

# Deciphering Distinct Genetic Risk Factors for FTLD-TDP Pathological Subtypes via Whole-Genome Sequencing

Corresponding Author: Professor Rosa Rademakers

**This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.**

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This is the largest genome-wide study (common and rare variant association) performed on FTLD-TDP cases. Interestingly, the authors describe common and rare genetic variants not only associated with FTLD-TDP but also with each FTLD-TDP subtype (A, B, and C). Lastly, they perform additional relevant analyses such as eTWAS, PWAS, sTWAS, and genetic colocalization, among others. This study describes novel loci associated with FTLD-TDP and its subtypes and highlights different genetic architectures associated with them.

Major comments:

- 1) Could you please provide the ALS and FTD diagnostic criteria used to classify patients as bvFTD/ALS or svPPA?
- 2) In this context, I have a major concern regarding the inclusion of svPPA as FTLD-type C. Three of the patients clinically diagnosed with svPPA are MAPT mutation carriers (2 p.P301L and 1 p.Q336H) (suppl. Table 20). In addition, not all patients with the right sided variant of svPPA could be confidently classified as FTLD-TDP C. In a recent study, only 64% of these patients were classified as FTLD-TDP C, while the rest were tau-MAPT, FTLD-TDP A, or B (Ulugut et al. Acta Neuropathol Commun. 2021 Aug 3;9(1):131). Therefore, given that three MAPT mutation carriers included in this study were diagnosed with svPPA and previous results show that almost 40% of right temporal variant of svPPA is not FTLD-type C, I am concerned that this might have increased the heterogeneity of the FTLD-type C group. Could you identify patients with the right-sided variant of svPPA so that you can remove them and repeat the analyses for the FTLD-type C subtype?
- 3) Patients with the C9orf72 repeat expansion or LOF GRN mutation carriers were removed prior to WGS. Why was not the patient with a LOF mutation in GRN (p.W304X; Suppl. Table 20) identified in the Sanger sequencing approach? In addition, why was this patient not removed after finding the LOF mutation in GRN, as was done with the other GRN mutation carriers?
- 4) LOF mutations in TBK1 are considered pathogenic. Why were patients harboring these mutations not removed as were those carrying the C9orf72 repeat expansion or LOF GRN mutations? In fact, 6 out of the 9 LOF TBK1 mutations were identified in Phase I.
- 5) Are any of the patients carrying a pathogenic variant in known genes (shown in suppl. Table 20) harboring a rare variant in VIPR1, RBPJL or L3MBTL1?
- 6) I am not sure whether TNIP1 is associated with FTLD-TDP as it does not reach statistical significance. TNIP1 is associated with FTLD-TDP B (which I found more interesting), and after meta-analyses with Dementia-seq, with FTLD. However, as authors describe, Dementia-seq includes patients clinically diagnosed with FTD (both tau and TDP43). Please, could you provide a detailed explanation?
- 7) The shared signal in TNIP1 (rs871269) has already been reported between ALS and ADRD (Wainberg M, Andrews SJ, Tripathy SJ. Shared genetic risk loci between Alzheimer's disease and related dementias, Parkinson's disease, and amyotrophic lateral sclerosis. Alzheimers Res Ther. 2023 Jun 16;15(1):113). Furthermore, figure 5 shows the genetic colocalization between UNC13A or TNIP1 loci in FTLD-TDP (meta-analyses) and ALS or ADRD. However, this group is neither FTLD-TDP All (as shown in the figure) nor FTLD-TDP as indicated in the figure legend, as it includes patients from Dementia-seq who were clinically diagnosed with FTD.

8) Please, provide genomic p-value inflation for all analyses.

9) In the gene prioritization strategy, why do monocytes and macrophages score in the eQTL-GWAs integration domain?

10) How do you explain that TMEM106B has a negative Z-score (-5.06) in eTWAS and a positive Z-score (5.61) in PWAS?

11) Figure 4 shows the enrichment of genes and cell types in each FTLD subtype. I am not sure if this figure is clear to the reader. In my opinion, it does not adequately reflect the enrichment scores (color) or the p-values (size). The color scale is confusing; I suggest using red at one extreme and yellow in the middle of the scale. Additionally, non-significant p-values are not easily distinguishable from statistically significant ones. For example, although statistically significant, the size of the circle for 'small intestine terminal ileum' is almost the same as other tissues in FTLD-TDP type C (not significant). Similarly, 'Astro' in FTLD-type A represents a p-value of 0.41. Without checking the supplementary tables, readers might find the circles confusing. The figure legend states that 'Asterisks denote brain regions or cell types enriched with FDR  $P < 0.05$ .' However, I do not see any asterisks, which could help the reader interpret the figure. I would suggest depicting non-significant p-values with a dot

Minor comments:

1) Although the authors correctly mention that accurately predicting the underlying neuropathological substrate is challenging during life, recent studies have shown that plasma extracellular vesicle tau and TDP43 levels can discriminate between TDP43 and tau with area under the curve values  $>0.9$ . Although beyond the scope of this study, this background information may be of interest to the reader (Chatterjee M, et al. Nat Med. 2024 Jun;30(6):1771-1783).

2) Regarding the contamination of samples, how was the Freemix score calculated? Using VerifyBamID2 tool? Does a freemix score of 4 corresponds to a 0.04% of contamination?

3) In Figure 1C (Manhattan plot of the FTLD-TDP B patients vs Controls), TNIP1 label corresponds to RCL1, whereas TNIP1 should be placed on chromosome 5.

4) The latest GWAs performed in FTD patients (Manzoni C, et al. Am J Hum Genet. 2024 Jul 11;111(7):1316-1329.) also provide evidence of UNC13A association in FTD. It may be of interest to the reader to cite this recently published result in the discussion section.

5) I find the gene CDC20B interesting in the FTLD TDP All rare variant analysis, which shows 5 rare variants in the group of cases and none in the controls (see supplementary table 17), with a p-value of  $9e-5$ . Could you please indicate which mutations are and to which group the carriers belong?

Reviewer #2

(Remarks to the Author)

Pottier and colleagues report the results of a genome-wide association study in a whole genome sequencing population composed of 985 FTLD-TDP patients and 3,153 controls. They confirm UNC13A as the most significant risk factor and identified TNIP1 as novel FTLD-TDP risk factor. They further performed a subgroup analysis in the 3 main FTLD-TDP neuropathological subgroups (FTLD-TDP A, B and C) and identified 7 novel risk loci: FARP2, TINAG, MZT1, GRN (FTLD-TDP A), TNIP1, PDS5B (FTLD-TDP B) and C19orf52 (FTLD-TDP C). Moreover, they performed a gene-based burden analysis and identified 3 genes harbouring rare variants contributing to FTLD-TDP risk (VIPR1, RBPJL und L3MBTL1).

The study is impressive and the experimental design highly meritorious.

I have only minor comments:

- Please, shortly describe the main 3 FTLD-TDP subtypes (A, B and C). Is there any correlation between TDP-43 aggregate foci in these diverse FTLD-TDP subtypes and peculiar enrichment of risk loci in different brain areas and neuronal phenotypes?

Reviewer #3

(Remarks to the Author)

The author provides a comprehensive and insightful explanation of the molecular mechanisms underlying FTD. The observed strong correlations between pathological subtypes, known causal mutations, and clinical phenotypes suggest that each subtype may represent a distinct pathway or mechanism leading to TDP-43 misfolding and accumulation. Investigating the biological and biochemical basis of these subtypes, their associations with various clinical manifestations, and their progression throughout the disease are critical research areas.

To validate previous findings and identify new genetic risk factors, the authors significantly increased their sample size, doubling both cases and controls. The GWAS utilized advanced techniques such as gene prioritization, enrichment analyses, and co-localization, leading to the discovery of novel FTLD-TDP risk loci specific to different pathological subtypes.

New loci associated with the disease are as follows:

#### FTLD-TDP A:

TINAG (rs138698596)

MZT1 (rs138959102)

FARP2 (rs886815)

#### FTLD-TDP B:

A genome-wide significant association was found at the UNC13A locus (rs12973192). This lead variant is in linkage disequilibrium (LD) with rs8111424 ( $D'=1$ ;  $r^2=0.43$ ) identified in the FTLD-TDP All analysis.

Additionally, three new genome-wide significant risk loci were identified:

TNIP1 (rs871269)

RCL1 (rs7674221)

PDS5B (rs5277499)

#### FTLD-TDP C:

C19orf52 (also known as TIMM29, rs576561313)

#### Minor Concerns and Recommendations:

Figure 2: The results of gene prioritization are currently presented in a confusing manner. I recommend consolidating these results for clarity, combining similar points into a cohesive summary, and potentially moving detailed data to the supplementary materials. This approach will help highlight the most impactful findings and reduce redundancy.

Figure 3: The GO Terms across Groups figure needs clarification, particularly regarding the overlapping genes. This information should be made more explicit.

Figure 5: The role of SNP rs871269 in the genes TNIP1 and ADRD requires further clarification. It is not clear how this SNP affects these genes.

Line 327: The statement on line 327 should be revised to clearly specify which genes overlap between different diseases and how this overlap is defined.

Lines 392-394: These lines need to be better articulated to clearly convey the intended message.

By addressing these points, the presentation of the results and their implications will be clearer and more effective

#### Reviewer #4

##### (Remarks to the Author)

In general, the authors present an interesting sequencing-based analysis of a low frequency but highly medically relevant form of dementia whose genetic risk factors remain poorly understood. However, there are a number of points throughout the paper where increased clarity/information for the reader is needed, and I have some methods/technical concerns as well.

1. My largest comment is that I am unclear on the similarity of methods for cases versus controls for sequencing and variant calling. I would much prefer to see joint recalling of data for cases and controls if possible (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9765323/>) but understand that this may not be feasible at this stage of the analysis. However this limitation needs to be mentioned, and many more details on control sample selection and control WGS methods are needed. Details could also be provided on average number of de novo variants per sample, total # of common variants in the post-QC dataset, how variant QC was conducted, etc. Are metrics/methods similar by WGS source? My understanding is the "Variant level quality control" section is for targeted FTLD case sequencing only, right?

2. Why were "non-Caucasian ethnicity" samples removed? How many individuals was this? Was genetic similarity assessed? Would also recommend avoiding the term "Caucasian" unless that was the only term used in participant self-report data (possible but I'm guessing something like White/Caucasian more likely used), as recommended by <https://www.nature.com/articles/d41586-021-02288-x> for example.

3. "The first FTLD-TDP genome-wide association study (GWAS) identified the TMEM106B locus (rs1990622), supporting lysosomal dysfunction in FTLD-TDP; however, this signal was strongly driven by GRN mutation carriers included in that study."

This statement needs more explanation for a general reader... which GRN carriers? What condition/subtypes is GRN known to associate with? What is the relationship between this TMEM106B locus and GRN? Is it fair to call GRN a new locus in Table 2 (is rs5848 distinct from the GRN variants used to remove cases prior to analysis)? More context for "C9orf72 repeat expansion and LOF mutations in GRN" being removed is needed for a general readership.

4. It should be easier for the reader to look at prioritized genes only at genome-wide significant loci. I'm not sure how I feel about inclusion of subthreshold, nonreplicated loci anyways- at the very least should be clear 1. How subthreshold defined and 2. What genes are from these subthreshold loci. Terminology such as "locus A4" does not add to clarity unless used consistently across tables (including Table 2).
5. In rare variant analysis is this the MAF for "any coding variant that met our filtering criteria" (i.e. most individuals do not have a variant that meets filtering criteria in these genes)? How many variants contribute to the tests in Table 3?
6. What do you mean by a "GWAS of rare variants" in the introduction? This should be clarified for the reader. It needs to be made clear to the reader how common and rare are defined. (I think you only ran your GWAS on variants with a MAF >1% in either patients or controls (which seems sensible due to power concerns) but this is buried in methods and makes the "GWAS of rare variants" piece extra unclear.
7. This paragraph on replication is important but should be supported by a look-up table, perhaps in the supplement, for the previously reported lead variants- what does not replicated mean? Is there any residual signal? What is the % sample overlap between this effort and prior ones? (quite high, right? So hard to really consider this a true replication effort, which should be mentioned... would recommend including a "new sample only" analysis to see what actually replicates in independent samples). Could a look-up table for all common variant GWAS hits for FTD (which I assume has mostly FTLD-TDP cases, even if pathology not known??) also be provided (and sample overlap be described)? (<https://pubmed.ncbi.nlm.nih.gov/24943344/> for example) Comparisons with Dementia-seq study (phs001963.v1.p1) findings also seem reasonable to include. Replication of rare variant gene-based signals should also be attempted. "In prior studies, besides UNC13A, common variants in the HLA and DPP6 loci were reported to be associated with FTLD11,13 511 . The HLA-DR5 locus was identified as associated with clinical 512 FTLD but was not replicated in phase I of our International FTLD-TDP WGS Consortium. HLA-DQA2 and DPP6 loci were reported as overall FTLD-TDP risk loci in phase I11 513 but were 514 not replicated in the current study."
8. What is meant by "corrected p-values" in methods? FDR? Bonferroni?
9. Are the 13 PCs used the top 13 PCs? (I think so, but this should be clarified in paragraph describing generation of PCs)
10. Minor point, but any evidence of inflation in common variant/rare variant analysis? Would be nice to see QQ plots.
11. specificity analysis (coloc PPH4=20.1%, for p12=10e-06)- do you mean you used different coloc priors in a sensitivity analysis? Would reword.
12. Could genetic correlation analyses between FTLD and ALS/ADRD/other neurodegenerative diseases be presented? Seems potentially quite informative and pretty easy. Or has this already been done with existing GWAS summary stats? Would add context to your discussion of UNC13A locus sharing between ALS and FTLD? Could also strengthen your claims about FTLD-TDP C having a distinct genetic architecture- versus just relying on top signals, in this underpowered analysis... statements like this one need more evidence, in my opinion "The recognition that individual FTLD-TDP subtypes could in fact be considered separate diseases with distinct pathomechanisms has significant implications for the design of clinical trials and therapeutic interventions."
13. two CpGs for TMEM106B-Assuming assignment of CpGs to genes is just based on proximity, not eQTM analysis or similar? This might be worth noting as a limitation.
14. This statement doesn't make sense- "we identified exome-wide significant association with the burden of common variants in GRN and TMEM106B". It is unusual to do an exome-wide analysis for common variants- more methods are needed. How did you ensure statistical independence of included common variants in these aggregate tests? You have TWAS results, is that what is meant? Should be clarified.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have addressed or answered most of the points raised by this reviewer. They have improved the discussion, including the limitations of the study, and made the appropriate changes to some figures and tables in the manuscript. Overall, this is a highly relevant study in the field of FTD genetics.

However, there is still one concern that requires clarification. In one of my comments, I requested that genomic inflation values should be provided for all analyses. Based on the response to my comment, the authors reported inflation values in the legend of Figure 1, which corresponds to the "Genome-wide association study on common variants". Could you please provide the inflation values associated with the rare-variant analyses?

----

I have carefully reviewed the responses to Reviewer #4. The authors largely addressed the suggestions and concerns raised by him/her. Therefore, in my opinion, I would accept the study for publication. In fact, in my questions as Reviewer #1, I had requested the genomic inflation values for the rare variant analyses. After carefully reading the response to Reviewer #4, I see that these values are already included in this new version of the manuscript (supplementary figure 9). Although I recommend acceptance of the manuscript, I believe that the high genomic inflation values are a limitation and must be further discussed. In this regard, the reported genomic inflation values are quite high, especially in the case of FTLD-TDP-A ( $\lambda$ : 2.18), which could result in an increase of false positives.

That being said, according to the authors' explanations, they have corrected for the top 13 principal components, included only genes with a minimum of three carriers in the patient group, controlled for the possibility of population stratification or contamination, filtered out low-quality variants, and retained only variants with REVEL > 0.75. Thus, they have carefully addressed various factors that could increase the genomic inflation value. Additionally, it is well known that the use of tests such as SKAT may contribute to this issue, which the authors, as experts in genomic analyses, have likely already considered. I believe the genomic inflation, as stated by the authors, is related to the small sample size. It is difficult to overcome this limitation given that the sample size is inherent to the unique study groups included. In my opinion, this should be further discussed as a limitation.

In this new version, the authors include the following sentence: "Even though we detected some inflation in our association tests likely due to small cohort size, we did detect nine exome-wide significant signals within FTLD-TDP pathological subtypes (Table 3, Supplementary Table 19-22)." I would also suggest adding a comment on the possibility of an increased rate of false positives in the rare variant analyses while acknowledging that this limitation is difficult to overcome, as the sample size is inherently linked to the unique and well-characterized study groups included in this study, the largest in the field of FTLD.

Reviewer #2

(Remarks to the Author)

The authors have adequately addressed the comments. This revision has significantly improved the manuscript. I do not have further comments.

Reviewer #3

(Remarks to the Author)

This work provides an in-depth annotation of gene ontology pathways and identifies tissue- and cell-type enrichments associated with genome-wide significant risk loci. By offering a deeper genetic and mechanistic perspective on FTLD-TDP, this study marks a major advance in unraveling its complex landscape and its connections with related neurodegenerative diseases.

Despite its significant contributions to the field of FTLD and related dementias, several key aspects warrant further clarification:

- **Rare Variant Analysis:** The functional impact of the identified rare variants remains unclear, particularly given their relatively low occurrence in patients. Figure 6 suggests that some rare variants may even be more prevalent in controls, raising questions about their pathogenicity.
- **Discussion Length & Clarity:** The discussion is extensive but at times dispersive. A new figure or table summarizing key findings could improve clarity and impact.
- **Shared Variants in AD and FTD:** It would be valuable to understand how the authors interpret the presence of the same genetic variant in both AD and FTD. Do these variants drive overlapping mechanisms, or are they modulated by disease-specific factors? Further discussion on this aspect, potentially in light of findings such as those in <https://doi.org/10.1038/s41588-022-01024-z>, would strengthen the paper's contributions.

This study undoubtedly pushes the boundaries of our understanding of FTLD-TDP, but addressing these open questions could further enhance its impact and translational potential.

**Open Access** This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

## RESPONSES TO REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This is the largest genome-wide study (common and rare variant association) performed on FTLD-TDP cases. Interestingly, the authors describe common and rare genetic variants not only associated with FTLD-TDP but also with each FTLD-TDP subtype (A, B, and C). Lastly, they perform additional relevant analyses such as eTWAS, PWAS, sTWAS, and genetic colocalization, among others. This study describes novel loci associated with FTLD-TDP and its subtypes and highlights different genetic architectures associated with them.

Major comments:

1) Could you please provide the ALS and FTD diagnostic criteria used to classify patients as bvFTD/ALS or svPPA?

We apologize that we did not specify more clearly the criteria used to diagnose the clinical patient population used in our study. We used established criteria (where available) and have now specified more clearly our approach in the methods section of the manuscript, under the heading 'Samples'. A detailed discussion on the diagnosis of svPPA patients is also included in the discussion (see next comment).

"Patients were diagnosed according to established diagnostic criteria<sup>1-4</sup>; however, in the subgroup of patients with svPPA, in addition to including svPPA patients diagnosed according to Gorno-Tempini et al.<sup>2</sup>, patients previously diagnosed with semantic dementia according to Neary et al.<sup>4</sup>, and a few patients suspected to have the right temporal variant of FTD were included. For the latter group of patients diagnostic criteria are only now being developed<sup>5,6</sup>, we thus relied on the expert knowledge of the specialized dementia centers to identify these patients, recognizing the limitations of this approach, especially since this was not performed systematically in all contributing centers."

In this context, I have a major concern regarding the inclusion of svPPA as FTLD-type C. Three of the patients clinically diagnosed with svPPA are MAPT mutation carriers (2 p.P301L and 1 p.Q336H) (suppl. Table 20). In addition, not all patients with the right sided variant of svPPA could be confidently classified as FTLD-TDP C. In a recent study, only 64% of these patients were classified as FTLD-TDP C, while the rest were tau-MAPT, FTLD-TDP A, or B (Ulugut et al. Acta Neuropathol Commun. 2021 Aug 3;9(1):131). Therefore, given that three MAPT mutation carriers included in this study were diagnosed with svPPA and previous results show that almost 40% of right temporal variant of svPPA is not FTLD-type C, I am concerned that this might have increased the heterogeneity of the FTLD-type C group. Could you identify patients with the right-sided variant of svPPA so that you can remove them and repeat the analyses for the FTLD-type C subtype?

We understand that the possible inclusion of clinically diagnosed patients which in fact do not have FTLD-TDP type C underlying pathology in the analyses may have introduced heterogeneity and potentially obscured the identification of true FTLD-TDP type C risk factors. The link between bvFTD/ALS and FTLD-TDP type B is likely stronger than the link between svPPA and FTLD-TDP type C,

yet, we still felt that including these clinical cases was a unique opportunity to increase sample size and power to identify possible FTLD-TDP type C specific risk factors given that the majority of the svPPA cases (either left- or right predominant) will in fact have FTLD-TDP type C. Given that we have full genomic information on everyone, we are confident that the 3 *MAPT* carriers already mentioned are the only *MAPT* carriers and so we know that this did not have a major impact on our results. However, we understand that even without a genetic mutation other diagnoses are still possible. We thus consulted with the many clinical collaborators on this manuscript and there was consensus that the pathological diagnosis of FTLD-TDP type C could not be guaranteed in all svPPA cases, yet they also felt that both left-sided and right-sided svPPA could lead to other pathologies (e.g. Pick's disease, Alzheimer's disease, GGT) and excluding all right-sided svPPA cases as suggested would therefore not fully solve the problem. Moreover, the discussion on diagnostic criteria for 'right-sided svPPA' or right-temporal FTD is a topic of active discussion and research for which working groups have been established; however, consensus criteria are not yet available. We thus fully agree with the reviewer that – in time- separate analysis for left- and right-predominant svPPA cases would be of interest but without clear diagnostic criteria and with the difficulty that this would require the review of more than 280 files from patients seen at dozens of hospitals this is not something we could achieve for the current manuscript.

Therefore, to address this concern and determine whether indeed specific risk factors may have been missed, we performed in the revised version of our manuscript an additional GWAS using only the pathologically confirmed FTLD-TDP type C patients. We also performed all gene prioritization, cell- and tissue type enrichment and rare variant burden analysis in this pathological subgroup. This new analysis (named FTLD-TDP C\*) is now incorporated in the results and discussion, as well as in all relevant tables and figures of the manuscript. In line with the expectations of the reviewer, focusing on a purely pathological cohort of FTLD-TDP C patients indeed identified 4 novel genetic risk factors, which were not present when the svPPA cases were included. A new detailed discussion on these results with comparison of the contribution of the pathological cases as compared to the clinical cases is now included (**Supplementary Table 24**). Based on the available data, we strongly feel that the original GWAS including both FTLD-TDP C and svPPA patients provides the most robust results. However, replication in future studies will be required.

Here is the relevant section from the discussion: “Importantly, however, while often considered a sporadic FTLD subtype<sup>7-9</sup>, we implicate several genes and risk loci in FTLD-TDP C and uncover a potential role for mitochondrial membrane dysfunction and the notch signaling pathway. *C19orf52* (*TIMM29*), which mediates the import and insertion of multi-pass transmembrane proteins into the mitochondrial inner membrane, was identified as the first genome-wide significant risk locus for FTLD-TDP C. This locus was identified when svPPA patients were included in the GWAS, and both pathologically confirmed FTLD-TDP C and svPPA patients were found to contribute to the association (MAF FTLD-TDP C: 0.028, svPPA: 0.019, controls: 0.001; **Supplementary Table 24**). This finding illustrates the power we gained by including the clinically diagnosed patients, notwithstanding the fact that the pathological diagnosis will not be FTLD-TDP C in all. The inclusion of svPPA patients in our study is somewhat complicated by the evolving clinical definitions of patients affected by focal anterior temporal lobe (ATL) atrophy in recent years. Those with left-predominant ATL atrophy show severe anomia and verbal semantic deficits leading to diagnoses of svPPA and (previously) semantic dementia<sup>2,4</sup>. However, patients with right ATL atrophy have been

more challenging to fit into current diagnostic criteria and may have received diagnoses such as right-sided svPPA, right temporal variant of FTD or, most recently, semantic behavioral variant of FTD<sup>5,10-14</sup>. At the start of our study in 2019, this field was just starting to evolve and we allowed inclusion of left- and right-sided svPPA patients as both were thought to have predominant FTLD-TDP C pathology<sup>15</sup>. More recently, through large cohort studies and the establishment of an international Working Group<sup>6</sup>, systematic re-assessment of clinically diagnosed patients with svPPA and bvFTD is ongoing with an emphasis on recognizing and diagnosing patients with right-temporal atrophy. As criteria are still in development and our clinical patients were not yet systematically assessed in this new framework, we also performed an additional GWAS including only pathologically confirmed FTLD-TDP C patients (FTLD-TDP C\*). This analysis led to the identification of 4 novel genome-wide significant loci (*LRP1B*, *COL22A1*, *TRPC4* and *TMEM135*). While these loci remained nominally significant in the combined analysis, inspection of the results showed no contribution from the svPPA patients to the association (**Supplementary Table 24**). Even with a heterogeneous population of clinical patients one would have expected an increased frequency of risk alleles in patients as compared to controls, raising concern that these additional loci may represent type I errors. In fact, the rare-variant burden analyses in FTLD-TDP C also showed more hits when including the svPPA patients.”

And..

“For FTLD-TDP C, the relative contribution of patients with left- and right-predominant ATL atrophy to the observed associations should be assessed once clinical criteria for patients with right temporal atrophy are finalized.”

3) Patients with the C9orf72 repeat expansion or LOF GRN mutation carriers were removed prior to WGS. Why was not the patient with a LOF mutation in GRN (p.W304X; Suppl. Table 20) identified in the Sanger sequencing approach? In addition, why was this patient not removed after finding the LOF mutation in GRN, as was done with the other GRN mutation carriers?

The mutation in this patient was originally missed by Sanger sequencing but identified by WGS. All the other FTLD-TDP A participants that were positive based on Sanger sequencing prior to WGS were excluded from the study. We excluded these individuals since they represent a substantial proportion of FTLD-TDP A genetic causes. Because the p.W304X mutation was only discovered after sequencing, processing, quality control and PCA analyses, we decided to keep it as this is only one individual, and very likely has a minor or no effect on our findings. That individual does not carry any of the rare variants that we identified in this study.

To explain what happened with this one *GRN* carrier, we added a note to **Supplementary Table 26**.

“a: *GRN* loss-of-function mutation was missed by Sanger sequencing in the analysis prior to start of whole genome sequencing.”

4) LOF mutations in *TBK1* are considered pathogenic. Why were patients harboring these mutations not removed as were those carrying the C9orf72 repeat expansion or LOF *GRN* mutations? In fact, 6 out of the 9 LOF *TBK1* mutations were identified in Phase I.

We understand the reviewer's comment. There were in fact multiple reasons why *TBK1* mutation carriers were not removed similar to *C9ORF72* and *GRN* mutation carriers. First of all, for most samples included in Phase II sequencing, the *TBK1* mutation status was not known prior to WGS. So we would have had to remove them after sequencing, processing and QC. We obviously did know about the LOF *TBK1* carriers from Phase I but we decided to keep the Phase I cohort unchanged (as compared to the previous study) and instead use *TBK1* as a positive control in our burden analyses to ensure we are identifying relevant genes associated with FTLT. Following this approach, we are now able to consistently report an association of *TBK1* with FTLT-TDP also in this larger study, which provides consistency with our first publication and confidence for *TBK1* as FTLT-TDP risk gene.

5) Are any of the patients carrying a pathogenic variant in known genes (shown in suppl. Table 20) harboring a rare variant in *VIPR1*, *RBPJL* or *L3MBTL1*?

We thank the reviewer for this question. Even though oligogenic mechanisms have been reported for ALS (part of the FTLT disease spectrum), such events are extremely rare. In our cohort we did not find double mutants in the known genes and any of the new genes we identified. This is now indicated in the results section as follow:

"None of the rare variant carriers had a mutation in a known FTLT gene."

We also want to point out an error on the reporting of the associated genes with an increased burden of rare variants in our original submission. While preparing the revised manuscript we noticed that we mistakenly used a too strict genome-wide significant cut-off rather than the required exome-wide significance with p-value threshold of  $p < 2.5 \times 10^{-6}$  (Bonferroni correction for 20,000 genes) to identify the associated genes with an increased burden of rare variants. For the overall analysis and type B it was done correctly in our original submission, yet for type A and C we were too strict. Using the correct significance threshold, we identified rare variants in 2 additional genes associated with FTLT-TDP A (*C3AR1* and *SMG8*) and one additional gene for FTLT-TDP C (*ANO9*). We have updated the results and discussion to reflect these changes and also added additional Supplementary tables to show the detailed results (**Supplementary Tables 18-23**).

6) I am not sure whether *TNIP1* is associated with FTLT-TDP as it does not reach statistical significance. *TNIP1* is associated with FTLT-TDP B (which I found more interesting), and after meta-analyses with Dementia-seq, with FTLT. However, as authors describe, Dementia-seq includes patients clinically diagnosed with FTD (both tau and TDP43). Please, could you provide a detailed explanation?

Indeed during the discovery phase *TNIP1* is not genome-wide significant with a pvalue= $4.42 \times 10^{-07}$  in FTLT-TDP All versus controls. However, it is already genome wide significant in FTLT-TDP B versus controls (pvalue= $4.72 \times 10^{-08}$ ), this suggests that the signal detected in FTLT-TDP All is driven by the FTLT-TDP B group which includes pathologically confirmed FTLT-TDPB cases and clinical bvFTD/ALS patients. While performing the meta-analysis combining FTLT-TDP All and clinical cases from the Dementia-Seq cohort, the signal became even stronger and genome-wide significant. This is most certainly due to the presence of bvFTD/ALS patients in the Dementia-seq cohort. Unfortunately, no detailed clinical information from this cohort is available so we cannot assess this

accurately.

We added the following sentence in the results section:

“Interestingly, while the *TNIP1* signal in FTLD-TDP All was not genome-wide significant, it became significant when running the meta-analysis. The change in significance likely results from the inclusion of participants with bvFTD/ALS (with likely FTLD-TDP B pathology), as part of the DementiaSeq cohort.”

7) The shared signal in *TNIP1* (rs871269) has already been reported between ALS and ADRD (Wainberg M, Andrews SJ, Tripathy SJ. Shared genetic risk loci between Alzheimer's disease and related dementias, Parkinson's disease, and amyotrophic lateral sclerosis. *Alzheimers Res Ther.* 2023 Jun 16;15(1):113). Furthermore, figure 5 shows the genetic colocalization between *UNC13A* or *TNIP1* loci in FTLD-TDP (meta-analyses) and ALS or ADRD. However, this group is neither FTLD-TDP All (as shown in the figure) nor FTLD-TDP as indicated in the figure legend, as it includes patients from Dementia-seq who were clinically diagnosed with FTD.

In fact, the *TNIP1* signal was previously reported as different between ALS and ADRD<sup>16</sup>. In our case, we report a similar association between FTLD-TDP and ADRD. On the other hand, we performed colocalization analysis between the *TNIP1* signal from our FTLD association study and the ALS GWAS and report a weak correlation. This is in line with the colocalization analysis performed in Wainberg et al between ALS and ADRD that only reported a 0.9% colocalization. We added the following sentence to the discussion:

“This is in line with the recent study reporting independent *TNIP1* association signals in ALS and ADRD<sup>16</sup>”

Moreover, we apologize for the confusion as a result of the labels used in the figure and the legend. We now changed it into FTLD and made clear it includes samples with FTLD-TDP and clinical FTLD patients from Dementia-seq.

8) Please, provide genomic p-value inflation for all analyses.

We are now providing the inflation value for all analyses in the legend of figure 1.

9) In the gene prioritization strategy, why do monocytes and macrophages score in the eQTL-GWAS integration domain?

We thank the reviewer for this question. Similar to our previous pipeline for prioritizing risk genes for AD GWAS<sup>17</sup>, we have included 8 naïve state monocyte/macrophage eQTL catalogues, uniformly reprocessed by the eQTL Catalogue<sup>18</sup>, in the eQTL-GWAS integration domain for several reasons. Historically in early studies predating development of efficient brain microglia isolation or single-cell/nuclei RNA sequencing methods<sup>19,20</sup>, monocytes/macrophages were used as a proxy for microglia, as they are closely related to microglia, altogether sharing a myeloid lineage with immune functions. Even with the recently available primary microglia<sup>21,22</sup> and snRNA-seq<sup>23</sup> based ct-eQTL catalogues containing microglia ct-eQTLs, using these monocytes/macrophages eQTL catalogues provide several important added values. First and most importantly, there is a growing body of evidence converging on the fact that the peripheral immune cells, including monocytes/macrophages, infiltrate into central nervous system, especially in pathophysiological

neuroinflammatory conditions in brain<sup>24</sup>. This proposes a potentially important role of monocytes/macrophages for neurodegenerative brain diseases, including for FTLN<sup>25,26</sup>. Second, despite the recently available microglia ct-eQTL catalogues, the sample sizes for available monocytes/macrophage eQTL catalogues are still 1.85-fold higher on average compared to microglia ct-eQTL catalogues included in our study, including Fairfax monocyte eQTL study<sup>27</sup> that included 420 samples (see **Supplementary Table 5**). Therefore, these monocyte/macrophage eQTL catalogues can provide important insights with increased discovery power due to sample size. In fact, in our study the eQTL-GWAS coloc results (**Supplementary Table 7**) show that there were a total of 94 coloc hits ( $PP4 \geq 70\%$ ) for 56 genes, arising from integrating 5 distinct FTLN-TDP GWAS with 8 monocyte/macrophage eQTLs; meanwhile microglia ct-eQTL integration (using 7 such catalogues) resulted in identification of 22 coloc hits ( $PP4 \geq 70\%$ ) for 14 genes. When normalized per number of catalogues, this translates into a 3.7-fold increase in number of coloc hits and a 3.5-fold increase in number of unique genes nominated by eQTL-GWAS coloc hits. Importantly, 67.9% of the monocyte/macrophage eQTL coloc hits pointed out to tier 1 and tier 2 prioritized genes in all investigated loci across 5 FTLN-TDP GWAS analyses, despite being weighted with 2-fold lower scores (**Supplementary Table 3**) compared to microglia, for which 71.4% of the ct-eQTL coloc hits pointed out to tier 1 and tier 2 prioritized genes. The genes whose prioritization was contributed by monocyte/macrophage eQTL integration included *TMEM106B* tier 1 prioritized risk gene we highlighted in A\_S14 (FTLN-TDP A), which had monocyte eQTL coloc hits in two catalogues (**Supplementary Table 7**). Taken together, we strongly believe that using monocyte/macrophage eQTL catalogues contributes significantly to our analyses for gene prioritization and functional interpretation.

10) How do you explain that *TMEM106B* has a negative Z-score (-5.06) in eTWAS and a positive Z-score (5.61) in PWAS?

We thank the reviewer for raising this point. We are aware of the intriguing observation in our eTWAS/PWAS results that the protective FTLN-TDP A GWAS signal in the locus was associated with genetic upregulation of *TMEM106B* gene expression, but also with genetic downregulation of *TMEM106B* protein expression, in bulk brains (**Supplementary Table 12 and 14**). First of all, we attempted to study these results in more detail. The eTWAS hit was obtained in GTEx cortex, while the PWAS hit was obtained in ROSMAP and Banner dorsolateral prefrontal cortex (DLPFC) brain regions. Ideally we would have investigated both RNA and protein effects in ROSMAP DLPFC, in which we have both eTWAS and PWAS results with a large overlapping sample ( $n=560$  and  $n=376$  samples, respectively), however *TMEM106B* was not included in ROSMAP DLPFC eTWAS reference panel as it was not a heritable feature in this panel (heritability  $P = 0.2$ , where features with heritability  $P < 0.05$  were modelled; see Bellenguez et al.<sup>17</sup>). We next performed a detailed investigation of the variants contributing to the eTWAS hit in GTEx cortex (MASHR model in PrediXcan) and to the PWAS hit in ROSMAP DLPFC and Banner DLPFC, and between these rs3807865 was the most relevant one showing pleiotropic effects for *TMEM106B* gene and protein expression. In FTLN-TDP A GWAS (where the top-associated variant in *TMEM106B* locus was rs10281425 [ $P = 2 \times 10^{-7}$ ,  $\beta = -0.616$ ]), rs3807865 was the 5<sup>th</sup> most significant variant ( $P = 2.78 \times 10^{-7}$ ,  $\beta$  for allele A = -0.612), positioned within promoter region of *TMEM106B*. rs3807865 was a significant eQTL in GTEx cortex ( $P = 4 \times 10^{-5}$ ,  $\beta$  for protective allele A = +0.14) and significant pQTL in ROSMAP DLPFC ( $P = 1 \times 10^{-24}$ ,  $\beta$  for

protective allele  $A = -0.26$ ), therefore in line with overall eTAS and PWAS-derived insights we obtained for *TMEM106B*.

While at first glance these results may be counterintuitive, we believe there is a biological explanation for this as we also mentioned and cited already in Lines 492-495 of our manuscript. There is now a growing body of work showing the presence of amyloid-like fibrils composed of the C-terminal fragment of *TMEM106B* across neurodegenerative diseases, including FTLD-TDP. These insoluble aggregates arise at the cost of the soluble form of *TMEM106B* (which is the species likely detected by mass spectroscopy). *TMEM106B* risk alleles/haplotype are associated both with more expression of *TMEM106B* mRNA but also with the formation of insoluble aggregates. Specifically, in our recent study we showed that the number of *TMEM106B* risk haplotypes and insoluble *TMEM106B* C-terminal fragment (CTF) burden are positively correlated<sup>28</sup>. So we argue that our observations can be explained by the fact that *TMEM106B* risk alleles lead to more *TMEM106B* mRNA but less *TMEM106B* protein because mass spectrometry methods detect the soluble *TMEM106B* protein which will be lower when there are more fibrils.

11) Figure 4 shows the enrichment of genes and cell types in each FTLD subtype. I am not sure if this figure is clear to the reader. In my opinion, it does not adequately reflect the enrichment scores (color) or the p-values (size). The color scale is confusing; I suggest using red at one extreme and yellow in the middle of the scale. Additionally, non-significant p-values are not easily distinguishable from statistically significant ones. For example, although statistically significant, the size of the circle for 'small intestine terminal ileum' is almost the same as other tissues in FTLD-TDP type C (not significant). Similarly, 'Astro' in FTLD-type A represents a p-value of 0.41. Without checking the supplementary tables, readers might find the circles confusing. The figure legend states that 'Asterisks denote brain regions or cell types enriched with FDR  $P < 0.05$ .' However, I do not see any asterisks, which could help the reader interpret the figure. I would suggest depicting non-significant p-values with a dot.

We have now adapted the Figure 4 according to the reviewer's comments. The highest z score is now orange and the lowest blue. Each significant pvalue is depicted as an asterisk in the circle.

Minor comments:

1) Although the authors correctly mention that accurately predicting the underlying neuropathological substrate is challenging during life, recent studies have shown that plasma extracellular vesicle tau and TDP43 levels can discriminate between TDP43 and tau with area under the curve values  $>0.9$ . Although beyond the scope of this study, this background information may be of interest to the reader (Chatterjee M, et al. Nat Med. 2024 Jun;30(6):1771-1783).

We thank the reviewer for sharing this. We have now added this information in the introduction as follows:

“Extensive research is focused on the accurate prediction of the underlying FTLD neuropathology in clinically diagnosed FTLD patients, however despite some recent reports on the possible use of Tau and TDP43 levels in plasma extracellular vesicles, this remains a major diagnostic challenge.<sup>29</sup>”

2) Regarding the contamination of samples, how was the Freemix score calculated? Using VerifyBamID2 tool? Does a freemix score of 4 corresponds to a 0.04% of contamination?

Estimates of sample contamination using a sequence-only method (i.e., no external genotypes were available for comparison) were calculated using data on all autosomes with 1000 Genomes European array allele frequencies as reference. The estimated contamination value, FREEMIX, provides a sequence-only estimate of contamination on a 0-1 scale. The FREEMIX estimate uses excessive heterozygosity to estimate the level of contamination. Freemix score and its use is now described in more detail in the manuscript:

“contamination defined by a FREEMIX score above 0.04 (4% or more of non-reference bases are observed in reference sites) were removed. In more detail, estimates of sample contamination were calculated using data on all autosomes with 1000 Genomes European array allele frequencies as reference. The estimated contamination value, FREEMIX 1.3, provides a sequence-only estimate of contamination on a 0-1 scale using excessive heterozygosity. In addition, we identified 31 non-European White participants (Admixture EUR Probability < 0.7) which we removed from analysis because of the small sample size and to increase genetic background homogeneity.”

3) In Figure 1C (Manhattan plot of the FTLD-TDP B patients vs Controls), TNIP1 label corresponds to RCL1, whereas TNIP1 should be placed on chromosome 5.

We thank the reviewer for noting this and we made the appropriate change.

4) The latest GWAs performed in FTD patients (Manzoni C, et al. Am J Hum Genet. 2024 Jul 11;111(7):1316-1329.) also provide evidence of UNC13A association in FTD. It may be of interest to the reader to cite this recently published result in the discussion section.

We now added this information in the discussion section:

“This locus was also reported in a recent large GWAS focused on clinical FTLD patients and controls without reaching genome-wide significance ( $P=1.07 \times 10^{-3}$ ), possibly due to the heterogeneity of the underlying pathologies in this cohort with only a subset having TDP-43 pathology<sup>30</sup>. The same *UNC13A* locus was previously linked to ALS<sup>31</sup> and is known to modify the phenotype in ALS patients by increasing the risk of frontotemporal cortical atrophy and diminished cognitive performance, which are reminiscent of an FTLD clinical presentation<sup>32</sup>.”

5) I find the gene CDC20B interesting in the FLTD TDP All rare variant analysis, which shows 5 rare variants in the group of cases and none in the controls (see supplementary table 17), with a p-value of  $9e-5$ . Could you please indicate which mutations are and to which group the carriers belong?

Variants in *CDC20B* were identified in all groups of FTLD-TDP. Only 3 variants were detected in 5 different individuals. Please see the table below for the number of individuals carrying each variant:

| Variant ID      | cDNA       | Protein Change | FTLD-TDP A | FTLD-TDP B | FTLD-TDP C |
|-----------------|------------|----------------|------------|------------|------------|
| chr5:55124847:A | c.C1171T   | p.Q391X        | 1          | 1          |            |
| chr5:55125030:G | c.990-2A>C |                |            |            | 1          |
| chr5:55143511:G | c.486+2T>C |                | 1          |            | 1          |

Reviewer #2 (Remarks to the Author):

Pottier and colleagues report the results of a genome-wide association study in a whole genome sequencing population composed of 985 FTLD-TDP patients and 3,153 controls. They confirm UNC13A as the most significant risk factor and identified TNIP1 as novel FTLD-TDP risk factor. They further performed a subgroup analysis in the 3 main FTLD-TDP neuropathological subgroups (FTLD-TDP A, B and C) and identified 7 novel risk loci: FARP2, TINAG, MZT1, GRN (FTLD-TDP A), TNIP1, PDS5B (FTLD-TDP B) and C19orf52 (FTLD-TDP C). Moreover, they performed a gene-based burden analysis and identified 3 genes harbouring rare variants contributing to FTLD-TDP risk (VIPR1, RBPJL und L3MBTL1).

The study is impressive and the experimental design highly meritorious.

We thank the reviewer for these kind words.

I have only minor comments:

- Please, shortly describe the main 3 FTLD-TDP subtypes (A, B and C). Is there any correlation between TDP-43 aggregate foci in these diverse FTLD-TDP subtypes and peculiar enrichment of risk loci in different brain areas and neuronal phenotypes?

We now describe the main pathological FTLD-TDP subtype in the Introduction as follows:

“The main subtypes are types A, B and C. Type A includes patients with moderate to numerous TDP-43-immunoreactive neuronal cytoplasmic inclusions (NCIs) and short dystrophic neurites (DNs) concentrated mainly in the upper cortical layers II/III. Type B encompasses patients with moderate to numerous TDP-43-immunoreactive NCIs and sparse DNs distributed across all cortical layers. Type C is designated for patients where long dystrophic neurites are prevalent primarily in the upper cortices and NCIs are infrequent.”

Moreover, we identified distinct genetic risk factors for each of the subtypes and showed that the associated genes have unique expression profiles across cell types and brain regions. A possible next study could indeed focus on the question whether specific vulnerable cell subpopulations differentially express associated genes and whether this is correlated with the subtypes. This is very interesting but not something we have studied at this point. This would ideally be studied in single nuclei transcriptomic datasets, including those from FTLD patients. We are generating such dataset at the moment but answering this question is beyond the scope of this manuscript.

We did however, add a statement to the discussion to indicate that this type of future work could be of interest.

“Whether specific neuronal subpopulations differentially express associated genes could be the focus of future studies once single-nuclei transcriptomic datasets from FTLD-TDP patients and controls are available.”

Reviewer #3 (Remarks to the Author):

The author provides a comprehensive and insightful explanation of the molecular mechanisms underlying FTD. The observed strong correlations between pathological subtypes, known causal mutations, and clinical phenotypes suggest that each subtype may represent a distinct pathway or mechanism leading to TDP-43 misfolding and accumulation. Investigating the biological and biochemical basis of these subtypes, their associations with various clinical manifestations, and their progression throughout the disease are critical research areas.

To validate previous findings and identify new genetic risk factors, the authors significantly increased their sample size, doubling both cases and controls. The GWAS utilized advanced techniques such as gene prioritization, enrichment analyses, and co-localization, leading to the discovery of novel FTLD-TDP risk loci specific to different pathological subtypes. New loci associated with the disease are as follows:

FTLD-TDP A:

TINAG (rs138698596)

MZT1 (rs138959102)

FARP2 (rs886815)

FTLD-TDP B:

A genome-wide significant association was found at the UNC13A locus (rs12973192). This lead variant is in linkage disequilibrium (LD) with rs8111424 ( $D'=1$ ;  $r^2=0.43$ ) identified in the FTLD-TDP All analysis.

Additionally, three new genome-wide significant risk loci were identified:

TNIP1 (rs871269)

RCL1 (rs7674221)

PDS5B (rs5277499)

FTLD-TDP C:

C19orf52 (also known as TIMM29, rs576561313)

Minor Concerns and Recommendations:

Figure 2: The results of gene prioritization are currently presented in a confusing manner. I recommend consolidating these results for clarity, combining similar points into a cohesive summary, and potentially moving detailed data to the supplementary materials. This approach will help highlight the most impactful findings and reduce redundancy.

In response to this reviewer and reviewer #4 we have now eliminated the subthreshold loci from Figure 2 and added data from the FTLT-TDP C\* analysis. In supplemental figures, we are still providing the full overview so that interested readers can find it. We did opt to keep all categories in the figure that contribute to the weighted evidence category scores so that readers can immediately see what the evidence is for the contribution of each gene and why it was selected as Tier 1 or Tier 2 respectively.

Figure 3: The GO Terms across Groups figure needs clarification, particularly regarding the overlapping genes. This information should be made more explicit.

We understand the reviewers comment. We have now added a supplementary figure showing the overlap between the genes within the GO terms. We added the following sentence in the result section:

“Note that some genes belong to multiple significant GO terms; for instance, *GRN* is present in terms associated with FTLT-TDP A and B (**Supplementary Table 15, Supplementary Fig.8**).”

**Supplementary Figure 8. Gene overlap between the significant Gene Ontology terms.** Upset plot illustrating the overlap of different gene sets in the significant Gene Ontology terms. The vertical bars represent the total number of unique genes in each set, while the intersections indicate the number of genes shared among the sets. The horizontal axis lists the Gene Ontology terms sets, with each bar corresponding to the number of genes present in that specific intersection.

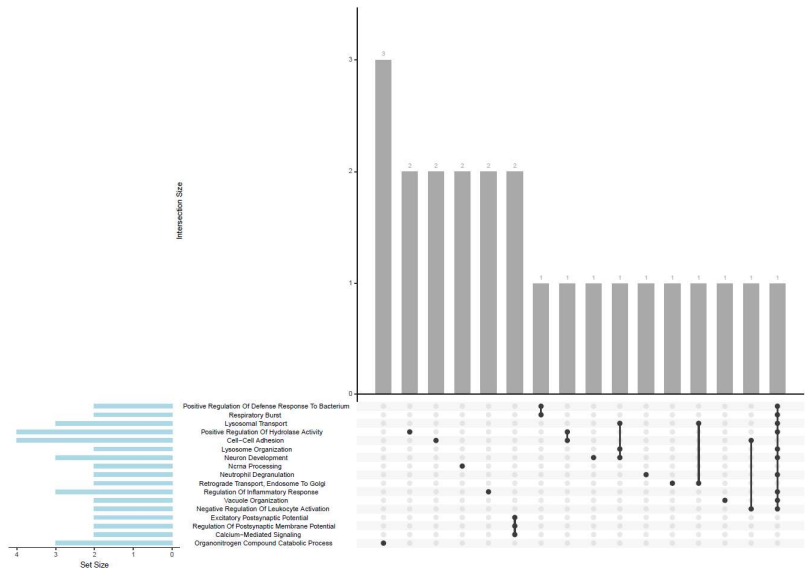


Figure 5: The role of SNP rs871269 in the genes *TNIP1* and ADRD requires further clarification. It is not clear how this SNP affects these genes.

We first want to clarify that ADRD refers to patients with Alzheimer's disease and related disorders. This is not a gene name. The variant rs871269 is located within the genomic region of the *TNIP1* gene. Because this same genomic region was previously implicated in both ALS and ADRD, we performed a colocalization analysis. We found it very interesting that our genetic finding in FTLT seems to be the same as what has been reported in ADRD but distinct from the genetic association reported in ALS. The recognition that there are two independent signals at this locus, each conferring risk to neurodegeneration, is in line with a recent study where the authors concluded using colocalization analysis that ADRD and ALS were unlikely to share a causal variant at this locus. We have now added this to our manuscript in the discussion.

"This is in line with the recent study reporting independent *TNIP1* association signals in ALS and ADRD<sup>16</sup>."

Line 327: The statement on line 327 should be revised to clearly specify which genes overlap between different diseases and how this overlap is defined.

We thank the reviewer for pointing this out. We used the function Cell Type from the FUMA website using the summary statistics from Bellenguez et al and Project Mine. The Cell Type module combines GWAS summary statistics with gene expression data across different cell types to identify whether specific cell types are enriched for the genes associated with the trait of interest. We selected genes that had a  $P < 10^{-5}$  for each trait. Therefore, the gene sets used are unique to each disease. We now added that description in the results section:

"Only loci with  $P < 10^{-5}$  for their respective traits were included."

"We also compared expression profiles of the genes at FTLT loci with expression of genes located at associated loci in Alzheimer's disease and related disorders (ADRD)<sup>17</sup> and ALS GWAS<sup>33</sup> using available summary statistics and restricting our query to loci with  $P < 10^{-5}$  for their respective traits."

Lines 392-394: These lines need to be better articulated to clearly convey the intended message.

We have rephrased this sentence to better convey our intended message.

"*UNC13A* variants thus enhance the effect of TDP-43 dysfunction, yet both ALS and FTLT-TDP are characterized by TDP-43 pathology, suggesting that additional genetic or environmental mechanisms must exist to explain the tissue specificity of the pathology and associated clinical phenotypes in individual patients. "

By addressing these points, the presentation of the results and their implications will be clearer and more effective

Reviewer #4 (Remarks to the Author):

In general, the authors present an interesting sequencing-based analysis of a low frequency but

highly medically relevant form of dementia whose genetic risk factors remain poorly understood. However, there are a number of points throughout the paper where increased clarity/information for the reader is needed, and I have some methods/technical concerns as well.

1. My largest comment is that I am unclear on the similarity of methods for cases versus controls for sequencing and variant calling. I would much prefer to see joint recalling of data for cases and controls if possible (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9765323/>) but understand that this may not be feasible at this stage of the analysis. However this limitation needs to be mentioned, and many more details on control sample selection and control WGS methods are needed. Details could also be provided on average number of de novo variants per sample, total # of common variants in the post-QC dataset, how variant QC was conducted, etc. Are metrics/methods similar by WGS source? My understanding is the “Variant level quality control” section is for targeted FTLD case sequencing only, right?

After reading this reviewer’s comment we realize that we poorly explained our quality control measures and methodology. We sincerely apologize as we now realize this created significant confusion and doubt on our analysis. The good news is that we already performed joint genotyping, we just failed to state this explicitly. The ‘Variant level quality control’ section is in fact referring to both FTLD case data and control data, which was applied together, as we performed joint-genotyping of the patients and controls. As recommended by the reviewer in the referenced article we included the intermediately processed data (gVCF) from the ADSP project to allow joint recalling of cases and controls together, thereby improving variant calling and harmonization. We apologize for not stating this more clearly in the initial submission.

We now have adapted the method section as follow:

“Fastq files for all patients and controls were transferred to Mayo Clinic and processed through the Mayo Genome GPS v4.0 pipeline in batches of up to 75 samples using the Burrows–Wheeler Aligner to map reads to the human reference sequence (GRCh38 build). Local realignment around indels and variant calling were performed using GATK HaplotypeCaller, followed by variant recalibration (VQSR) according to GATK best practice recommendations. Participants from the Mayo Clinic Biobank with a possible clinical diagnosis or family history of a neurodegenerative disorder were removed during analysis.”

“To enhance our study, we incorporated genomic variant call format (gVCF) files from 2,037 controls obtained from the Alzheimer’s Disease Sequencing Project (ADSP). gVCF files provide a comprehensive record of variant calls and reference positions, which are essential for accurate joint-genotyping. The gVCF files from ADSP controls were merged with our cohort’s gVCF files using the joint-genotyping approach implemented with the Genome Analysis Toolkit (GATK). By merging these gVCFs, we ensured all our patients and controls were analyzed together, allowing for a more robust comparison and reducing batch effects. Overall, we obtained genomic data on a total of 1,061 patients and 3,341 controls.”

2. Why were “non-Caucasian ethnicity” samples removed? How many individuals was this? Was genetic similarity assessed? Would also recommend avoiding the term “Caucasian” unless that

was the only term used in participant self-report data (possible but I'm guessing something like White/Caucasian more likely used), as recommended by <https://www.nature.com/articles/d41586-021-02288-x> for example.

We identified 31 FTLD participants from non-European ancestry (based on admixture analysis). These samples were removed because of the small sample size of these participants and to increase genetic homogeneity in our analyses. We have added the following sentence in the manuscript:

“In addition, we identified 31 non-European White participants (Admixture EUR Probability < 0.7) which we removed from analysis because of the small sample size and to increase genetic background homogeneity.”

3. “The first FTLD-TDP genome-wide association study (GWAS) identified the *TMEM106B* locus (rs1990622), supporting lysosomal dysfunction in FTLD-TDP; however, this signal was strongly driven by GRN mutation carriers included in that study.”

This statement needs more explanation for a general reader... which GRN carriers? What condition/subtypes is GRN known to associate with? What is the relationship between this *TMEM106B* locus and GRN?

Thank you to the reviewer for highlighting this point. *TMEM106B* was indeed identified as a risk factor for FTLD-TDP patients in the publication by van Deerlin et al. 2010<sup>34</sup>. Importantly, similarly to our study this GWAS focused on pathologically confirmed FTLD-TDP patients, yet, they also allowed inclusion of patients in which a pathogenic loss of function mutation in *GRN* was already identified. In fact, about a quarter of their cases were carrying a *GRN* mutation, and it was this subgroup of patients which contributed most to the observed association signal on the *TMEM106B* locus.

Moreover, all *GRN* mutation carriers exhibit FTLD-TDP type A at autopsy. We have now made this clarification in the main text.

More information on *TMEM106B* and the relationship to GRN is now provided in the introduction:

“A small number of autosomal dominant genes and risk factors associated with FTLD-TDP have been reported<sup>7,35-40</sup>. A quarter of FTLD-TDP patients are associated with repeat expansions in *C9orf72* which are mainly enriched in patients with FTLD-TDP B, while autosomal dominant loss-of-function mutations in *GRN* explain 10-15% of familial patients and invariably lead to FTLD-TDP A.<sup>41,42</sup> The first FTLD-TDP genome-wide association study (GWAS) identified the *TMEM106B* locus (rs1990622); however, this signal was strongly driven by the FTLD-TDP A patients with *GRN* mutations carriers included in that study<sup>34</sup>.”

Is it fair to call GRN a new locus in Table 2 (is rs5848 distinct from the GRN variants used to remove cases prior to analysis)?

Thank you for this comment, it has made us realize that some extra information is needed in the manuscript. The genetic association of the rs5848 variant in *GRN* is distinct from the causal loss-of-function mutations found in *GRN*. rs5848 is a common variant which we and others have shown can

influence *GRN* levels (decrease by ~10-20% in heterozygote individuals in minor allele carriers), but not to the same extent as loss-of-function mutations (decrease by 50% in heterozygote individuals). No individuals were removed from the GWAS analysis based on the presence of the common variant only based on the presence of loss-of-function mutations. In our previous GWAS, the *GRN* locus did not reach significance in the additive model, only in exploratory subsequent analysis under the recessive model. It is in this light that we considered the identification of *GRN* as bona fide risk locus, a novel finding in this manuscript. We have made changes to the discussion to make this clearer.

“The *GRN* association is driven by rs5848, a variant located in the 3’UTR of *GRN* which was shown to partially reduce plasma and CSF progranulin levels, independent of the presence of *GRN* loss-of-function mutations.<sup>43,44</sup>”

More context for “C9orf72 repeat expansion and LOF mutations in *GRN*” being removed is needed for a general readership.

We added the following sentence to the introduction:

“A quarter of FTLT-DTP patients are associated with repeat expansions in *C9orf72* which are mainly enriched in patients with FTLT-DTP B, while autosomal dominant loss-of-function mutations in *GRN* explain 10-15% of familial patients and invariably lead to FTLT-DTP A.<sup>41,42</sup>”

And the method section:

“Patients carrying repeat expansions in *C9orf72* or loss of function (LOF) mutations in *GRN*, both associated with autosomal dominant forms of FTLT-DTP, were removed prior to WGS.”

4. It should be easier for the reader to look at prioritized genes only at genome-wide significant loci. I’m not sure how I feel about inclusion of subthreshold, nonreplicated loci anyways- at the very least should be clear 1. How subthreshold defined

We thank the reviewer for that comment. We place significant emphasis in our manuscript on the two genome-wide significant loci that replicate in independent datasets: UNC13A and TNIP1.

To facilitate easier understanding of our results, we have updated Figure 2 to display only genome-wide significant signals and added the FTLT-DTP C\* results. Subthreshold signals are now located in **Supplementary Figures 3 to 7**.

Regarding the definition of subthreshold signals: we performed gene prioritization analysis using loci with a P-value of less than  $1 \times 10^{-3}$ . Genes were then classified into tiers based on evidence from publicly available datasets. Tier 1 genes were utilized for further biological insights through gene ontology analysis.

These clarifications are now reflected in the Results section as follows:

“In order to prioritize risk genes and identify possible biological mechanisms, we applied a range of variant annotation and molecular quantitative trait loci (QTL)-GWAS integration analyses as previously described<sup>17</sup> on all variants (**Supplementary Tables 3-14**).”

“Next, we performed gene ontology (GO) analyses on tier 1 prioritized genes, separately for each GWAS.”

Additionally, the Methods section now includes:

“We performed the gene prioritization and functional interpretation analyses for FTLD-TDP All and each FTLD-TDP pathological subtype separately by using the subtype-specific GWAS summary statistics on all variants.”

and 2. What genes are from these subthreshold loci. Terminology such as “locus A4” does not add to clarity unless used consistently across tables (including Table 2).

We kindly refer the reviewer to Supplementary Table 4 to see the gene list used as well as the locus number they are associated with. To increase clarity, we added the locus number in Table 2 and refer to the table in the result section:

“Genes included in the analyses are listed in **Supplementary Table 4.**”

5. In rare variant analysis is this the MAF for “any coding variant that met our filtering criteria” (i.e. most individuals do not have a variant that meets filtering criteria in these genes)? How many variants contribute to the tests in Table 3?

We apologize for the oversight in naming the columns in **table 3**. We actually refer to the cumulative minor allele frequency of the variants that pass our filtering criteria. We have now added the definition in the Table 3 legend and added the number of unique variants included in the test as well.

6. What do you mean by a “GWAS of rare variants” in the introduction? This should be clarified for the reader. It needs to be made clear to the reader how common and rare are defined. (I think you only ran your GWAS on variants with a MAF >1% in either patients or controls (which seems sensible due to power concerns) but this is buried in methods and makes the “GWAS of rare variants” piece extra unclear.

We thank the reviewer for highlighting this. We added the following text in the Introduction:

“Association study analyses of both common (minor allele frequency, MAF, >1%) and rare (MAF<1%) variants”

We also added the following text in the result section for clarity:

“Variants with MAF >1% in patients or controls were considered.”

“To identify genes carrying rare variants contributing to FTLD-TDP, we performed a burden test in genes with rare variants (MAF<1% in patients or controls; **Supplementary Fig.9**) likely to affect protein function.”

7. This paragraph on replication is important but should be supported by a look-up table, perhaps in the supplement, for the previously reported lead variants- what does not replicated mean?

Is there any residual signal? What is the % sample overlap between this effort and prior ones? (quite high, right? So hard to really consider this a true replication effort, which should be mentioned... would recommend including a “new sample only” analysis to see what actually replicates in independent samples).

We thank the reviewer for the interest raised by our findings. We are presenting a phase II association study with overlapping samples with Phase I as pointed out by the reviewer. 512 of the cases (out of 985) are from Phase I, and 838 of the controls (out of 2318) are from Phase II. Overall, Phase I constitutes 32% of the Phase II. We respectfully disagree with the reviewer and think that creating a table with new samples would confuse the reader between “confirmation” (Phase II compared to Phase I) and “replication” (Phase II compared to the independent Dementia-seq data).

We did however add a table (**Supplementary Table 2**) with an overview of the previously reported FTLD loci and their p-values in the current GWAS analyses.

Importantly, we were able to replicate our findings from the FTLD-TDP All vs controls association study using the completely independent DEMENTIA-SEQ cohort. Indeed, there is no overlap with DEMENTIA-SEQ as we removed duplicate samples between our Phase II and DEMENTIA-SEQ based on genomic information. We added that information in the method section:

“Identification of duplicate samples between our discovery dataset and the Dementia-seq data was performed and all duplicates were removed.”

And modified the following sentence in the discussion:

“We replicate in an independent cohort for the first time our previously reported GWAS signal at the *UNC13A* locus in FTLD-TDP patients<sup>39</sup>”

“We further establish and replicate in an independent cohort a novel genetic association between the *TNIP1* locus and FTLD-TDP.”

Could a look-up table for all common variant GWAS hits for FTD (which I assume has mostly FTLD-TDP cases, even if pathology not known??) also be provided (and sample overlap be described)? (<https://pubmed.ncbi.nlm.nih.gov/24943344/> for example). Comparisons with Dementia-seq study (phs001963.v1.p1) findings also seem reasonable to include. Replication of rare variant gene-based signals should also be attempted.

As mentioned above, we now added an additional Supplemental Table (**Supplementary Table 2**) summarizing common variant GWAS hits. The Dementia-seq study is not in the table as there were no genome-wide significant signals in that study.

We also added a paragraph in the results, explicitly summarizing the GWAS results in previously identified FTLD loci.

“We also specifically assessed evidence for association at previously reported FTLD risk loci. Besides the *UNC13A* locus which we previously identified in our Phase I study, none of the other reported risk loci were replicated at the genome-wide significance level (**Supplementary Table 2**).

Notably, *TMEM106B* rs1990622 almost reached the genome-wide significance threshold in the FTLD-TDP A subgroup ( $P=2.7 \times 10^{-7}$ , OR=0.80)."

8. What is meant by "corrected p-values" in methods? FDR? Bonferroni?

We added the following to the text in methods section:

"Corrected Bonferroni *P* are provided."

9. Are the 13 PCs used the top 13 PCs? (I think so, but this should be clarified in paragraph describing generation of PCs)

We clarified now in the text as follow:

"This analysis identified that the 13 top PCs were significantly associated with patient/control status."

And

"Sex and the first 13 PCs were included as covariates in the models."

10. Minor point, but any evidence of inflation in common variant/rare variant analysis? Would be nice to see QQ plots.

QQ plots for common variants have now been added in the supplementary data showing minor inflation (from 1.05-1.07). We also added QQ plots for the rare variants analyses, which as expected from the small sample size showed inflations (1.35-2.18). We added the following sentence to the result section:

"Even though we detected some inflation in our association tests likely due to small cohort size, we did detect nine exome-wide significant signals within FTLD-TDP pathological subtypes (**Table 3, Supplementary Table 19-22**)."

11. specificity analysis (coloc PPH4=20.1%, for p12=10e-06)- do you mean you used different coloc priors in a sensitivity analysis? Would reword.

We apologize for this , and indeed performed sensitivity analysis with different priors. We rephrased as follow:

"which was confirmed in a sensitivity analysis"

12. Could genetic correlation analyses between FTLD and ALS/ADRD/other neurodegenerative diseases be presented? Seems potentially quite informative and pretty easy. Or has this already been done with existing GWAS summary stats?

In the first submission of our manuscript we already included colocalization studies with the top loci, e.g. UNC13A and TNIP1. In response to the request from this reviewer we have now performed genetic correlation studies (using LD score regression) between FTLD-TDP and ALS and, separately, between FTLD-TDP and ADRD. We used the results from our largest FTLD-TDP cohort for this as LD score regression is less optimal in small sample size GWAS such as our FTLD-TDP subtypes. For ALS and ADRD the following datasets were used:

ALS GWAS 2021: [https://ftp.ebi.ac.uk/pub/databases/gwas/summary\\_statistics/GCST90027001-GCST90028000/GCST90027164/](https://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90027001-GCST90028000/GCST90027164/)

AD GWAS 2022: [https://ftp.ebi.ac.uk/pub/databases/gwas/summary\\_statistics/GCST90027001-GCST90028000/GCST90027158/](https://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90027001-GCST90028000/GCST90027158/)

The new analyses show significant genetic correlation between FTLD-TDP and ALS ( $r_g=0.88$ , standard error 0.2322;  $p=0.0002$ ) but not between FTLD-TDP and ADRD ( $r_g=0.22$ , standard error 0.22;  $p=0.31$ ), in line with the current view of ALS and FTLD-TDP as two extremes of a disease continuum.

Would add context to your discussion of UNC13A locus sharing between ALS and FTLD?

We added some more information about the UNCA13 locus in the discussion part as follow:

“This locus was also reported in a recent large GWAS focused on clinical FTLD patients and controls without reaching genome-wide significance ( $P=1.07 \times 10^{-3}$ ), possibly due to the heterogeneity of the underlying pathologies in this cohort with only a subset having TDP-43 pathology<sup>30</sup>. The same *UNC13A* locus was previously linked to ALS<sup>31</sup> and is known to modify the phenotype in ALS patients by increasing the risk of frontotemporal cortical atrophy and diminished cognitive performance, which are reminiscent of an FTLD clinical presentation<sup>32</sup>.”

Could also strengthen your claims about FTLD-TDP C having a distinct genetic architecture- versus just relying on top signals, in this underpowered analysis... statements like this one need more evidence, in my opinion “The recognition that individual FTLD-TDP subtypes could in fact be considered separate diseases with distinct pathomechanisms has significant implications for the design of clinical trials and therapeutic interventions.”

With regard to FTLD-TDP C we have added a lot of additional discussion to make sure the results are interpreted with caution. At the same time, we do clearly identify distinct risk factors for each of the pathological subtypes. We feel that our findings contribute to the growing knowledge base, also supported by recent cryo-EM data, that indeed the different FTLD-TDP pathological subtypes are the result of distinct pathomechanism with a unique pathological TDP-43 filament in each subtype. We have reworded the end of the discussion to put our findings more into this context.

“Importantly, by enriching in neuropathologically confirmed patients and substantially increasing our cohort size, we uncovered distinct genetic aetiologies for each of the three main FTLD-TDP pathological subtypes. Our findings align with recent data obtained from cryo-electron microscopy which identified distinct homomeric TDP-43 filaments in FTLD-TDP A and B and heteromeric amyloid filaments of ANXA11 and TDP-43 in FTLD-TDP C<sup>45-47</sup>. The recognition of individual FTLD-TDP

subtypes as potentially distinct diseases with unique pathomechanism may have important implications for the design of clinical trials and therapeutic interventions.”

13. two CpGs for TMEM106B-Assuming assignment of CpGs to genes is just based on proximity, not eQTM analysis or similar? This might be worth noting as a limitation.

We derived this signal through a methylation QTL signal that is indeed associated with methylation at CpG sites within or near *TMEM106B*. As also explained in the **Supplementary Table 9** legend, we assigned CpGs to genes using BECon<sup>48</sup> platform, which has performed the annotation of CpGs based on an algorithm that considers their positioning with respect to transcription start sites (TSS) of genes. If the BECon annotation was not available, the CpGs were mapped to the nearest gene based on GENCODEv24.

We added the following sentence in the method section related to mQTL-GWAS integrative analyses and also for haQTL-GWAS integrative analyses:

“Of note, for mQTL-GWAS integrative analyses, the CpGs were assigned to genes using BECon annotations<sup>48</sup> whenever available, and if not, the CpGs were mapped to the nearest genes based on GENCODE v24; while for haQTL-GWAS integrative analyses, the histone acetylation peaks were assigned to the nearest genes based on GENCODE v24 as well.”.

14. This statement doesn’t make sense- “we identified exome-wide significant association with the burden of common variants in GRN and TMEM106B”. It is unusual to do an exome-wide analysis for common variants- more methods are needed. How did you ensure statistical independence of included common variants in these aggregate tests? You have TWAS results, is that what is meant? Should be clarified.

We actually used the MAGMA algorithm to perform gene based association study of common variants. MAGMA is a tool designed to analyze associations at the gene level by aggregating SNP p-values (summary statistics) while accounting for linkage disequilibrium between SNPs. We added the following sentence in the method section of the manuscript:

“MAGMA (v1.6) was used to perform gene-based analysis of common variants using summary statistics. MAGMA is a tool designed to analyze associations at the gene level by aggregating SNV p-values while accounting for linkage disequilibrium.”.

## REFERENCES

1. Rascovsky, K. *et al.* Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* **134**, 2456-77 (2011).
2. Gorno-Tempini, M.L. *et al.* Classification of primary progressive aphasia and its variants. *Neurology* **76**, 1006-14 (2011).
3. Brooks, B.R. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors. *J Neurol Sci* **124 Suppl**, 96-107 (1994).
4. Neary, D. *et al.* Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* **51**, 1546-54 (1998).
5. Younes, K. *et al.* Right temporal degeneration and socioemotional semantics: semantic behavioural variant frontotemporal dementia. *Brain* **145**, 4080-4096 (2022).
6. Ulugut, H. *et al.* Clinical recognition of frontotemporal dementia with right anterior temporal predominance: A multicenter retrospective cohort study. *Alzheimers Dement* **20**, 5647-5661 (2024).
7. Ferrari, R. *et al.* Frontotemporal dementia and its subtypes: a genome-wide association study. *Lancet Neurol* **13**, 686-99 (2014).
8. Rohrer, J.D. *et al.* The heritability and genetics of frontotemporal lobar degeneration. *Neurology* **73**, 1451-6 (2009).
9. Snowden, J., Neary, D. & Mann, D. Frontotemporal lobar degeneration: clinical and pathological relationships. *Acta Neuropathol* **114**, 31-8 (2007).
10. Acocella, G. & Billing, B.H. The effect of rifamycin SV on pigment excretion in rats. *Gastroenterology* **49**, 526-30 (1965).
11. Snowden, J.S. *et al.* Semantic dementia and the left and right temporal lobes. *Cortex* **107**, 188-203 (2018).
12. Josephs, K.A. *et al.* Two distinct subtypes of right temporal variant frontotemporal dementia. *Neurology* **73**, 1443-50 (2009).
13. Kumfor, F. *et al.* On the right side? A longitudinal study of left- versus right-lateralized semantic dementia. *Brain* **139**, 986-98 (2016).
14. Chan, D. *et al.* The clinical profile of right temporal lobe atrophy. *Brain* **132**, 1287-98 (2009).
15. Borghesani, V. *et al.* Regional and hemispheric susceptibility of the temporal lobe to FTLDTDP type C pathology. *Neuroimage Clin* **28**, 102369 (2020).
16. Wainberg, M., Andrews, S.J. & Tripathy, S.J. Shared genetic risk loci between Alzheimer's disease and related dementias, Parkinson's disease, and amyotrophic lateral sclerosis. *Alzheimers Res Ther* **15**, 113 (2023).
17. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet* **54**, 412-436 (2022).
18. Kerimov, N. *et al.* A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat Genet* **53**, 1290-1299 (2021).
19. Novikova, G. *et al.* Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. *Nat Commun* **12**, 1610 (2021).
20. Huang, K.L. *et al.* A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. *Nat Neurosci* **20**, 1052-1061 (2017).
21. Young, A.M.H. *et al.* A map of transcriptional heterogeneity and regulatory variation in human microglia. *Nat Genet* **53**, 861-868 (2021).

22. Lopes, K.P. *et al.* Genetic analysis of the human microglial transcriptome across brain regions, aging and disease pathologies. *Nat Genet* **54**, 4-17 (2022).
23. Bryois, J. *et al.* Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat Neurosci* **25**, 1104-1112 (2022).
24. Spiteri, A.G., Wishart, C.L., Pamphlett, R., Locatelli, G. & King, N.J.C. Microglia and monocytes in inflammatory CNS disease: integrating phenotype and function. *Acta Neuropathol* **143**, 179-224 (2022).
25. Hartnell, I.J. *et al.* Glial reactivity and T cell infiltration in frontotemporal lobar degeneration with tau pathology. *Brain* **147**, 590-606 (2024).
26. Sirkis, D.W., Bonham, L.W., Karch, C.M. & Yokoyama, J.S. Immunological signatures in frontotemporal lobar degeneration. *Curr Opin Neurol* **32**, 272-278 (2019).
27. Fairfax, B.P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949 (2014).
28. C, T.V. *et al.* C-terminal TMEM106B fragments in human brain correlate with disease-associated TMEM106B haplotypes. *Brain* **146**, 4055-4064 (2023).
29. Chatterjee, M. *et al.* Plasma extracellular vesicle tau and TDP-43 as diagnostic biomarkers in FTD and ALS. *Nat Med* **30**, 1771-1783 (2024).
30. Manzoni, C. *et al.* Genome-wide analyses reveal a potential role for the MAPT, MOBP, and APOE loci in sporadic frontotemporal dementia. *Am J Hum Genet* **111**, 1316-1329 (2024).
31. van Es, M.A. *et al.* Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat Genet* **41**, 1083-7 (2009).
32. Placek, K. *et al.* UNC13A polymorphism contributes to frontotemporal disease in sporadic amyotrophic lateral sclerosis. *Neurobiol Aging* **73**, 190-199 (2019).
33. Project Min, E.A.L.S.S.C. Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. *Eur J Hum Genet* **26**, 1537-1546 (2018).
34. Van Deerlin, V.M. *et al.* Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. *Nat Genet* **42**, 234-9 (2010).
35. Baker, M. *et al.* Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature* **442**, 916-9 (2006).
36. Cruts, M. *et al.* Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. *Nature* **442**, 920-4 (2006).
37. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245-56 (2011).
38. Renton, A.E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257-68 (2011).
39. Pottier, C. *et al.* Genome-wide analyses as part of the international FTLD-TDP whole-genome sequencing consortium reveals novel disease risk factors and increases support for immune dysfunction in FTLD. *Acta Neuropathol* **137**, 879-899 (2019).
40. Pottier, C. *et al.* Whole-genome sequencing reveals important role for TBK1 and OPTN mutations in frontotemporal lobar degeneration without motor neuron disease. *Acta Neuropathol* **130**, 77-92 (2015).
41. Mackenzie, I.R. *et al.* The neuropathology of frontotemporal lobar degeneration caused by mutations in the progranulin gene. *Brain* **129**, 3081-90 (2006).
42. Mackenzie, I.R. & Neumann, M. Subcortical TDP-43 pathology patterns validate cortical FTLD-TDP subtypes and demonstrate unique aspects of C9orf72 mutation cases. *Acta Neuropathol* **139**, 83-98 (2020).

43. Swift, I.J. *et al.* A systematic review of progranulin concentrations in biofluids in over 7,000 people-assessing the pathogenicity of GRN mutations and other influencing factors. *Alzheimers Res Ther* **16**, 66 (2024).
44. Nicholson, A.M. *et al.* Progranulin protein levels are differently regulated in plasma and CSF. *Neurology* **82**, 1871-8 (2014).
45. Arseni, D. *et al.* Heteromeric amyloid filaments of ANXA11 and TDP-43 in FTLD-TDP type C. *Nature* **634**, 662-668 (2024).
46. Arseni, D. *et al.* Structure of pathological TDP-43 filaments from ALS with FTLD. *Nature* **601**, 139-143 (2022).
47. Arseni, D. *et al.* TDP-43 forms amyloid filaments with a distinct fold in type A FTLD-TDP. *Nature* **620**, 898-903 (2023).
48. Edgar, R.D., Jones, M.J., Meaney, M.J., Turecki, G. & Kobor, M.S. BECon: a tool for interpreting DNA methylation findings from blood in the context of brain. *Transl Psychiatry* **7**, e1187 (2017).

## REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have addressed or answered most of the points raised by this reviewer. They have improved the discussion, including the limitations of the study, and made the appropriate changes to some figures and tables in the manuscript. Overall, this is a highly relevant study in the field of FTD genetics.

However, there is still one concern that requires clarification. In one of my comments, I requested that genomic inflation values should be provided for all analyses. Based on the response to my comment, the authors reported inflation values in the legend of Figure 1, which corresponds to the “Genome-wide association study on common variants”. Could you please provide the inflation values associated with the rare-variant analyses?

----

I have carefully reviewed the responses to Reviewer #4. The authors largely addressed the suggestions and concerns raised by him/her. Therefore, in my opinion, I would accept the study for publication. In fact, in my questions as Reviewer #1, I had requested the genomic inflation values for the rare variant analyses. After carefully reading the response to Reviewer #4, I see that these values are already included in this new version of the manuscript (supplementary figure 9). Although I recommend acceptance of the manuscript, I believe that the high genomic inflation values are a limitation and must be further discussed. In this regard, the reported genomic inflation values are quite high, especially in the case of FTLD-TDP-A ( $\lambda$ : 2.18), which could result in an increase of false positives.

That being said, according to the authors' explanations, they have corrected for the top 13 principal components, included only genes with a minimum of three carriers in the patient group, controlled for the possibility of population stratification or contamination, filtered out low-quality variants, and retained only variants with  $REVEL > 0.75$ . Thus, they have carefully addressed various factors that could increase the genomic inflation value. Additionally, it is well known that the use of tests such as SKAT may contribute to this issue, which the authors, as experts in genomic analyses, have likely already considered. I believe the genomic inflation, as stated by the authors, is related to the small sample size. It is difficult to overcome this limitation given that the sample size is inherent to the unique study groups included. In my opinion, this should be further discussed as a limitation.

In this new version, the authors include the following sentence: “Even though we detected some inflation in our association tests likely due to small cohort size, we did detect nine exome-wide significant signals within FTLD-TDP pathological subtypes (Table 3, Supplementary Table 19-22).” I would also suggest adding a comment on the possibility of an increased rate of false positives in the rare variant analyses while acknowledging that this limitation is difficult to overcome, as the sample size is inherently linked to the unique and well-characterized study groups included in this study, the largest in the field of FTLD.

Response: We thank the reviewer for these comments. We agree that the apparent inflation, despite our effort to reduce potential false positive association, may have led to false positive findings. We appreciate the suggestion to state this more clearly and have now added the following sentence in the discussion:

“Novel genes with a significant increased cumulative frequency of rare variants in specific FTLD-TDP subtypes as compared to controls were also identified, and require confirmation in future studies. We acknowledge that limited sample sizes in these studies may have led to false positive findings; yet, this limitation is inherently linked to the unique and well-characterized study groups included in this study, the largest in the field of FTLD.”

In the results, we also made more clear that we carefully selected the variants that we included in the rare variant analysis (also in response to Reviewer #3)

“Only variants that are frameshift (insertion/deletion/block substitution), stopgain, stoploss and splicing SNVs (jointly defined as LOF variants), and non-synonymous SNVs with a REVEL pathogenicity prediction score above 0.75<sup>1</sup> were included.”

Reviewer #2 (Remarks to the Author):

The authors have adequately addressed the comments. This revision has significantly improved the manuscript. I do not have further comments.

Response: we thank the reviewer for taking the time to re-review our manuscript.

Reviewer #3 (Remarks to the Author):

This work provides an in-depth annotation of gene ontology pathways and identifies tissue- and cell-type enrichments associated with genome-wide significant risk loci. By offering a deeper genetic and mechanistic perspective on FTLD-TDP, this study marks a major advance in unraveling its complex landscape and its connections with related neurodegenerative diseases.

Despite its significant contributions to the field of FTLD and related dementias, several key aspects warrant further clarification:

- Rare Variant Analysis: The functional impact of the identified rare variants remains unclear, particularly given their relatively low occurrence in patients. Figure 6 suggests that some rare variants may even be more prevalent in controls, raising questions about their pathogenicity.

Response: We understand the point made by reviewer 3. Indeed, some of the variants we identified are present in the control population as depicted in figure 6, however, the cumulative minor allele frequency of all the variants included in each of these genes is increased in FTLD-TDP subtypes as compared to controls. We inferred pathogenicity of the variants based on their effect on protein coding sequence and prediction score. We describe it in the result section but have now added details in the result section to ensure there is no confusion in what type of variants we are looking at:

“Only variants that are frameshift (insertion/deletion/block substitution), stopgain, stoploss and splicing SNVs (jointly defined as LOF variants), and non-synonymous SNVs with a REVEL pathogenicity prediction score above 0.75<sup>1</sup> were included.”

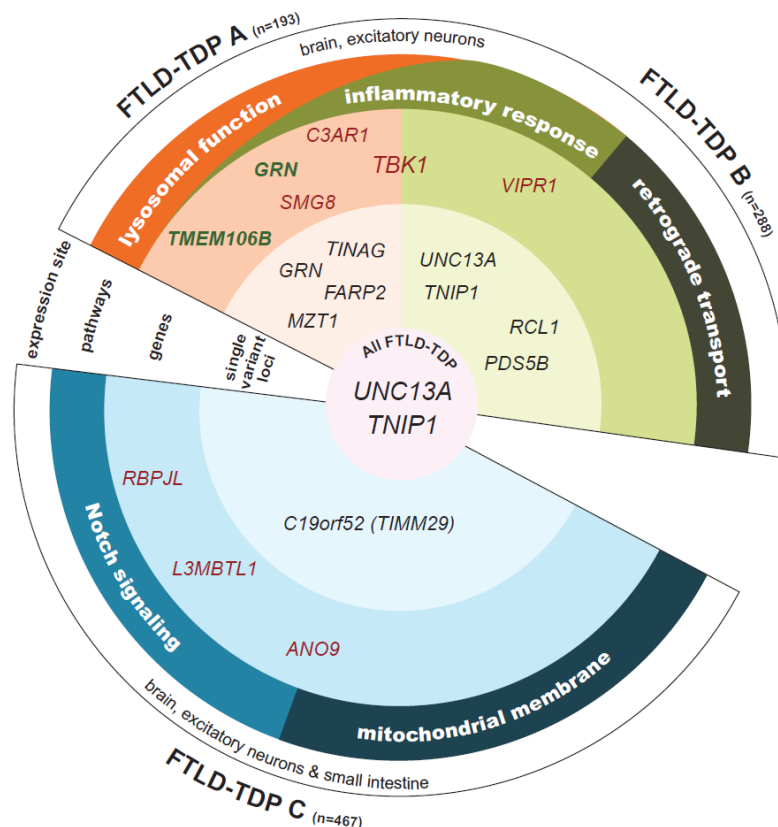
We also adapted the legend of figure 6 to include the total number of FTLD-TDP patients and controls in each analysis to make it clearer that despite the fact that absolute numbers may be higher in controls, the frequency of these variants is significantly lower as compared to patients.

“Total number of subjects included in the analyses were FTLD-TDP A (n=193), FTLD-TDP B (n=288), FTLD-TDP C (n=467) and FTLD-TDP C\* (n=199) and controls (n=3,153).”

• Discussion Length & Clarity: The discussion is extensive but at times dispersive. A new figure or table summarizing key findings could improve clarity and impact.

Response: We understand that the discussion is quite extensive. However, we feel that it puts into context the vast amount of new findings that we are reporting. We already provided a summary figure (see below), which summarized all our findings, including common variant, rare variant as well as gene ontology and enrichment analysis. We had mentioned the figure number (Fig 7) in the discussion but without clear reference to its content. We have now referenced the figure more clearly in the first paragraph of the discussion to guide readers to this figure to follow the discussion.

“A graphical summary of the significant findings across FTLD-TDP subtypes can be found in (Fig. 7).



• Shared Variants in AD and FTD: It would be valuable to understand how the authors interpret

the presence of the same genetic variant in both AD and FTD. Do these variants drive overlapping mechanisms, or are they modulated by disease-specific factors? Further discussion on this aspect, potentially in light of findings such as those in <https://doi.org/10.1038/s41588-022-01024-z>, would strengthen the paper's contributions.

Response : We thank the reviewer for this comment. The paper mentioned by this reviewer was already included in our discussion. It is indeed the ADRD dataset used in this paper that we analyzed for genetic overlap with FTLD-TDP. We have expanded on this finding now in our discussion in response to the reviewer's comment:

"In fact, we observed a shared signal at this locus between our FTLD-TDP GWAS and the recent large ADRD GWAS<sup>2</sup>, whereas independent *TNIP1* association signals were reported for ALS and ADRD<sup>3</sup>. The genetic overlap between FTLD-TDP and AD at this locus is of interest and suggests that *TNIP1*, an important ubiquitin-binding adaptor protein regulating cell death and innate immune responses through NF-kb activation<sup>4-6</sup>, modifies a disease process shared by AD and FTLD-TDP. The idea of a continuum of neurodegenerative disorders in which common pathological mechanisms are involved is further supported by recent GWAS<sup>2,7</sup>."

## References:

1. Ioannidis, N.M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* **99**, 877-885 (2016).
2. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet* **54**, 412-436 (2022).
3. Wainberg, M., Andrews, S.J. & Tripathy, S.J. Shared genetic risk loci between Alzheimer's disease and related dementias, Parkinson's disease, and amyotrophic lateral sclerosis. *Alzheimers Res Ther* **15**, 113 (2023).
4. Oshima, S. *et al.* ABIN-1 is a ubiquitin sensor that restricts cell death and sustains embryonic development. *Nature* **457**, 906-9 (2009).
5. Su, Z. *et al.* ABIN-1 heterozygosity sensitizes to innate immune response in both RIPK1-dependent and RIPK1-independent manner. *Cell Death Differ* **26**, 1077-1088 (2019).
6. Zhou, J. *et al.* A20-binding inhibitor of NF-kappaB (ABIN1) controls Toll-like receptor-mediated CCAAT/enhancer-binding protein beta activation and protects from inflammatory disease. *Proc Natl Acad Sci U S A* **108**, E998-1006 (2011).
7. Shade, L.M.P. *et al.* GWAS of multiple neuropathology endophenotypes identifies new risk loci and provides insights into the genetic risk of dementia. *Nat Genet* **56**, 2407-2421 (2024).