medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Title page

Title

FTLD targets brain regions expressing recently evolved genes

Short title

Transcriptomics of FTLD vulnerability

Authors

Lorenzo Pasquini[1,2*], Felipe L. Pereira[1], Sahba Seddighi[3], Yi Zeng[4], Yongbin Wei[5], Ignacio Illán-Gala[1,6,7], Sarat C. Vatsavayai[1], Adit Friedberg[1,6], Alex J. Lee[1], Jesse A. Brown[1], Salvatore Spina[1], Lea T. Grinberg[1,8], Daniel W. Sirkis[1], Luke W. Bonham[1,9], Jennifer S. Yokoyama[1,9], Adam L. Boxer[1], Joel H. Kramer[1], Howard J. Rosen[1], Jack Humphrey[10], Aaron D. Gitler[4], Bruce L. Miller[1], Katherine S. Pollard[11,12,13,14], Michael E. Ward[3], William W. Seeley[1,8*]

Affiliations

[1]Department of Neurology, Memory and Aging Center, University of California, San Francisco, CA, USA

[2]Department of Neurology, Neuroscape, University of California, San Francisco, CA, USA

[3]National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA

[4]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

[5]School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China

[6]Global Brain Health Institute, University of California, San Francisco, San Francisco, CA, USA and Trinity College Dublin, Dublin, Ireland

[7]Department of Neurology, Hospital de la Santa Creu i Sant Pau, Biomedical Research Institute, Universitat Autònoma de Barcelona, Barcelona, Catalunya, Spain

[8]Department of Pathology, University of California, San Francisco, CA, USA

[9]Department of Radiology, University of California, San Francisco, CA, USA

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

[10]Nash Family Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[11]Gladstone Institute of Data Science and Biotechnology, San Francisco, CA, USA

[12]Institute for Human Genetics, University of California San Francisco, San Francisco, CA, USA

[13]Department of Epidemiology & Biostatistics and Bakar Institute for Computational Health Sciences, University of California San Francisco, San Francisco, CA, USA

[14]Chan Zuckerberg Biohub, San Francisco, CA, USA

Author Contributions

L.P. and W.W.S. conceived the study. L.P., J.A.B., A.J.L., and F.L.P. designed and developed the methodologies. L.P., I.I.-G., A.F., and F.P. analyzed the data. W.W.S., S.C.V., S.Se., L.T.G., and S.Sp. provided patient and genetic data. L.P. and W.W.S. wrote the paper. L.P., W.W.S., Y.W., J.H., J.S.Y., L.B., and D.W.S. interpreted the findings. All authors revised the paper and provided critical feedback. A.L.B, J.H.K, H.J.R., B.L.M., A.D.G., M.E.W., K.S.P. and W.W.S. supervised the study.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

In frontotemporal lobar degeneration (FTLD), pathological protein aggregation is associated with a decline in human-specialized social-emotional and language functions. Most disease protein aggregates contain either TDP-43 (FTLD-TDP) or tau (FTLD-tau). Here, we explored whether FTLD targets brain regions that express genes containing human accelerated regions (HARs), conserved sequences that have undergone positive selection during recent human evolution. To this end, we used structural neuroimaging from patients with FTLD and normative human regional transcriptomic data to identify genes expressed in FTLD-targeted brain regions. We then integrated primate comparative genomic data to test our hypothesis that FTLD targets brain regions expressing recently evolved genes. In addition, we asked whether genes expressed in FTLD-targeted brain regions are enriched for genes that undergo cryptic splicing when TDP-43 function is impaired. We found that FTLD-TDP and FTLD-tau subtypes target brain regions that express overlapping and distinct genes, including many linked to neuromodulatory functions. Genes whose normative brain regional expression pattern correlated with FTLD cortical atrophy were strongly associated with HARs. Atrophy-correlated genes in FTLD-TDP showed greater overlap with TDP-43 cryptic splicing genes compared with atrophy-correlated genes in FTLD-tau. Cryptic splicing genes were enriched for HAR genes, and vice versa, but this effect was due to the confounding influence of gene length. Analyses performed at the individual-patient level revealed that the expression of HAR genes and cryptically spliced genes within putative regions of disease onset differed across FTLD-TDP subtypes. Overall, our findings suggest that FTLD targets brain regions that have undergone recent evolutionary specialization and provide intriguing potential leads regarding the transcriptomic basis for selective vulnerability in distinct FTLD molecular-anatomical subtypes.

Keywords

Frontotemporal lobar degeneration, cryptic exon, human accelerated regions, TDP-43, tau, gene expression

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

Frontotemporal dementia (FTD) refers to a group of clinical syndromes linked to underlying frontotemporal lobar degeneration (FTLD) pathology. These FTD syndromes present with deficits in recently evolved social-emotional and language functions[1,2], resulting from degeneration within networks anchored by frontal, insular, and anterior temporal brain regions[3,4]. The most common FTD syndromes are the behavioral variant (bvFTD) characterized by social-emotional deficits[2]; semantic variant primary progressive aphasia (svPPA), which presents with loss of knowledge related to words, objects, and emotions[1,5]; nonfluent variant PPA (nfvPPA), characterized by speech production difficulties[1,5]; corticobasal syndrome, an asymmetric akinetic-rigid syndrome associated with a progressive loss of limb controls[4]; and progressive supranuclear palsy-Richardson syndrome (PSP-RS), associated with oculomotor deficits and gait instability[4]. FTLD pathology, in turn, can be divided into three major molecular classes based whether the neuronal and glial inclusions are composed of tau (FTLD-tau), TDP-43 (FTLD-TDP) or FUS/EWS/TAF15 (FET) family proteins[6]. Each major molecular class can be further divided into distinct subtypes based on the pathomorphology and distribution of the inclusions across regions, layers, and cell types. Factors driving the selective predilection of each FTLD subtype for distinct, yet overlapping, anatomical structures remain unknown.

Comparative genomic studies have the potential to shed light on the relationship between brain evolution and the erosion of human-specialized functions in FTLD. Novel genome-wide studies have identified a unique set of loci that are conserved across mammalian evolution, indicating important biological roles, but significantly diverged in the human lineage, suggesting changes to those roles in humans compared to chimpanzees and other primates[7,8]. These short sequences, referred to as human accelerated regions (HARs), are commonly located in non-coding DNA, often near genes associated with transcription and DNA binding[9]. Several studies have linked HARs to the emergence of distinctive human traits such as the opposable thumb[8], language, and social behavior[10], and mutations in HAR genes are observed in neuropsychiatric[11,12] and neurodevelopmental disorders[7].

Because FTLD is pathologically heterogeneous, it is possible that FTLD-TDP and FTLD-tau intersect with brain evolution through distinct yet overlapping mechanisms. The role of protein aggregation has been a major focus of research, but in FTLD-TDP, TDP-43 aggregation

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

is almost always accompanied, and may be preceded, by loss of nuclear TDP-43 expression[13–16]. TDP-43, encoded by *TARDBP*, is an RNA-binding protein predominantly expressed in the nucleus of healthy neurons[17–19], where it plays diverse roles in transcription regulation. One role is to suppress incorporation of cryptic exons, short stretches of non-conserved intronic RNA that are normally removed during pre-mRNA splicing[20–23]. Upon TDP-43 loss-of-function, cryptic exons are spliced into messenger RNAs, often introducing premature stop codons or frame shifts that result in nonsense mediated decay and loss of normal protein expression[24–26]. To date, more than 1,000 potential genes subject to cryptic splicing (CS) have been identified, yet, despite some compelling candidates[17,24], it remains unknown which CS events, if any, are most relevant to FTLD-TDP pathogenesis. Crucially, although TDP-43 is a highly conserved protein, CS genes show almost no overlap when comparing mice to humans[20], suggesting that TDP-43 may have adopted a unique regulatory role across the human evolutionary lineage.

Here, we combined neuroimaging, neuropathological, normative regional brain transcriptomic, comparative genomic, and cell biological data to identify genes whose expression pattern, in the healthy brain, correlates with cortical atrophy patterns in FTLD and to test the hypothesis that HAR genes are expressed in patterns resembling FTLD atrophy. We further explored associations between expression patterns for TDP-43 CS genes and FTLD-associated atrophy and the overlap between HAR and CS genes.

Methods

Participants

We searched the University of California, San Francisco (UCSF) Neurodegenerative Disease Brain Bank database for participants autopsied between 2008 and 2020 who had at least one high quality structural MRI scan prior to death; this search yielded 322 participants. From this pool, we included patients with a primary neuropathological diagnosis of FTLD-TDP (Types A-C) or FTLD-tau (corticobasal degeneration [CBD] or Pick's disease). Although PSP is a common subtype of FTLD-tau, we omitted it here because it is characterized by a predominantly subcortical atrophy pattern not well suited to our transcriptomic data (see below). Patients were excluded if they had: (*i*) Braak neurofibrillary tangle stage[27] >3, (*ii*) Alzheimer's disease neuropathological change[28] >intermediate, (*iii*) Lewy body disease[29] >brainstem

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

predominant, or (*iv*) major territorial ischemic infarcts or intracranial hemorrhages[30]. Based on these stringent criteria, 28 patients with TDP-A, 35 with TDP-B, 29 with TDP-C, 45 with CBD, and 27 with Pick's disease were included in the study. In accordance with the declaration of Helsinki, patients or their surrogates provided written informed consent prior to participation, including consent for brain donation. The UCSF Committee on Human Research approved the study.

All patients received clinical diagnoses by a multidisciplinary team following thorough neurological, neuroimaging, and neuropsychological assessments. Clinical severity was assessed using the Mini Mental State Examination[31] and the Clinical Dementia Rating scale total and sum of boxes scores, using a version of the Clinical Dementia Rating scale adapted for FTD[32]. Patient demographics, as well as clinical and neuropathological data, are provided in **Table 1**. Race and ethnicity, handedness, and years of education were self-reported by patients or surrogates.

Neuropathology

Postmortem research-oriented autopsies were performed at the UCSF Neurodegenerative Disease Brain Bank. Neuropathological diagnoses were made following consensus diagnostic criteria[6,29,33] based on standard histological and immunohistochemical methods[34,35].

Structural Neuroimaging

Acquisition and preprocessing. Because patients were evaluated over an 18-year interval, MRI scans were obtained on different scanners over time. For 74 patients, T1-weighted magnetization prepared rapid gradient echo (MPRAGE) MRI sequences were acquired at the UCSF Neuroscience Imaging Center, either on a 3T Siemens Tim Trio (n=67) or a 3T Siemens Prisma Fit scanner (n=7). Both scanners had similar acquisition parameters (sagittal slice orientation; slice thickness = 1.0 mm; slices per slab = 160; in-plane resolution = 1.0x1.0 mm; matrix = 240x256; repetition time = 2,300 ms; inversion time = 900 ms; flip angle = 9°), although echo time slightly differed (Trio: 2.98 ms; Prisma: 2.9 ms). For the remaining 87 patients, MRI scans were obtained at the San Francisco Veterans Affairs Medical Center using MPRAGE sequences acquired either on a Siemens 1.5 Tesla Magnetom scanner (n=73; voxel resolution 1.0 × 1.0 × 1.5 mm; TR = 10 ms; TE = 4 ms; TI = 300 ms; flip angle = 15) or on a Siemens 4 Tesla Magnetom scanner (n=14; voxel resolution 1.0 × 1.0 × 1.0 mm;

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

TR = 2.3 ms; TE = 3 ms; TI = 950 ms; flip angle = 7). The average interval between scanning date and autopsy was 4.8 +/- 2.6 years.

All structural MR images underwent a voxel-based morphometry analysis[36] after being visually inspected for motion and scanning artifacts. Structural images were segmented in gray matter, white matter, and cerebrospinal fluid and normalized to MNI space using SPM12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/). Gray matter images were modulated by dividing the tissue probability values by the Jacobian of the warp field. The resulting images of voxel resolution 2.0 x 2.0 x 2.0 mm were smoothed with an isotropic Gaussian kernel with a full width at half maximum of 8 mm.

W-score maps. To generate participant-specific atrophy maps, the smoothed gray matter images were transformed to W-score maps. W-score maps are voxel-wise statistical maps that reflect levels of atrophy for each individual after adjustment for relevant covariates[37,38]. This approach is particularly useful when performing analyses with heterogenous samples since gray matter atrophy estimates are adjusted for demographical and methodological nuisance covariates, including age, gender, scanner type, and total intracranial volume[38]. The W-score model used for this study was published[39]; it involved 397 healthy older controls and included age at MRI, sex, years of education, handedness, scanner type, and total intracranial volume as covariates. Briefly, multiple regression analyses at the voxel-wise level were performed on the normative dataset to estimate gray matter intensity as adjusted for age, sex, years of education, handedness, MRI scanner, and total intracranial volume. This model was then used to estimate the expected gray matter intensity for a patient by using their demographical data. The discrepancy between the raw and the expected maps is used to derive the W-score map:

$$W = \frac{(Raw - Expected)}{SD\varepsilon}$$

where *Raw* is the raw value of a voxel from the smoothed image of a patient; *Expected* is the expected value for the voxel of that specific patient based on the healthy control model; and $SD\varepsilon$ is the standard deviation of the residuals from the healthy control model.

Gene expression

Allen Human Brain Atlas (AHBA). Normative human brain gene expression data were derived from microarray data available through the AHBA (http://human.brain-

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

map.org/static/download). As described in detail elsewhere[40], tissue samples were extracted across both hemispheres from two human brain donors, as well as the left hemisphere of four additional donors, totaling 3,702 tissue samples. Microarray analysis quantified gene expression across 58,692 probes, providing an estimate of the relative expression of 20,734 genes within the tissue samples[41]. The publicly available toolbox abagen (https://github.com/rmarkello/abagen)[42] was then used for: *(i)* updating probe-to-gene annotations using the latest available data; *(ii)* data filtering, where expression values that do not exceed background are removed; *(iii)* probe selection, which, for genes indexed by multiple probes, involves selecting a single representative measure to represent the expression of that gene across all donor brains; *(iv)* sample assignment, where tissue samples from the AHBA were mapped to 273 parcels of the Brainnetome brain regional parcellation atlas (https://atlas.brainnetome.org/)[43]. This atlas was selected because it was derived using functional and structural anatomy and connectivity, and it includes paired homologous bilateral regions. Steps *i-iv* were followed by *(v)* normalization of expression measures to account for inter-individual differences and outlying values; *(vi)* gene-set filtering, to remove genes that are inconsistently expressed across the six brains; and *(vii)* averaging gene expression across donors. This procedure resulted in 15,655 genes (***Supplementary Data***) and their normative expression in 273 brain parcels. Finally, we removed subcortical regions (cerebellum, thalamus, basal ganglia, and hippocampus) due to substantial differences in their gene expression values when compared to cortical parcels[44], yielding a 202 regions x 15,655 genes matrix (***Supplementary Figure S1***, ***Supplementary Data***).

HAR genes. HARs were taken from a comparative genomic analysis that identified, from a list of several recent publications, loci with accelerated divergence in humans when compared to chimpanzees[7]. A total of 2,737 HARs were identified, representing 2,164 unique HAR-associated genes[7]. Of these 2,164 HAR genes, 1,373 were identified as sufficiently expressed in the brain based on the AHBA brain-expressed gene dataset (probes used if they exceeded background signal for more than 50% of all samples) and used in our analyses, referred to simply as HAR genes[45] (***Supplementary Data***).

CS genes. Three prior studies contributed to the list of TDP-43-dependent CS genes[24–26]. These studies were chosen because they identified CS targets either based on FTLD-TDP human *postmortem* tissue data[24] or iPSC-derived neuronal cell lines[25,26], omitting studies based

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

on non-neuronal cells. For the tissue-based study, splicing analyses were performed on RNA-sequencing data from TDP-43-positive and TDP-43-negative neuronal nuclei isolated from frontal cortices of seven patients with FTLD-TDP. This approach identified 66 CS genes[24], of which 63 genes exceeded the brain expression threshold imposed on the AHBA data. In a separate study, RNA sequencing was performed on human iPSC-derived cortical-like neurons, in which TDP-43 expression was reduced using Clustered Regularly Interspaced Short Palindromic Repeats interference (CRISPRi)[25]. Differential splicing and expression analyses based on this dataset followed by splice junction evaluation via visual inspection of sashimi plots revealed 107 genes with putative CS events, of which 97 also exceeded the brain expression threshold in the AHBA dataset. Finally, a third study performed a comprehensive analysis of differential splicing events in TDP-43 depleted versus control iPSC-derived neurons to develop a high-quality neuronal cryptic exon atlas[26]. This analysis revealed 233 CS genes, of which 201 exceeded the brain expression threshold in the AHBA brain-expressed gene dataset, as previously defined. By combining the two iPSC-derived neuronal CS gene sets, 216 unique CS genes were identified and combined with the *postmortem* tissue-derived CS genes to yield a final brain-expressed CS gene set of 257 unique genes (***Supplementary Data***).

Overlap of gene lists. We used Venn diagrams to visualize the overlap between distinct gene lists. To test whether gene lists overlapped to a degree higher than expected by chance, we used bootstrap-based tests. In this procedure, for example, a gene list of the same number of genes as the list of HAR genes is randomly drawn 5,000 times from the AHBA dataset of brain-expressed genes. For each of these randomly selected bootstrap gene lists, we then quantify the overlap with the CS gene list, resulting in a distribution of overlaps between CS genes and randomly selected gene lists. We next count how often the random overlap was equal to or higher than the observed overlap between the HAR and CS gene lists and derive a *p* value by calculating the proportion of all bootstrap gene lists with an overlap exceeding the observed overlap.

Spatial correlation analyses. We derived group-averaged, voxel-wise W-score maps for each FTLD pathological subtype (TDP-A, TDP-B, TDP-C, CBD, and Pick's disease). For each subtype-averaged map, we then averaged the W-score values of voxels contained in each cortical Brainnetome parcel to derive a vector reflecting the level of regional cortical atrophy in

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

each FTLD subtype. These FTLD subtype-specific atrophy vectors were then separately correlated with the cortical gene expression levels derived from the AHBA brain-expressed gene dataset using Pearson's correlation coefficients. This procedure enabled us to assess the spatial similarity between each FTLD subtype atrophy pattern and the cortical expression levels of 15,655 brain-expressed genes from the AHBA[44,46,47].

Because spatial autocorrelation inherent to neuroimaging data can inflate *p*-values in brain map analyses[48], we corrected for this autocorrelation by applying a bootstrap-based approach (***Supplementary Figure S2***), in which 5,000 surrogate maps that preserve the autocorrelation properties of the FTLD atrophy maps were generated using the toolbox BrainSMASH (https://brainsmash.readthedocs.io/en/latest/approach.html). We then derived autocorrelation-corrected *p*-values by counting how often the correlation between surrogate atrophy maps and each gene map was equal to or higher than the true correlation and divided this number by the total number of bootstraps. Only False Discovery Rate (FDR) adjusted *p*-values < 0.05 were considered significant. This procedure enabled us to identify genes with spatial expression patterns resembling the FTLD atrophy maps.

To remove genes displaying spurious correlations to FTLD-TDP atrophy maps from subsequent analyses, we assessed how applying a threshold based on the absolute correlation value would affect the list of selected genes significantly correlating with FTLD-TDP atrophy maps. We applied absolute correlation thresholds between 0-0.5 in steps of 0.05 and derived an index assessing how each threshold impacted the ratio between the number of genes uniquely associated with each FTLD-TDP atrophy map and the number of genes that significantly correlated with at least two FTLD-TDP subtype atrophy maps:

$$Uniqueness_{index} = \frac{(Unique_{TDP-A} + Unique_{TDP-B} + Unique_{TDP-C})}{(Shared_{TDP-AB} + Shared_{TDP-BC} + Shared_{TDP-AC} + Shared_{TDP-ABC})}$$

This uniqueness index was used to identify 0.2 as the optimal absolute correlation threshold to better identify subtype-specific correlating genes while at the same time preserving longer gene lists. Genes displaying correlation values between 0.2 and -0.2 were hence considered as not being sufficiently correlated with FTLD-TDP atrophy maps and discarded

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

from further analyses regardless of the associated *p*-value. We then assessed the overlap between gene lists representing FTLD atrophy-correlated, HAR, and CS genes using the previously described bootstrap approach.

Gene set enrichment analysis. Gene set enrichment analyses are computational methods that query large genomic databases to determine whether an *a priori* defined set of genes shows statistically significant associations with known biological pathways. We ran five separate gene set enrichment analyses[49] leveraging the correlation values (FDR adjusted $p < 0.05$) between brain genes and atrophy maps of FTLD-TDP and FTLD-tau subtypes (one analysis for each subtype). We performed gene set enrichment analyses using the *clusterProfiler* package in R software v4.2.1 for Gene Ontology[50] and for Kyoto Encyclopedia of Genes and Genomes (KEGG)[51]. We then focused on terms and pathways with a Holm adjusted *p*-value of < 0.05 shared by a majority of FTLD subtypes. Correlation matrices were used to reflect the strength of correlation between genes associated with various KEGG terms and FTLD-TDP and FTLD-tau subtype atrophy patterns (FDR adjusted $p < 0.05$).

Graph network. Over the course of the study, we identified a set of genes that correlated with FTLD-TDP atrophy and overlapped with the previously introduced CS and HAR gene lists. Graph theoretical approaches were applied to this list of 23 CS-HAR genes correlating with FTLD-TDP atrophy to generate topographical representations of gene expression networks[52]. The regional expression values of selected genes were correlated with each other to generate a gene-to-gene regional co-expression matrix. This matrix was binarized at a Pearson's correlation coefficient threshold of 0.3 and the number of surviving edges between genes was counted for each gene of interest to derive a measure of nodal degree, reflecting the level of connectedness between a specific gene and each other gene selected for the analysis. Resulting data were rendered as a graph.

Gene length. Standard *BiomatRt* code in R (https://useast.ensembl.org/info/data/biomart/index.html) leveraging the widely used Ensembl genome browser (https://ensemblgenomes.org/) was used to derive the full coordinates of exonic and intronic sequences of brain-expressed genes used in our analyzes (BSgenome.Hsapiens.UCSC.hg38). This information was used to estimate the length of HAR

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

genes used in our analyses and to identify a set of 1,353 length-matched non-HAR genes through nearest neighbor matching using the *matchit* library in R.

Disease epicenters. We used an approach previously applied by our group and others[53–56] to identify patient-tailored epicenters. This method defines an epicenter as the brain region whose normative functional (task-free) connectivity map most closely resembles the patient's atrophy W-map. Normative connectivity maps were derived using task-free functional MRI data from a cohort of 75 healthy older subjects to generate a library of 194 intrinsic functional connectivity maps from seed regions of the Brainnetome atlas spanning the entire cerebral cortex. Characterization of the normative sample and methodological details are presented elsewhere[53]. We then compared each patient's atrophy W-map with this library to select the seed that generated a functional connectivity map showing the highest spatial (Pearson) correlation to the subject-specific atrophy W-map. Subtype-specific frequency maps were used to quantify the spatial distribution of epicenters in FTLD-TDP, since group-level atrophy in FTLD-TDP was associated with both CS and HAR genes. Finally, we used Wilcoxon signed-rank tests to assess whether the normative expression of CS and HAR genes within disease epicenters differed across FTLD-TDP subtypes. For each gene, epicenters' gene expression levels were then averaged for each subtype to derive a cross-fold change score reflecting higher or lower gene expression in one group versus the other.

Results

FTLD subtypes target brain regions expressing shared and distinct genes

We first sought to explore the relationship between FTLD atrophy topography and normative brain regional gene expression. We included the three major FTLD-TDP subtypes (A-C) and the two major sporadic FTLD-tau subtypes, CBD and Pick's disease, with predominant cortical involvement. First, we sought to replicate the known atrophy patterns associated with these disorders using *in vivo* brain MRI data from patients with an autopsy-based diagnosis[57,58]. Structural MRI volumes were segmented into gray matter tissue probability maps and further transformed to W-score maps[37,38], where higher W-scores reflect more severe voxel-wise atrophy for each patient, adjusted for demographic variables. FTLD subtype specific cortical atrophy

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

maps were derived by averaging W-score maps within each group (**Figure 1A-E**, left panels), revealing a more dorsal frontal and insular pattern in TDP-A; a more ventral pattern in TDP-B, with less severe atrophy overall; an anterior temporal-predominant pattern in TDP-C; a milder, predominantly frontal atrophy pattern in CBD; and a severe frontal and insular pattern of atrophy in Picks' disease.

We next leveraged normative gene expression microarray data from the AHBA to identify genes whose cortical expression pattern resembles cortical gray matter atrophy characteristic of each FTLD subtype. Starting with a matrix of 15,655 genes and their regional expression in 202 cortical brain parcels (***Supplementary Figure S1***), we performed spatial correlation analyses to associate group-level gray matter atrophy patterns with normative regional expression levels of brain-expressed genes. We used a bootstrap-based test, correcting for the spatial autocorrelations inherent to neuroimaging data[48], to delineate genes significantly correlated with atrophy in each FTLD subtype (FDR adjusted $p < 0.05$, **Figure 1A-E**, red bars in the right panels). To remove spurious correlations from further analyses, we only considered absolute correlation values ≥ 0.2, based on a uniqueness index maximizing the number of genes uniquely associated with each FTLD subtype (***Supplementary Figure S2***). This procedure identified unique and shared genes correlating with atrophy across FTLD subtypes. TDP-C showed the most uniquely correlated genes (3,715), followed by TDP-A (661 genes) and TDP-B (228 genes). A total of 992 genes were shared between all FTLD-TDP subtypes (**Figure 1F**). Among FTLD-tau, CBD showed only 27 uniquely correlated genes, due to the mild frontal atrophy found in this sample highly resembling the inherent frontal-posterior spatial autocorrelation gradient inherent to neuroimaging data[48]. CBD shared 119 genes with Pick's disease, which yielded a total of 5,434 uniquely correlated genes for FTLD-tau (**Figure 1G**). When pooling both FTLD-TDP and FTLD-tau together, 3,137 genes were uniquely correlated with at least one TDP subtype, 441 with at least one tau subtype, and 5,139 were shared among TDP-43 and tau FTLD subtypes (**Figure 1H**).

To better understand the biological function of FTLD atrophy-correlated genes, we performed gene set enrichment analyses[49] against gene ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG)[50]. While no gene ontology was considered enriched, the gene set enrichment analysis revealed only one KEGG pathway, associated with neuroactive ligand-receptor interactions (has04080)[51], shared across FTLD-TDP and FTLD-Pick subtypes. This

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

KEGG pathway consisted of common and unique associations between FTLD atrophy-correlated genes and major classes of neuromodulatory and homeostatic neurotransmitter systems (**Figure 2**). Genes associated with acetylcholine receptors did not, for the most part, correlate with atrophy in any FTLD subtype (**Figure 2A**), as expected based on the lack of a cholinergic deficit in FTLD[59]. Conversely, genes associated with receptors within the epinephrine/norepinephrine, opioid, somatostatin, and neuropeptide Y systems correlated across most FTLD subtypes (**Figure 2B, F-H**). Some receptor systems showed more specific associations with FTLD subtypes, as in the case of serotonin and dopamine to TDP-C (**Figure 2C, D**), or oxytocin to TDP-B, TDP-C, and Pick's disease (**Figure 2E)**. The lack of associations for CBD is likely attributable to a low number of genes associated with atrophy in this subtype and the high spatial autocorrelation of its atrophy map.

Genes correlated with FTLD atrophy are enriched for HAR genes

We identified 1,373 brain-expressed HAR genes derived from a comparative genomic study of conserved loci with elevated divergence in humans versus chimpanzees and other mammals (**Figure 3A, *Supplementary Materials***)[7]. We next assessed the overlap between brain-expressed HAR genes and genes whose expression patterns correlated with atrophy in FTLD. To determine whether these overlaps occurred at a rate greater than chance, we used a bootstrap-based test (***Supplementary Figure S3***), selecting random gene sets from the pool of brain-expressed genes. Atrophy-correlated genes were enriched for HAR genes in both FTLD-TDP (808 genes, $p < 0.0005$, **Figure 3B**) and FTLD-tau (560 genes, $p < 0.0005$, **Figure 3C**), with a significantly greater overlap observed for FTLD-TDP atrophy-correlated genes ($\chi^2 = 89.59$, $p < 0.0001$). HAR genes correlating with at least one FTLD subtype tended to be either unique to FTLD-TDP (283 genes) or shared between FTLD-TDP and FTLD-tau (525 genes) (***Supplementary Figure S4A***), with only 35 HAR genes correlating uniquely with at least one FTLD-tau subtype.

FTLD-TDP atrophy-correlated genes are linked to TDP-43 CS genes

Recent studies have identified a corpus of genes that undergo CS upon TDP-43 loss-of-function[24–26] (**Figure 4A**). Given the strong enrichment of FTLD-TDP atrophy-correlated genes for HAR genes, and that humans show non-conserved patterns of TDP-43 related CS, we

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

hypothesized that TDP-43 CS genes might be enriched among genes correlated with FTLD-TDP (but not FTLD-tau) atrophy. To explore this possibility, we consolidated published and unpublished TDP-43 CS gene lists, filtering for human brain-expressed CS genes identified using human cell models (216 genes)[25,26] or patient brain tissues (63 genes[24]; ***Supplementary Table S1***), resulting in 257 unique CS genes (***Supplementary Materials***). 146 genes were shared between FTLD-TDP atrophy-correlated and CS genes, with this overlap trending towards significance ($p = 0.10$, **Figure 4B**). To evaluate the biological relevance of the observed trend, we compared the overlap between CS genes and FTLD-TDP atrophy-correlated genes to the overlap between CS genes and FTLD-tau atrophy-correlated genes. We identified 88 FTLD-tau correlated CS genes; in contrast to FTLD-TDP, however, this overlap did not approach significance ($p = 0.71$, **Figure 4C**). The proportion of CS genes overlapping with FTLD-TDP atrophy-correlated genes was significantly higher than the proportion of CS genes overlapping with FTLD-tau atrophy-correlated genes ($\chi^2 = 26.39$, $p < 0.0001$). To control these analyses for the smaller number of FTLD-tau atrophy-correlated genes, we randomly re-sampled a single list of 5,580 FTLD-TDP atrophy-correlated genes (matching the number of FTLD-tau atrophy-correlated genes), which revealed a significant overlap of 106 genes shared with CS genes (***Supplementary Figure S5***, $p < 0.05$), suggesting that the differential overlap between FTLD-TDP and FTLD-tau atrophy-correlated genes with CS genes is not merely a reflection of gene list length. CS genes correlated with at least one FTLD subtype were either unique to FTLD-TDP (63 genes) or shared between FTLD-TDP and FTLD-tau (83 genes) (***Supplementary Figure S4B***). These finding suggest that although both FTLD-TDP and FTLD-tau target brain regions expressing recently evolved genes, FTLD-TDP exhibits a closer relationship to regions that express TDP-43 CS genes.

HAR genes overlap with TDP-43 CS genes

Among the TDP-43 CS genes, 37 were also HAR genes (**Figure 5A, *Supplementary Materials***), an overlap that was higher than expected by chance, as demonstrated by distributions derived from bootstrap-based tests, where the overlap between randomly selected brain-expressed gene lists and HAR genes ($p < 0.005$) or CS genes ($p < 0.0005$) was repeatedly assessed and compared to the observed overlap between HAR genes and CS genes (**Figure 5B-C**). Identified CS-HAR genes included genes linked to social behavior and language (e.g., *ERC2*,

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

KIF26B, *CBLN2*, *SEMA6D*)[60–64] , neurodevelopment, and neuropsychiatric diseases (e.g., *CBLN2SPRY3*, *CAMTA1*, *LRP8*)[65–69]. Some CS-HAR genes identified (e.g., *CBLN2*, *ICA1*, *PTPRT, and PTPRD*) have been shown to accurately distinguish FTLD-TDP *postmortem* cortex from non-neurological disease control tissue based on the frontal/temporal cortex expression of cryptic exons[26]. CS-HAR genes tended to be longer than other brain-expressed genes ($t(39) = 4.74$; $p < 0.0001$), but their length fell within the range found for 85% of brain expressed genes (***Supplementary Figure S6A***). CS-HAR genes contained mostly intronic HAR sequences (76%, ***Supplementary Figure S6B***), followed by intergenic (21%) and exonic (3%) sequences, and cryptic splicing events affecting CS-HAR genes mainly included cryptic exon incorporation[20] (71%, ***Supplementary Figure S6C***), consistent with other (non-HAR) CS genes. We performed control analyses assessing the overlap between CS genes and a set of non-HAR genes matched for length to HAR genes (***Supplementary Figure S7A***). This approach yielded an overlap of 42 genes (***Supplementary Figure S7B***, $\chi^2 = 0.36$, $p = 0.55$), suggesting that gene size is a confounding factor driving the overlap between CS and HAR genes. Therefore, although HAR and TDP-43 CS genes overlap, our analyses do not provide evidence for a special, gene-length independent relationship between HAR and TDP-43 CS genes.

We next sought to better understand relationships between FTLD atrophy-correlated genes and the 37 CS-HAR genes (**Figure 5A**). Of these 37 genes, 23 overlapped with FTLD-TDP atrophy-correlated genes ($p = 0.17$, **Figure 5D**) and 14 with FTLD-tau atrophy-correlated genes ($p = 0.46$, **Figure 5E**), but these overlaps were not greater than expected by chance. Nonetheless, the proportion of CS-HAR genes overlapping with FTLD-TDP atrophy-correlated genes was significantly higher than the proportion of CS-HAR genes overlapping with FTLD-tau atrophy-correlated genes ($\chi^2 = 4.38$, $p < 0.05$). CS-HAR genes correlating with at least one FTLD subtype tended to be either unique to FTLD-TDP (10 genes) or shared between FTLD-TDP and FTLD-tau (13 genes) (***Supplementary Figure S4C***). Accordingly, to generate hypotheses, we examined the relationships among the 23 FTLD-TDP atrophy-associated CS-HAR genes in terms of their co-expression relationships in the healthy brain and their links to specific FTLD-TDP subtypes (**Figure 6**). While some genes correlated negatively with atrophy (e.g., *PTPRD*), suggesting lower normative expression among FTLD-TDP atrophied regions, other genes correlated positively (e.g., *ERC2*), reflecting higher expression among targeted regions. Some FTLD-TDP subtype-specific associations emerged, such as *PTPRD* with TDP-A,

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

RHOBTB3 with TDP-B, or *ERC2* with TDP-C, while other genes, such as *RALGAPA2*, *ICA1*, and *SHISA9*, were shared among all FTLD-TDP subtypes.

Normative expression of CS and HAR genes varies across FTLD-TDP subtypes and patient epicenters

The preceding analyses used group-level methods to identify brain-expressed HAR and CS genes associated with cortical gray matter atrophy in FTLD subtypes, but there is also prominent anatomical heterogeneity within each subtype[70], leading us to explore a more individualized, patient-tailored approach. Neurodegenerative diseases begin within specific vulnerable and network-anchoring regions, which we have termed "epicenters" [53,55], before spreading to other areas along large-scale brain network connections[4,55,71]. In previous work, we introduced a method for identifying patient-specific epicenters, which enhanced network-based predictions of longitudinal atrophy progression in FTLD[53,55]. This approach uses task-free fMRI acquired in a normative sample to estimate each brain region's connectivity to all other brain regions (***Supplementary Figure S8A***) and correlates these connectivity maps with an individual patient's atrophy pattern to identify the region (i.e., epicenter) whose normative connectivity most resembles the atrophy pattern[53,55,56]. Based on this approach, we identified a single best-fit epicenter for each patient, focusing on FTLD-TDP given its specific relationship to cryptic splicing (**Figure 7A**). Frequency maps revealed common epicenters in TDP-A, including the pregenual and subgenual anterior cingulate cortices and the anterior insula, with the anterior cingulate being the most common epicenter. The left anterior insula was the most common epicenter in TDP-B, which showed additional epicenters in the anterior cingulate and temporal pole. In TDP-C, the left entorhinal cortex was the most common epicenter, with additional epicenters in adjacent temporal polar regions (***Supplementary Figure S8B-D***). While many epicenters were shared across subtypes (e.g., the left anterior cingulate cortex), several were associated with just one FTLD-TDP subtype, such as the left inferior frontal gyrus in TDP-A, the left angular gyrus in TDP-B, or the left fusiform gyrus with TDP-C (**Figure 7A**). Overall, despite each FTLD-TDP subtype providing 7-10 uniquely associated epicenters (**Figure 7B**), a substantial proportion of patients shared epicenters across subtypes (**Figure 7C**), suggesting that any influence of gene expression on subtype-associated regional vulnerability is not likely to be deterministic.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

To formulate hypotheses regarding potential HAR- and CS-related genes and pathways whose expression might influence the onset of FTLD-TDP in different regions across different FTLD-TDP subtypes, we then compared the normative expression of HAR genes or TDP-43 CS genes across FTLD-TDP subtype epicenters by using Wilcoxon signed-rank tests (FDR adjusted $p < 0.05$). Resulting p values were plotted against the cross-fold change, reflecting higher or lower normative gene expression in epicenters of one FTLD-TDP subtype versus another (**Figure 7D-I**). These analyses revealed genes more highly expressed in regional epicenters found in TDP-B compared to TDP-A, such as CS gene *ERC2* and HAR gene *TFEC* (**Figure 7D-E**), and genes more highly expressed in epicenters for TDP-A versus TDP-B, such as HAR genes *LRRC4C* and *SDK1* (**Figure 7E**). When comparing TDP-A to TDP-C, the CS gene *STMN2* was significantly more expressed in regional epicenters found in TDP-A, while the CS gene *SEMA3D* was more expressed in regional epicenters found in TDP-C (**Figure 7F**). The HAR gene *LMO1* was significantly more expressed in epicenters of TDP-C, while the HAR gene *TCF7L2* was more expressed in epicenters of TDP-A (**Figure 7G**). When comparing TDP-B to TDP-C, the CS gene *CBLN2* and the HAR gene *METTL4* showed significantly higher normative expressions among disease epicenters derived from TDP-B, while the CS gene *ONECUT1* and the HAR gene *PAX8* were more highly expressed in epicenters of TDP-C (**Figure 7H-I**).

Discussion

The integrative anatomical biology data presented here identified overlapping and distinct genes associated with cortical atrophy in FTLD-TDP and FTLD-tau. Both gene sets were enriched for HAR genes, providing novel comparative transcriptomic evidence that FTLD, regardless of the underlying molecular pathology, targets brain regions expressing genes that have undergone recent evolution as humans diverged from chimpanzees. Human studies leveraging chromatin interaction profiles in the fetal and adult cortex have revealed that HAR elements converge on specific cell types and laminae involved in cerebral cortical expansion[72]. These findings echo neuroimaging studies showing that the normative expression of HAR genes in the cortex is highest among brain networks that have undergone rapid evolutionary expansion in humans[45]. These recently expanded higher-order networks overlap with systems supporting language and social-emotional functions, the same systems that undergo early and progressive neurodegeneration in FTLD[70]. The normative expression of atrophy-associated HAR and non-

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

HAR genes was either positively or negatively associated with FTLD atrophy. Genes showing a positively associated gradient are likely to be highly expressed, in the healthy brain, in regions undergoing FTLD atrophy. Dependence on these genes could contribute to FTLD regional vulnerability, especially if protein expression is diminished, for example by TDP-43 loss-of-function and associated mis-splicing. Conversely, genes displaying a negative gradient are probably expressed at lower levels, in the healthy brain, in FTLD-atrophied regions, with neurodegeneration further compromising already low levels of expression. FTLD atrophy-associated genes were characterized by common and unique associations with neuromodulatory and homeostatic transmitter systems affected in FTLD[73]. Dopaminergic, norepinephrinergic, and serotonergic imbalances are common in FTD, while the acetylcholine system remains relatively intact[59]. Common antidepressants, such as selective serotonin reuptake inhibitors, have been shown to improve the behavioral symptoms of FTD [59], reinforcing the notion of neuromodulatory imbalances in FTLD. Our analyses further revealed that FTLD-TDP atrophy-correlated genes are slightly enriched for genes that are targets of TDP-43 cryptic splicing repression, significantly more so than FTLD-tau atrophy-correlated genes. This observation, though somewhat tentative, suggests a potential FTLD-TDP-specific mechanism whereby incipient TDP-43 loss-of-function focuses vulnerability on specific brain networks that express high levels of TDP-43 CS genes.

Several lines of evidence derived from animal[74,75] and cellular models[76], as well as human neuroimaging[53,55] and tissue studies[30], support the notion that neuropathological changes start within vulnerable brain regions, often referred to as epicenters, before spreading via network pathways[77,78]. In this context, our study not only shows that the normative expression of HAR and CS genes resembles atrophy patterns in FTLD-TDP but also that the normative expression of these genes within putative epicenters differs among FTLD-TDP subtypes. Our findings raise the possibility that the regional expression of specific HAR and TDP-43 CS genes may predispose local circuits to cultivate specific FTLD-TDP subtypes, each characterized by a distinct TDP-43 aggregation pattern[79]. After accumulating within vulnerable epicenters, TDP-43 aggregates could then spread to other brain regions via functional and structural connectivity pathways. Histopathological studies suggest that neurodegenerative diseases may start within susceptible cell-types harbored within vulnerable brain regions[4,14,80,81]. For example, von Economo neurons and fork cells have been shown to be selectively targeted by TDP-43 depletion

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

and inclusions in bvFTD due to FTLD-TDP[82]. These large, elongated neuron-types are almost exclusively found within the anterior cingulate and frontoinsula of large-brained, highly social mammals[83] – including humans, greater apes, elephants, and cetaceans – raising the possibility that these neurons contribute to circuits linked to the emergence of complex social behaviors[83]. Further, a recent neuroimaging genomics study revealed that atrophied regions in FTD genetic mutation carriers express astrocyte and endothelial cell-related genes [47]. Future histopathological studies could shed light on how expression levels of HAR and CS genes within vulnerable neuronal, such as von Economo neurons, interact with glial cells and contribute to the emergence of distinct FTLD-TDP subtypes.

Our findings should be interpreted in light of several methodological considerations. First, our study utilized normative brain transcriptomic data derived from the AHBA, which is based on bulk tissue microarray data from the brains of six donors, and both hemispheres were studied in only two of six. Future studies leveraging transcriptomic data derived from larger normative samples are needed to confirm the genetic correlates of regional gray matter atrophy in FTLD. Further, our study focused on cortical gene expression, omitting subcortical areas despite the relevance of the striatum, thalamus, and other subcortical areas in FTLD[84]. Gene expression levels from bulk tissue microarray dramatically differ depending on whether these are extracted from cortical or subcortical areas, with most studies, including ours, analyzing either one set of structures or the other[42,45]. Second, the overlap we observed between HAR and CS genes, though interesting, appears to be driven primarily by gene length. This makes sense given that longer genes may be implicated in multiple processes including brain development and neurodegenerative diseases[85,86]. Further studies are needed to assess whether altered sequences within HAR genes mechanistically interact with regions regulated or bound by TDP-43. Finally, our study focused on CS genes identified in FTLD-TDP human *postmortem* tissue data[24] or iPSC-derived neuronal cell lines[25,26], omitting those identified in HeLa cell lines[87]. Although this choice was intended to enhance biological confidence regarding the neural relevance of these genes, future studies may expand the TDP-43 “cryptic spliceosome” to explore a more exhaustive gene set.

Data availability

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

In house Python 2.7.16 (https://www.python.org/), R 4.1.2 (https://www.r-project.org/), and MATLAB R2021a (https://www.mathworks.com/products/matlab.html) scripts were used for the analyses. W-score maps averaged for each FTLD subtype are publicly available on NeuroVault[88]. The gene lists used to generate the findings are available as ***Supplementary Data***. Further data that support the findings of this study are available on request from the corresponding author. Correspondence should be directed to Lorenzo Pasquini or William W. Seeley.

Acknowledgements

We thank the participants and their families for their invaluable contributions to neurodegeneration research. We thank Sean Whalen for useful feedback on data analysis.

Funding

This work was supported by NIH grants K99-AG065457 and R00-AG065457 to LP; P30AG062422, P01AG019724, U01AG057195, and U19AG063911 (WWS), as well as the Rainwater Charitable Foundation and the Bluefield Project to Cure FTD (WWS); GBHI ALZ UK-21-720973 and AACSF-21-850193 as well as JR20/0018 and PI21/00791 to IIG; 2021-A-023-FEL to AF; Gladstone Institutes and R01MH123178 to KSP.

Competing interests

The authors declare no competing interests.

References

1. Gorno-Tempini ML, Hillis AE, Weintraub S, et al. Classification of primary progressive aphasia and its variants. 2011;02:1-10. doi: 10.1212/WNL.0b013e31821103e6
2. Rascovsky K. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain*. 2011;6(4):234-237. doi:10.17925/ENR.2011.06.04.234
3. Schroeter ML, Raczka K, Neumann J, von Cramon DY. Neural networks in frontotemporal dementia—A meta-analysis. *Neurobiol Aging*. 2008;29(3):418-426. doi:https://doi.org/10.1016/j.neurobiolaging.2006.10.023
4. Seeley WilliamW. Mapping Neurodegenerative Disease Onset and Progression. *Cold Spring Harbor Perspective in Biology*. 2017;1389(2017). doi:10.1126/science.1123511

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

5. Gorno-Tempini ML. Cognition and Anatomy in Three Variants of Primary Progressive Aphasia. *Annals Neurology*. 2004;119(11):2658-2666. doi:10.1016/j.clinph.2008.06.019.Brain-Computer

6. Mackenzie IRA, Neumann M, Baborie A, et al. A harmonized classification system for FTLD-TDP pathology. *Acta Neuropathol*. 2011;122(1):111-113. doi:10.1007/s00401-011-0845-8

7. Doan RN, Bae B Il, Cubelos B, et al. Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior. *Cell*. 2016;167(2):341-354.e12. doi:10.1016/j.cell.2016.08.071

8. Prabhakar S, Visel A, Akiyama JA, et al. Human-Specific Gain of Function in a Developmental Enhancer. *Science (1979)*. 2008;321(5894):1346-1350. doi:10.1126/science.1159974

9. Pollard KS, Salama SR, King B, et al. Forces Shaping the Fastest Evolving Regions in the Human Genome. *PLoS Genet*. 2006;2(10):e168. doi:10.1371/JOURNAL.PGEN.0020168

10. Levchenko A, Kanapin A, Samsonova A, Gainetdinov RR. Human accelerated regions and other human-specific sequence variations in the context of evolution and their relevance for brain development. *Genome Biol Evol*. 2018;10(1):166-188. doi:10.1093/gbe/evx240

11. Cheung JP, Tubbs JD, Sham PC. Extended gene set analysis of human neuro-psychiatric traits shows enrichment in brain-expressed human accelerated regions across development. *Schizophr Res*. 2022;246:148-155. doi:https://doi.org/10.1016/j.schres.2022.06.023

12. Guardiola-Ripoll M, Fatjó-Vilas M. A Systematic Review of the Human Accelerated Regions in Schizophrenia and Related Disorders: Where the Evolutionary and Neurodevelopmental Hypotheses Converge. *Int J Mol Sci*. 2023;24(4). doi:10.3390/ijms24043597

13. Vatsavayai SC, Yoon SJ, Gardner RC, et al. Timing and significance of pathological features in *C9orf72* expansion-associated frontotemporal dementia. *Brain*. Published online 2016:aww250. doi:10.1093/brain/aww250

14. Nana AL, Sidhu M, Gaus SE, et al. Neurons selectively targeted in frontotemporal dementia reveal early stage TDP-43 pathobiology. *Acta Neuropathol*. 2019;137(1). doi:10.1007/s00401-018-1942-8

15. Iguchi Y, Katsuno M, Niwa J ichi, et al. Loss of TDP-43 causes age-dependent progressive motor neuron degeneration. *Brain*. 2013;136(5):1371-1382. doi:10.1093/brain/awt029

16. Yang C, Wang H, Qiao T, et al. Partial loss of TDP-43 function causes phenotypes of amyotrophic lateral sclerosis. *Proceedings of the National Academy of Sciences*. 2014;111(12):E1121-E1129. doi:10.1073/pnas.1322641111

17. Melamed Z, López-Erauskin J, Baughn MW, et al. Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. *Nat Neurosci*. 2019;22(2):180-190. doi:10.1038/s41593-018-0293-z

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

18. Klim JR, Williams LA, Limone F, et al. ALS-implicated protein TDP-43 sustains levels of STMN2, a mediator of motor neuron growth and repair. *Nat Neurosci*. 2019;22(2):167-179. doi:10.1038/s41593-018-0300-4

19. Polymenidou M, Lagier-Tourenne C, Hutt KR, et al. Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat Neurosci*. 2011;14(4):459-468. doi:10.1038/nn.2779

20. Ling JP, Pletnikova O, Troncoso JC, Wong PC. TDP-43 repression ofnonconserved cryptic exons is compromised in ALS-FTD. *Science*. 2015; 349(6248):650-655. doi: 10.1126/science.aab0983

21. Donde A, Sun M, Ling JP, et al. Splicing repression is a major function of TDP-43 in motor neurons. *Acta Neuropathol*. 2019;138(5):813-826. doi:10.1007/s00401-019-02042-8

22. Sun M, Bell W, LaClair KD, et al. Cryptic exon incorporation occurs in Alzheimer's brain lacking TDP-43 inclusion but exhibiting nuclear clearance of TDP-43. *Acta Neuropathol*. 2017;133(6):923-931. doi:10.1007/s00401-017-1701-2

23. Jeong YH, Ling JP, Lin SZ, et al. Tdp-43 cryptic exons are highly variable between cell types. *Mol Neurodegener*. 2017;12(1):13. doi:10.1186/s13024-016-0144-x

24. Ma XR, Prudencio M, Koike Y, et al. TDP-43 represses cryptic exon inclusion in the FTD–ALS gene UNC13A. *Nature*. 2022;603(7899):124-130. doi:10.1038/s41586-022-04424-7

25. Brown AL, Wilkins OG, Keuss MJ, et al. TDP-43 loss and ALS-risk SNPs drive mis-splicing and depletion of UNC13A. *Nature*. 2022;603(7899):131-137. doi:10.1038/s41586-022-04436-3

26. Seddighi S, Brown AL, Wilkins OG, et al. Mis-spliced transcripts generate de novo proteins in TDP-43-related ALS/FTD. *Bioarxiv*. 2023. doi: https://doi.org/10.1101/2023.01.23.525149

27. Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol*. 1991;82(4):239-259. Accessed April 9, 2019. http://www.ncbi.nlm.nih.gov/pubmed/1759558

28. Montine TJ, Phelps CH, Beach TG, et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta Neuropathol*. 2012;123(1):1-11. doi:10.1007/s00401-011-0910-3

29. McKeith IG, Dickson DW, Lowe J, et al. Diagnosis and management of dementia with Lewy bodies: Third report of the DLB consortium. *Neurology*. 2005;65(12):1863-1872. doi:10.1212/01.wnl.0000187889.17253.b1

30. Kim EJ, Hwang JHL, Gaus SE, et al. Evidence of corticofugal tau spreading in patients with frontotemporal dementia. *Acta Neuropathol*. 2019;(0123456789). doi:10.1007/s00401-019-02075-z

31. Folstein MF, Folstein SE MP. Mini-mental state. A grading the cognitive state of patiens for the clinician. *J Psychiatr Res*. 1975;12:189-198. doi:10.1016/0022-3956(75)90026-6

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

32. Knopman DS, Kramer JH, Boeve BF, et al. Development of methodology for conducting clinical trials in frontotemporal lobar degeneration. *Brain*. 2008;131(Pt 11):2957-2968. doi:10.1093/brain/awn234

33. MacKenzie IRA, Neumann M, Bigio EH, et al. Nomenclature and nosology for neuropathologic subtypes of frontotemporal lobar degeneration: An update. *Acta Neuropathol*. 2010;119(1):1-4. doi:10.1007/s00401-009-0612-2

34. Kim EJ, Sidhu M, Gaus SE, et al. Selective frontoinsular von economo neuron and fork cell loss in early behavioral variant frontotemporal dementia. *Cerebral Cortex*. 2012;22(2):251-259. doi:10.1093/cercor/bhr004

35. Tartaglia MC, Sidhu M, Laluz V, et al. Sporadic corticobasal syndrome due to FTLD-TDP. *Acta Neuropathol*. 2010;119(3):365-374. doi:10.1007/s00401-009-0605-1

36. Ashburner J, Friston KJ. Voxel-based morphometry - The methods. *Neuroimage*. 2000;11(6 I):805-821. doi:10.1006/nimg.2000.0582

37. La Joie R, Visani A V., Baker SL, et al. Prospective longitudinal atrophy in Alzheimer's disease correlates with the intensity and topography of baseline tau-PET. *Sci Transl Med*. 2020;12(524):1-13. doi:10.1126/scitranslmed.aau5732

38. Pasquini L, Nana AL, Toller G, et al. Salience Network Atrophy Links Neuron Type-Specific Pathobiology to Loss of Empathy in Frontotemporal Dementia. *Cerebral Cortex*. Published online 2020:1-13. doi:10.1093/cercor/bhaa119

39. Friedberg A, Pasquini L, Diggs R, et al. Prevalence, Timing, and Network Localization of Emergent Visual Creativity in Frontotemporal Dementia. *JAMA Neurol*. 2023;80(4):377-387. doi:10.1001/jamaneurol.2023.0001

40. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. 2012;489(7416):391-399. doi:10.1038/nature11405

41. Arnatkevic□iūtė A, Fulcher BD, Fornito A. A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage*. 2019;189:353-367. doi:10.1016/j.neuroimage.2019.01.011

42. Markello RD, Arnatkeviciute A, Poline JB, Fulcher BD, Fornito A, Misic B. Standardizing workflows in imaging transcriptomics with the abagen toolbox. Jbabdi S, Makin TR, Jbabdi S, Burt J, Hawrylycz MJ, eds. *Elife*. 2021;10:e72129. doi:10.7554/eLife.72129

43. Fan L, Li H, Zhuo J, et al. The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cerebral Cortex*. 2016;26(8):3508-3526. doi:10.1093/cercor/bhw157

44. Brown JA, Lee AJ, Pasquini L, Seeley WW. A dynamic gradient architecture generates brain activity states. *Neuroimage*. 2022;261:119526. doi:https://doi.org/10.1016/j.neuroimage.2022.119526

45. Wei Y, de Lange SC, Scholtens LH, et al. Genetic mapping and evolutionary analysis of human-expanded cognitive networks. *Nat Commun*. 2019;10(1):1-11. doi:10.1038/s41467-019-12764-8

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

46. Vogel JW, Joie R La, Grothe MJ, et al. A molecular gradient along the longitudinal axis of the human hippocampus informs large-scale behavioral systems. *Nat Commun*. Published online 2020:1-17. doi:10.1038/s41467-020-14518-3

47. Altmann A, Cash DM, Bocchetta M, et al. Analysis of brain atrophy and local gene expression in genetic frontotemporal dementia. *Brain Commun*. 2020;2(2):1-13. doi:10.1093/braincomms/fcaa122

48. Burt JB, Helmer M, Shinn M, Anticevic A, Murray JD. Generative modeling of brain maps with spatial autocorrelation. *Neuroimage*. 2020;220:117038. doi:https://doi.org/10.1016/j.neuroimage.2020.117038

49. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. 2005;102(43):15545-15550. doi:10.1073/pnas.0506580102

50. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25-29. doi:10.1038/75556

51. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28(1):27-30. doi:10.1093/nar/28.1.27

52. Bassett DS, Sporns O. Network neuroscience. *Nat Neurosci*. 2017;20(3). doi:10.1038/nn.4502

53. Brown JA, Neuhaus J, Miller BL, et al. Patient-Tailored, Connectivity-Based Forecasts of Spreading Brain Atrophy. *Neuron*. Published online 2019:1-13. doi:10.1016/j.neuron.2019.08.037

54. Brown JA, Hua AY, Trujllo A, et al. Advancing functional dysconnectivity and atrophy in progressive supranuclear palsy. *Neuroimage Clin*. 2017;16(August):564-574. doi:10.1016/j.nicl.2017.09.008

55. Zhou J, Gennatas ED, Kramer JH, Miller BL, Seeley WW. Predicting Regional Neurodegeneration from the Healthy Brain Functional Connectome. *Neuron*. 2012;73(6):1216-1227. doi:10.1016/j.neuron.2012.03.004

56. Franzmeier N, Dewenter A, Frontzkowski L, et al. Patient-centered connectivity-based prediction of tau pathology spread in Alzheimer's disease. *Sci Adv*. 2023;6(48):eabd1327. doi:10.1126/sciadv.abd1327

57. Perry DC, Brown JA, Possin KL, et al. Clinicopathological correlations in behavioural variant frontotemporal dementia. *Brain*. 2017;140(12):3329-3345. doi:10.1093/brain/awx254

58. Rohrer JD, Geser F, Zhou J, et al. TDP-43 subtypes are associated with distinct atrophy patterns in frontotemporal dementia. *Neurology*. 2010;75(24):2204. doi:10.1212/WNL.0b013e318202038c

59. Huey ED, Putnam KT, Grafman J. A systematic review of neurotransmitter deficits and treatments in frontotemporal dementia. *Neurology*. 2006;66(1):17 LP - 22. doi:10.1212/01.wnl.0000191304.55196.4d

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

60. Satoh Y, Endo S, Nakata T, et al. ERK2 Contributes to the Control of Social Behaviors in Mice. *The Journal of Neuroscience*. 2011;31(33):11953 LP - 11967. doi:10.1523/JNEUROSCI.2349-11.2011

61. Scekic-Zahirovic J, Sanjuan-Ruiz I, Kan V, et al. Cytoplasmic FUS triggers early behavioral alterations linked to cortical neuronal hyperactivity and inhibitory synaptic defects. *Nat Commun*. 2021;12(1):3028. doi:10.1038/s41467-021-23187-9

62. Seigneur E, Wang J, Dai J, Polepalli J, Südhof TC. Cerebellin-2 regulates a serotonergic dorsal raphe circuit that controls compulsive behaviors. *Mol Psychiatry*. 2021;26(12):7509-7521. doi:10.1038/s41380-021-01187-x

63. Ercan-Sencicek AG, Davis Wright NR, Sanders SJ, et al. A balanced t(10;15) translocation in a male patient with developmental language disorder. *Eur J Med Genet*. 2012;55(2):128-131. doi:https://doi.org/10.1016/j.ejmg.2011.12.005

64. Thomas T, Perdue M V, Khalaf S, et al. Neuroimaging genetic associations between SEMA6D, brain structure, and reading skills. *J Clin Exp Neuropsychol*. 2021;43(3):276-289. doi:10.1080/13803395.2021.1912300

65. Quesnel-Vallières M, Dargaei Z, Irimia M, et al. Misregulation of an Activity-Dependent Splicing Network as a Common Mechanism Underlying Autism Spectrum Disorders. *Mol Cell*. 2016;64(6):1023-1034. doi:https://doi.org/10.1016/j.molcel.2016.11.033

66. Ning Z, McLellan AS, Ball M, et al. Regulation of SPRY3 by X chromosome and PAR2-linked promoters in an autism susceptibility region. *Hum Mol Genet*. 2015;24(25):7450. doi:10.1093/hmg/ddv448

67. Shinawi M, Coorg R, Shimony JS, Grange DK, Al-Kateb H. Intragenic CAMTA1 deletions are associated with a spectrum of neurobehavioral phenotypes. *Clin Genet*. 2015;87(5):478-482. doi:https://doi.org/10.1111/cge.12407

68. Xiao X, Yu H, Li J, et al. Further evidence for the association between LRP8 and schizophrenia. *Schizophr Res*. 2020;215:499-505. doi:https://doi.org/10.1016/j.schres.2017.05.002

69. Shibata M, Pattabiraman K, Lorente-Galdos B, et al. Regulation of prefrontal patterning and connectivity by retinoic acid. *Nature*. 2021;598(7881):483-488. doi:10.1038/s41586-021-03953-x

70. Kim EJ, Vatsavayai S, Seeley WW. Neuropathology of dementia. In: *The Behavioral Neurology of Dementia*. Cambridge University Press; 2016:94-122. doi:10.1017/9781139924771.008

71. Brettschneider J, Del Tredici K, Lee VMY, Trojanowski JQ. Spreading of pathology in neurodegenerative diseases: a focus on human studies. *Nat Rev Neurosci*. 2015;16(2):109-120. doi:10.1038/nrn3887

72. Won H, Huang J, Opland CK, Hartl CL, Geschwind DH. Human evolved regulatory elements modulate genes involved in cortical expansion and neurodevelopmental disease susceptibility. *Nat Commun*. 2019;10(1):2396. doi:10.1038/s41467-019-10248-3

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

73. Premi E, Dukart J, Mattioli I, et al. Unravelling neurotransmitters impairment in primary progressive aphasias. *Hum Brain Mapp*. 2023;44(6):2245-2253. doi:https://doi.org/10.1002/hbm.26206

74. Porta S, Xu Y, Restrepo CR, et al. Patient-derived frontotemporal lobar degeneration brain extracts induce formation and spreading of TDP-43 pathology in vivo. *Nat Commun*. 2018;9(1):4220. doi:10.1038/s41467-018-06548-9

75. de Calignon A, Polydoro M, Suárez-Calvet M, et al. Propagation of Tau Pathology in a Model of Early Alzheimer's Disease. *Neuron*. 2012;73(4):685-697. doi:https://doi.org/10.1016/j.neuron.2011.11.033

76. Westergard T, Jensen BK, Wen X, et al. Cell-to-Cell Transmission of Dipeptide Repeat Proteins Linked to C9orf72-ALS/FTD. *Cell Rep*. 2016;17(3):645-652. doi:https://doi.org/10.1016/j.celrep.2016.09.032

77. Raj A, Kuceyeski A, Weiner M. Article A Network Diffusion Model of Disease Progression in Dementia. *Neuron*. 2011;73(6):1204-1215. doi:10.1016/j.neuron.2011.12.040

78. Vogel JW, Iturria-Medina Y, Strandberg OT, et al. Spread of pathological tau proteins through communicating neurons in human Alzheimer's disease. *Nat Commun*. 2020;11(1):2612. doi:10.1038/s41467-020-15701-2

79. Arseni D, Chen R, Murzin AG, et al. TDP-43 forms amyloid filaments with a distinct fold in type A FTLD-TDP. *Nature*. 2023;620(7975):898-903. doi:10.1038/s41586-023-06405-w

80. Fu H, Hardy J, Duff KE. Selective vulnerability in neurodegenerative diseases. *Nat Neurosci*. 2018;21(10):1350-1358. doi:10.1038/s41593-018-0221-2

81. Bonham LW, Steele NZR, Karch CM, et al. Protein network analysis reveals selectively vulnerable regions and biological processes in FTD. *Neurol Genet*. 2018;4(5):e266. doi:10.1212/nxg.0000000000000266

82. Gami-Patel P, van Dijken I, van Swieten JC, et al. Von Economo neurons are part of a larger neuronal population that are selectively vulnerable in C9orf72 frontotemporal dementia. *Neuropathol Appl Neurobiol*. 2019;45(7):671-680. doi:https://doi.org/10.1111/nan.12558

83. Allman JM, Tetreault NA, Hakeem AY, et al. The von Economo neurons in the frontoinsular and anterior cingulate cortex. *Ann N Y Acad Sci*. 2011;1225(1):59-71. doi:10.1111/j.1749-6632.2011.06011.x

84. Lee SE, Sias AC, Kosik EL, et al. Thalamo-cortical network hyperconnectivity in preclinical progranulin mutation carriers. *Neuroimage Clin*. 2019;22(March):101751. doi:10.1016/j.nicl.2019.101751

85. Lopes I, Altab G, Raina P, de Magalhães JP. Gene Size Matters: An Analysis of Gene Length in the Human Genome . *Frontiers in Genetics*. 2021;12. https://www.frontiersin.org/articles/10.3389/fgene.2021.559998

86. Soheili-Nezhad S, van der Linden RJ, Olde Rikkert M, Sprooten E, Poelmans G. Long genes are more frequently affected by somatic mutations and show reduced expression in

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Alzheimer's disease: Implications for disease etiology. *Alzheimer's & Dementia*. 2021;17(3):489-499. doi:https://doi.org/10.1002/alz.12211

87. Humphrey J, Emmett W, Fratta P, Isaacs AM, Plagnol V. Quantitative analysis of cryptic splicing associated with TDP-43 depletion. *BMC Med Genomics*. 2017;10(1):1-17. doi:10.1186/s12920-017-0274-1

88. Gorgolewski KJ, Varoquaux G, Rivera G, et al. NeuroVault.Org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Front Neuroinform*. 2015;9(APR):1-9. doi:10.3389/fninf.2015.00008

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Figures, Legends, and Tables

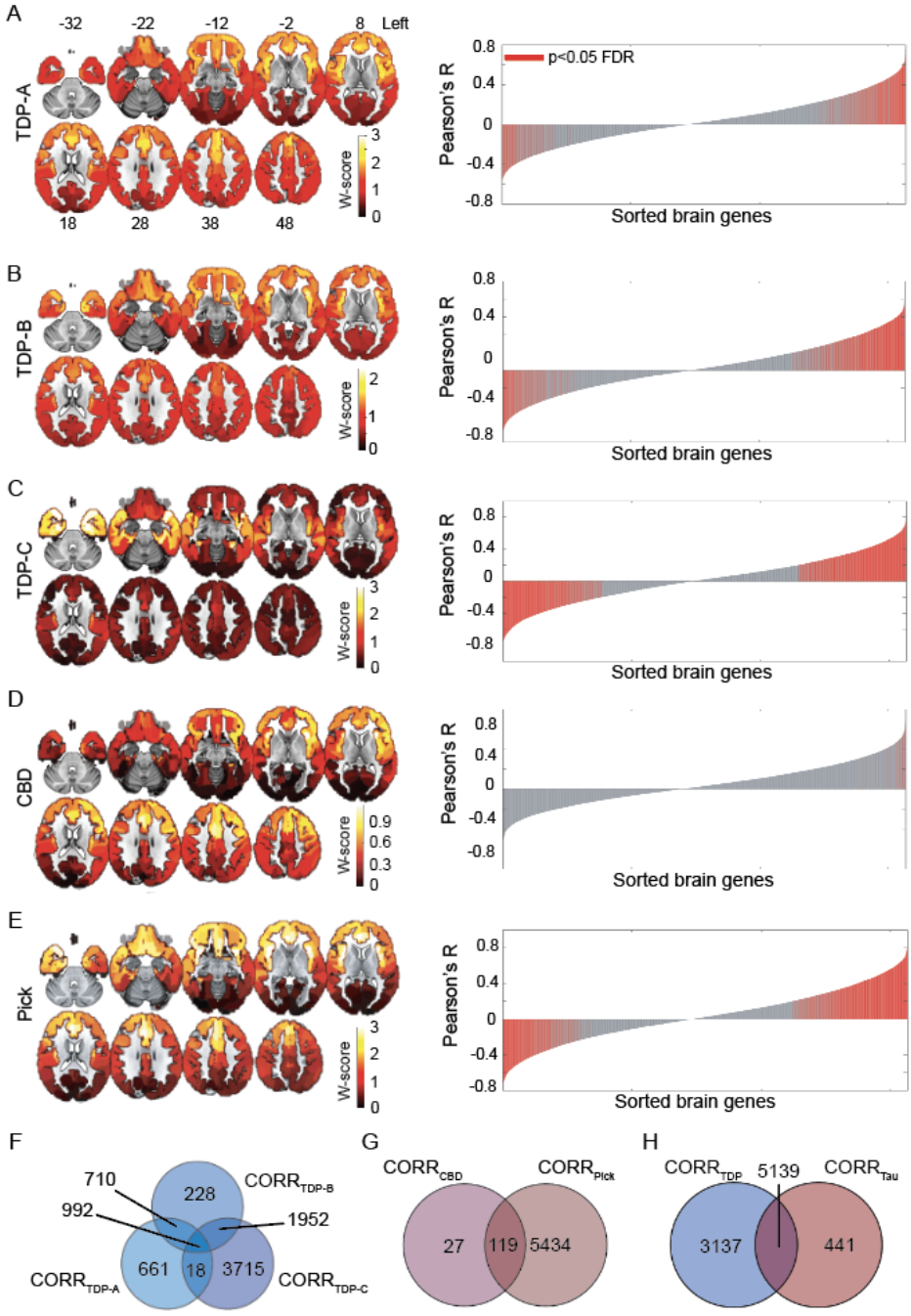
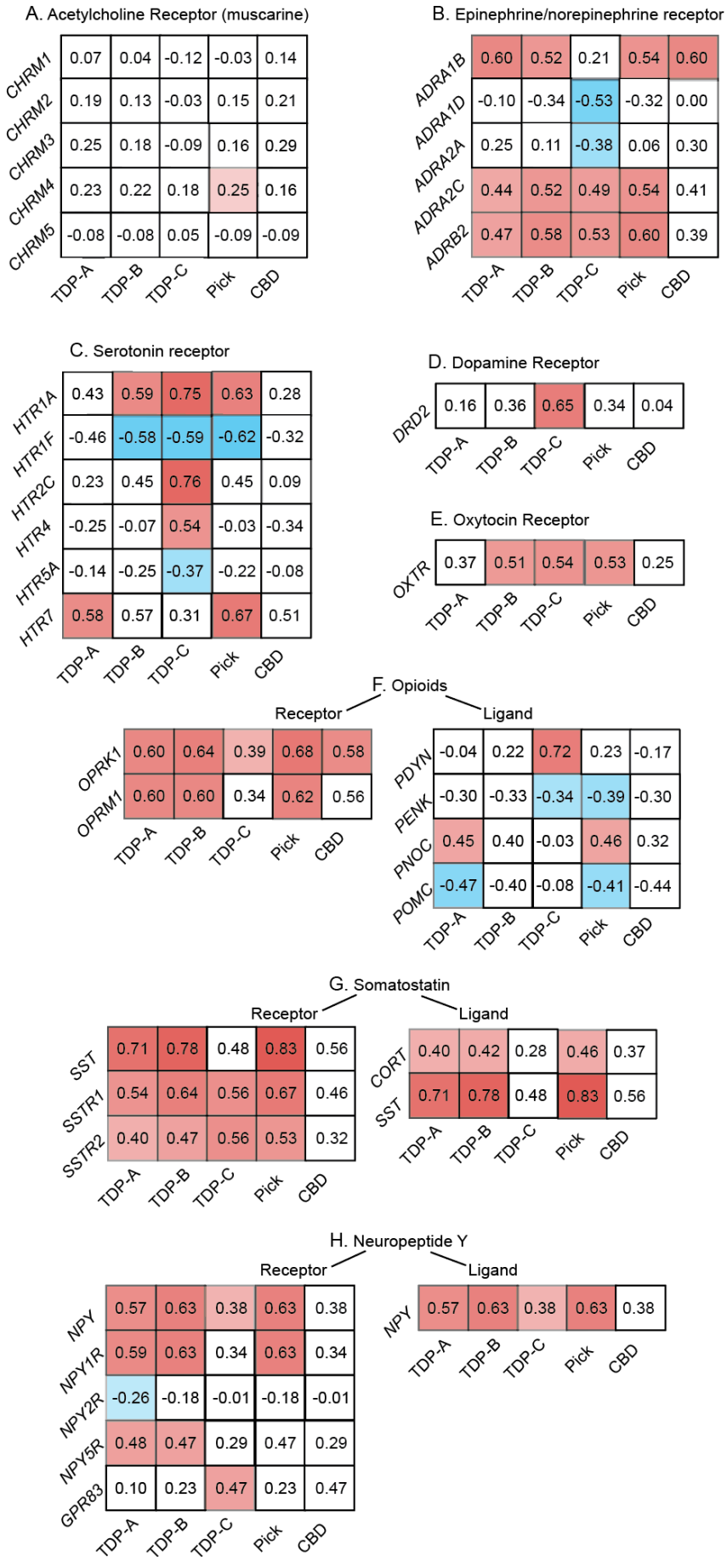


Figure 1. Regional gene expression correlating with atrophy in FTLD subtypes. (A-E). Maps on the left show group-averaged W-score maps for each FTLD subtype; warmer colors reflect greater cortical gray matter atrophy. Bar plots on the right show the Pearson correlation

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

coefficients between the regional gene expression and regional atrophy in FTLD subtypes; red bars indicate significantly correlated genes, FDR adjusted p <0.05. **(F)** Several genes were shared among FTLD-TDP subtypes, although a conspicuous number of genes was uniquely correlated to each subtype, particularly to FTLD-TDP-C. **(G)** While FTLD-Pick showed many uniquely correlated genes, most genes correlating with atrophy in FTLD-CBD were shared with FTLD-Pick. **(H)** FTLD-TDP had many uniquely correlated genes, while a large proportion of genes correlating with atrophy in FTLD-tau were shared with FTLD-TDP. $CORR_{FTLD\text{-}subtype}$ indicates lists of genes correlating with that specific FTLD subtype.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 2. Genes correlating with atrophy in FTLD are enriched for neuromodulatory terms. Gene set enrichment analyses using the spatial correlation values between brain gene expression levels and FTLD atrophy severity revealed a KEGG neuroligand receptor interaction pathway in common among most FTLD subtypes. Whereas genes associated with acetylcholine neurotransmission **(A)**, a system often preserved in FTLD, did, for the most part, not correlate with FTLD atrophy, shared and unique associations were observed between FTLD subtypes and genes associated with epinephrine **(B)**, serotonin **(C)**, dopamine **(D)**, oxytocin **(E)**, opioids **(F)**, somatostatin **(G)**, and neuropeptide Y receptor and ligand systems **(H)**. Matrices reflect the strength and sign of spatial correlation between genes associated with ligand and receptor systems and atrophy maps in FTLD subtypes. White squares reflect genes whose correlation with FTLD atrophy maps did not reach significance based on the bootstrap test correcting for spatial autocorrelation (FDR adjusted $p < 0.05$).

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

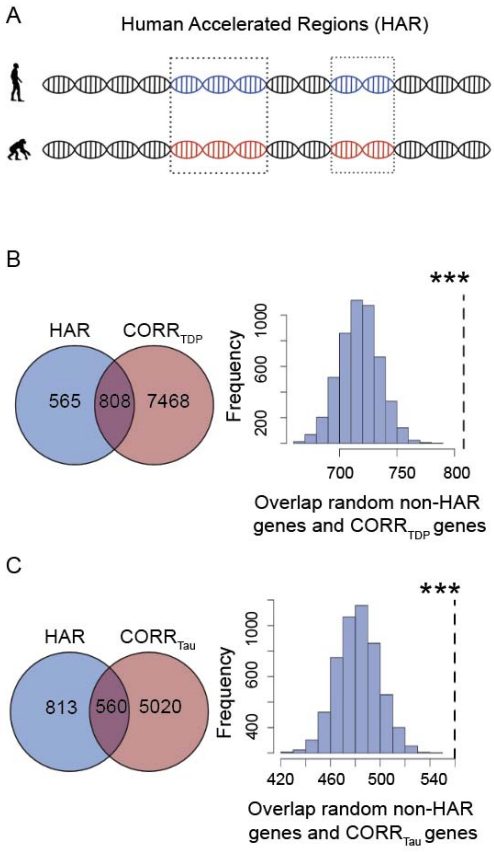


Figure 3. Overlap between FTLD atrophy correlated genes and HAR genes and CS genes. **(A)** HARs are conserved genomic loci which have undergone accelerated divergence in the human evolutionary lineage. **(B)** Genes correlating with atrophy in FTLD-TDP had 808 genes in common with HAR genes, **(C)** while genes correlating with atrophy in FTLD-tau had 560 genes in common with HAR genes. Both overlaps were unlikely to occur by chance when compared to a background set of brain-expressed genes. ***$p < 0.0005$. Panel A adapted with permission from[7,45].

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

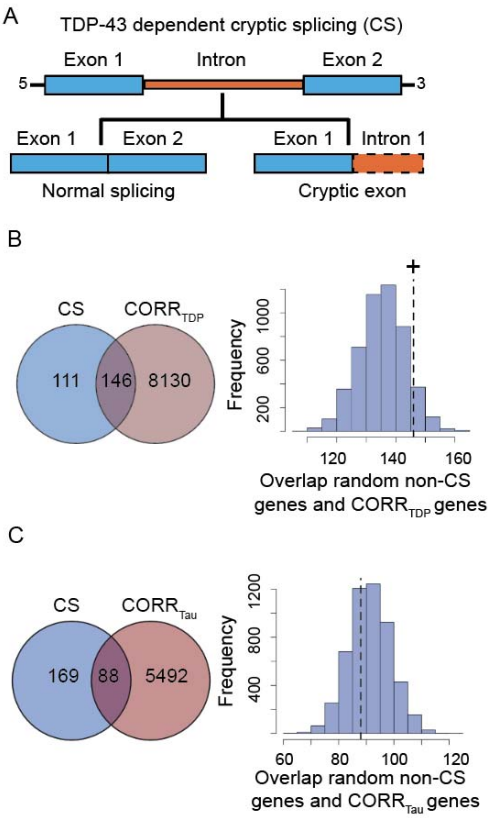


Figure 4. Overlap between FTLD atrophy correlated genes and HAR genes and CS genes. **(A)** CS genes are genes that incorporate novel intronic RNA into mature mRNA when TDP-43 is knocked down experimentally or depleted from diseased neuronal nuclei. **(B)** 146 genes were shared between CS genes and genes correlating with atrophy in FTLD-TDP. This overlap reached trend significance ($p = 0.10$). **(C)** 88 genes were shared between CS genes and genes correlating with atrophy in FTLD-tau, with this overlap being far from significant ($p = 0.71$). $^{+}p < 0.1$.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

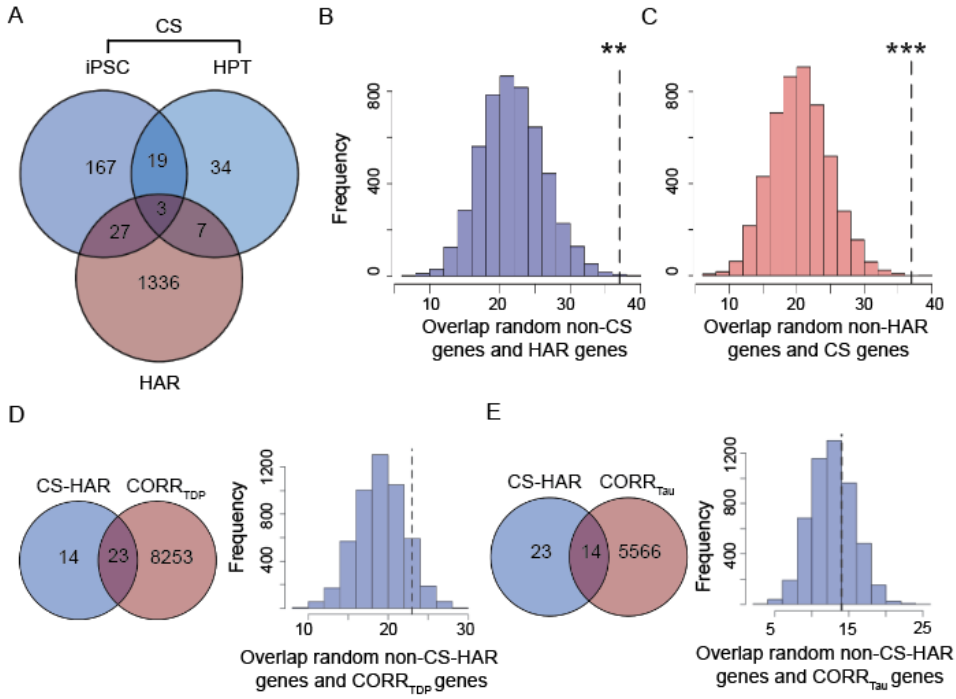


Figure 5. HAR and CS genes overlap. (A) Comparing the HAR and CS gene lists, including those derived from iPSC-derived neurons and from human patient tissues (HPT), 37 total genes were shared. Bootstrap-based testing, **(B-C)** was used to determine whether gene sets overlapped at a rate greater than expected by chance. First, we compared the overlap between random gene lists of the same number of genes as the CS genes with the HAR gene list **(B)**. Second, we compared the overlap between random gene lists of same number of genes as the HAR genes with the CS gene list **(C)**. Both tests showed that the overlap between CS genes and HAR genes is far greater than expected by chance. **(D)** 23 genes were shared between CS-HAR genes and genes correlating with atrophy in FTLD-TDP. This overlap was higher than expected by chance but did not reach nominal significance ($p = 0.17$). **(E)** 14 genes were shared between CS-HAR genes and genes correlating with atrophy in FTLD-tau, with this overlap not approaching significance ($p = 0.46$). $^{**}p < 0.005$; $^{***}p < 0.0005$.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

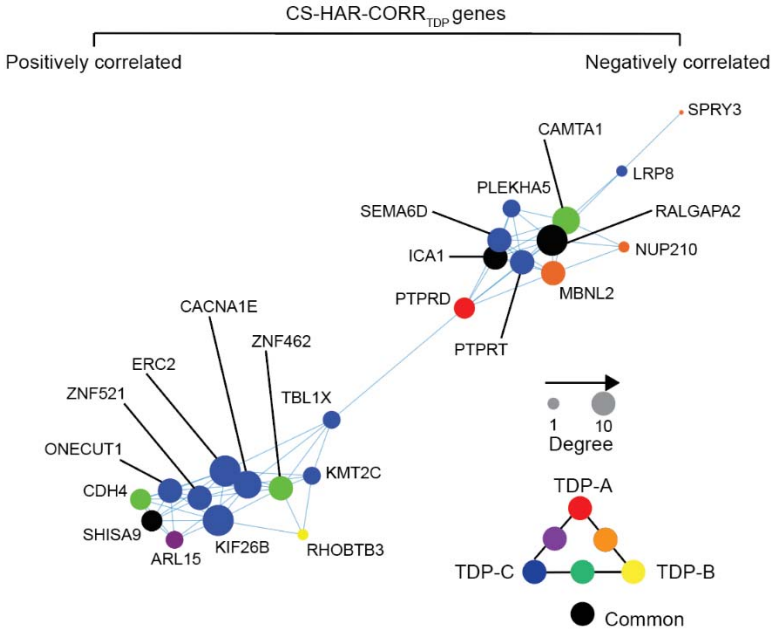


Figure 6. Co-expression network of CS-HAR genes associated with FTLD-TDP atrophy. The regional expression covariance of CS-HAR genes associated with FTLD-TDP atrophy was used to generate a topographical network of gene expression. The size of each circle reflects the centrality of each gene in the network, while the color of the circles reflects which FTLD-TDP subtype whose atrophy pattern correlated with the gene's expression pattern.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

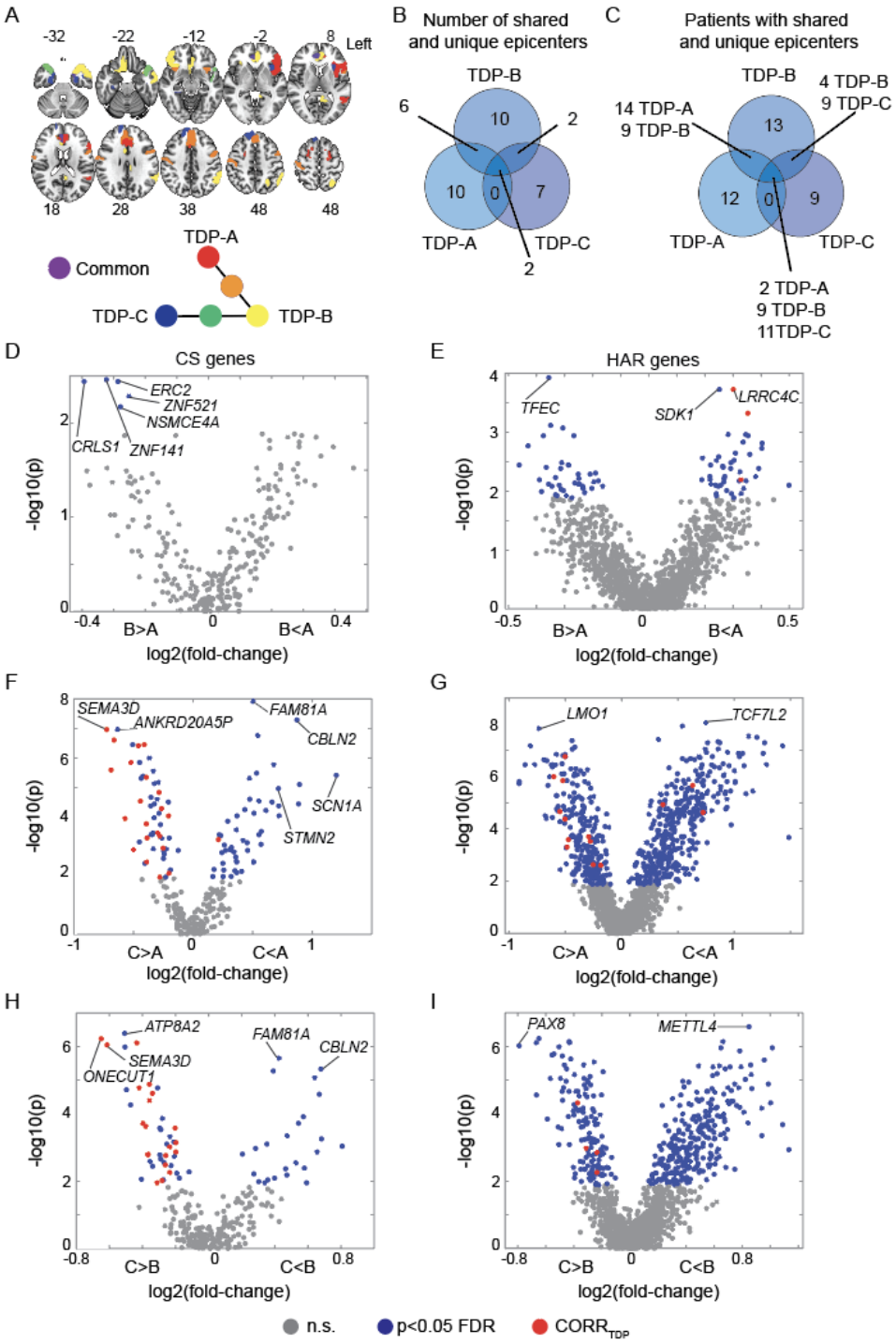


Figure 7. Normative expression of CS and HAR genes in FTLD-TDP disease epicenters. (A) Neuroanatomical distribution of FTLD-TDP subtype unique and shared epicenters. **(B)** Number of epicenters unique or shared across FTLD-TDP subtypes. **(C)** Number of patients with FTLD-

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

TDP having unique or shared epicenters. Epicenters' normative expression of CS genes compared between **(D)** TDP-A and TDP-B, **(E)** TDP-A and TDP-C, and **(F)** TDP-B and TDP-C. Epicenters' normative expression of HAR genes compared between **(G)** TDP-A and TDP-B, **(H)** TDP-A and TDP-C, **(I)** TDP-B and TDP-C. Grey circles depict genes with normative expression within epicenters that do not significantly (n.s.) differ across FTLD-TDP subtypes, blue circles depict genes with normative expression within epicenters that significantly differ (FDR adjusted $p < 0.05$) across FTLD-TDP subtypes, red circles depict genes with normative expression within epicenters that significantly differ (FDR adjusted $p < 0.05$) across FTLD-TDP subtypes and also positively correlate with atrophy in the FTLD-TDP subtype where they are more expressed.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297687; this version posted October 28, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Pathological diagnoses	TDP-A	TDP-B	TDP-C	CBD	PSP
N	28	35	29	45	27
Age at MRI scan in years	63.1 (6.7)	62.1 (9.1)	64.3 (7.4)	64.5 (6.7)	64.4 (7.1)
Scan-death interval in years	4.7 (3.2)	2.6 (2.2)	7.4 (2.8)	3.2 (1.6)	5.1 (3.4)
Sex (female:male)	16:12	15:20	11:18	25:20	11:16
Handedness (ambidextrous:left:right)	0:2:26	0:2:33	0:3:26	0:0:45	0:4:23
Education in years	16.1 (3.4)	16.1 (3.0)	16.6 (3.0)	16.2 (2.9)	16.0 (2.7)
Race (Asian:Black:White:n.a.)	1:0:24:3	0:0:34:1	0:0:28:1	4:1:36:4	2:0:22:3
Ethnicity (Hispanic/Latino:n.a.)	0:13	2:12	1:3	1:16	2:6
Scanner type (1.5T:Trio:Prisma:4T)	13:12:0:3	9:16:3:7	21:5:0:3	16:24:4:1	14:10:0:3
Total intracranial volume in liters	1.4 (0.1)	1.4 (0.1)	1.5 (0.2)	1.5 (0.1)	1.5 (0.2)
Clinical Dementia Rating (CDR)	1.1 (0.8)	1.4 (0.8)	0.7 (0.6)	0.8 (0.6)	0.8 (0.6)
CDR Sum of Boxes	6.1 (4.3)	7.4 (4.2)	3.8 (3.0)	4.1 (3.6)	5.3 (3.4)
MMSE	22.4 (7.2)	24.2 (5.4)	23.3 (7.2)	24.3 (5.8)	22.8 (8.0)
Clinical diagnoses					
Alzheimer's disease-type dementia	0	1	0	2	1
Behavioral variant FTD	16	30	2	11	16
Corticobasal syndrome	3	0	0	15	2
Mild cognitive impairment	0	1	1	0	0
Nonfluent variant PPA	4	1	0	11	6
Posterior supranuclear palsy syndrome	0	0	0	6	0
Semantic variant PPA	5	2	26	0	2

Table 1. Pathological, clinical, and demographical characteristics of patients with FTLD.

FTD = frontotemporal dementia; MMSE = Mini Mental State Examination; n.a. = not available;

PPA = primary progressive aphasia