TDP-43

https://doi.org/10.1038/s41586-022-04670-9

Received: 25 September 2021

Accepted: 22 March 2022

Published online: 28 March 2022

Check for updates

Yi Xiao Jiang[1,2,5], Qin Cao[1,2,3,5], Michael R. Sawaya[1,2], Romany Abskharon[1,2], Peng Ge[1,2], Michael DeTure[4], Dennis W. Dickson[4], Janine Y. Fu[1], Rachel R. Ogorzalek Loo[1], Joseph A. Loo[1] & David S. Eisenberg[1,2]✉

Frontotemporal lobar degeneration (FTLD) is the third most common neurodegenerative condition after Alzheimer's and Parkinson's diseases[1]. FTLD typically presents in 45 to 64 year olds with behavioural changes or progressive decline of language skills[2]. The subtype FTLD-TDP is characterized by certain clinical symptoms and pathological neuronal inclusions with TAR DNA-binding protein (TDP-43) immunoreactivity[3]. Here we extracted amyloid fibrils from brains of four patients representing four of the five FTLD-TDP subclasses, and determined their structures by cryo-electron microscopy. Unexpectedly, all amyloid fibrils examined were composed of a 135-residue carboxy-terminal fragment of transmembrane protein 106B (TMEM106B), a lysosomal membrane protein previously implicated as a genetic risk factor for FTLD-TDP[4]. In addition to TMEM106B fibrils, we detected abundant non-fibrillar aggregated TDP-43 by immunogold labelling. Our observations confirm that FTLD-TDP is associated with amyloid fibrils, and that the fibrils are formed by TMEM106B rather than TDP-43.

Pathological deposits of amyloid proteins are associated with more than 50 systemic and neurodegenerative diseases[5,6], and include amyloid-β and tau in Alzheimer's disease, α-synuclein in Parkinson's disease, and amylin (also known as hIAPP) in type II diabetes[7]. Recent advances in cryo-electron microscopy (cryo-EM) techniques have enabled near-atomic resolution of structures of amyloid fibrils extracted from patients with these diseases[8–13]. Structural information from ex vivo fibrils offers insights into the molecular signatures of disease and can guide the design of therapeutic interventions intended to prevent, delay or reverse the aggregation of proteins in amyloid disorders.

FTLD causes presenile dementia in around 81 out of every 100,000 people between the ages of 45 and 64[14]. FTLD presents clinically as disorders of social behaviour and language ability[2]. The major subtype of FTLD is characterized by neuronal inclusions containing TDP-43, termed FTLD-TDP, which accounts for around 50% of all FTLD cases[15]. FTLD-TDP is further classified into types A to E according to morphology and neuroanatomical distribution of TDP-43 inclusions[16,17]: each type is associated with diverse clinical symptoms[18]. Although the defining neuropathological feature of FTLD-TDP is immunoreactivity for deposits of ubiquitinated, hyperphosphorylated TDP-43[3,19], amyloid fibrils in FTLD-TDP brains have not been studied structurally. Here we extracted amyloid fibrils from four donors diagnosed with FTLD-TDP types A, B, C and D, and determined 12 near-atomic resolution structures by cryo-EM. Using atomic model building, mass spectrometry and western blot, we confirmed that these fibrils are formed by TMEM106B, a protein previously identified as a genetic risk factor for FTLD-TDP[4]. Our study reveals the amyloidogenic nature of TMEM106B in FTLD-TDP, and focuses attention on a protein not previously associated with amyloid conditions[20–22].

Fibril extraction from human brains

The Mayo Clinic Brain Bank contributed frozen brain tissues of patients with FTLD-TDP (40 donors) and age-matched, non-FTLD-TDP controls (8 donors) (Extended Data Fig. 1a and Extended Data Table 1). We inspected the detergent-insoluble fractions of these samples by negative stain transmission election microscopy. In most donors with FTLD-TDP (38 out of 40 donors), we observed amyloid fibrils as well as non-fibrillar aggregates (donors 1–4 in Fig. 1a and 34 other donors in Extended Data Fig. 2). By contrast, no fibrils were observed in the eight non-FTLD-TDP donors (donor 5 in Fig. 1a and 7 other donors in Extended Data Fig. 2). We then selected four donors neuropathologically confirmed to be FTLD-TDP types A, B, C and D (donors 1–4, Table 1, Extended Data Fig. 3) for further study. We collected cryo-EM data on samples from each of the four donors and determined amyloid fibril structures (Fig. 2, Table 2, Extended Data Figs. 4–6 and Extended Data Table 2). Three fibril polymorphs—PM1, PM2 and PM3—were identified in each cryo-EM dataset (Extended Data Fig. 4a–c), with similar distributions in all four FTLD-TDP donors (Fig. 2 and Extended Data Fig. 4d).

Identification of the fibril-forming protein

We used our highest resolution cryo-EM map (2.9 Å, from donor 1, PM1) to begin model building. We initially presumed that the fibrils were composed of TDP-43; however, extensive incompatibility between the side chain density in the map and the sequence of TDP-43 invalidated this presumption. To identify the protein in the fibrils, we constructed query sequences by selecting amino acids that best fit the PM1 density

[1]Departments of Chemistry and Biochemistry and Biological Chemistry, UCLA-DOE Institute, and Molecular Biology Institute, UCLA, Los Angeles, CA, USA. [2]Howard Hughes Medical Institute, UCLA, Los Angeles, CA, USA. [3]Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Ministry of Education, Shanghai Jiao Tong University, Shanghai, China. [4]Mayo Clinic, Jacksonville, FL, USA. [5]These authors contributed equally: Yi Xiao Jiang, Qin Cao. ✉e-mail: david@mbi.ucla.edu

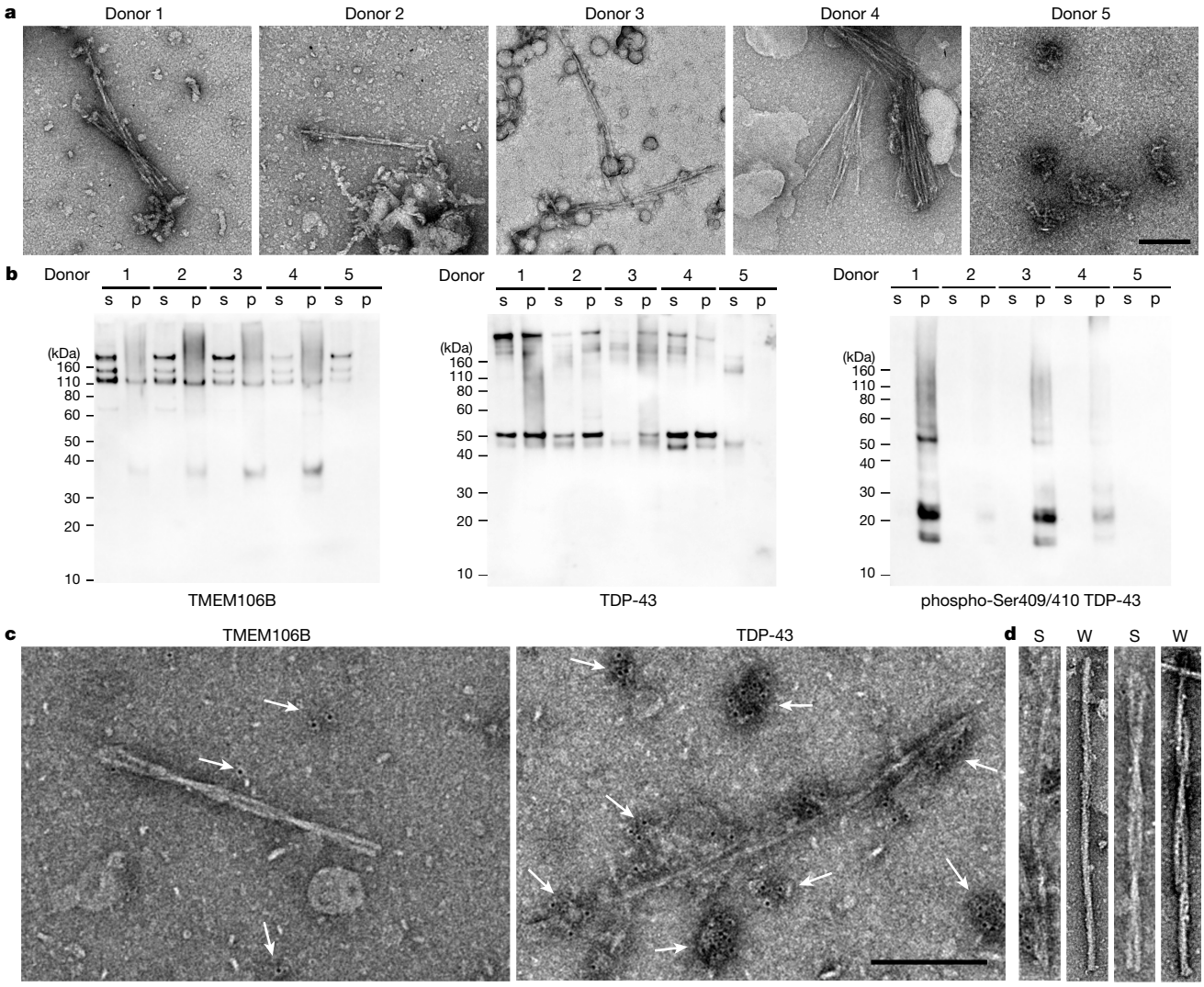


Fig. 1 | Characterization of TMEM106B and TDP-43 in FTLD-TDP brain extracts. a, Negative-stain transmission electron microscopy (TEM) images of sarkosyl-insoluble fractions from FTLD-TDP donors 1 to 4 and a non-FTLD-TDP donor (donor 5). Scale bar, 200 nm. **b**, Western blots of sarkosyl-soluble (s) and sarkosyl-insoluble (p) fractions from donors 1 to 5 probed with TMEM106B (left), TDP-43 (middle) or phospho-Ser409/410 TDP-43 (right) antibody. The original, uncropped blots are shown in Supplementary Fig. 3. **c**, Representative negative-stain TEM images of the sarkosyl-insoluble fraction from FTLD-TDP donor 1 showing immunogold labelling with TMEM106B (left) and TDP-43 (right) antibodies. White arrows label immunogold beads, showing low-density, probably nonspecific binding of TMEM106B antibody (left) and high-density, non-fibrillar TDP-43 aggregates (right). Scale bar, 200 nm. **d**, Comparison of fibrils extracted from FTLD-TDP donor 1 using the sarkosyl-based protocol (S; same fibrils as in **c**) and a water-based protocol (W) imaged by negative-stain TEM. PM1 (left two images) and PM2 or PM3 (right two images; PM2 and PM3 are difficult to distinguish by negative-stain TEM). Fibrils extracted from both protocols exhibit similar morphology.

without regard to their similarity to known protein sequences. We built two models with opposite directionality, thereby obtaining two query sequences (Extended Data Fig. 7 and Methods). A search for human proteins similar to these two queries revealed only one hit: residues 121–254 of TMEM106B, a lysosomal transmembrane protein.

Several observations lend confidence that TMEM106B is indeed the protein building block of these fibrils: (1) the sequence of TMEM106B fits the map well, leaving no unexplained density (Fig. 2); (2) density protruding from Asn145, Asn151, Asn164 and Asn183 (Extended Data Fig. 8a, b) accounts for the known *N*-linked glycosylation sites of TMEM106B[23] (Fig. 3a); (3) the only density that connects two side chains is accounted for by a disulfide bond between Cys214 and Cys253 (Fig. 2); and (4) a previous study revealed that genetic variants of TMEM106B are associated with FTLD-TDP[4]. No other protein could explain these features so completely. The models for PM2 and PM3 of donor 1 were built using rigid body fits of the PM1 model into the density maps, with minor adjustments and refinement applied subsequently.

Cryo-EM maps of all three polymorphs make it clear that the fibrils are composed of a proteolytic fragment rather than the full-length TMEM106B. The amino terminus of the core, Ser120, is buried by residues Tyr158, Val160, His239 and Glu241, leaving no space for residues preceding Ser120 (Figs. 2, 3b). Consequently, proteolytic cleavage between residues Arg119 and Ser120 must precede fibril formation (Fig. 3a, Discussion). By contrast, the C terminus of the core, Gly254, is solvent-exposed, allowing the possibility that the remaining C-terminal 20 residues are attached but disordered. Notably, the range of the ordered core (residues 120–254) coincides closely with the definition of the lumenal domain of TMEM106B (Fig. 3a), which is known to be released from the lysosomal membrane by proteolytic cleavage at an undetermined location in the vicinity of Ser120[24].

Table 1 | Information on FTLD-TDP and non-FTLD-TDP donors

Donor ID	Gender	Age	Braak	PathDx (type)	FHx	Brain region	Distribution (%)			TMEM106B gene
							PM1	PM2	PM3	
Donor 1	Male	65	2	FTLD-TDP (C)	False	MF	78.1	17.8	4.1	Heterozygous T185S
Donor 2	Female	76	3	FTLD-TDP (B)	False	MF	84.9	10.0	5.1	Heterozygous T185S
Donor 3	Male	86	0	FTLD-TDP (A)	False	MF	85.3	12.8	1.8	Wild type
Donor 4	Female	64	3	FTLD-TDP (D)	True	MF	85.3	10.7	4.0	Heterozygous T185S
Donor 5	Male	59	0	Normal	False	MF	n.a.	n.a.	n.a.	n.a.

PathDx, pathological diagnosis; FHx, family history; MF, medial frontal gyrus; n.a., not available.

Stable fold of TMEM106B fibrils

Protofilaments of all three polymorphs from donor 1 exhibit the same fold (Fig. 2, bottom left) despite their different types of bundling. A single protofilament constitutes PM1, and a pair of protofilaments constitute PM2 and PM3 and intertwine with distinct symmetries: C_2 in PM2 and pseudo-2_1 in PM3. Moreover, the protofilament interfaces of PM2 and PM3 differ. In PM2, side chains of Arg180 point towards residual density at the centre of the interface, whereas PM3 features a hydrophobic interaction between pairs of Met207 and Tyr209 residues (Fig. 2 and Extended Data Fig. 9). The only notable structural differences between the folds in the three polymorphs are confined to the PM3 protofilament interface—the rotamers of Glu206, Met207, Tyr209 and Tyr211 in PM3 vary from those in PM1 and PM2 (Fig. 2 and Extended Data Fig. 9c).

The conserved fold of TMEM106B contains 18 β-strands (Fig. 3b), varying in length (from 3 to 15 residues) and curvature. Notably, 18 β-strands exceeds the number typically associated with amyloid fibrils[25]; this abundance of β-strands contributes to an unusually large and stable fibril core. Indeed, these 18 β-strands mate together in pairs, forming multiple steric zippers that strongly stabilize the fibril (Supplementary Fig. 1). Compared with the average of the 63 known amyloid fibril structures (in 2021), the 'golf course' fold of TMEM106B is nearly three times more stable according to solvation energy estimates[25] (−62 versus −22 kcal mol^{-1} per chain). These crude estimates of energetic stability may explain the fibrils' resistance to 2% sarkosyl and 1% sodium dodecyl sulfate (SDS) during extraction. This stability suggests that formation of TMEM106B fibrils in the brain is irreversible, which is consistent with their possible pathogenic roles in FTLD-TDP.

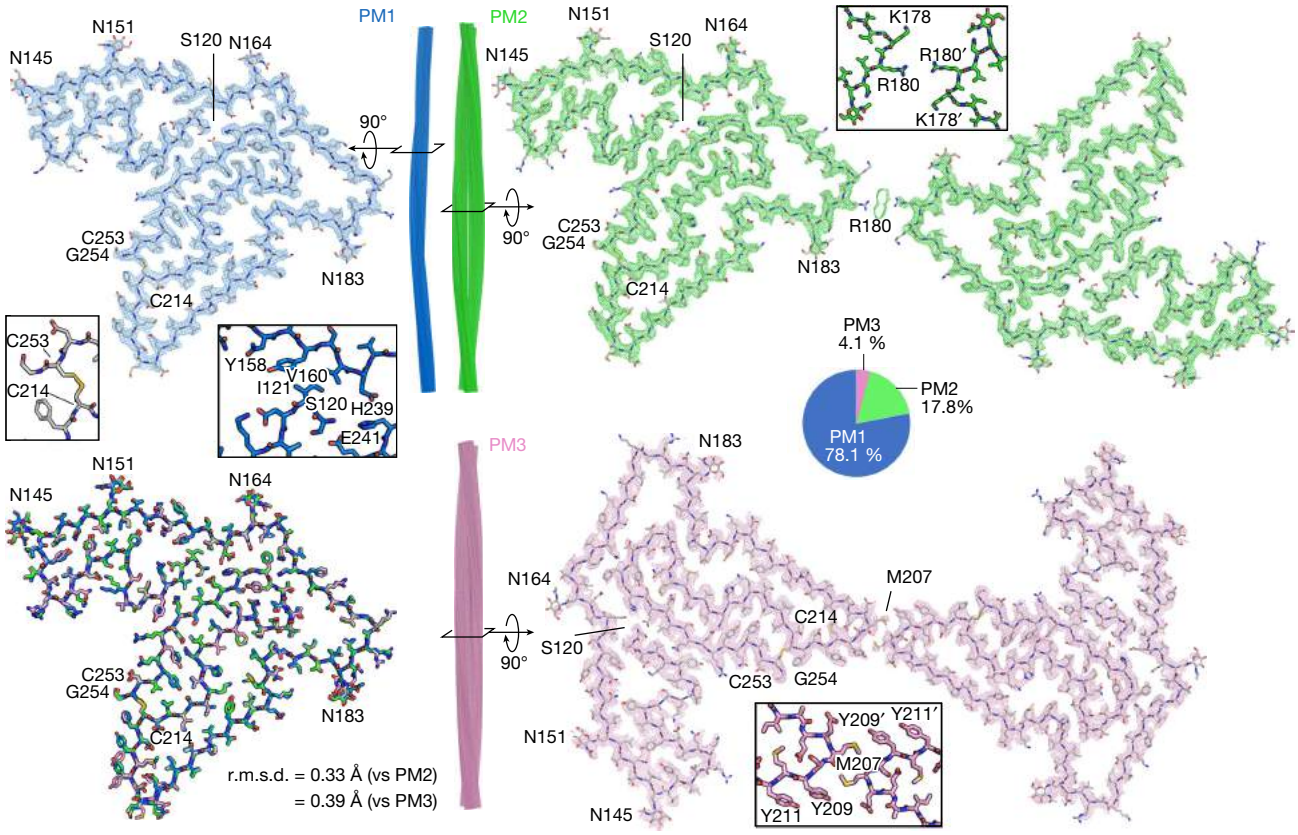


Fig. 2 | Cryo-EM structures of TMEM106B fibrils from FTLD-TDP donor 1. Cryo-EM maps and atomic models of one cross-sectional layer of PM1 (blue, top left), PM2 (green, top right) and PM3 (pink, bottom right). Side view of fibril reconstructions of PM1, PM2 and PM3 (middle). Superimposition of a single chain from PM1, PM2 and PM3 (bottom left). Enlarged views of the C214–C253 disulfide bond, N-terminal Ser120 in PM1, and dimer interfaces of PM2 and PM3 are shown as insets. The distribution of PM1, PM2 and PM3 polymorphs in the donor 1 dataset is shown as a pie chart.

Clues to the origin of the TMEM106B fibrils are given by our observation that four uncomplemented acid side chains face inward in the fibril core structure: Glu163, Glu205, Asp212 and Glu246 (Fig. 3b). Such uncomplemented, buried negative charges are rare among pathogenic amyloids, as electrostatic repulsion between closely spaced molecules with the identical charge would weaken the fibril. However, given that TMEM106B is a lysosomal protein and this is a lumenal domain, the pH of the environment might be low enough to protonate and neutralize some inward-facing acids. Neutralization of these negative charges may help to overcome the barrier to fibril nucleation.

FTLD-TDP donors share a conserved fold

Our study of four patients with FTLD-TDP contributes to mounting evidence that protein conformation is conserved among patients diagnosed with the same amyloid disease. In fact, it has been proposed that protein conformation could form the basis for classifying disease[26]. Accordingly, we found that the golf course fold of TMEM106B is conserved among all four FTLD-TDP donors and constitutes the basis of all three polymorphs. However, we do observe a small but significant shift in the protofilament–protofilament interface of PM2 among donors. The interface from donors 1 and 4 (types C and D) is centred on Lys178, whereas the interfaces from donors 2 and 3 (types B and A) are centred on Arg180 (Extended Data Fig. 6 and Extended Data Fig. 9a, b). It is unclear whether these structural differences are linked to the different FTLD-TDP subtypes or other patient attributes, such as age. PM2 accounts for around 10% of TMEM106B fibrils in all four donors

Genotyping suggests that donor 3 contains wild-type *TMEM106B*, whereas donors 1, 2 and 4 harbour heterozygous Thr185Ser variants in their *TMEM106B* gene (Table 1). Mass spectrometry identified a peptide with the Thr185Ser variant in donor 1 (Extended Data Fig. 10 and Extended Data Table 3). The Thr185Ser variant was detected in a genome-wide association study of patients with FTLD-TDP[4], and a follow-up study suggested that the Thr185Ser variant protects against FTLD-TDP[27]. The Ser185 isoform is associated with lower protein expression and is more rapidly degraded than the wild-type Thr185 isoform, which may contribute to its protective mechanism[28]. The observation that all four donors possess fibrils with the same fold suggests that this fold is compatible with wild-type TMEM106B protein; however, we cannot confirm whether the heterozygous donor's fibrils contain only wild-type TMEM106B, only variant TMEM106B, or a mixture of both. Our maps confirm that residue 185 is within the protofilament core (Fig. 2a and Extended Data Fig. 8c), but even the highest-resolution cryo-EM map (donor 1 PM1) is insufficient to distinguish among Thr, Ser or an average of both sidechains (Extended Data Fig. 8c). Since the Thr/Ser185 side chain faces the solvent with no contributing interactions in the fibril structure, we believe that the fibril fold observed here can accommodate both wild-type and Thr185Ser TMEM106B proteins.

Table 2 | Cryo-EM data collection, refinement and validation statistics of FTLD-TDP donor 1

	PM1 (EMD-24953, PDB 7SAQ)	PM2 (EMD-24954, PDB 7SAR)	PM3 (EMD-24955, PDB 7SAS)
Data collection and processing			
Magnification	×81,000	×81,000	×81,000
Voltage (kV)	300	300	300
Electron exposure (e^-/Å^2)	36	36	36
Defocus range (μm)	0.5-5.0	0.5-5.0	0.5-5.0
Pixel size (Å)	1.1	1.1	1.1
Symmetry imposed	C_1	C_2	C_1
Helical rise (Å)	4.91	4.905	2.452
Helical twist (°)	−0.42	179.59	179.79
Initial particle images (no.)	883,793	883,793	883,793
Final particle images (no.)	60,806	16,883	9,203
Map resolution (Å)	2.9	3.2	3.7
FSC threshold	0.143	0.143	0.143
Map resolution range (Å)	200-2.9	200-3.2	200-3.7
Refinement			
Initial model used (PDB code)	De novo	7SAQ	7SAQ
Model resolution (Å)	3.1	3.4	4.3
FSC threshold	0.5	0.5	0.5
Model resolution range (Å)	200-3.1	200-3.4	200-4.3
Map sharpening *B* factor (Å^2)	124	93	98
Model composition			
Nonhydrogen atoms	5,710	11,420	11,420
Protein residues	675	1,350	1,350
Ligands	20	40	40
B factors (Å^2)			
Protein	7.10	6.57	29.53
Ligand	11.64	10.61	52.32
R.m.s. deviations			
Bond lengths (Å)	0.005	0.004	0.004
Bond angles (°)	0.793	0.732	0.818
Validation			
MolProbity score	2.2	2.3	2.4
Clashscore	11.9	15.1	18.6
Poor rotamers (%)	0	0	0
Ramachandran plot			
Favoured (%)	87.2	88.6	88.0
Allowed (%)	12.8	11.4	12.0
Disallowed (%)	0	0	0

TMEM106B and TDP-43 in patient extracts

To verify that the protein in FTLD-TDP fibrils is indeed TMEM106B, we performed western blotting of patient extracts with a C-terminal TMEM106B antibody. We observed a TMEM106B-positive band at around 35 kDa (along with other high molecular weight bands; Methods) in sarkosyl-insoluble fractions from most of the FTLD-TDP donors that we examined, but not in extracts from the non-FTLD-TDP donors (Fig. 1b and Extended Data Fig. 1b), consistent with our observations by electron microscopy that TMEM106B fibrils are present only in patients with FTLD-TDP (Fig. 1a and Extended Data Fig. 2). These results support the hypothesis that the aggregation of TMEM106B is associated with FTLD-TDP. The 35-kDa band could be the fibril-forming cleavage product (TMEM106B 120–274), a 155-residue peptide with five glycans. To further confirm the presence of TMEM106B, we performed mass spectrometry on gel slices around the 35 kDa region and detected peptide fragments from only the lumenal domain (Extended Data Fig. 10 and Extended Data Table 3), consistent with our cryo-EM maps.

We attempted immunogold labelling of patient-extracted fibrils using three different TMEM106B antibodies (Methods), but none of them labelled the fibrils (Fig. 1c, left). This was anticipated, since the epitopes of the antibodies are either cleaved before fibril formation or buried inside the fibril core. It may also explain why TMEM106B aggregation has not been detected by immunohistochemical staining of patient brains. Future work is needed to develop antibodies that probe TMEM106B fibrils, either by binding to the fibril core, or by recognizing residues 255–274 of TMEM106B, which are not included in the fibril core and may form an accessible fuzzy coat.

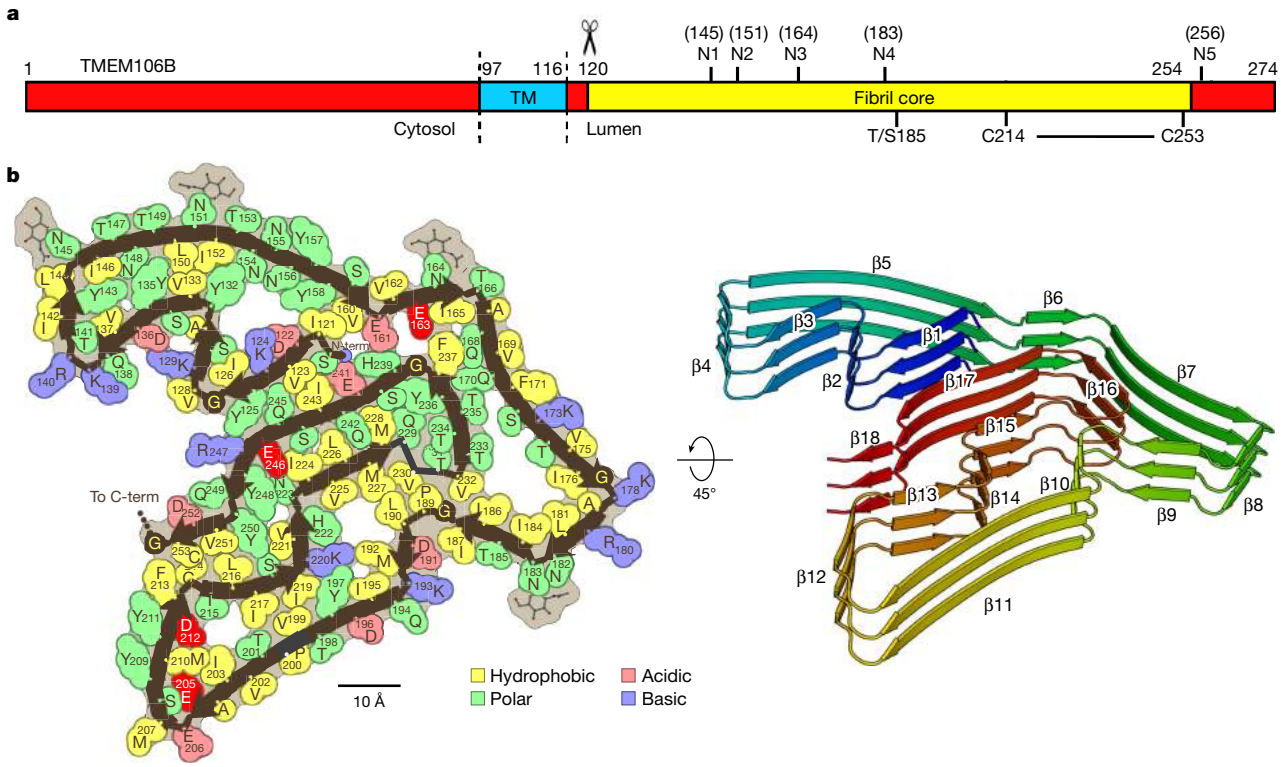


Fig. 3 | The conserved golf-course-like fold of TMEM106B fibrils. **a**, Schematic of TMEM106B showing the cytosolic, transmembrane (TM) and lumenal domains, as well as the fibril core of the structures observed here. Labelled are the five known glycosylation sites (N1–N5), the heterozygous Thr185Ser variant and the C214–C253 disulfide bond. **b**, Space-filling representation (left) and cartoon model (right) of the conserved protofilament core represented by PM1 from FTLD-TDP donor 1. Uncomplemented acidic residues buried inside the fibril core are highlighted in bright red. N-term, N terminus; C-term, C terminus.

To verify the presence of TDP-43 aggregates, we also probed patient extracts with TDP-43 antibodies. Immunoblotting revealed that the FTLD-TDP-associated, phosphorylated form of TDP-43[29] is present with TMEM106B in the sarkosyl-insoluble fractions of FTLD-TDP donors but not the non-FTLD-TDP control (Fig. 1b). These results are consistent with a previous study[30] and suggest that TDP-43 forms aggregates in patients with FTLD-TDP. Immunogold labelling of sarkosyl-insoluble fractions with the TDP-43 antibody revealed that non-fibrillar aggregates were present and identified them as TDP-43 (Fig. 1c, right). Fibrils were also detected but were not labelled; these were presumably composed of TMEM106B.

Discussion

A decade ago, TMEM106B was identified as a genetic risk factor for FTLD-TDP in a genome-wide association study; ever since, its role in the disease has remained unknown. Our finding of TMEM106B amyloid fibrils in ex vivo patient brains gives us a firm, new grasp on this association. Indeed, TMEM106B amyloid appears to be a signature of FTLD-TDP as evidenced by its presence in four different FTLD-TDP types, both with and without the Thr185Ser variant, and in both familial and sporadic cases (Table 1). Each of the three polymorphic structures we observed from each patient reveals TMEM106B folded in the same golf-course-like fold. This discovery directs new focus on TMEM106B and the role that its aggregation may have on the aggregation of TDP-43, previously recognized as "the major disease protein"[31]. It remains unknown whether TMEM106B fibrils are pathogenic and, if so, whether their pathogenicity arises from a gain or loss-of-function. Alternatively, they may be benign by-products downstream of an undiscovered, primary pathological pathway. Understanding these possible pathogenic roles of TMEM106B fibrils will require studies from a larger cohort.

Our structures reveal that proteolytic cleavage happens prior to, and is essential for, formation of the core of TMEM106B fibrils (Figs. 2, 3a)—this suggests that protease inhibition might offer a strategy for FTLD-TDP intervention. TMEM106B undergoes lumenal domain shedding, in which the amyloid core-forming C-terminal fragment is released within the lumen of the lysosome by a resident protease—the cleavage site and the identity of the protease is still unknown[24]. Our structural evidence that cleavage occurs between Arg119 and Ser120 (Supplementary Fig. 2) is consistent with the propensity of many known proteases to cleave between arginine and serine residues. The specificity of the TMEM106B cleavage site required to enable fibril growth would seem to be even greater than that of amyloid precursor protein[32], another single-pass transmembrane protein whose proteolytic cleavage by secretases is involved in the growth of amyloid fibrils in Alzheimer's disease. In the case of β-amyloid, cleavage at various sites leads to amyloid formation, whereas in TMEM106B, burial of its cleaved terminus in the fibril core constrains amyloid-competent fragments to those cleaved between Arg119 and Ser120.

Evidence of a ligand-binding site in the TMEM106B PM2 polymorph identifies another potential factor affecting fibril formation. The interface between PM2 protofilaments features positively charged residues (Lys178 or Arg180) that point towards an unmodelled residual density (Extended Data Fig. 9b). This positively charged environment suggests that the ligand bound here is negatively charged. Our observation that this ligand accommodates two slightly different protofilament arrangements of PM2 (Extended Data Fig. 9a) is consistent with the small size of the interface and the dominance of electrostatic forces in its stabilization. Of note, residual densities are also found in similar positions in PM1 and PM3 (Extended Data Fig. 8a), and may correspond to the same ligand. However, the resolution of the residual densities does

not allow us to identify the ligand. We cannot rule out the possibility that the density corresponds to sarkosyl—which is negatively charged at pH 7.5—that may have been incorporated during fibril extraction. If fibril formation is sensitive to the presence of this ligand, then ligand withdrawal may inhibit their formation. Ligand-binding sites have similarly been discovered in pathogenic fibrils of tau[13] and α-synuclein[12].

Our experiments show that aggregates of TDP-43 in our samples of FTLD-TDP are amorphous rather than amyloid-like. FTLD-TDP has been defined by inclusions of TDP-43[15], and TDP-43-immunolabelled fibrils have been isolated from patients with FTLD-TDP[33]. However, sarkosyl-insoluble TDP-43 extracted from patients with FTLD-TDP exhibits non-fibrillar rather than fibrillary morphology, shown both here (Fig. 1) and in a previous study[30]. Our cryo-EM structures suggest that the amyloid fibrils that we imaged contain only TMEM106B; however, we cannot exclude the possibility that TDP-43 forms fibrils in patients with FTLD-TDP, and that our extraction method did not capture TDP-43 fibrils that were present. We believe that this is unlikely because we observed no fibrils under electron microscopy in the other fractions throughout our fibril-extraction protocol, and the sarkosyl and SDS treatment should not dissolve pathogenic irreversible fibrils. TDP-43 fibrils have recently been characterized in ALS with FTLD-TDP type B[34], a related neurodegenerative disease. It is uncertain whether the presence of TDP-43 fibrils in those cases stems from the different neuropathological condition or from different extraction procedures. In our western blot and immunogold-labelling experiments, we found that TDP-43 aggregates can be co-extracted with TMEM106B fibrils from patients with FTLD-TDP. It is unclear whether TDP-43 aggregates and TMEM106B fibrils co-localize in the patients' brains, and whether a cross-seeding mechanism is involved in FTLD-TDP pathogenesis.

Conclusion

Our structures of patient-derived TMEM106B fibrils reveal protein cleavage followed by amyloid formation of the C-terminal fragment of TMEM106B. This finding may refocus pathogenic studies of FTLD-TDP and perhaps other neurodegenerative diseases to include TMEM106B.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04670-9.

1. Mohandas, E. & Rajmohan, V. Frontotemporal dementia: an updated overview. *Indian J. Psychiatry* **51**, S65–S69 (2009).
2. Neary, D. et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* **51**, 1546–1554 (1998).
3. Neumann, M. et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
4. 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).
5. Chiti, F. & Dobson, C. M. Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annu. Rev. Biochem.* **86**, 27–68 (2017).
6. Benson, M. D. et al. Amyloid nomenclature 2020: update and recommendations by the International Society of Amyloidosis (ISA) nomenclature committee. *Amyloid Int. J. Exp. Clin. Invest.* **27**, 217–222 (2020).
7. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188–1203 (2012).
8. Fitzpatrick, A. W. P. et al. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).
9. Falcon, B. et al. Structures of filaments from Pick's disease reveal a novel tau protein fold. *Nature* **561**, 137–140 (2018).
10. Falcon, B. et al. Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. *Nature* **568**, 420–423 (2019).
11. Kollmer, M. et al. Cryo-EM structure and polymorphism of Aβ amyloid fibrils purified from Alzheimer's brain tissue. *Nat. Commun.* **10**, 4760 (2019).
12. Schweighauser, M. et al. Structures of α-synuclein filaments from multiple system atrophy. *Nature* **585**, 464–469 (2020).
13. Zhang, W. et al. Novel tau filament fold in corticobasal degeneration. *Nature* **580**, 283–287 (2020).
14. Ratnavalli, E., Brayne, C., Dawson, K. & Hodges, J. R. The prevalence of frontotemporal dementia. *Neurology* **58**, 1615–1621 (2002).
15. Goldman, J. S. et al. Frontotemporal dementia: genetics and genetic counseling dilemmas. *Neurologist* **10**, 227–234 (2004).
16. Mackenzie, I. R. A. et al. A harmonized classification system for FTLD-TDP pathology. *Acta Neuropathol.* **122**, 111–113 (2011).
17. Lee, E. B. et al. Expansion of the classification of FTLD-TDP: distinct pathology associated with rapidly progressive frontotemporal degeneration. *Acta Neuropathol.* **134**, 65–78 (2017).
18. Lashley, T., Rohrer, J. D., Mead, S. & Revesz, T. Review: an update on clinical, genetic and pathological aspects of frontotemporal lobar degenerations. *Neuropathol. Appl. Neurobiol.* **41**, 858–881 (2015).
19. Hasegawa, M. et al. Phosphorylated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Ann. Neurol.* **64**, 60–70 (2008).
20. Lu, R.-C., Wang, H., Tan, M.-S., Yu, J.-T. & Tan, L. TMEM106B and APOE polymorphisms interact to confer risk for late-onset Alzheimer's disease in Han Chinese. *J. Neural Transm.* **121**, 283–287 (2014).
21. Rutherford, N. J. et al. TMEM106B risk variant is implicated in the pathologic presentation of Alzheimer disease. *Neurology* **79**, 717–718 (2012).
22. Vass, R. et al. Risk genotypes at TMEM106B are associated with cognitive impairment in amyotrophic lateral sclerosis. *Acta Neuropathol.* **121**, 373–380 (2011).
23. 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).
24. Brady, O. A., Zhou, X. & Hu, F. Regulated intramembrane proteolysis of the frontotemporal lobar degeneration risk factor, TMEM106B, by signal peptide peptidase-like 2a (SPPL2a). *J. Biol. Chem.* **289**, 19670–19680 (2014).
25. Sawaya, M. R., Hughes, M. P., Rodriguez, J. A., Riek, R. & Eisenberg, D. S. The expanding amyloid family: structure, stability, function, and pathogenesis. *Cell* **184**, 4857–4873 (2021).
26. Shi, Y. et al. Structure-based classification of tauopathies. *Nature* **598**, 359–363 (2021).
27. Cruchaga, C. et al. Association of TMEM106B gene polymorphism with age at onset in granulin mutation carriers and plasma granulin protein levels. *Arch. Neurol.* **68**, 581–586 (2011).
28. Nicholson, A. M. et al. TMEM106B p.T185S regulates TMEM106B protein levels: implications for frontotemporal dementia. *J. Neurochem.* **126**, 781–791 (2013).
29. Inukai, Y. et al. Abnormal phosphorylation of Ser409/410 of TDP-43 in FTLD-U and ALS. *FEBS Lett.* **582**, 2899–2904 (2008).
30. Laferrière, F. et al. TDP-43 extracted from frontotemporal lobar degeneration subject brains displays distinct aggregate assemblies and neurotoxic effects reflecting disease progression rates. *Nat. Neurosci.* **22**, 65–77 (2019).
31. Neumann, M. et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
32. O'Brien, R. J. & Wong, P. C. Amyloid precursor protein processing and Alzheimer's Disease. *Annu. Rev. Neurosci.* **34**, 185–204 (2011).
33. Nonaka, T. et al. Prion-like properties of pathological TDP-43 aggregates from diseased brains. *Cell Rep.* **4**, 124–134 (2013).
34. Arseni, D. et al. Structure of pathological TDP-43 filaments from ALS with FTLD. *Nature* **601**, 139–143 (2022).

Methods

Post-mortem human brain samples

The Brain Bank for Neurodegenerative Disorders at Mayo Clinic Florida provided post-mortem brain tissues from patients with neuropathologically confirmed FTLD-TDP. Information on human donors is provided in Table 1. Autopsies were performed after consent by the next of kin or someone with legal authority to grant permission. The brain bank operates under protocols approved by the Mayo Clinic Institutional Review Board.

Immunohistochemistry staining

Post-mortem brains were immersion-fixed in 10% formalin, and sections of brain were embedded in paraffin, cut on a microtome at 5 µm thickness and mounted on positively charged glass slides. Sections were dried overnight and used for immunohistochemistry staining. Paraffin-embedded brain sections were deparaffinized in xylene, and rehydrated through a series of ethanol solutions, followed by washing in deionized H_2O. Antigen retrieval was performed by steaming slides in deionized H_2O or Tris-EDTA (DAKO), pH 9.0 for 30 min followed by a 5-min incubation in DAKO Peroxidase Block (DAKO, catalogue (cat.) no. S2001) to block endogenous peroxidase activity. Slides were blocked with DAKO Protein Block Serum-Free (DAKO, cat. no. X0909) for 1 h, and incubated with Anti TAR DNA-Binding Protein 43 (TDP-43) phospho-Ser409/410 monoclonal antibody (Cosmo Bio USA, cat. no. CAC-TIP-PTD-M01, lot no. 11-9-20) diluted 1:1,000 for 45 min. After washing, sections were incubated for 30 min in DAKO Envision-Plus System HRP Labelled Polymer Anti-Mouse (DAKO, cat. no. K4001). Peroxidase labelling was visualized with the Liquid DAB+ Substrate Chromogen System (DAKO, cat. no. K3468). Glass-mounted tissue sections were viewed with an Olympus BX51 and microscope images were captured with Olympus DP73 digital camera.

Fibril extraction from FTLD-TDP patient brains

Frozen brain tissues were weighed, diced into small pieces, and resuspended in 10 ml g^{-1} homogenization-solubilization (HS) buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1 mM EDTA, 1 mM dithiothreitol) supplemented with 1:100 (v/v) Halt protease inhibitor (Thermo Scientific). Resuspended tissue was homogenized using a Polytron homogenizer (Thomas Scientific) for 45 s and mixed with 10 ml g^{-1} of 4% (w/v) *N*-lauroyl-sarcosine (sarkosyl, Sigma), 2 U μl^{-1} home-made benzonase and 4 mM $MgCl_2$. Mixed solution was incubated at 37 °C with constant shaking at 300 r.p.m. for 45 min. Sample solution was then mixed with 10 ml g^{-1} ice-cold HS buffer with 0.5% (w/v) sarkosyl, and centrifuged at 3,000*g* at 4 °C for 5 min. The supernatant was extracted and centrifuged at 21,000*g* for 30 min. The supernatant was discarded, and the pellet was resuspended in 1.5 ml g^{-1} Tris buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl) and incubated at 4 °C overnight. After incubation, the solution was centrifuged at 6,000*g* for 5 min. The supernatant was transferred into a new tube, mixed with 1% (w/v) SDS (Invitrogen), and gently rotated at room temperature for 15 min. The solution was then sonicated for 3 min and centrifuged at 6,000*g* for 10 min. The supernatant was centrifuged at 21,000*g* for 30 min, and the pellet was resuspended in 10 µl g^{-1} Tris buffer, sonicated for 3 min and used for EM. Approximately 1 g of brain tissue from each patient was used for cryo-EM study.

Fibrils were also extracted using ice-cold water following a previously described protocol[35]. In brief, brain tissues were diced and resuspended in 4 ml g^{-1} of Tris-calcium buffer (20 mM Tris, pH 8.0, 138 mM NaCl, 2 mM $CaCl_2$, 0.1% NaN_3), then centrifuged at 3,100*g* at 4 °C for 5 min. The supernatant was collected and the Tris-calcium buffer wash was repeated four more times. After the fifth wash, the pellet was resuspended in Tris-calcium buffer with collagenase and incubated at 37 °C with constant shaking at 500 r.p.m. overnight, then centrifuged at 3,100*g* at 4 °C for 30 min. The pellet was resuspended in 4 ml g^{-1} Tris-EDTA buffer (20 mM Tris, pH 8.0, 140 mM NaCl, 10 mM EDTA, 0.1% NaN_3), then centrifuged at at 3,100*g* at 4 °C for 5 min. The supernatant was collected and the Tris-EDTA buffer wash was repeated nine more times. After the tenth wash, the pellet was resuspended in 2 ml g^{-1} ice-cold water, then centrifuged at 3,100*g* at 4 °C for 5 min. The fibril-containing supernatant was collected and the ice-cold water extraction was repeated nine more times. Fibrils extracted using sarkosyl and ice-cold water exhibited similar morphologies when examined by negative-stain electron microscopy (Fig. 1d), suggesting that sarkosyl does not induce the fibril formation during the extraction process. Fibril samples extracted using sarkosyl were more abundant and contained less contaminants, thus were used for cryo-EM study.

Negative-stain transmission electron microscopy

Carbon-coated formvar support films (400-mesh) mounted on copper grids (Ted Pella) were glow-discharged for 30 s before sample preparation. Three microlitres of sample solution was applied to the grids and incubated for 3 min, then excess sample solution was blotted off using filter paper. Grids were stained with 3 µl of 2% uranyl acetate (Electron Microscopy Sciences) for 1 min, washed with an additional 3 µl of 2% uranyl acetate and air-dried for 2 min. The grids were imaged using a Tecnai T12 transmission electron microscope (FEI).

Cryo-EM data collection and processing

To prepare the cryo-EM grids, we applied 2.6 µl of sample solution onto Quantifoil 1.2/1.3 200 mesh electron microscope grids glow-discharged for 2 min before use. Grids were plunge frozen into liquid ethane using a Vitrobot Mark IV (FEI). Cryo-EM data were collected on a Titan Krios transmission electron microscope (FEI) equipped with a K3 Direct Detection Camera (Gatan), operated with 300 kV acceleration voltage and energy filter width of 20 eV. Super-resolution movies were collected with a nominal physical pixel size of 1.1 Å per pixel (0.55 Å per pixel in super-resolution movie frames) and a dose per frame of ~1 $e^- Å^{-2}$. A total of 36 frames with a frame rate of 12 Hz were taken for each movie, resulting in a final dose of ~36 $e^- Å^{-2}$ per image. Manual data collection was performed using Leginon software package[36]. Cryo-EM datasets of four FTLD-TDP donors were collected and processed separately.

CTF estimation was performed using CTFFIND4[37]. For the dataset of donor 1, particle picking was performed using CrYOLO[38], trained with ~150 micrographs that we picked manually using e2helixboxer.py from EMAN2[39]. Particle extraction, 2D classification, helical reconstruction, and 3D refinement were performed in RELION[40,41]. Particles were extracted using an inter-box distance of 102.4 Å and a box size of 1,024 pixels scaled down to a 432-pixel box size. Two-dimensional classification with tau_fudge 2 was performed with all particles, and we found that all identifiable classes could be grouped into one of three fibril polymorphs, which we named PM1 (78.1% of all particles), PM2 (17.8% of all particles) and PM3 (4.1% of all particles). No other fibril morphologies were identified in this dataset. Particles from each polymorph were selected and used for 3D classification with $K = 3$ (for PM1) or $K = 1$ (for PM1 and PM2), using a Gaussian cylinder as the initial model. The best 3D classes were used as the initial model for subsequent 3D classifications with smaller box size particles. We re-extracted particles from all micrographs using an inter-box distance of 32 Å and a box size of 686 pixels scaled down to a 320 pixels box size. We manually selected 686-pixel particles for each polymorph based on the 2D classification of 1,024-pixel particles. For example, if a 1,024-pixel particle was classified as PM1 in 2D classification, then all 686-pixel particles extracted from that particular filament were selected for the PM1 polymorph. These manually selected particles were used for 2D classification for each polymorph separately, and the best 2D classes were selected for 3D classification. Two rounds (for PM1) or one round (for PM2 and PM3) of $K = 3$ 3D classification was performed. Particles from the best 3D class for each polymorph were re-extracted using an inter-box distance of 32 Å and a box size of 320 pixels (no scaling)

and used for high-resolution 3D refinement. To improve resolution, additional 2D classifications were performed for PM1 and PM2, with tau_fudge starting at 2 and increased incrementally to 8 in the final iterations. The 2D classes with clear 4.8 Å separation were selected for further 3D refinement. High-resolution gold-standard refinement was performed for each polymorph and the initial near-atomic resolution maps with refined helical parameters were generated. CTF refinement and Bayesian polishing was performed, and the final reconstructions were generated by one or two rounds of additional golden-standard refinement. The resolution of each reconstruction was estimated using the 0.143 Fourier shell correlation (FSC) resolution cut-off. See Table 2 for data collection and processing statistics for donor 1.

For the datasets of donor 2–4, similar strategies were applied for data collection and processing. All three datasets were processed independently up to and throughout 2D classification, with no information from other dataset introduced. 2D classification of all particles suggested that these three datasets also contain PM1, PM2, and PM3, with similar morphology and distribution as donor 1 (Extended Data Fig. 4), and no other identifiable fibril morphologies. During initial 3D reconstruction, we processed the donor 2 dataset independently, whereas for donors 3 and 4 we used the maps generated from the donor 1 dataset as initial model to expedite data processing. We believe these reconstructions were not biased because we used a 30 Å low-pass-filter on the initial model so that any higher-resolution information beyond 30 Å would originate from the dataset and not the initial model, and the final map of each polymorph was near-atomic (3.5–5.3 Å). After initial 3D reconstruction, a similar strategy was applied for each dataset to generate the final maps. See Extended Data Table 2 for data collection and processing statistics for donors 2–4.

Atomic model building

Our first model building efforts focused on PM1 of donor 1. We attempted to model the sequence of TDP-43 in the map using two methods: manual building with Coot[42] and automatic building using phenix.sequence_from_map[43]. No satisfactory model could be made with the TDP-43 sequence; multiple sidechains were inconsistent with densities even taking into consideration potential post-translational modifications such as phosphorylation of serine residues (the best model is shown in Extended Data Fig. 7a).

We then applied an unbiased strategy for model building: we built two polyalanine backbones with opposite directionality of N- and C-termini. For each model, we mutated the residues to amino acids that best fit the side chain densities (Extended Data Fig. 7a). The sequences of two resulting models were used as queries in BLAST to identify the protein in our maps. We restricted the BLAST database to human sequences (NCBI:txid9606), knowing that the brain samples that produced the fibrils were human. No other parameters of the search were altered from the default values. TMEM106B stood out as the only significant match to our query sequence.

We obtained five hits in total, but all five hits were sequences of the same protein, TMEM106B (including the protein labelled as hypothetical protein FLJ11273) (Extended Data Table 4). The percent identity of the hits was low (~16%), but the coverage was high, covering 125 of the 134 residues in the query (93%) without gaps or insertions in the alignment. All of these TMEM106B hits score above the default expect threshold value of 10. The 'expect' value that we obtained for the highest scoring hit, TMEM106B, was 0.003. This value means the number of times that a match as good or better would occur by chance is 0.003 in a database of this size. We interpret this value to indicate that the match is significant. If we increase the permissiveness of the expect threshold to higher values, we find the next best hit (6th) yields an expect value of 224, indicating this sequence (nuclear receptor subfamily 2 group E member) is an insignificant match to our query.

The TMEM106B sequence was threaded onto the initial model with phenix.sequence_from_map. The TMEM106B model fit the density better than the query model (Extended Data Fig. 7b). We extended the model to include five layers by applying the helical symmetry operators of the map and the model was refined with phenix.real_space_refine[44]. The orientations of mainchain oxygen atoms were manually adjusted to form inter-layer hydrogen bonds. The model was refined again with phenix.real_space_refine and the final model was validated using MolProbity[45]. The model was built and adjusted with Coot[42].

For PM2 and PM3 of donor 1, we fit a single layer of PM1 model into PM2 or PM3 maps with rigid body fit, and manually adjusted the model with Coot. A five-layer model (containing ten chains) was generated with the helical symmetry of PM2 or PM3. Model refinement was performed with phenix.real_space_refine and the final model was validated using MolProbity[46]. No major conformational changes were observed during model refinement.

For all polymorphs of donor 2–4, we did not build models de novo. Instead, we fit a single chain of each donor 1 polymorph into corresponding donor 2–4 maps as a rigid body (Extended Data Fig. 6). For PM2 and PM3, a one-layer model was generated by applying their helical symmetry. All models fit the maps well, suggesting that TMEM106B forms fibrils with the same fold in different FTLD-TDP patients.

Standard free energy calculation

Standard free energy values were estimated using a solvation energy algorithm as described previously[47]. We note that our free energy calculations neglect the contribution of glycans, as we cannot confirm the composition or conformation of the sugar groups.

Patient genotyping

Genomic DNA was extracted from patient brain tissues by the UCLA Technology Center for Genomics and Bioinformatics. Primers were designed to amplify the 7 exonic DNA fragments encoding TMEM106B (hg38 chr7:12,214,811–12,231,975, exons 3 to 9); primers were synthesized by Integrated DNA Technologies. PCR amplification was performed using Phusion High-Fidelity DNA Polymerase (New England Biolabs) and the PCR products were sequenced by Genewiz. For donor 3, sequencing results suggest the genotype of TMEM106B is wild type. For donors 1, 2 and 4, the sequencing results indicate a heterozygous Thr185Ser variant: at position hg38 chr7:12,229,791 in exon 6 of the TMEM106B gene, there is equal detection of the wild-type cytosine nucleotide and the variant guanine nucleotide. This single nucleotide polymorphism would induce a point mutation from wild-type Thr185 (ACC) to variant Ser185 (AGC). From the 50–50 distribution of cytosine and guanine nucleotides measured in the sequencing chromatogram, we conclude that one allele contains the wild-type gene and the other harbours the variant.

Western blotting

The supernatant and pellet from the first 21,000*g* centrifugation step of fibril extraction were used as samples for immunoblotting. Sample solutions were mixed with SDS–PAGE loading dye containing 2 M urea and 1 M β-mercaptoethanol, sonicated for 10 min in ice water, then boiled at 100 °C for 10 min. Samples were loaded onto a NuPAGE 4–12%, Bis-Tris, 1.0 mm, 12-well Mini Protein Gel (Invitrogen) and electrophoresis was performed at 200 V for 30 min. Proteins were wet transferred onto 0.2 μm nitrocellulose membranes (Bio-Rad) by application of a constant 35 V overnight, in transfer buffer consisting of 25mM Tris, pH8.3, 192mM glycine, 20% (w/v) methanol. Membranes were incubated with gentle rocking in 5% (w/v) Blocking-Grade Blocker (milk, Bio-Rad) in 1× Tris-buffered saline, with 0.1% (v/v) Tween20 (TBST). Membranes were incubated for 1 h with TMEM106B antibody (Novus Biologicals, cat. no. NBP1-91311, lot no. QC18333-42825) diluted 1:300, TDP-43 monoclonal antibody (Proteintech, cat. no. 60019-2-IG, lot no. 10011784) diluted 1:500, or anti TAR DNA-binding protein 43 phospho-Ser409/410 monoclonal antibody (Cosmo Bio, cat. no. CAC-TIP-PTD-M01, lot no. 11-9-20) diluted 1:300 in 2% milk in TBST. Membranes were washed three times

in TBST with gentle rocking for 5 min. Then, membranes were incubated for 1 h with either goat anti-rabbit IgG HRP (Invitrogen, cat. no. A27036, lot no. 2116291) or goat anti-mouse IgG HRP (Abcam, cat. no. ab205719, lot no. GR3271082-2) diluted 1:4,000 in 2% milk in TBST. Membranes were washed three times in TBST with gentle rocking for 5 min. Pierce ECL Plus Western Blotting Substrate (Thermo Scientific) was applied to membranes. Membranes were imaged using an Azure 600 (Azure Biosystems). Western blotting was repeated three times independently with similar results.

In addition to the ~35 kDa band on the TMEM106B western blot (Fig. 1b, left), bands with molecular weights (>110 kDa) higher than what is expected for TMEM106B were observed in the sarkosyl-soluble (3 bands) and insoluble fractions (1 band). These bands could correspond to different cleavage products or different post translational modification states of TMEM106B. The high molecular weights could be due to TMEM106B oligomerization or association with other macromolecules to form SDS-resistant complexes. Although the identify of these species cannot be confirmed, the conclusion of this study is unaffected as the ~35kDa band has been validated by mass spectrometry.

Immunogold labelling

Four-hundred-mesh carbon coated copper grids were glow discharged for 30 s before sample preparation. Three microlitres of sample solution was applied to the grids and incubated for 3 min, then excess sample solution was blotted off using filter paper. Blocking buffer (phosphate buffered saline, pH 7.4, 0.1% w/v gelatin) was applied to the grids and incubated for 10 min; excess solution was blotted off. TMEM106B antibody specific for residues 2–53 (Atlas Antibodies, cat. no. HPA058342, lot no. 22721), residues 204–253 (Novus Biologicals, cat. no. NBP1-91311, lot no. QC18333-42825) or residues 218-252 (antibodies-online, cat. no. ABIN6578799, lot no. SA160811DF), or TDP-43 Monoclonal Antibody (Proteintech, cat. no. 60019-2-IG, lot no. 10011784) diluted 1:100 in blocking buffer was applied to the grids and incubated for 30 min; excess solution was blotted off. Grids were washed five times with blocking buffer, and excess liquid was blotted off between each wash. 6 nm Colloidal Gold AffiniPure Goat Anti-Mouse IgG (Jackson ImmunoResearch Laboratories, cat. no. 115-195-146, lot no. 150545) or 6 nm Colloidal Gold AffiniPure Goat Anti-Rabbit IgG (Jackson ImmunoResearch Laboratories, cat. no. 111-195-144, lot no. 146470) diluted 1:8 in blocking buffer was applied to the grids and incubated for 30 min; excess solution was blotted off. Grids were washed five times with distilled water, and excess liquid was blotted off between each wash. Finally, grids were stained with 3 µl of 2% uranyl acetate for 1 min, washed with an additional 3 µl of 2% uranyl acetate and air-dried for 10 min. The grids were imaged using a Tecnai T12 electron microscope.

Mass spectrometry

The pellet from the first 21,000*g* centrifugation step of fibril extraction was analysed by SDS-PAGE as previous described. Gels were stained by InstantBlue Coomassie Protein Stain (Abcam) and gel bands were excised for mass spectrometry analysis. Protein digestion and peptide identification were adapted from a previously described protocol[48,49]. In brief, proteins trapped in gel bands were reduced with 10 mM dithiothreitol (Sigma) at 60 °C for 1 h, alkylated with 50 mM iodoacetamide (Sigma) at 45 °C for 45 min in the dark, digested with 200 ng trypsin (Promega) at 37 °C overnight and then 100 ng endoproteinase GluC (New England BioLabs) was added for another overnight incubation at 25 °C. Peptides were extracted from the gel bands in 50% acetonitrile/49.9% water/ 0.1% trifluoroacetic acid (TFA) and cleaned with C18 StageTip[50] before mass spectrometry analysis. Digested peptides were separated on an EASY-Spray column (25 cm × 75 µm internal diameter, PepMap RSLC C18, 2 µm, Thermo) connected to an U3000 RSLCnano HPLC (Thermo) and eluted using a gradient of 3%–32% acetonitrile in 0.1% formic acid and a flow rate of 300 nl min^{-1} (total time 45 min). Tandem mass spectra were collected in a data-dependent manner with an Orbitrap Exploris 480 mass spectrometer (Thermo) interfaced to a nano-ESI source (Thermo). Raw MS/MS data were analysed using Mascot (version 2.5) and files were searched against the UniProt human database (as of 19 December 2018; 20,387 entries) supplemented with common laboratory contaminants. The search parameters were: enzyme specificity: trypsin and GluC; maximum number of missed cleavages: 6, precursor mass tolerance, 10 ppm; product mass tolerance, 0.02 Da; variable modifications included cysteine carbamidomethylation and methionine oxidation. Peptide-spectrum matches (PSMs) were filtered to 1% false discovery rate using the target–decoy strategy. Additional searches were performed applying identical parameters to Mascot's error-tolerant algorithm. The mass spectrometry proteomics data has been deposited to the ProteomeXchange Consortium via MassIVE partner repository with the dataset identifier PXD029876.

Mass spectrometric analysis identified peptide LNNISIIGPLDMK (corresponding to TMEM106B 181–193) from the sarkosyl-insoluble fraction of donor 1. The Thr185Ser variant is present in this peptide, suggesting the variant protein is present in the fibrils. The peptide also contains the N184 glycosylation site without a glycan modification. Detection of this unglycosylated peptide does not contradict the glycosylation of N184 suggested by the cryo-EM structures. The type and size of the glycan could not be determined, thus it was much more difficult to identify the glycosylated peptide using mass spectrometry and only the unglycosylated peptide was detected.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Cryo-EM maps and atomic models of proteins from FTLD-TDP donor 1 have been deposited in the Protein Data Bank (PDB) and the Electron Microscopy Data Bank (EMDB) with accession codes 7SAQ and EMD-24953 for PM1, 7SAR and EMD-24954 for PM2, and 7SAS and EMD-24955 for PM3, respectively. Mass spectrometry data have been deposited to the ProteomeXchange Consortium via the MassIVE partner repository with the dataset identifier PXD029876. Any other relevant data are available from the corresponding author upon reasonable request.

Code availability

Energetic calculations were performed using custom written software; the code is available at https://doi.org/10.5281/zenodo.6321286.

35. Schmidt, M. et al. Cryo-EM structure of a transthyretin-derived amyloid fibril from a patient with hereditary ATTR amyloidosis. *Nat. Commun.* **10**, 5008 (2019).
36. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
37. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
38. Wagner, T. et al. Two particle-picking procedures for filamentous proteins: SPHIRE–crYOLO filament mode and SPHIRE-STRIPER. *Acta Crystallogr. D* **76**, 613–620 (2020).
39. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
40. He, S. & Scheres, S. H. W. Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).
41. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
42. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
43. Terwilliger, T. C. Automated side-chain model building and sequence assignment by template matching. *Acta Crystallogr. D* **59**, 45–49 (2003).
44. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
45. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
46. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
47. Cao, Q., Boyer, D. R., Sawaya, M. R., Ge, P. & Eisenberg, D. S. Cryo-EM structure and inhibitor design of human IAPP (amylin) fibrils. *Nat. Struct. Mol. Biol.* **27**, 653–659 (2020).

48. Thevis, M., Ogorzalek Loo, R. R. & Loo, J. A. In-gel derivatization of proteins for cysteine-specific cleavages and their analysis by mass spectrometry. *J. Proteome Res.* **2**, 163–172 (2003).
49. McConnell, S. A. et al. Protein labeling via a specific lysine–isopeptide bond using the pilin polymerizing sortase from *Corynebacterium diphtheriae*. *J. Am. Chem. Soc.* **140**, 8420–8423 (2018).
50. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).

Acknowledgements We thank Z. H. Zhou for the use of Electron Imaging Center for Nanomachines (EICN) instruments supported by the NIH (1S10RR23057 and IS10OD018111), NSF (DBI-1338135) and California NanoSystems Institute at UCLA. We thank S. Scheres and M. Goedert for discussion of TMEM106B antibodies, and L. Salwinski for evaluation of the significance of sequence matching. The authors acknowledge NIH AG 054022, NIH AG061847, DOE DE-FC02-02ER63421, NIH GM103479, NIH GM007185, NSF DGE-1650604, ADRC (P30 AG062677), Einstein Aging Study (P01 AG003949), ALLFTD (U19 AG063911), C9ORF72 P01 (P01 NS084974), and the National Facility for Translational Medicine (Shanghai) for support.

Author contributions Y.X.J. and Q.C. extracted fibrils from the material donated by patients with FTLD-TDP, prepared cryo-EM grids, and collected and processed cryo-EM data. Y.X.J., Q.C. and M.R.S. built the atomic models. P.G. assisted in cryo-EM data collection. Y.X.J. and R.A. performed western blotting and immunolabelling assays. Y.X.J. performed patient genotyping. J.Y.F., R.R.O.L. and J.A.L. performed mass spectrometry. M.D. and D.W.D. prepared frozen brain samples and performed immunohistochemistry staining. Q.C. and M.R.S. performed solvation energy calculations. All authors analysed the results and wrote the manuscript. D.S.E. supervised the project.

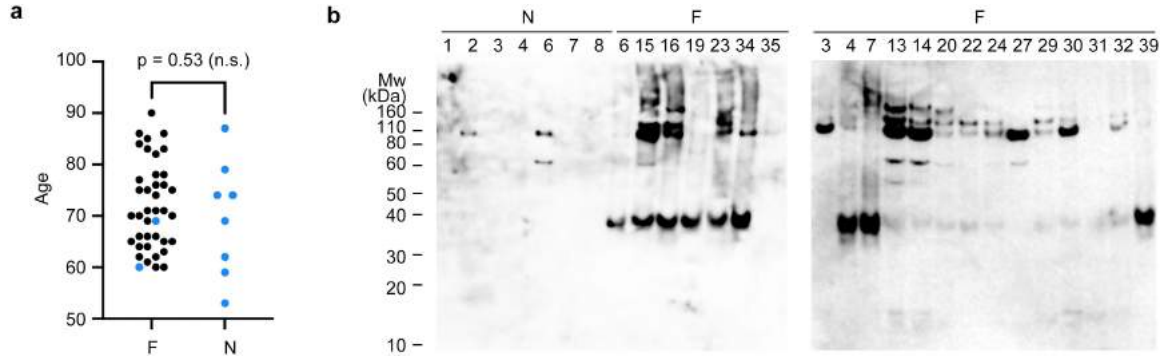
Competing interests D.S.E. is an advisor and equity shareholder in ADRx, Inc. The other authors declare no competing interests.

Additional information

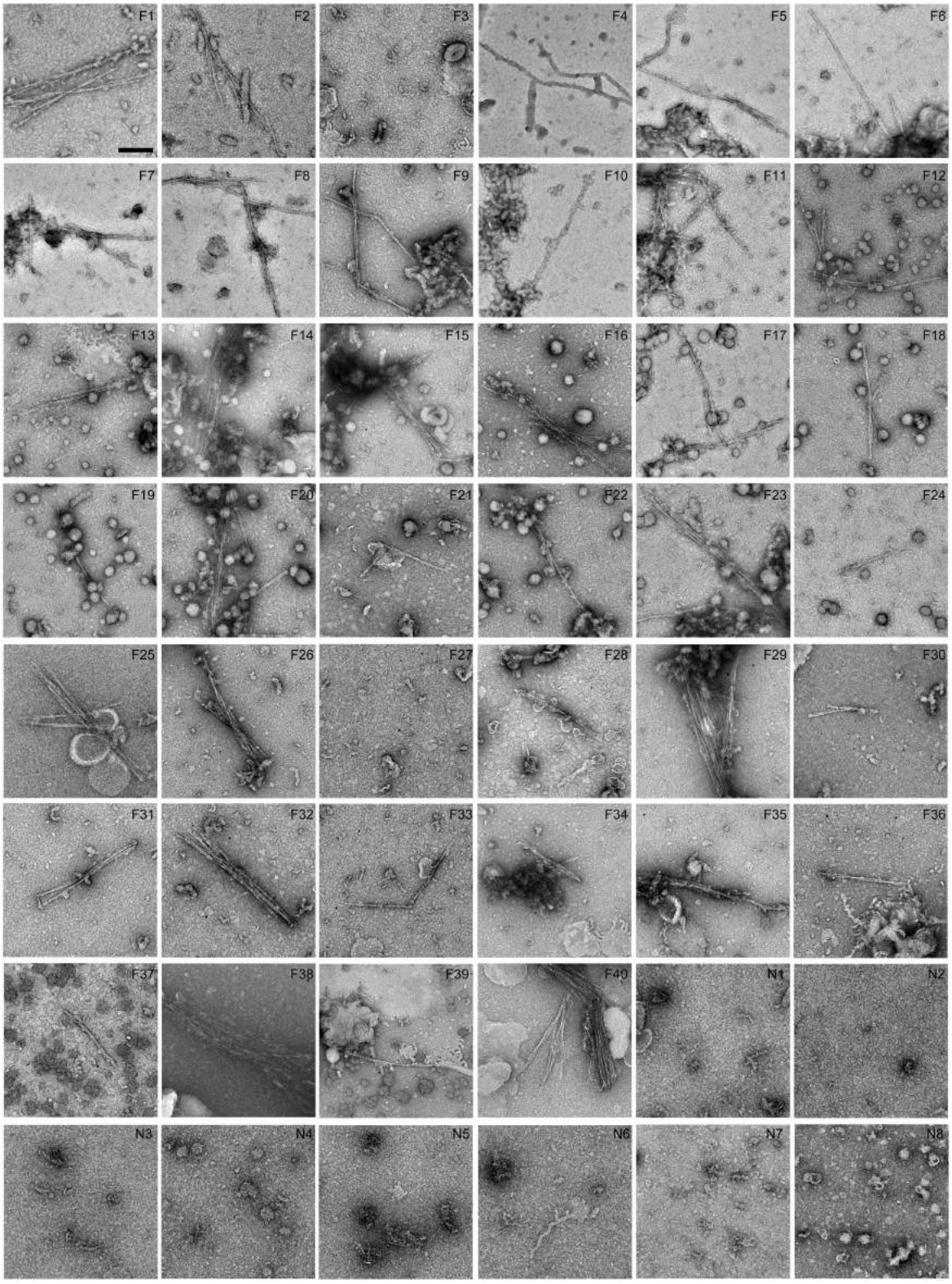
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04670-9.

Correspondence and requests for materials should be addressed to David S. Eisenberg.

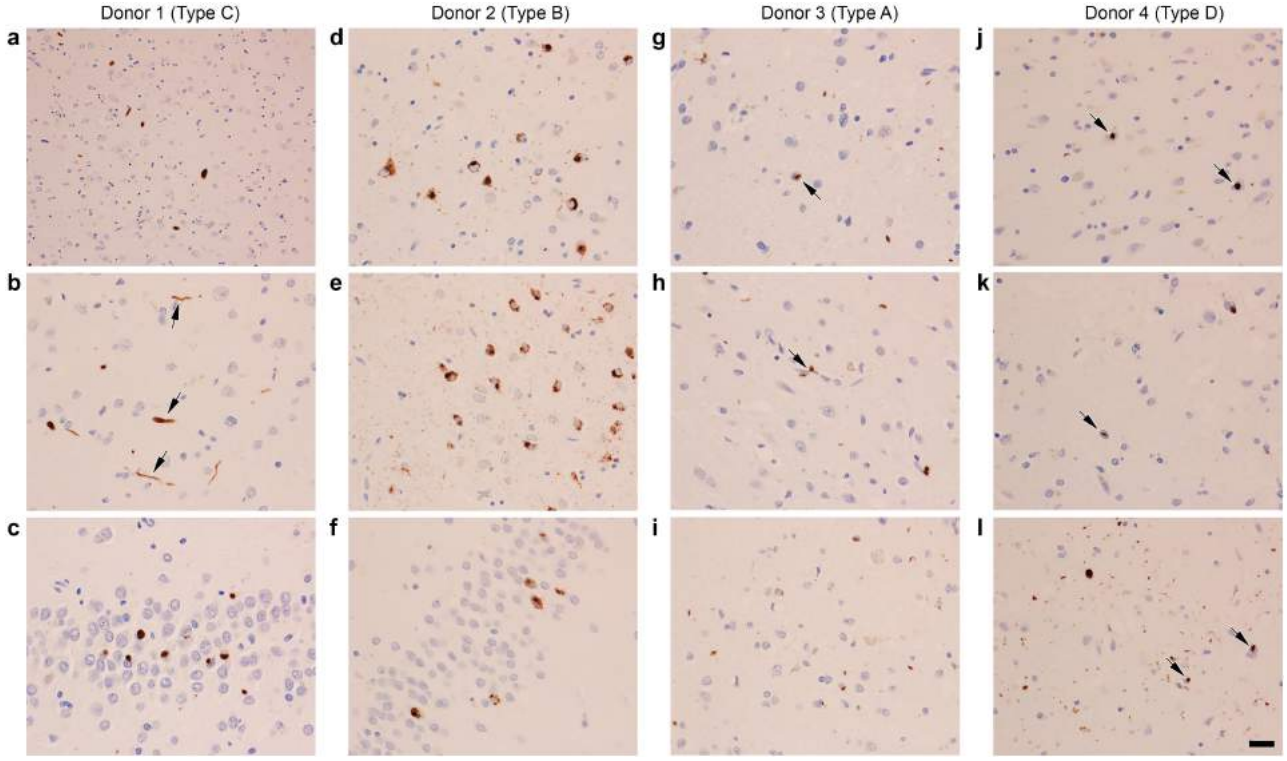
Peer review information *Nature* thanks Henning Stahlberg and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



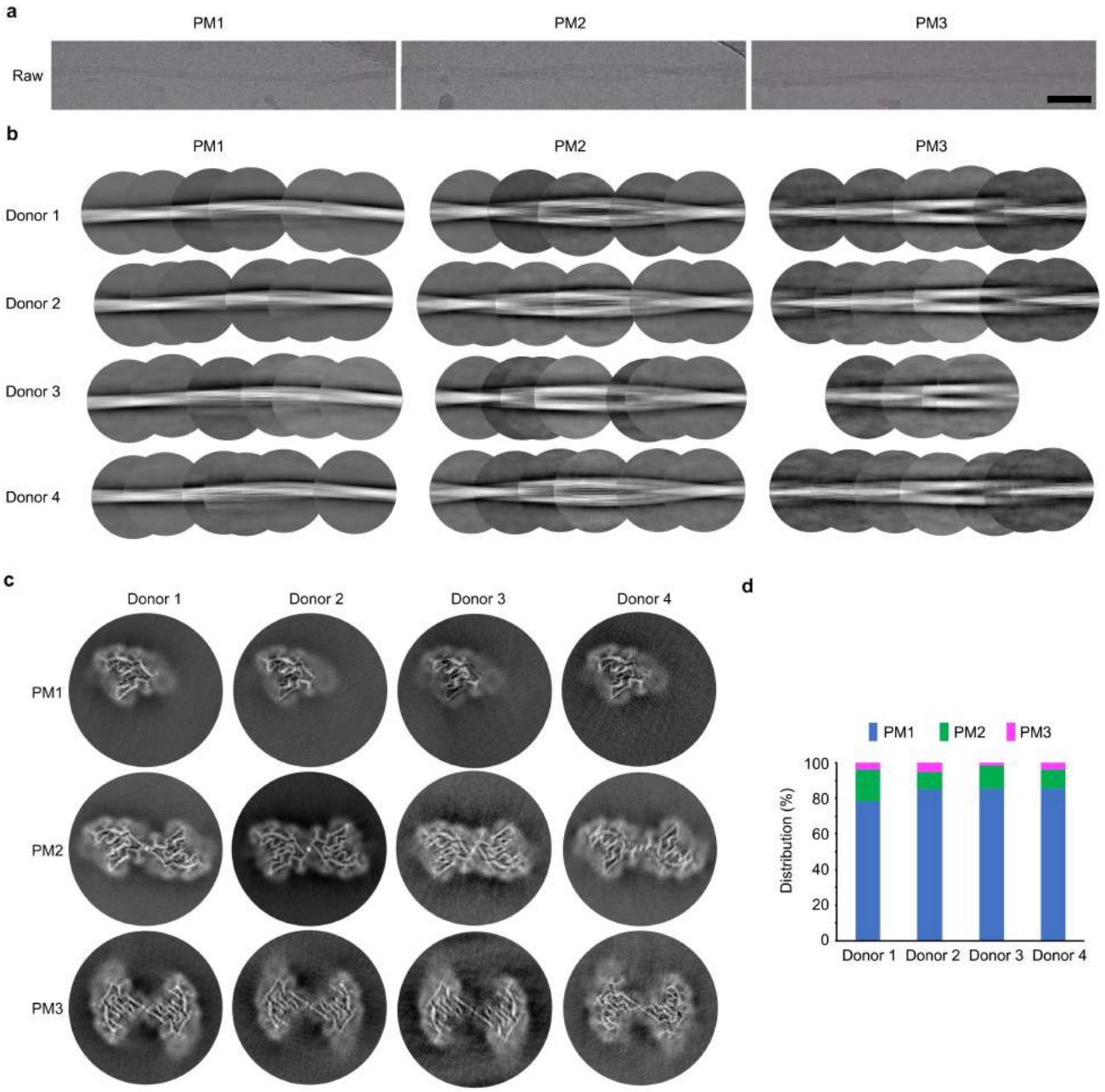
Extended Data Fig. 1 | Comparison of FTLD-TDP and non-FTLD-TDP donors. **a**, Age distribution of FTLD-TDP donors (F, n = 40) and non-FTLD-TDP controls (N, n = 8). Donors with and without fibrils detected under negative stain EM are coloured black and blue, respectively. P-value of 0.53 (n.s., not significant) from an unpaired, two-tailed t-test suggests that the presence of fibrils is disease-dependent but not age-dependent. **b**, Western blots of sarkosyl-insoluble fractions from FTLD-TDP (F) and non-FTLD-TDP donors (N) probed by TMEM106B antibody. The ~35 kDa TMEM106B-positive band was found in none of the non-FTLD-TDP donors (N5 shown in Fig. 1b as donor 5) and all of the FTLD-TDP donors except F35, F3, and F27. P-value of less than 0.0001 was obtained from an unpaired, two-tailed t-test comparing presence of the ~35kDa band (value of "1" for present, "0" for absent) in non-FTLD-TDP donors (0 out of 8, n = 8) and FTLD-TDP donors (22 out of 25, n = 25). Fibrils were not observed in F3 and F27 by negative stain EM, consistent with the western blot. Fibrils were observed in F35, which suggests that western blot may not always be accurate in detecting TMEM106B aggregation. Western blot membranes were prepared in parallel and exposed with equal time. The original, uncropped blots are shown in Supplementary Fig. 3.



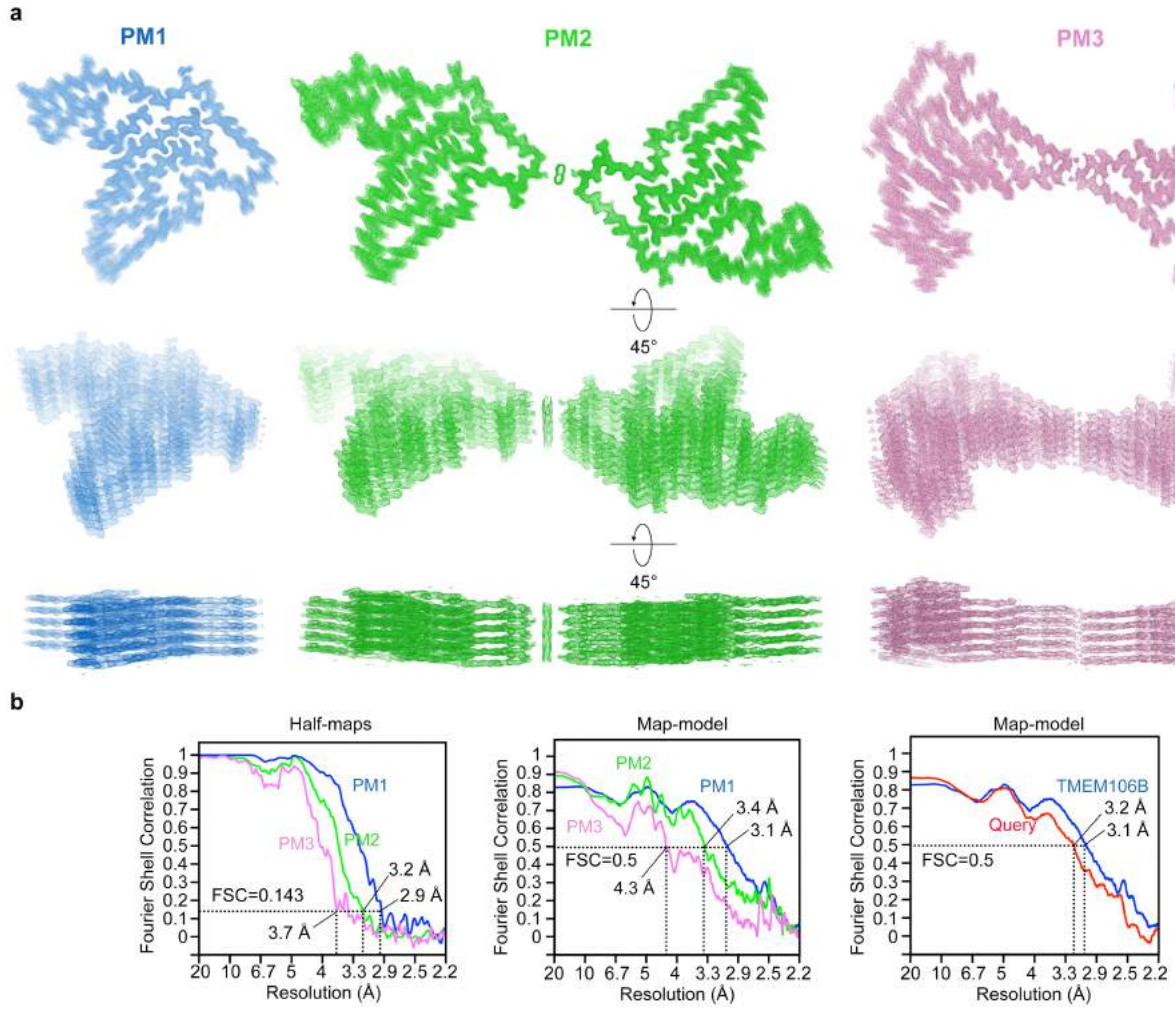
Extended Data Fig. 2 | Fibril screen of FTLD-TDP and non-FTLD-TDP donors by negative stain EM. Negative stain EM images of sarkosyl-insoluble fractions from all donors. Scale bar 200 nm. Donors 1–5 (F26, F36, F17, F40 and N5, respectively) are also shown in Fig. 1a. Fibrils with similar morphologies were found in all FTLD-TDP donors except F3 and F27. No fibrils were found in any non-FTLD-TDP donors. P-value of less than 0.0001 was obtained from an unpaired, two-tailed t-test comparing fibrils detected by EM (value of 1 for present, 0 for absent) in FTLD-TDP donors (38 out of 40, n = 40) and non-FTLD-TDP donors (0 out of 8, n = 8).



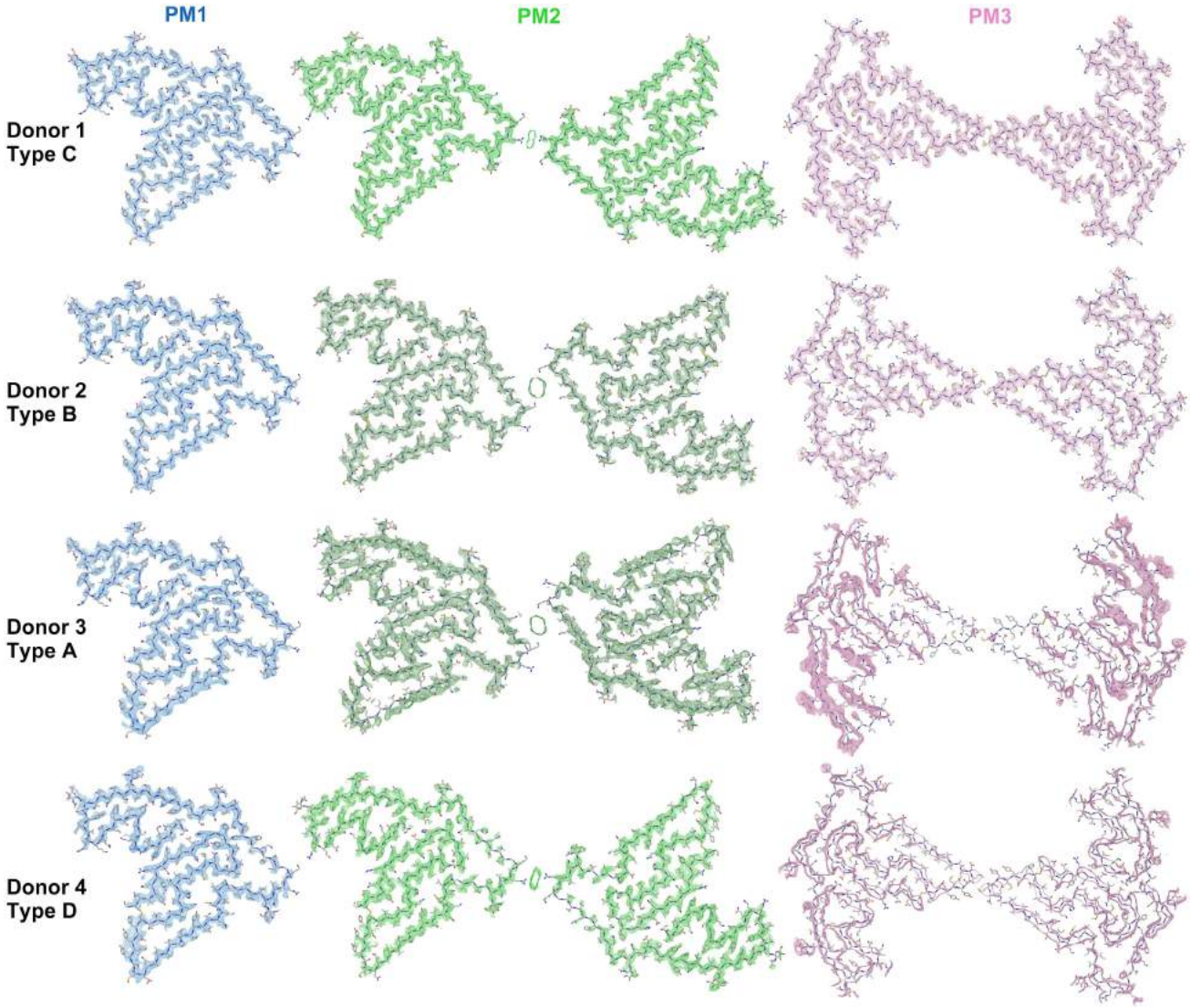
Extended Data Fig. 3 | Neuropathological diagnosis of donors 1–4 as FTLD-TDP. Immunohistochemistry staining using a phosphor-Ser409/410 TDP-43 antibody was performed for brain sections from donors 1–4. All four donors were confirmed to be FTLD-TDP cases, representing the 4 subtypes A to D, respectively. For all figures, scale bar 20 µm. **a**, **b**, **c**, Donor 1 is FTLD-TDP type C, with long, thick neurites (arrows in **b**) and 'Pick-body like NCI' in dentate fascia (**c**). **d**, **e**, **f**, Donor 2 is FTLD-TDP type B, displaying characteristic granular cytoplasmic NCI in cortex (**d**), hippocampus (**e**) and dentate fascia (**f**). **g**, **h**, **i**, Donor 3 is FTLD-TDP type A, exhibiting small dense neuronal cytoplasmic inclusions (NCI), sparse neuronal intranuclear inclusions (NII, arrow in **g**), and perivascular glial inclusions (arrow in **h**). **j**, **k**, **l**, Donor 4 is FTLD-TDP type D, shown by frequent NII (arrows in **j**, **k** and **l**) and small NCI and neurites. **a**, **b**, **d**, **e**, **g**, **h**, **i**, **j**, **k**, **l** are brain sections from the temporal cortex. **c** and **f** are brain sections from dentate fascia.



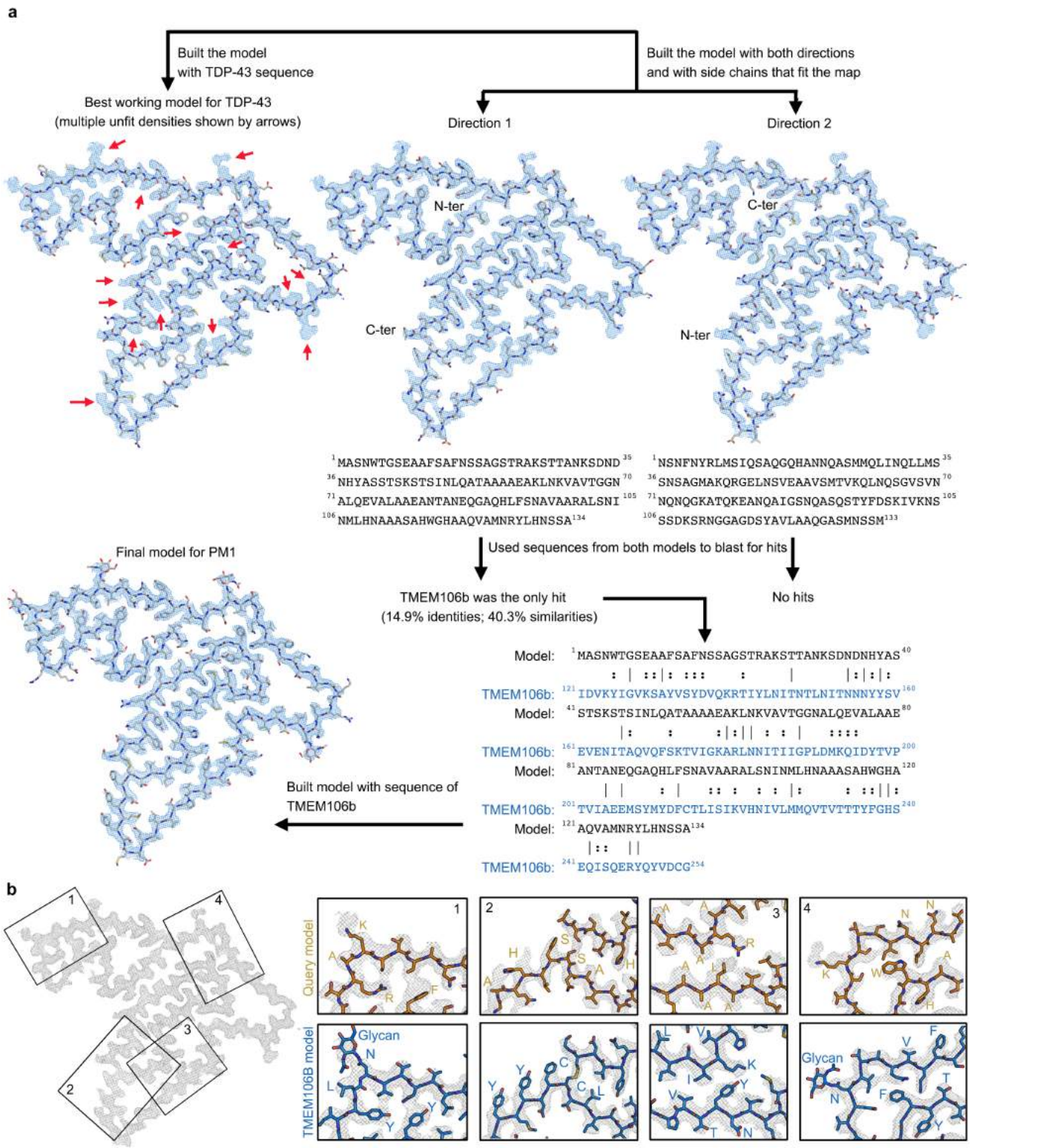
Extended Data Fig. 4 | Cryo-EM data processing of amyloid fibrils from FTLD-TDP donors 1 to 4. a, Representative micrographs of PM1-3 from FTLD-TDP donor 1. Scale bar 500 Å. **b**, Representative 2D classes of PM1-3 from FTLD-TDP donors 1 to 4. 2D classes are stitched together to show a full cross-over of each morphology. **c**, 3D reconstructions of PM1-3 from the four FTLD-TDP donors. **d**, Distributions of the three fibril polymorphs in the four FTLD-TDP donors.



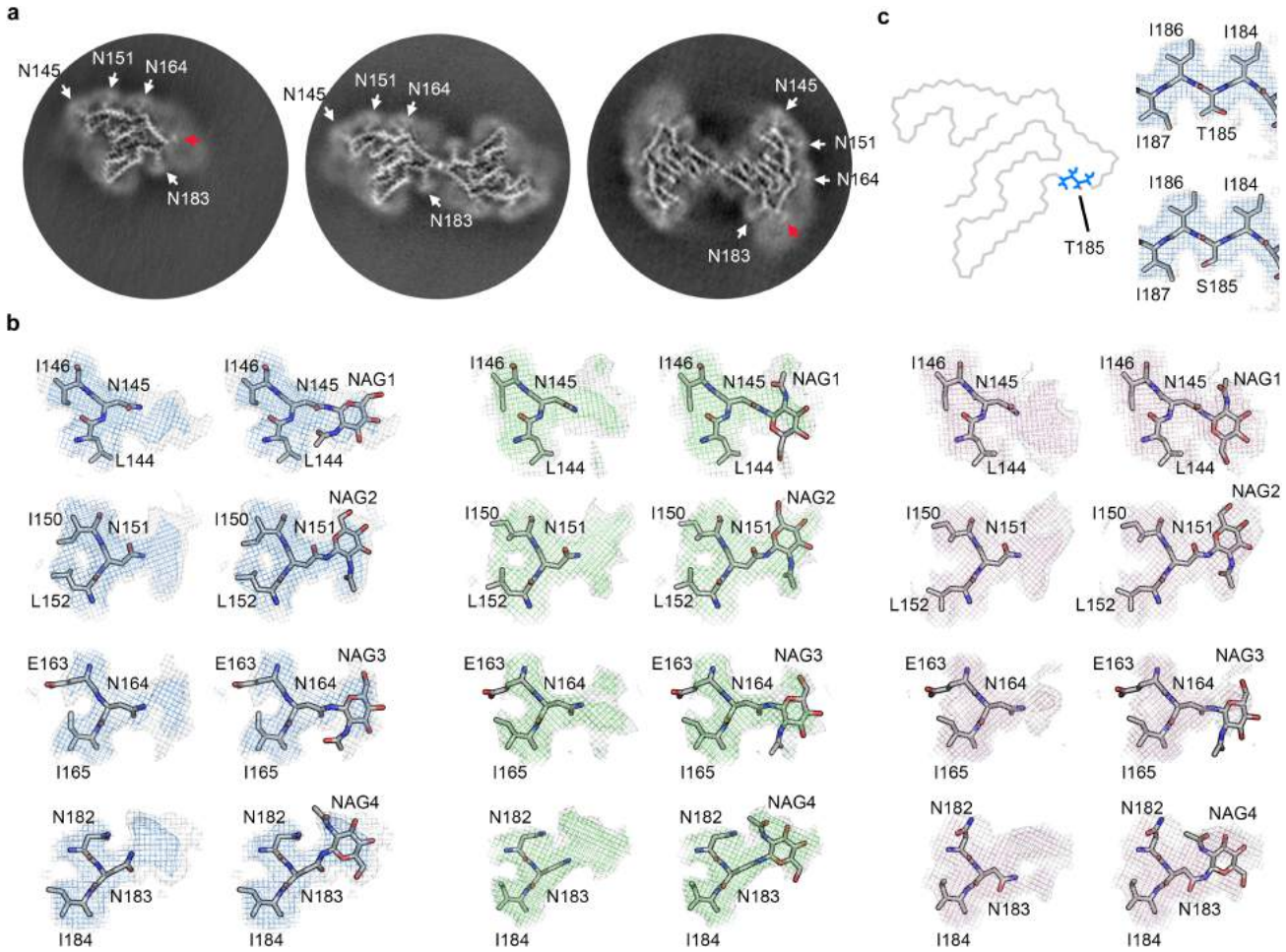
Extended Data Fig. 5 | Cryo-EM maps and FSC curves of FTLD-TDP donor 1. **a**, Views of the cryo-EM maps from FTLD-TDP donor 1 with five layers shown. **b**, FSC curves between two half-maps (left) and the cryo-EM reconstruction and refined atomic model (middle) of each polymorph PM1 (blue), PM2 (green), and PM3 (pink). FSC curves between cryo-EM reconstruction and the query model (red, Direction 1 chain of Extended Data Fig. 5a) and the atomic model of PM1 from FTLD-TDP donor 1 (blue) are compared on the right.



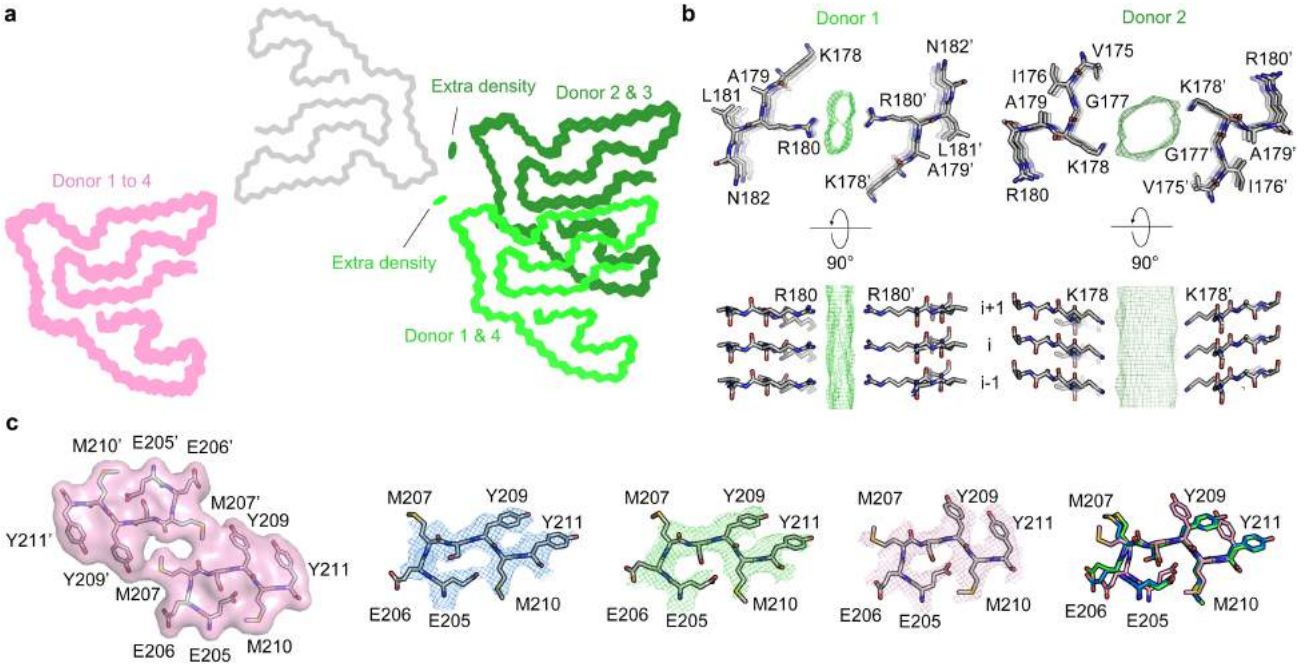
Extended Data Fig. 6 | Cryo-EM maps of TMEM106B fibrils from FTLD-TDP donors 1 to 4. The models of PM1-3 from donor 1 were rigid body fitted into the maps of PM1-3 from donors 2 to 4. These cryo-EM structures reveal that the polymorphs from all four donors share the same protofilament fold. Two subtypes of PM2 are exhibited by donors 1 and 4 (light green) and donors 2 and 3 (dark green), respectively.



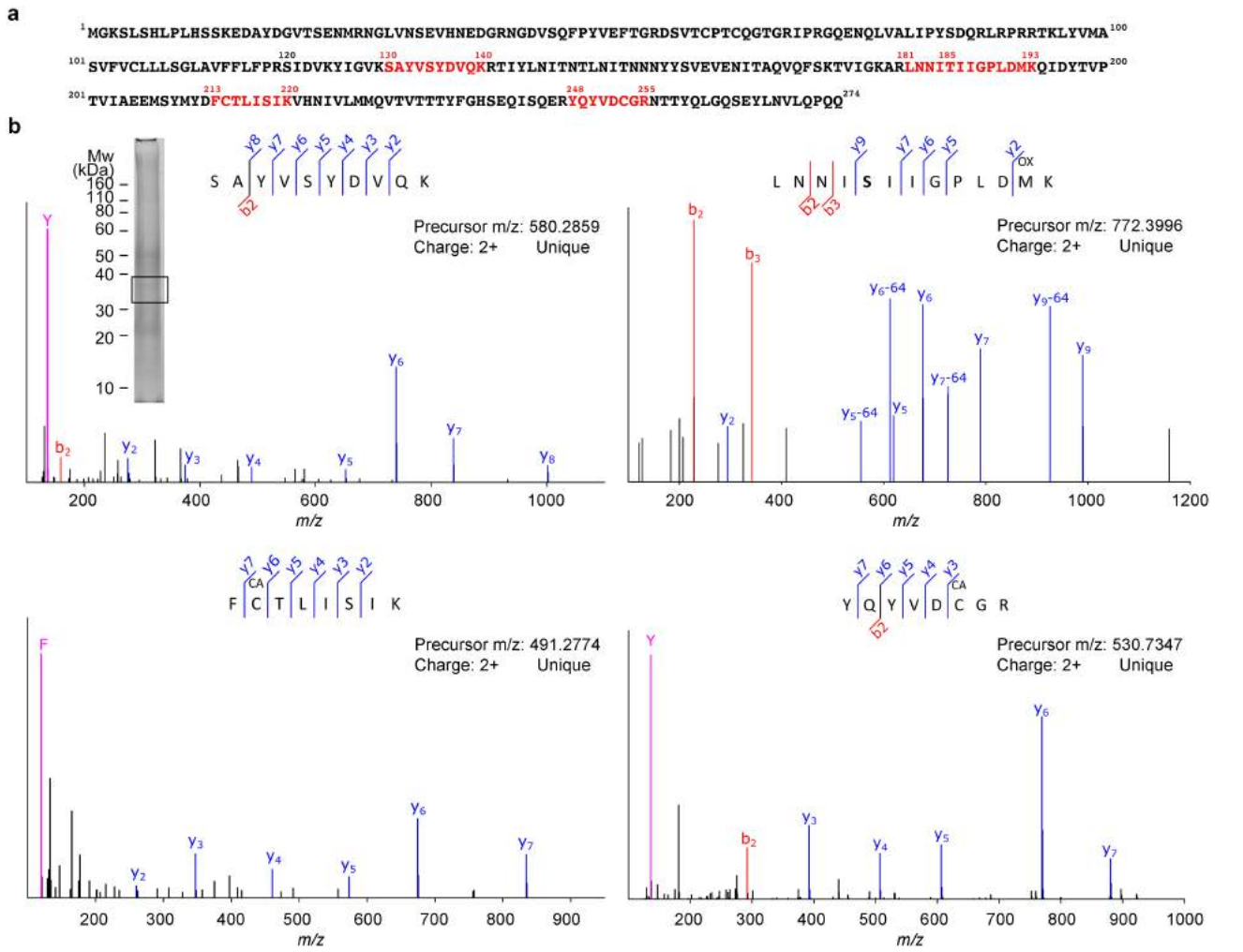
Extended Data Fig. 7 | Identification of the TMEM106B molecule by atomic model building. a, Atomic model building flowchart for PM1 of FTLD-TDP donor 1. In the sequence alignment (bottom right), lines indicate identical residues and two dots indicate similar residues. **b**, Comparison of query model (Direction 1 chain, orange) with final model built with TMEM106B sequence (blue). Four representative regions of the cryo-EM map are shown. Residues that show clear differences in side chain density fitting are labelled in both models.



Extended Data Fig. 8 | Evidence of glycosylation and genetic polymorphism in maps of TMEM106B fibrils. a, Three-dimensional reconstructions of PM1-3 from FTLD-TDP donor 1. White arrows point to the four glycosylation sites within the fibril core. Red arrows point to the residual densities outside the fibril core in PM1 and PM3, which may correspond to the binding of the same undefined, negatively charged ligand present in the PM2 dimer interface (Extended Data Fig. 9). **b**, Maps and models of the four glycosylation sites with or without the sugar group in PM1 (left), PM2 (middle), and PM3 (right) from FTLD-TDP donor 1. **c**, Position of Thr/Ser185 genetic polymorphism in the conserved fibril fold (left) and the map and model of the Thr/Ser185 environment (right, represented by PM1 from donor 1).



Extended Data Fig. 9 | Diverse dimer interfaces of PM2 and PM3 of TMEM106B fibrils. a, Dimer arrangements of PM2 and PM3 from FTLD-TDP donors 1 to 4. PM2 and PM3 from all donors are aligned at chain A (grey, represented by PM2 of donor 1). Chain B of PM2 (light green for donors 1 and 4, dark green for donors 2 and 3) and PM3 (pink) from each donor is shown. Residual densities in the PM2 dimer interfaces are shown as green ovals. The dimer arrangement of PM3 is consistent among all donors, whereas there are two subtypes of dimer arrangements for PM2. **b**, Atomic model and the residual density in the dimer interface of donor 1 (left, represents donors 1 and 4) or donor 2 (right, represents donors 2 and 3). In donors 1 and 4, Arg180 from each protofilament is on the opposite sides of an extra density in the middle of the PM2 dimer interface; in donors 2 and 3, the dimer interface is shifted to Lys178. Although two slightly different dimer interfaces were observed, we consider PM2 in all four FTLD-TDP donors to be the same morphology because of the similarity in dimer formation (see Discussion). **c**, Comparison of the residues near the PM3 interface (far left) from PM1 (blue), PM2 (green), PM3 (pink), and the superimposition of those residues from PM1-3 (far right). PM1-3 are all represented by FTLD-TDP donor 1.



Extended Data Fig. 10 | Detection of TMEM106B peptides by mass spectrometry. a, Sequence map of TMEM106B. Sequences in red indicate unique peptides detected by LC-MS/MS from excised gel bands. **b**, Fragmentation spectra (MS/MS) of detected peptides from TMEM106B. Detected fragment ions (b, y, and immonium ions) are labelled accordingly. The peptide modifications methionine oxidation (OX) and cysteine carbamidomethylation (CA) were observed. SDS-PAGE gel of sarkosyl-insoluble fraction of donor 1 shown as an insert. Box indicates the gel region, which corresponds to the ~35kDa band from the TMEM106B western blot (Fig. 1b), excised for LC-MS/MS analyses.

Extended Data Table 1 | Demographics and inclusion properties for all FTLD-TDP and non-FTLD-TDP donors

Original ID	Gender	Age	Braak	PathDx (type)	FHx	Brain region	Fibrils under EM	Band in western blot	Donor ID reported
F1	Male	66	2	FTLD-TDP (A)	FALSE	MF	Yes	n.a.	
F2	Male	65	2.5	FTLD-TDP (A)	FALSE	MF	Yes	n.a.	
F3	Male	60	3	FTLD-TDP (A)	FALSE	MF	No	No	
F4	Female	71	2	FTLD-TDP (A)	TRUE	MF	Yes	Yes	
F5	Male	71	2.5	FTLD-TDP (A)	TRUE	MF	Yes	n.a.	
F6	Male	61	1	FTLD-TDP (A)	FALSE	MF	Yes	Yes	
F7	Male	60	2	FTLD-TDP (A)	TRUE	MF	Yes	Yes	
F8	Female	84	2.5	FTLD-TDP (A)	TRUE	MF	Yes	n.a.	
F9	Male	76	1	FTLD-TDP (A)	FALSE	MF	Yes	n.a.	
F10	Female	78	1.5	FTLD-TDP (A)	FALSE	MF	Yes	n.a.	
F11	Female	85	1	FTLD-TDP (A)	FALSE	MF	Yes	n.a.	
F12	Male	63	0	FTLD-TDP (A)	TRUE	MF	Yes	n.a.	
F13	Male	86	2.5	FTLD-TDP (A)	FALSE	MF	Yes	Yes	
F14	Male	70	1	FTLD-TDP (A)	TRUE	MF	Yes	Yes	
F15	Female	83	3	FTLD-TDP (A)	TRUE	MF	Yes	Yes	
F16	Male	60	1	FTLD-TDP (A)	TRUE	MF	Yes	Yes	
F17	Male	86	0	FTLD-TDP (A)	FALSE	MF	Yes	Yes	Donor 3
F18	Female	90	2.5	FTLD-TDP (A)	FALSE	MF	Yes	n.a.	
F19	Female	64	2	FTLD-TDP (A)	TRUE	MF	Yes	Yes	
F20	Male	83	3	FTLD-TDP (A)	FALSE	MF	Yes	Yes	
F21	Female	71	0	FTLD-TDP (C)	FALSE	MF	Yes	n.a.	
F22	Male	70	0.5	FTLD-TDP (C)	FALSE	MF	Yes	Yes	
F23	Male	75	0	FTLD-TDP (C)	FALSE	MF	Yes	Yes	
F24	Male	74	2.5	FTLD-TDP (C)	FALSE	MF	Yes	Yes	
F25	Female	75	2.5	FTLD-TDP (C)	FALSE	MF	Yes	n.a.	
F26	Male	65	2	FTLD-TDP (C)	FALSE	MF	Yes	Yes	Donor 1
F27	Male	66	3	FTLD-TDP (C)	TRUE	MF	No	No	
F28	Male	75	2	FTLD-TDP (C)	FALSE	MF	Yes	n.a.	
F29	Male	65	1	FTLD-TDP (C)	FALSE	MF	Yes	Yes	
F30	Male	78	1	FTLD-TDP (C)	TRUE	MF	Yes	Yes	
F31	Female	69	3	FTLD-TDP (C)	FALSE	MF	Yes	Yes	
F32	Male	66	0	FTLD-TDP (C)	FALSE	MF	Yes	Yes	
F33	Female	69	2	FTLD-TDP (C)	TRUE	MF	Yes	n.a.	
F34	Female	82	3	FTLD-TDP (C)	TRUE	MF	Yes	Yes	
F35	Male	70	2.5	FTLD-TDP (B)	TRUE	MF	Yes	No	
F36	Female	76	3	FTLD-TDP (B)	FALSE	MF	Yes	Yes	Donor 2
F37	Female	62	0	FTLD-TDP (B)	FALSE	MF	Yes	n.a.	
F38	Female	62	1	FTLD-TDP (B)	TRUE	MF	Yes	n.a.	
F39	Male	77	2	FTLD-TDP (B)	TRUE	MF	Yes	Yes	
F40	Female	64	3	FTLD-TDP (D)	TRUE	MF	Yes	Yes	Donor 4
N1	Male	69	2	Normal	No	MF	No	No	
N2	Male	62	2	VaD	Yes	MF	No	No	
N3	Male	79	1	VaD	No	MF	No	No	
N4	Female	87	2	VaD	No	MF	No	No	
N5	Male	59	0	Normal	No	MF	No	No	Donor 5
N6	Male	53	1	PART	No	MF	No	No	
N7	Female	74	2	PART	No	MF	No	No	
N8	Male	74	3	PART	No	MF	No	No	

PathDx, pathological diagnosis; FHx, family history; VaD, vascular dementia; PART, primary age-related tauopathy; MF, medial frontal gyrus; band in western blot, detection of the ~35 kDa TMEM106B antibody-positive band in the sarkosyl-insoluble fraction; n.a., not available. All donors were numbered by original ID (F1-40 for FTLD-TDP cases and N1-8 for non-FTLD-TDP cases) during initial fibril screen. F26, F36, F17, F64 and N5 were selected for further study and renumbered as donors 1–5 in the main text.

Extended Data Table 2 | Cryo-EM data collection and processing statistics of FTLD-TDP donors 2–4

	Donor 2			Donor 3			Donor 4		
	PM1	PM2	PM3	PM1	PM2	PM3	PM1	PM2	PM3
Magnification	×81,000	×81,000	×81,000	×81,000	×81,000	×81,000	×81,000	×81,000	×81,000
Voltage (kV)	300	300	300	300	300	300	300	300	300
Electron exposure (e^{-}/Å^2)	37	37	37	36	36	36	36	36	36
Defocus range (μm)	1.0-5.1	1.0-5.1	1.0-5.1	0.5-5.1	0.5-5.1	0.5-5.1	0.8-4.9	0.8-4.9	0.8-4.9
Pixel size (Å)	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
Symmetry imposed	C_1	C_2	C_1	C_1	C_2	C_1	C_1	C_2	C_1
Helical rise (Å)	4.896	4.9	2.446	4.897	4.906	2.452	4.897	4.904	2.4281
Helical twist (°)	-0.413	179.619	179.807	-0.413	179.6	179.8	-0.413	179.605	179.81
Initial particle images	438,393	438,393	438,393	419,107	419,107	419,107	228,432	228,432	228,432
Final particle images	38,894	24,420	5,317	26,533	15,587	5,335	14,407	13,482	5,798
Map resolution (Å)	3.8	4.0	4.6	3.6	4.1	5.3	3.5	3.6	5.0
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	200-3.8	200-4.0	200-4.6	200-3.6	200-4.1	200-5.3	200-3.5	200-3.6	200-5.0

Extended Data Table 3 | TMEM106B peptides identified by mass spectrometry

Sequence	Modifications	Protein Group Accessions	Position in TMEM106B	# PSMs	IonScore	E-value	m/z (Da)	Δ m/z (ppm)
SAYVSYDVQK		Q9NUM4	130-140	3	39	0.0044	580.28595	+1.34
LNNISIIGPLDMK[a]	M12(Oxidation)	Q9NUM4	181-193	—	—	—	722.39960	+6.35
FCTLISIK	C2(Carbamidomethyl)	Q9NUM4	213-220	3	59	2.1e-005	491.27747	+3.89
YQYVDCGR	C6(Carbamidomethyl)	Q9NUM4	248-255	3	34	0.0025	530.73291	+6.41

[a]Peptide identified from Mascot's error-tolerant algorithm.

Extended Data Table 4 | BLAST search results

Description	Scientific Name	Max Score	Total Score	Query Cover	E-value	Percent Identity	Accession Length	Accession
TMEM106B protein	Homo sapiens	39.3	39.3	93%	0.003	16.00%	274	AAH39741.1
transmembrane protein 106B	Homo sapiens	39.3	39.3	93%	0.003	16.00%	274	NP_001127704.1
transmembrane protein 106B, isoform CRA_b	Homo sapiens	39.3	39.3	92%	0.003	16.13%	285	EAW93639.1
TMEM106B protein	Homo sapiens	38.9	38.9	93%	0.004	16.00%	314	AAH28108.1
hypothetical protein FLJ11273 variant	Homo sapiens	37.0	37.0	92%	0.019	16.13%	274	BAD96983.1
nuclear receptor subfamily 2 group E member 1 isoform b	Homo sapiens	25.4	25.4	50%	224	32.89%	385	NP_003260.1