



# Data-driven neuropathological staging and subtyping of TDP-43 proteinopathies

Alexandra L. Young,<sup>1,2,†</sup> Jacob W. Vogel,<sup>3,4,†</sup> John L. Robinson,<sup>5,6</sup>  
Corey T. McMillan,<sup>7,8</sup> Rik Ossenkoppele,<sup>4,9,10</sup> David A. Wolk,<sup>8</sup> David J. Irwin,<sup>7,8,11</sup>  
Lauren Elman,<sup>8</sup> Murray Grossman,<sup>7,8</sup> Virginia M.-Y. Lee,<sup>5,6</sup> Edward B. Lee<sup>5,6</sup>  
and Oskar Hansson<sup>4,12</sup>

<sup>†</sup>These authors contributed equally to this work.

See Hickman and Scholz (<https://doi.org/10.1093/brain/awad175>) for a scientific commentary on this article.

TAR DNA-binding protein-43 (TDP-43) accumulation is the primary pathology underlying several neurodegenerative diseases. Charting the progression and heterogeneity of TDP-43 accumulation is necessary to better characterize TDP-43 proteinopathies, but current TDP-43 staging systems are heuristic and assume each syndrome is homogeneous. Here, we use data-driven disease progression modelling to derive a fine-grained empirical staging system for the classification and differentiation of frontotemporal lobar degeneration due to TDP-43 (FTLD-TDP,  $n = 126$ ), amyotrophic lateral sclerosis (ALS,  $n = 141$ ) and limbic-predominant age-related TDP-43 encephalopathy neuropathologic change (LATE-NC) with and without Alzheimer's disease ( $n = 304$ ).

The data-driven staging of ALS and FTLD-TDP complement and extend previously described human-defined staging schema for ALS and behavioural variant frontotemporal dementia. In LATE-NC individuals, progression along data-driven stages was positively associated with age, but negatively associated with age in individuals with FTLD-TDP. Using only regional TDP-43 severity, our data driven model distinguished individuals diagnosed with ALS, FTLD-TDP or LATE-NC with a cross-validated accuracy of 85.9%, with misclassifications associated with mixed pathological diagnosis, age and genetic mutations. Adding age and SuStain stage to this model increased accuracy to 92.3%. Our model differentiates LATE-NC from FTLD-TDP, though some overlap was observed between late-stage LATE-NC and early-stage FTLD-TDP. We further tested for the presence of subtypes with distinct regional TDP-43 progression patterns within each diagnostic group, identifying two distinct cortical-predominant and brainstem-predominant subtypes within FTLD-TDP and a further two subcortical-predominant and corticolimbic-predominant subtypes within ALS. The FTLD-TDP subtypes exhibited differing proportions of TDP-43 type, while there was a trend for age differing between ALS subtypes. Interestingly, a negative relationship between age and SuStain stage was seen in the brainstem/subcortical-predominant subtype of each proteinopathy. No subtypes were observed for the LATE-NC group, despite aggregating individuals with and without Alzheimer's disease and a larger sample size for this group. Overall, we provide an empirical pathological TDP-43 staging system for ALS, FTLD-TDP and LATE-NC, which yielded accurate classification. We further demonstrate that there is substantial heterogeneity amongst ALS and FTLD-TDP progression patterns that warrants further investigation in larger cross-cohort studies.

- 1 Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, SE5 8AF, UK
- 2 Centre for Medical Image Computing, Department of Computer Science, University College London, London, WC1V 6LJ, UK
- 3 Department of Clinical Sciences, SciLifeLab, Lund University, SE-222 42 Lund, Sweden

Received January 31, 2023. Revised March 27, 2023. Accepted April 16, 2023. Advance access publication May 8, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of the Guarantors of Brain.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

- 4 Clinical Memory Research Unit, Lund University, SE-222 42 Lund, Sweden
- 5 Penn Center for Neurodegenerative Disease Research, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA
- 6 Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA
- 7 Penn Frontotemporal Degeneration Center, Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
- 8 Department of Neurology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA
- 9 Alzheimer Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, 1081 HZ Amsterdam, The Netherlands
- 10 Amsterdam Neuroscience, Neurodegeneration, 1081 HV Amsterdam, The Netherlands
- 11 Digital Neuropathology Laboratory, Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
- 12 Memory Clinic, Skåne University Hospital, SE-205 02 Malmö, Sweden

Correspondence to: Alexandra Young  
 Centre for Neuroimaging Sciences  
 De Crespigny Park, Camberwell, London SE5 8AF, UK  
 E-mail: alexandra.young@kcl.ac.uk

Correspondence may also be addressed to: Jacob Vogel  
 Skånes universitetssjukhus, VE Minnessjukdomar  
 205 02 Malmö, Sweden  
 E-mail: jacob.vogel@med.lu.se

**Keywords:** machine learning; neuropathological staging; limbic-predominant age-related TDP-43 encephalopathy; amyotrophic lateral sclerosis; frontotemporal lobar degeneration

## Introduction

The accumulation of TAR DNA-binding protein 43 (TDP-43) underlies a variety of syndromes, being the primary cause of many cases of frontotemporal dementia (FTD) and most cases of amyotrophic lateral sclerosis (ALS). More recently TDP-43 accumulation has been identified as a frequent feature of Alzheimer's disease and advanced ageing, a phenomenon described as limbic-predominant age-related TDP-43 encephalopathy neuropathological change (LATE-NC).<sup>1</sup> Neuropathological staging systems have been derived by neuropathologists for each proteinopathy that describe the stereotypical spatiotemporal progression pattern of TDP-43 accumulation.<sup>1–5</sup> However, the manual derivation of these staging systems limits their spatiotemporal resolution and relies on the assumption that there is a single common spatiotemporal progression of TDP-43 for each syndrome. This limitation is of particular importance given that the mapping between TDP-43 pathology and clinical syndrome is imperfect.<sup>6,7</sup>

Frontotemporal dementia is clinically and pathologically heterogeneous,<sup>8</sup> and recent work has recognized systematic variation in pathological features of TDP-43 underlying FTD.<sup>9</sup> ALS also exhibits substantial clinical heterogeneity<sup>10</sup>; whilst ALS is primarily characterized as a neuromuscular disorder, approximately 20% of ALS cases also have cognitive impairment,<sup>11</sup> and/or a typical FTD-syndrome. The concurrence of ALS and FTD suggest the possibility of these disorders constituting two extremes of an ALS-FTD spectrum disorder (ALS-FTD),<sup>12</sup> supported by observations that C9orf72 mutations can lead to either phenotype, or a mix of both.<sup>13,14</sup>

LATE-NC is a recently proposed terminology describing the commonly observed phenomenon of TDP-43 deposition mostly within medial temporal lobe structures in elderly adults.<sup>1</sup> Unlike FTD and ALS, the clinical syndrome that is associated with LATE-NC (termed 'LATE') currently does not have a corresponding

set of criteria for a clinical diagnosis. Although LATE is associated with memory impairment, discriminating whether cognitive impairment is due to LATE rather than other conditions is challenging without neuropathological evidence, particularly since it commonly co-occurs with Alzheimer's disease, which is also associated with memory impairment. LATE-NC is defined neuropathologically by a stereotypical distribution of TDP-43 in older adults.<sup>1,4,5</sup> However, the definition and description of this common neuropathological phenomenon is still being debated, and the terms LATE and LATE-NC have not met full consensus.<sup>15</sup>

Much is still unknown about the origin, progression and variability of TDP-43 aggregation as a whole, as well as within respective neurodegenerative diseases. For example, while many argue that there are clear clinical and pathological distinctions denoting distinct entities, there remains debate surrounding whether late-stage LATE-NC can be discriminated from early stage frontotemporal lobar degeneration due to TDP-43 (FTLD-TDP) given overlap of regions that are affected.<sup>15,16</sup> Similarly, the regional origin of FTLD-TDP is still unknown, with both the amygdala<sup>3</sup> and fronto-insular or anterior cingulate<sup>17</sup> proposed as candidate pathological epicentres.

A potential solution to capture the progression and spatial heterogeneity of TDP-43 proteinopathies is the use of data-driven disease progression modelling, which facilitates the probabilistic reconstruction of sequential disease progression patterns from cross-sectional data in a similar manner to traditional staging schema. However, in contrast to the traditional method of constructing neuropathological staging systems by hand, disease progression modelling can account for uncertainty in the level of pathology in each region and can enable the reconstruction of more complex progression patterns with large numbers of regions.<sup>18</sup> The Subtype and Stage Inference (SuStaIn)<sup>19</sup> algorithm combines disease progression modelling with clustering to further enable the

identification of subgroups of individuals (subtypes) with distinct disease progression patterns. This enables the probabilistic assignment of an individual to a subtype based on their subgroup, and stage based on their position along the inferred progression pattern for that subgroup. While previously applied to imaging datasets,<sup>19–22</sup> SuStain has recently been adapted to accommodate scored data often used in pathological assessment.<sup>23</sup>

Here we use SuStain to assess the progression and heterogeneity of TDP-43 accumulation in FTLD-TDP, ALS and LATE-NC across 21 brain regions. We first estimate a single progression pattern for each group to enable comparison with existing TDP-43 staging schema. Next, we demonstrate the potential utility of disease progression models for the three-way classification of FTLD-TDP, ALS and LATE-NC. We further explore whether there is heterogeneity in the progression of TDP-43 within FTLD-TDP, ALS and LATE-NC by modelling multiple progression patterns within each group, and we test for differences between the resulting subgroups.

## Materials and methods

### Dataset

Neuropathological samples with TDP-43 pathology were taken from the Center for Neurodegenerative Disease Research Brain Bank at the University of Pennsylvania, for which the tissue preparation, staining and immunohistochemistry procedures have been described previously.<sup>24</sup> The TDP-43 immunostain used in this study was 1D3 (rat anti-p409/410 TDP-43). Both sporadic cases and mutation carriers were included (Table 1 and Supplementary Table 1).

**Table 1 Demographic information for all individuals included in this study, stratified by TDP-43 diagnosis**

	ALS	FTLD-TDP	LATE-AD+	LATE-AD–
n	141	126	235	69
Age at onset	59.5 (11.0)	61.0 (9.0)	70.2 (9.0)	68.8 (10.0)
Age at death	63.7 (10.6)	68.5 (10.0)	80.9 (8.4)	80.9 (8.7)
Disease duration	4.3 (4.3)	7.3 (4.1)	10.3 (4.3)	11.9 (8.0)
% Female	40.4%	45.2%	62.1%	31.9%
Braak stage	1.7 (1.7)	2.0 (1.6)	5.8 (0.7)	3.3 (1.4)
% APOE4 carrier	25.0%	24.6%	63.3%	33.8%
% APOE2 carrier	12.9%	18.2%	5.2%	8.8%
% Mutation carrier	19.9%	46.8%	4.2%	5.8%
% Co-pathology	76.6%	88.9%	42.7%	55.1%
% AD phenotype	0.0%	12.7%	77.4%	8.7%
% FTDS phenotype	1.4%	81.7%	9.4%	15.9%
% MND phenotype	96.5%	0.0%	0.0%	1.4%
% LBD phenotype	0.7%	0.8%	4.3%	62.3%
% Other phenotype	1.4%	4.8%	9.0%	11.6%
PMI	12.6 (7.9)	12.8 (7.3)	12.6 (7.1)	11.5 (6.7)

Entries are mean (standard deviation) for the group, unless indicated to be percentages.

AD = Alzheimer's disease; ALS = amyotrophic lateral sclerosis; Co-pathology = demonstrates at least one secondary co-pathology (other than LATE); FTDS = frontotemporal dementia spectrum; FTLD-TDP = frontotemporal lobar degeneration due to TDP-43; LATE-AD+ = Alzheimer's individuals with amygdalar but not medullar TDP-43 pathology; LATE-AD– = non-Alzheimer's individuals with amygdalar but not medullar TDP-43 pathology; LBD = Lewy body disease; MND = motor neuron disease; PMI = post-mortem interval.

For each brain, up to 21 regions were sampled comprising the amygdala, hippocampal dentate gyrus, hippocampal cornu ammonis (CA)/subiculum, entorhinal cortex, anterior cingulate, superior/middle temporal gyrus, middle frontal gyrus, angular gyrus, occipital cortex, thalamus, globus pallidus, caudate/putamen, substantia nigra, midbrain, locus coeruleus, upper pons, cerebellum, medulla, orbitofrontal cortex, motor cortex and spinal cord (cervical spinal cord  $\alpha$ -motor neurons in lamina 9) (Supplementary Fig. 1). Each region was assigned a score according to a semi-quantitative rating scale based on the extent of TDP-43 inclusions, where 0 = non-detectable, 0.5 = sparse, 1 = mild, 2 = moderate and 3 = severe (Fig. 1). The reliability of these scores was verified by independent expert neuropathologists. All individuals with any evidence of TDP-43 pathology in the brain were included in this study.

Three data subsets—those with FTLD-TDP, ALS and LATE-NC—were extracted for disease progression modelling (Table 1 and Supplementary Table 1). The inclusion criteria for the FTLD-TDP and ALS subsets were a primary neuropathological diagnosis of FTLD-TDP<sup>9,25</sup> and ALS,<sup>12</sup> respectively. Four ALS patients with a diagnosis of 'ALS-Other', which includes ALS without TDP-43 proteinopathy, and one ALS patient with a diagnosis of 'ALS-Dementia' were excluded from this study. The inclusion criterion for the LATE-NC subset was a secondary or tertiary neuropathological diagnosis of LATE-NC,<sup>1</sup> with (LATE-AD+) or without (LATE-AD–) a primary neuropathological diagnosis of Alzheimer's disease neuropathologic change.<sup>26</sup> A secondary neuropathological diagnosis of LATE-AD+ was made if individuals showed a primary diagnosis of Alzheimer's disease (using NIA-AA criteria), did not have a secondary diagnosis of FTLD or ALS, and also had evidence of TDP-43.<sup>16</sup> The criteria for a neuropathological diagnosis of LATE-AD– were the absence of another TDP-43 neuropathological diagnosis (i.e. ALS, FTLD-TDP, or corticobasal degeneration), a TDP-43 score of 1 or more in the amygdala, and a TDP-43 score of less than 1 in the medulla (a region with near 100% sampling rate across diagnoses). These inclusion criteria were the sole criteria for inclusion in the study. In particular, we included individuals regardless of their TDP-type (A–E), and mutation carriers (e.g. GRN, C9orf72) and ALS-FTD cases were included. In total there were 126 individuals with FTLD-TDP, 141 with ALS and 304 with LATE-NC that were included in the SuStain modelling (Table 1 and Supplementary Table 1).

Age at death and estimated age at symptom onset were recorded for all individuals. Note that age of symptom onset refers to the age that primary symptoms were reported to manifest, which may or may not be symptoms directly related to TDP-43 neuropathology. Disease duration was calculated as the difference between age at symptom onset and age at death. Sex and interval between death and post-mortem examination was also recorded for all patients. APOE genotype (ALS: 99%; FTLD-TDP: 100%; LATE-NC: 98%) and Braak tau stage (ALS: 95%; FTLD-TDP: 99%; LATE-NC: 97%) were also recorded for nearly all individuals using methods that have been previously described.<sup>24</sup> TDP-43 type was also available for FTLD-TDP patients using published criteria.<sup>9</sup> TDP-43 type was missing for two FTLD-TDP patients.

### Disease progression modelling using SuStain

Disease progression modelling was performed using Ordinal SuStain.<sup>23,27</sup> The SuStain algorithm combines clustering and disease progression modelling to estimate subtypes with distinct progression patterns from cross-sectional data. The Ordinal version of SuStain is specifically designed for modelling ratings data such as



neuropathological ratings. Ordinal SuStaIn describes disease progression as a sequence in which different regions reach different scores, requiring as input the probability that a given score is observed in each region. In addition to learning the sequence in which different regions reach different scores, the algorithm quantifies the uncertainty in that sequence using Markov Chain Monte Carlo sampling. The uncertainty in the sequence is influenced by several factors including the number of samples available to estimate the sequence at different stages, the heterogeneity in the sequence between individuals and the signal to noise ratio of the biomarker measurements (neuropathological ratings in this case). We modelled scores of 1, 2 and 3 in Ordinal SuStaIn, translating the scores of 0.5, 1, 2 and 3 to probabilities by evaluating a normal distribution around that score with a standard deviation of 0.5, and normalizing by the sum of the probabilities of each score (Supplementary Table 2). Note that, for example, a score of 0.5 ('sparse') corresponds to an equal probability of the score being 0 or 1. Missing neuropathological ratings were modelled as having an equal probability of having each score. Regions with a significant proportion of missing data (missing in more than 25% of subjects) were excluded when running Ordinal SuStaIn. For FTLD-TDP, the orbitofrontal cortex was excluded due to missingness, giving a total of 20 regions. For ALS, the occipital cortex, locus coeruleus and orbitofrontal cortex were excluded, with 18 regions remaining. For LATE-NC, the motor cortex and spinal cord were excluded, leaving 19 regions. As each SuStaIn stage corresponds to a new region reaching a new score, there were  $n = 20 \times 3 = 60$  stages for FTLD-TDP,  $n = 18 \times 3 = 54$  stages for ALS and  $n = 19 \times 3 = 57$  stages for LATE-NC.

Ordinal SuStaIn was first used to estimate a single progression pattern for each group, enabling direct comparison with previous staging systems. Next, Ordinal SuStaIn was used to estimate multiple progression patterns for each group. The subtype progression patterns were cross-validated using 10-fold cross-validation and the optimal number of progression patterns (number of subtypes) within each group was chosen using the cross-validation information criterion, which balances model accuracy with model complexity.<sup>19</sup> For each progression pattern, individuals were first assigned to a subtype (in cases where the number of subtypes was greater than one) and were then assigned to a stage, based on maximum likelihood.<sup>19</sup> The SuStaIn stage can be thought of as a proxy for progression along a pathological trajectory. Subtypes were evaluated for crossover events by plotting the distribution of subtype probability across stages. Subtype probability distributions that crossed chance levels (in this case, 50% for both ALS and FTLD-TDP given  $k = 2$  subtypes) after Stage 1 were tagged as evidence for a crossover event. Crossover events suggest the possibility of discontinuity between variation observed before and after the event. Brain schematics in Figs 1, 5 and 6 were generated using: [https://github.com/AllenInstitute/hba\\_brain\\_schematic](https://github.com/AllenInstitute/hba_brain_schematic).

### Classification using SuStaIn

Three-way classification of FTLD-TDP, ALS and LATE-NC was initially performed by directly using the outputs of SuStaIn with a single progression pattern per group. SuStaIn was used to assign each individual to a group (FTLD-TDP, ALS or LATE-NC) and stage by evaluating the probability each individual belonged to each stage of each group's progression pattern. The group and stage combination that had the highest probability was chosen as the predicted diagnosis for that individual. Individuals that had a most probable stage of 0 for their most probable SuStaIn-based diagnosis were labelled as 'Unclassified'. Note that probabilities were computed

out-of-sample for each individual (using 10-fold cross-validation when evaluating within-diagnosis) and took into account uncertainty in the progression pattern by integrating over the Markov Chain Monte Carlo samples of the progression pattern output by SuStaIn. Unclassified Stage 0 individuals were excluded before classification.

### Classification using logistic regression

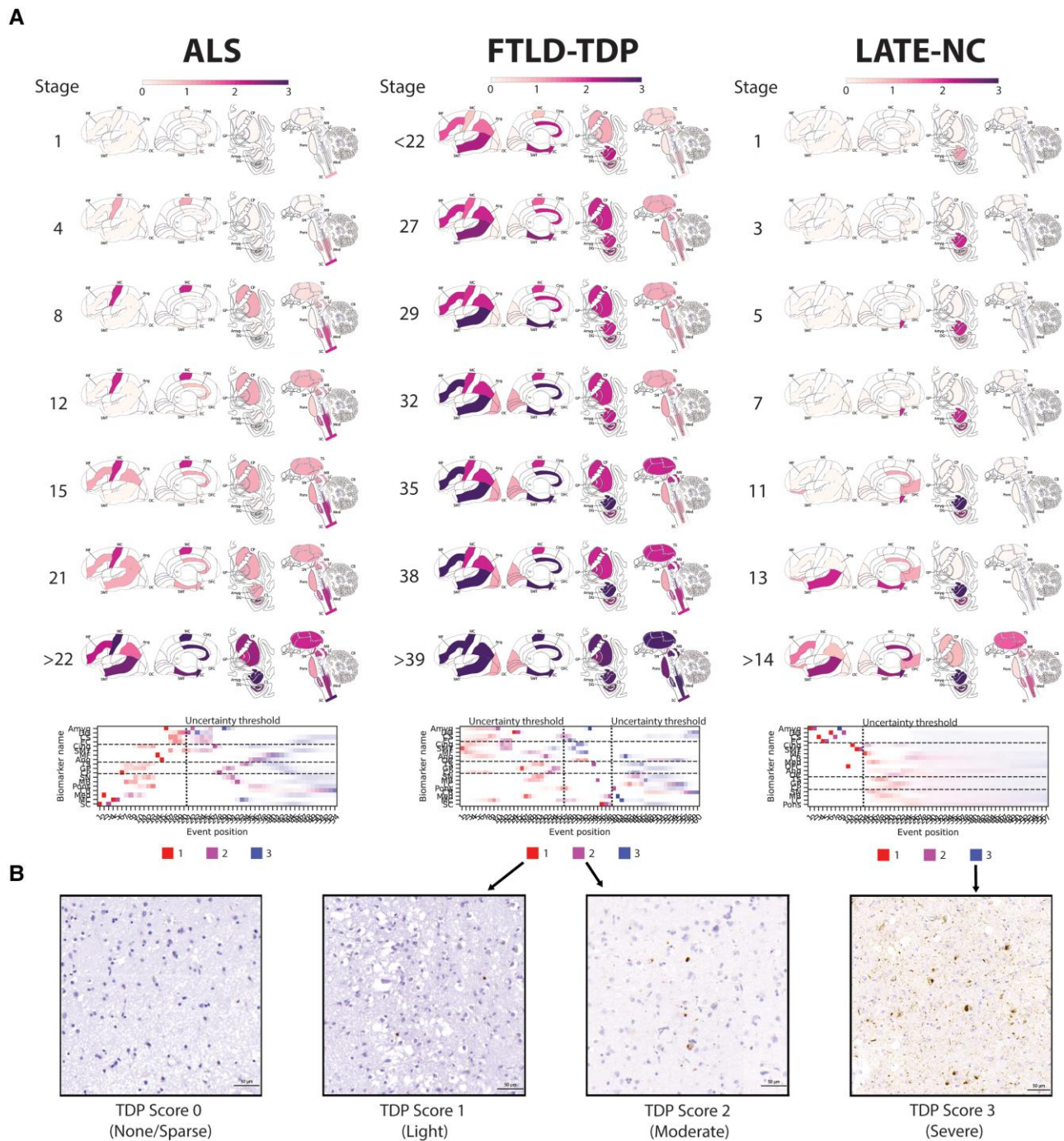
A second predictive diagnostic model was evaluated using a combination of SuStaIn and logistic regression. Progression patterns for each group (ALS, FTLD-TDP, LATE-NC) were used to assess the probability that each individual fit within each progression pattern. Similar to above, these probabilities were calculated by evaluating the probability of the maximum likelihood stage within the progression pattern for each diagnosis. The three probabilities, along with maximum likelihood SuStaIn stage and age at death, were used as independent features in a multinomial logistic regression three-way classification model using an L2 penalty and Limited-memory BFGS ('lbfgs') solver. The sample was divided into training (80%) and testing (20%) sets. In the training set, 10-fold cross-validation was used to optimize over a grid of 10 linearly spaced values between  $1 \times 10^{-4}$  and  $1 \times 10^4$  for the hyperparameter 'C'. Performance was evaluated by recording accuracy, precision, recall and F1 score, both overall and class-weighted, of predictions in the left out test set. This process was repeated 100 times to establish confidence intervals around the measurements.

### Comparison of SuStaIn-derived progression patterns to previously described staging systems

Previously described staging systems were used to stage each individual based on regional presence of TDP-43 pathology. For ALS staging, we used the four-stage system described in Brettschneider et al.<sup>2</sup> For FTLD-TDP, we used the four-stage system described in Brettschneider et al.<sup>3</sup> Given that this system was derived based specifically on patients with behavioural variant FTD (bvFTD), we describe staging across both the whole group of FTLD-TDP patients and across patients with a primary clinical diagnosis of bvFTD only. Since there are multiple proposed staging schemes for LATE-NC, we use three different proposed LATE-NC staging systems, as described by Nelson et al.<sup>1</sup> For each staging system, composite regions of interest were created for each stage based on averaging the TDP-43 scores of the regions falling within that stage (Supplementary Fig. 2). A patient was assigned to a stage if they showed a TDP-43 score equal to or greater than 1 for that stage as well as all previous stages in the regime. If a patient showed 'out of order' staging (e.g. score >1 in Stages I and III but not II), they were assigned to an 'Unclassifiable' category. Once all patients were staged using their group specific staging scheme, this stage was compared to the data-driven stage assigned to that patient. This approach allowed us to evaluate how much the data-driven SuStaIn stages resembled previously described manual staging systems of these various proteinopathies.

### Statistical analysis

Correlations were used to describe univariate relationships between SuStaIn stage and age, disease progression and total TDP-43 pathology, the latter calculated as the sum of all regional TDP-43 scores across an individual. Independent sample t-tests were used to assess differences between stage classifiable and unclassifiable patients (see 'Comparison of SuStaIn-derived



**Figure 1 Progression patterns of TDP-43 proteinopathies.** (A) Inferred trajectory of regional TDP-43 progression based on individuals with a primary pathological diagnosis of ALS (left), a primary pathological diagnosis of FTL-D-TDP (middle), or a secondary or tertiary pathological diagnosis of LATE-NC. Colours within each brain region represent the cumulative sum of probabilities (0–1) summed across three stages of severity, light pathology (1) in red, moderate pathology (2) in purple and severe pathology (3) in blue (see key at the bottom of each column). Only stages with reduced event ordering uncertainty (e.g. those represented by at least five individual donors) are shown. Brain schematics were generated using: [https://github.com/AllenInstitute/hba\\_brain\\_schematic](https://github.com/AllenInstitute/hba_brain_schematic). Below the brains are positional variance diagrams. Each box represents the degree of certainty that a given brain region (y-axis) has reached a given severity stage (red, light; purple, moderate; blue, severe) at a given SuStaIn stage (x-axis). (B) Representative micrographs for TDP-43 scores of 0, 1, 2 and 3, respectively. ALS = amyotrophic lateral sclerosis; Amyg = amygdala; Ang = angular gyrus; CB = cerebellum; Cing = anterior cingulate; CP = caudate/putamen; CS = CA1/subiculum; DG = dentate gyrus; EC = entorhinal cortex; FTL-D-TDP = frontotemporal lobar degeneration due to TDP-43; GP = globus pallidus; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathologic change; LC = locus coeruleus; MB = midbrain; MC = motor cortex; MF = middle frontal gyrus; Med = medulla; OC = occipital cortex; OFC = orbitofrontal cortex; SC = spinal cord; SMT = superior and middle temporal gyrus; SN = substantia nigra; TS = thalamus.

progression patterns to previously described staging systems' section) in the probability of assignment to a patient's diagnostic progression pattern (see 'Classification using SuStaIn' section).

Ordinary least squares general linear models and  $\chi^2$  tests were used compare misclassified to correctly classified patients from the machine learning classification analysis. To explore regional



differences between subtypes, general linear models were fit using each region as the dependent variable, subtype as the independent variable and SuStain stage as a nuisance covariate (to ensure relationships were not due to differences in disease progression). Benjamini–Hochberg false discovery rate was applied across all regions to correct for multiple comparisons. Multivariate logit maximum likelihood estimate regression models were used to investigate other subtype differences. Models were constructed using subtype as the dependent variable and the following as simultaneous independent variables: Age at symptom onset, age at death, sex, post-mortem interval, APOE E2 carriage, APOE E4 carriage, Braak stage and SuStain stage. For the FTLD-TDP model only, TDP-43 type was also included as a covariate. Owing to low numbers of individuals expressing TDP-43 type D ( $n = 2$ ), these individuals were excluded. In addition, two ALS individuals were assigned to ‘Stage 0’ by SuStain, indicating they did not have sufficient TDP-43 pathology to be subtyped, and were excluded from the statistical analyses. Both the ALS ( $n = 130$ ) and the FTLD-TDP ( $n = 115$ ) models were fitted on a sample of subjects with complete data for all variables.

### Data availability

Source code for the Ordinal SuStain algorithm is available at <https://github.com/ucl-pond/pySuStain>. All reasonable requests for access to neuropathological data used in this study will be considered by the Penn Neurodegeneration Data Sharing Committee. Applications can be completed using an online request form (<https://www.pennbindlab.com/data-sharing>).

## Results

### Progression of LATE-NC

Assuming a common progression pattern across all individuals with LATE-NC (Fig. 1), SuStain identified that TDP-43 deposition was initially confined to the amygdala region (SuStain Stages 1–2). This was expected given our definition of LATE-NC required TDP-43 pathology in the amygdala. During SuStain Stages 3–9, TDP-43 deposition spread to the hippocampal CA/subiculum, followed by the entorhinal cortex and then to the hippocampal dentate gyrus. Subsequently, at SuStain Stages 10–13, TDP-43 deposition emerged in the orbitofrontal cortex, followed by the anterior cingulate and then the superior/middle temporal gyrus. Moderate to severe TDP-43 pathology was present in all medial temporal regions before the appearance of TDP-43 pathology in cortical regions at SuStain Stage 10. Beyond SuStain Stage 13 there were few samples available, and the progression pattern had high uncertainty.

### Progression of ALS

Under the assumption of a single common progression pattern (Fig. 1), SuStain estimated that TDP-43 deposition in ALS began in the spinal cord, before progressing to the medulla and motor cortex in SuStain Stages 2–5. Subsequently, in SuStain Stages 6–13, there was TDP-43 deposition in the caudate/putamen, thalamus, globus pallidus, midbrain, substantia nigra, pons and anterior cingulate. In SuStain Stages 14–15, TDP-43 progressed to the middle frontal and angular gyri. By SuStain Stage 21, TDP-43 was found in all regions sampled except the cerebellum, including regions of the medial temporal lobe.

### Progression of FTLD-TDP

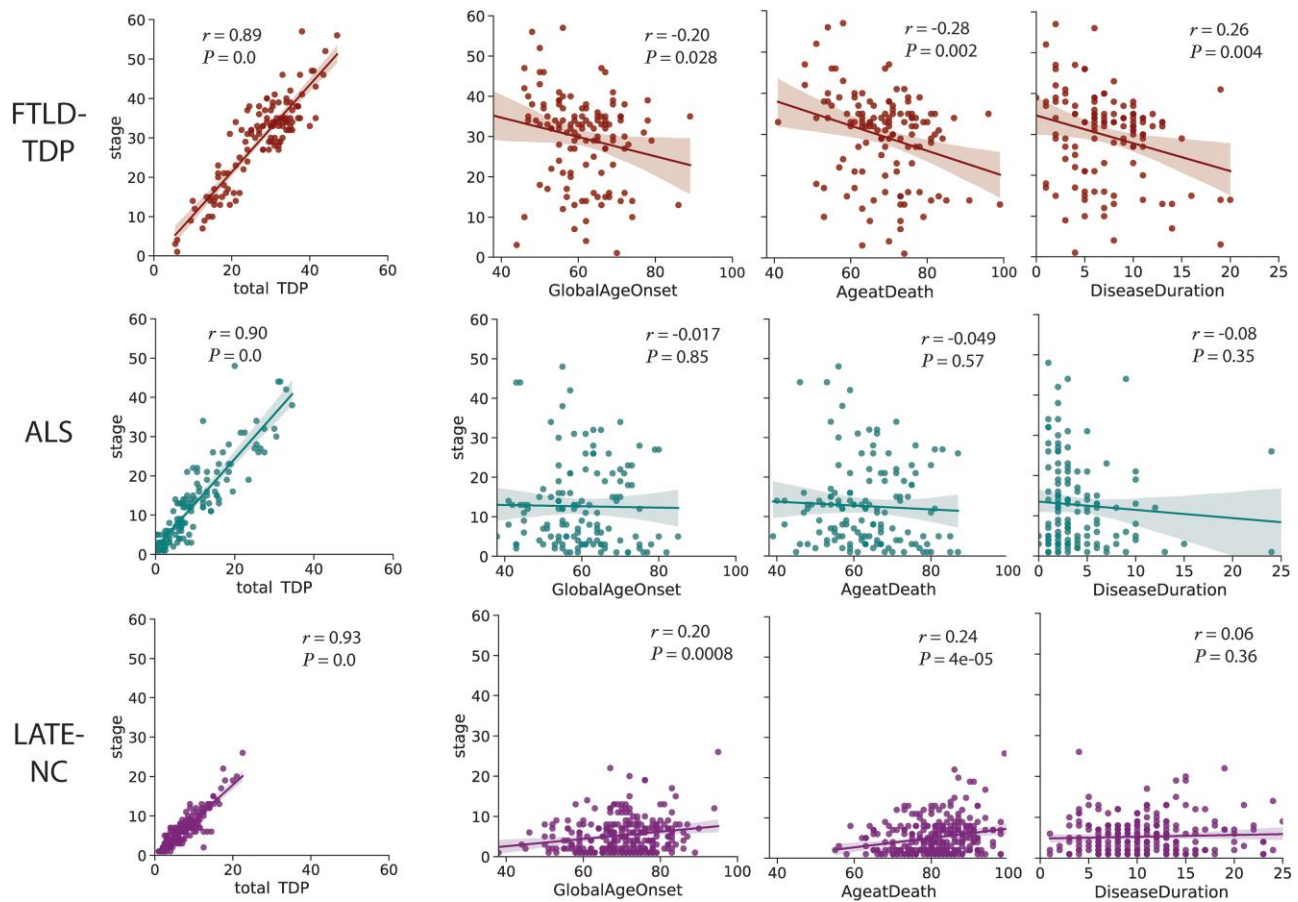
Assuming a single common progression pattern for FTLD-TDP (Fig. 1), SuStain identified that there was high uncertainty in the initial stages of TDP-43 progression in FTLD-TDP due to a lack of individuals with early stage TDP-43. The progression pattern was under-sampled and had high uncertainty before SuStain Stage 27. While uncertainty was too high to draw confident inference, SuStain inferred that the amygdala, entorhinal cortex, anterior cingulate and superior/middle temporal gyrus were the first regions to display moderate TDP-43 pathology, between SuStain Stages 10–13. When uncertainty became lower at SuStain Stage 27, there was already moderate TDP-43 deposition throughout regions sampled from the medial temporal lobe, basal ganglia and cerebral cortex, with the exception of the occipital lobe, and light pathology throughout all subcortical regions sampled. At SuStain Stages 28–32, TDP-43 deposition in the superior/middle temporal gyrus, anterior cingulate, entorhinal cortex and middle frontal became severe, while mild TDP-43 deposition was seen in the occipital cortex. By SuStain Stage 38, TDP-43 burden became severe in the amygdala, moderate in the thalamus, midbrain and medulla, and also appeared in the spinal cord. Beyond SuStain Stage 38, model uncertainty became too high for confident interpretation, though the locus coeruleus and cerebellum appeared to be the latest regions to be affected.

### Relationship between SuStain stage and age of onset, age at death and disease duration

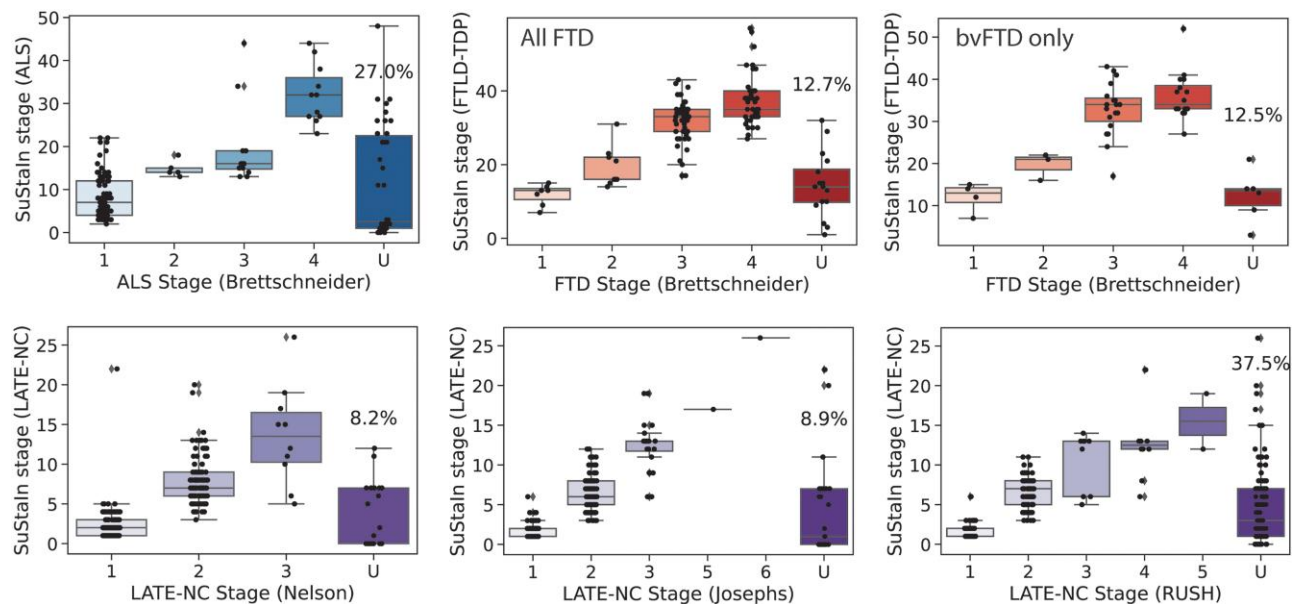
Figure 2 shows the relationship between SuStain stage and total TDP-43 pathology, age of onset, age at death and disease duration in each condition. As expected and by design, SuStain stage was nearly collinear with total TDP-43 pathology, making it a convenient proxy for overall pathologic progression. FTLD-TDP had an inverse relationship between SuStain stage and global age of onset, age at death, and disease duration. In contrast the LATE-NC group had a positive relationship between SuStain stage and global age of onset and age at death but no relation with disease duration. There was no significant relationship between SuStain stage and age or disease duration in ALS; however, see ALS subtype results below.

### Comparison with previous staging schema

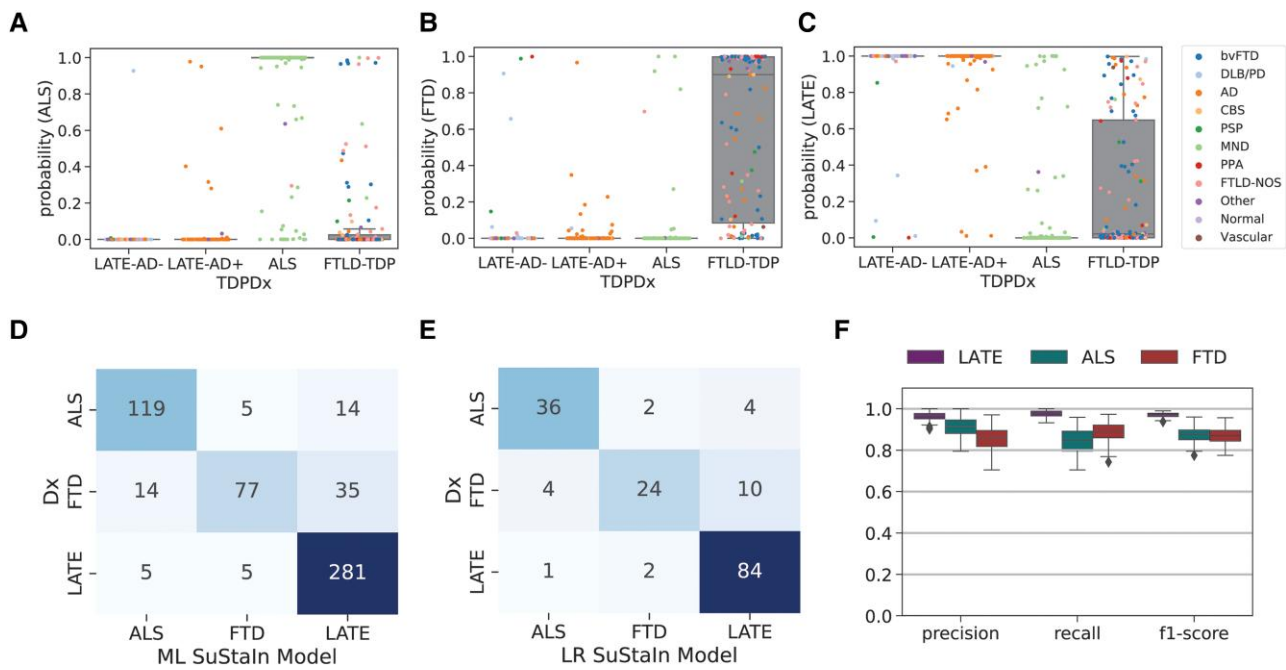
SuStain stage showed a good correspondence with existing staging schema (Fig. 3), whilst providing additional granularity. Individuals were considered unclassifiable on existing staging schema if some regions had evidence of TDP-43 pathology (score greater than 1) that was consistent with a late stage, but other regions lacked evidence of TDP-43 pathology consistent with the earlier stages of the staging scheme. Notably, for all prior staging systems, unclassifiable individuals occurred at low, moderate and high levels of pathology. Owing to its probabilistic formulation, SuStain was able to stage individuals that were otherwise unclassifiable using existing staging schema. This probabilistic information could also be used to identify individuals that did not fit the staging system well. Indeed, compared to classifiable individuals, unclassifiable individuals had a lower (cross-validated) probability of belonging to the dominant progression pattern of their diagnostic group for the ALS staging system ( $t = 2.60$ ,  $P = 0.01$ ) and two of three LATE-NC staging systems (Nelson:  $t = 1.46$ ,  $P = 0.14$ ; Josepchs:  $t = 5.84$ ,  $P < 0.001$ ; Rush:  $t = 2.85$ ,  $P = 0.0046$ ), though not for the FTLD-TDP staging system (full sample:  $t = -0.99$ ,  $P = 0.32$ ; bvFTD only:  $t = -1.26$ ,  $P = 0.21$ ).



**Figure 2 Relationships between age and disease progression.** The relationship between total TDP-43 pathology, age of disease onset, age at death and disease duration with total TDP-43 pathologic burden (as measured with SuStain stage) across all three diagnostic categories. SuStain stage was strongly related with total TDP-43 pathology across all diagnostic groups. LATE-NC showed a significant positive relationship between SuStain stage and age (at onset and at death), whereas FTL-D-TDP showed a significant negative relationship with age and disease duration. ALS = amyotrophic lateral sclerosis; FTL-D-TDP = frontotemporal lobar degeneration due to TDP-43; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathologic change.



**Figure 3 Comparison between SuStain-inferred TDP-43 pathological progression and previous staging systems.** Individuals are staged based on various proposed staging criteria (Supplementary Fig. 2). TDP-43 pathology was averaged within regions belong to each stage. If an individual shows TDP-43 pathology (1+) in an advanced stage before showing TDP-43 pathology in a previous stage, they are considered to be 'out-of-stage', represented by 'U' for 'unclassifiable'. SuStain stage generally showed good correspondence with previous pathological staging systems. ALS = amyotrophic lateral sclerosis; (bv)FTD = (behavioural variant) frontotemporal dementia; FTL-D-TDP = frontotemporal lobar degeneration due to TDP-43; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathologic change.



**Figure 4 Pathology-based disease classification.** All three diagnostic models were applied to all individuals regardless of TDP-type or genetic group, and the probability of the maximum likelihood stage was recorded. This probability represents a proxy for how well the individual's regional TDP-43 pattern fit the model trajectory for LATE-NC/ALS/FTLD-TDP. Box plots in A–C show the distribution of probabilities for each of the three diagnostic models (x-axis), stratified by pathological diagnosis (TDP Dx on the x-axis), for (A) the ALS model, (B) the FTLN-TDP model and (C) the LATE-NC model. Note that each graph (A–C) includes a probability for every subject. Note also that probabilities were derived using 10-fold cross-validation for within-diagnosis assessments (e.g. ALS cases tested using the ALS model) to avoid over-fitting. Each individual is coloured in accordance with their clinical diagnosis. Generally, individuals showed a high probability in models trained on their diagnosis, and a low probability in others. (D) A confusion matrix showing agreement between pathological diagnosis and maximum likelihood subtype model. True pathological diagnosis labels are represented on the y-axis, while predicted labels are shown on the x-axis. Individual pathological profiles tended to agree best with models fit to their clinical diagnostic group, but this was not true in all cases. (E) A logistic regression model was trained on the probabilities, plus maximum likelihood SuStain stage and age at death, using 100 iterations of train/test splits. The confusion matrix shows the average agreement between pathological diagnosis and predicted subtype in the 100 left-out test groups. (F) Distribution of classification statistics for performance of the maximum likelihood model across the 100 train/test splits, stratified by pathological diagnosis. See [Supplementary Table 3](#) for further statistics and comparison with the maximum likelihood model. AD = Alzheimer's disease; ALS = amyotrophic lateral sclerosis; bvFTD = behavioural variant frontotemporal dementia; DLB/PD = dementia with Lewy bodies/Parkinson's disease; CBS = corticobasal syndrome; Dx = diagnosis; FTD = frontotemporal dementia; FTLN-NOS = frontotemporal lobar degeneration not otherwise specified; LATE-AD+ = Alzheimer's individuals with amygdalar but not medullar TDP-43 pathology; LATE-AD- = non-Alzheimer's individuals with amygdalar but not medullar TDP-43 pathology; MND = motor neuron disease; PPA = primary progressive aphasia; PSP = progressive supranuclear palsy.

### Three-way classification of FTLN-TDP, ALS and LATE-NC

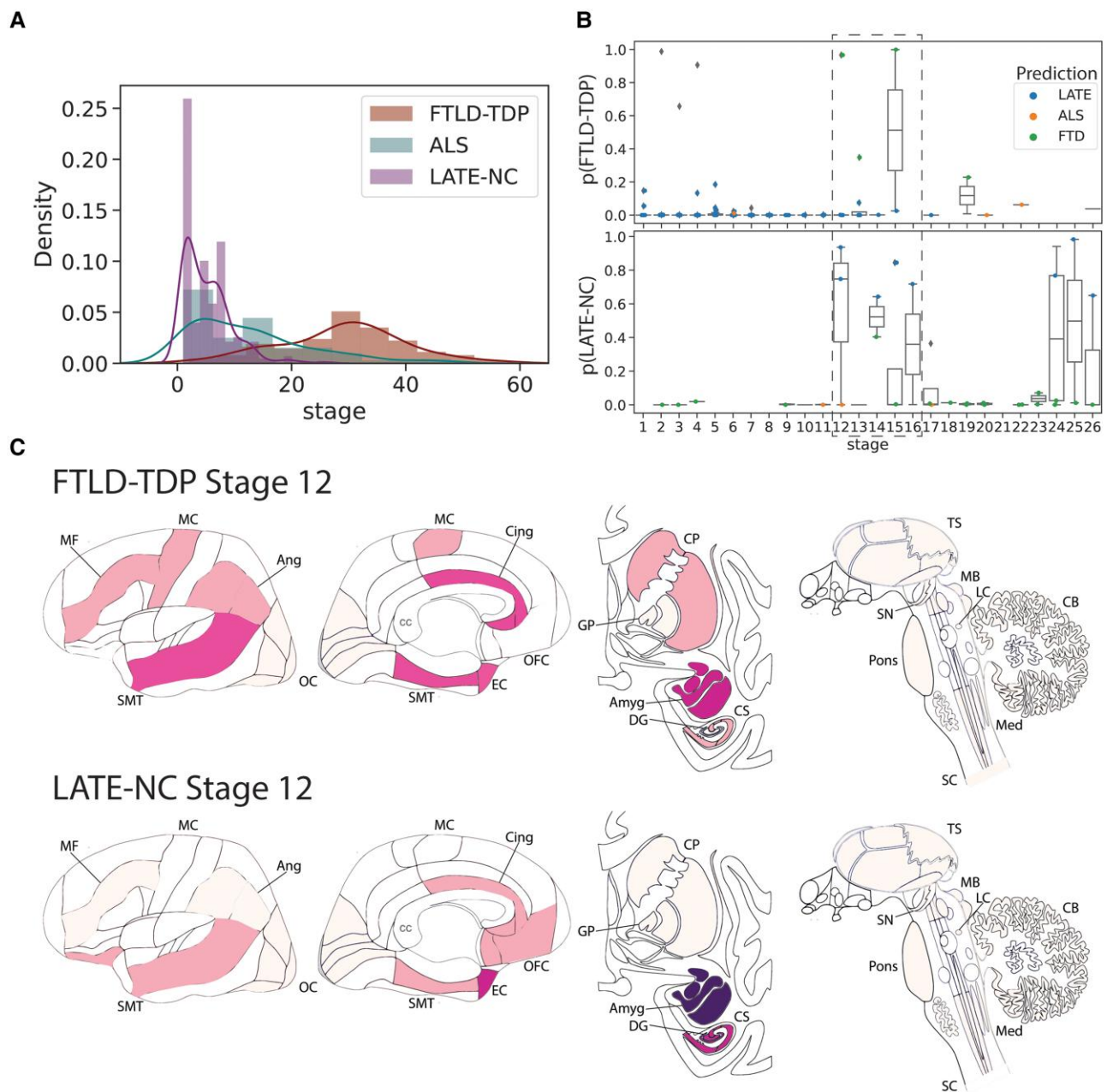
Using SuStain, each participant was assigned a probability of belonging to the typical ALS, FTLN-TDP or LATE-NC progression ([Fig. 4A–C](#)), regardless of TDP-type or genetic group. Most cases featured a high probability of their true pathological diagnosis. In fact, a winner-takes-all classification using only these probabilities resulted in a three-way classification accuracy of 85.9% (balanced accuracy = 81.3%; [Fig. 4D](#) and [Supplementary Table 3](#)), which was hampered most by a relatively low recall in FTLN-TDP diagnosis. Specifically, many FTLN-TDP cases were misclassified as LATE-NC ([Fig. 4D](#)), and the probability of a LATE-NC diagnosis varied widely among FTLN-TDP cases ([Fig. 4B](#)). After further optimizing this model using logistic regression and including SuStain stage and age at death as features, out-of-sample three-way classification accuracy improved to 92.3% (balanced accuracy = 90.3%; [Fig. 4E](#) and [F](#) and [Supplementary Table 3](#)). Diagnostic tests of the decision boundaries suggested this improvement involved allowing higher ALS probability thresholds in the context of lower LATE-NC thresholds ([Supplementary Fig. 3A](#)). Most misclassifications across both the

maximum likelihood and the logistic regression models involved FTLN-TDP patients misclassified as ALS or LATE-NC, and many of these misclassifications were high confidence ([Fig. 4B](#) and [Supplementary Fig. 3A](#)). Interestingly, cases having both ALS and FTLN-TDP as primary and secondary diagnosis (in either order) tended to have intermediate probabilities of both ALS and FTLN-TDP ([Supplementary Fig. 3B](#)). Many of the misclassifications were consistent with secondary underlying pathology, such as FTLN-TDP and ALS; for further case studies of misclassified individuals see [Supplementary material](#) and [Supplementary Fig. 4](#).

### Differentiating LATE-NC from FTLN-TDP

Despite SuStain issuing good classification of LATE-NC from FTLN-TDP, the very low number of early stage FTLN-TDP cases or late-stage LATE-NC cases ([Fig. 5A](#)) made it difficult to examine the point where these pathological entities could potentially intercept. However, sufficient cases were present at SuStain Stage 12 of both diagnostic groups to examine overlap of modelled SuStain's progression patterns with a tolerable degree of certainty. Qualitatively, at Stage 12, SuStain projects FTLN-TDP

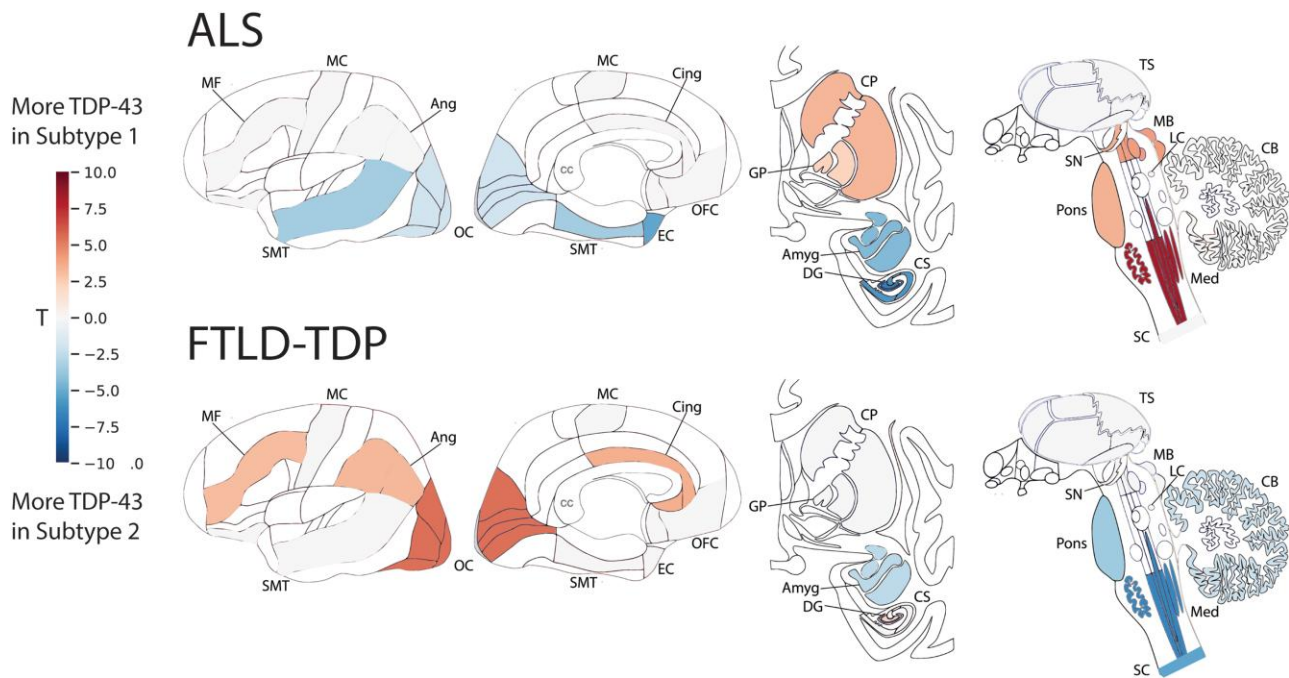




**Figure 5 Comparing LATE-NC and FTLD-TDP.** (A) Histograms showing the proportion of individuals assigned to each stage, separately for FTLD-TDP, ALS and LATE-NC models/individuals. Nearly all LATE-NC individuals were assigned to Stages 16 or lower, while nearly all FTLD-individuals were assigned to Stages 16 or higher. (B) SuStaIn tends to confuse LATE-NC and FTLD-TDP, particularly the stages of the crossover noted in A. The top plot represents the probability of being classified as FTLD-TDP among LATE-NC individuals, whereas the bottom plot represents the probability of being classified as LATE-NC for FTLD-TDP individuals. For both sets of patients, probability of misclassification is very low, except between Stages 12–16. (C) Comparison of FTLD-TDP and LATE-NC at Stage 12, with colours representing the amount of pathology using the same scale as Fig. 1, ranging from white (no pathology) to red (light pathology) to purple (moderate pathology) to blue (severe pathology). Brain schematics were generated using: [https://github.com/AllenInstitute/hba\\_brain\\_schematic](https://github.com/AllenInstitute/hba_brain_schematic). ALS = amyotrophic lateral sclerosis; Amyg = amygdala; Ang = angular gyrus; CB = cerebellum; Cing = anterior cingulate; CP = caudate/putamen; CS = CA1/subiculum; DG = dentate gyrus; EC = entorhinal cortex; FTD = frontotemporal dementia; FTLD-TDP = frontotemporal lobar degeneration due to TDP-43; GP = globus pallidus; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathologic change; LC = locus coeruleus; MB = midbrain; MC = motor cortex; MF = middle frontal gyrus; Med = medulla; OC = occipital cortex; OFC = orbitofrontal cortex; SC = spinal cord; SMT = superior and middle temporal gyrus; SN = substantia nigra; TS = thalamus.

patients to have greater cortical burden of TDP-43 pathology, particularly in the anterior cingulate, as well as sparse TDP-43 burden in the basal ganglia. In contrast, Stage 12 LATE-NC cases were projected to have greater (more severe) TDP-43 burden in the medial temporal lobe (Fig. 5C). However, among LATE-NC

cases, a brief increase of FTLD probability was seen around LATE-NC Stages 12–16, and a similar phenomenon of higher LATE-NC probability was seen in FTLD-TDP cases also around Stages 12–16 (Fig. 5B), indicating some degree of overlap between these two entities around these stages.



**Figure 6** Subtypes of primary TDP-43 proteinopathies. T-maps showing regions that are significantly different between ALS subtypes (top) and FTLD-TDP subtypes (bottom), after controlling for SuStaIn stage and multiple comparisons. A positive t-value (red) indicates more severe TDP-43 pathology in Subtype 1 and a negative t-value (blue) indicates more severe TDP-43 pathology in Subtype 2. Brain schematics were generated using: [https://github.com/AllenInstitute/hba\\_brain\\_schematic](https://github.com/AllenInstitute/hba_brain_schematic). ALS = amyotrophic lateral sclerosis; Amyg = amygdala; Ang = angular gyrus; CB = cerebellum; Cing = anterior cingulate; CP = caudate/putamen; CS = CA1/subiculum; DG = dentate gyrus; EC = entorhinal cortex; FTLD-TDP = frontotemporal lobar degeneration due to TDP-43; GP = globus pallidus; LC = locus coeruleus; MB = midbrain; MC = motor cortex; MF = middle frontal gyrus; Med = medulla; OC = occipital cortex; OFC = orbitofrontal cortex; SC = spinal cord; SMT = superior and middle temporal gyrus; SN = substantia nigra; TS = thalamus.

### Heterogenous pathological progression patterns of TDP-43 proteinopathies

For each neuropathologic diagnostic category, SuStaIn was refit allowing estimation of multiple progression patterns. For ALS, SuStaIn identified two subtypes of TDP-43 progression (Fig. 6 and Supplementary Figs 5 and 6). The first subtype, which we labelled as subcortical predominant, had greater TDP-43 deposition in brainstem and subcortical regions, namely in the medulla, pons, basal ganglia, midbrain and substantia nigra. The second subtype, which we labelled as corticolimbic predominant, had greater TDP-43 deposition in cortical and medial temporal regions, namely the amygdala, hippocampus, superior/middle temporal gyrus, entorhinal cortex and the occipital cortex (Fig. 6 and Supplementary Fig. 6). No evidence was found for putative crossover events between the two subtypes after Stage 1 (Supplementary Fig. 5).

When allowing SuStaIn to estimate multiple progression patterns for FTLD-TDP, SuStaIn identified two subtypes with distinct TDP-43 progression (Fig. 6 and Supplementary Figs 5 and 7). The first subtype, which we labelled as cortical predominant, had greater TDP-43 deposition in the angular gyrus, occipital cortex, superior/middle temporal gyrus, middle frontal gyrus, anterior cingulate and dentate gyrus. The second subtype, which we labelled as brainstem predominant, had greater TDP-43 deposition in the medulla, pons, spinal cord, cerebellum and amygdala (Fig. 6 and Supplementary Fig. 7). No evidence for a subtype crossover event was observed (Supplementary Fig. 5).

Despite aggregating a diverse set of individuals with and without Alzheimer's disease, a single progression pattern best described the TDP-43 progression pattern for the  $n = 304$  participants with a

secondary or tertiary neuropathological diagnosis of LATE-NC (Supplementary Fig. 5).

### Clinical characteristics of FTLD-TDP subtypes

FTLD-TDP pathological subtypes were assessed for differences in demographic, pathological and genetic variables (Table 2). Compared to Subtype 2 (brainstem predominant), Subtype 1 (cortical predominant) participants were more likely to present with TDP-43 type C pathology, while Subtype 2 was more likely to present with TDP-43 type B pathology and type E pathology, after adjusting for other covariates. There was also a significant interaction between SuStaIn subtype and SuStaIn stage on age at death ( $t = -2.85$ ;  $P = 0.005$ ) and disease duration ( $t = -3.26$ ,  $P = 0.001$ ), such that Subtype 2 individuals showed a stronger negative relationship between SuStaIn stage and age at death (Supplementary Fig. 8). Individuals with a secondary diagnosis of ALS were more likely to be Subtype 2 (85.7% of seven individuals with secondary neuropathological evidence of ALS versus 52.9% of 119 individuals without secondary neuropathological evidence of ALS), but this trend did not reach statistical significance ( $P = 0.1928$ ).

### Clinical characteristics of ALS subtypes

Similarly, differences in various clinical variables were tested between ALS Subtype 1 (subcortical predominant) and ALS Subtype 2 (corticolimbic predominant). Two confounding factors significantly discriminated between the two subtypes (Table 2): Subtype 1 had a longer post-mortem interval and also tended to have less overall brain pathology (e.g. lower SuStaIn stage) compared with

Table 2 Comparison between regional TDP-43 subtypes of FTLD-TDP and ALS

	S1	S2	coef	P	Versus B	P	Versus C	P	Versus E	P
<b>FTLD-TDP</b>										
n	57	69								
Age at onset	60.5 (9.1)	61.5 (9.0)	−0.00	0.89						
Age at death	68.3 (9.3)	68.6 (10.7)								
Disease duration	7.6 (2.8)	7.1 (4.9)	0.04	0.45						
% Female	57.9%	52.2%	0.10	0.80						
Braak stage	2.0 (1.4)	2.0 (1.8)	−0.13	0.58						
% APOE4 carrier	26.3%	23.2%	0.20	0.67						
% APOE2 carrier	17.5%	18.8%	−0.21	0.70						
% TDP-A	33.3%	37.7%			0.64	0.23	−0.68	0.19	2.06	0.09
% TDP-B	24.6%	34.8%					1.32	0.03	1.42	0.24
% TDP-C	36.8%	17.4%							2.74	0.03
% TDP-E	1.8%	8.7%								
SuStain stage	30.1 (10.3)	27.8 (11.3)	−0.03	0.13						
PMI	12.4 (6.9)	13.1 (7.6)	−0.01	0.84						
<b>ALS</b>										
n	78	61								
Age at onset	57.4 (11.0)	61.5 (10.2)	−0.04	0.09						
Age at death	62.0 (10.3)	65.2 (10.3)								
Disease duration	4.6 (4.5)	3.9 (4.2)	0.02	0.66						
% Female	61.5%	57.4%	0.27	0.53						
Braak stage	0.8 (0.8)	0.9 (0.9)	−0.06	0.82						
% APOE4 carrier	28.6%	21.3%	0.46	0.31						
% APOE2 carrier	15.6%	9.8%	0.66	0.38						
SuStain stage	10.9 (9.4)	15.4 (10.7)	−0.05	0.01						
PMI	14.3 (8.7)	10.6 (6.1)	0.07	0.02						

The top and bottom of the table represent the output for FTLD-TDP and ALS, respectively, of a single model including all covariates. Entries are mean (standard deviation) for a subtype, unless indicated to be percentages, coefficients, or P-values.

ALS = amyotrophic lateral sclerosis; FTLD-TDP = frontotemporal lobar degeneration due to TDP-43; PMI = post-mortem interval; S1 = Subtype 1; S2 = Subtype 2; SuStain = Subtype and Stage Inference.

Subtype 2. Covarying for these and other variables, there was a non-significant trend ( $P=0.09$ ) for Subtype 1 being associated with a younger onset age. There was a significant SuStain subtype  $\times$  SuStain stage interaction on symptom onset age ( $t=2.56$ ,  $P=0.01$ ) and age at death ( $t=2.72$ ,  $P=0.007$ ), where Subtype 1 showed a stronger negative relationship (Supplementary Fig. 8).

## Discussion

In this study we employed data-driven disease progression modelling to derive an empirical system for staging and differentiating three major TDP-43 proteinopathies: ALS, FTLD-TDP and LATE-NC. The data driven staging schema derived from disease progression modelling corroborated previously described staging systems, whilst revealing additional detail in the progression patterns of each TDP-43 proteinopathy. Individuals diagnosed with ALS, FTLD-TDP or LATE-NC could be distinguished with high accuracy based on their probability of belonging to the learnt progression pattern for each proteinopathy. Most misclassified cases appeared to be driven by features of more than one TDP-43 diagnostic entity intermingling, as well as mutations and presence of co-pathologies. Our study further identified considerable heterogeneity in the progression of ALS and FTLD-TDP across individuals, but the progression of LATE-NC was found to be remarkably homogeneous.

Previous staging systems have been based on the delineation of a smaller, coarser set of stages based on the overall severity of pathology. In contrast, our data-driven staging systems were based on probabilistic inference of the sequential progression of TDP-43, enabling us to infer fine-grained progression patterns and to stage

individuals probabilistically. Where previously described staging systems resulted in a substantial proportion of ‘unclassifiable’ individuals that did not precisely match the staging system, our data-driven method was able to probabilistically assign individuals to their best matching stage, accounting for the variability and uncertainty in the TDP-43 progression pattern for each particular TDP-43 proteinopathy. The probabilistic nature of the SuStain algorithm also allowed us to build proteinopathy classifiers by comparing an individual’s regional pattern to that of the dominant pattern of a given proteinopathy. This resulted in a robust and fully automated classification of diseases based only on an array of regional TDP-43 frequency scores plus age at death. The combined staging and classification could be useful for pathologists in classifying unusual cases and for establishing consistent quantification of individual TDP-43 progression for research studies.

Previous studies have described considerable pathological, clinical and biomarker heterogeneity in both ALS<sup>28–30</sup> and FTLD-TDP,<sup>31,32</sup> and the two syndromes often co-occur within an individual.<sup>12</sup> Here we identified two major neuropathological subtypes in ALS and a further two in FTLD-TDP. The ALS subtypes broadly split into a subcortical predominant subtype with greater involvement of the spinal cord, medulla, pons and the caudate/putamen, and a corticolimbic predominant subtype with greater involvement of the temporal lobe, amygdala and hippocampus. Previous studies have similarly identified a subgroup of ALS cases with greater hippocampal and cortical involvement.<sup>33,34</sup> There is some controversy as to whether TDP-43 pathology in ALS originates in the motor cortex or the spinal cord. While our subtype analysis revealed subgroup differences in the relative timing of regional



onset of motor cortex pathology, SuStain inferred the spinal cord as the first susceptible region for both ALS subtypes.

The FTLD-TDP patterns identified in this study divided into a brainstem predominant subtype with greater subcortical and brainstem involvement, and a cortical predominant subtype with greater cortical involvement. This is consistent with previous studies identifying subcortical involvement in FTD subgroups,<sup>35,36</sup> with the brainstem predominant subtype having greater subcortical involvement. The brainstem predominant subtype was more likely to present with TDP-43 type B and E pathology, whilst the cortical predominant subtype was more likely to present with TDP-43 type C pathology. This aligns with previous work demonstrating that TDP-43 type B is associated with FTD-ALS phenotypes,<sup>32,37–39</sup> with the subcortical and brainstem involvement in the brainstem predominant subtype being closer to the pattern of TDP-43 progression in ALS. As with ALS, further studies in larger samples may reveal further subdivision of the subtypes.

SuStain stage was positively correlated with age in LATE-NC, in agreement with previous studies.<sup>1</sup> However, in FTLD-TDP SuStain stage was found to be negatively correlated with age. Interestingly, when allowing for subtypes, both ALS and FTLD-TDP exhibited subtypes with an increased amount of subcortical pathology and an inverse relationship between age and pathology, suggesting that age associations may have been driven by subgroup associations for FTLD-TDP and negated by subgroup associations for ALS. The factors leading to positive and negative relationships between pathology and age are complex and incompletely understood. A positive association with age might be produced by a slow progressing (typically late onset) pathology, which is often present alongside other pathologies and is unlikely to be the primary cause of death. A negative association with age, in contrast, may be produced by a fast progressing (often early onset) pathology, with a higher burden of pathology being found in individuals with faster progression who consequently die at a younger age. This idea is supported by previous studies that have found inverse correlations between the primary pathology and age, such as between age and tau deposition in Alzheimer's disease.<sup>20,40,41</sup> An alternative explanation for an inverse correlation between disease stage and age is that more neuronal death in older (or more progressed) subjects leads to lower proportions of TDP-43 pathology being stained on histopathological slides. Some evidence for this phenomenon is presented by individuals with long-lasting ALS, who demonstrate decreased TDP-43 staining due to neuronal death of TDP-43-vulnerable neurons.<sup>42,43</sup> One final explanation for an inverse correlation with age is that younger subjects may have higher neurocognitive reserve and are therefore able to function under a greater burden of pathology.<sup>44</sup>

We identified a single progression pattern for LATE-NC, suggesting that the pattern of TDP-43 progression in LATE-NC is relatively stereotypical, or at least spread around a single major mode. We did not find evidence in the present study that the progression of LATE-NC differed between individuals with or without Alzheimer's disease, although our sample was dominated by individuals with Alzheimer's disease and larger studies will be required to confirm this. A recent study by Cykowski et al.<sup>45</sup> identified subtypes of LATE-NC with distinct TDP-43 inclusion morphologies. Our study does not necessarily contradict this study as we did not consider the morphology of the TDP-43 inclusions when deriving subtypes and stages. It is also worth noting that we did not consider very early 'sparse' pathology, whereas early pathology was found to be more heterogeneous in the study by Cykowski et al.<sup>45</sup>

The regions affected by LATE-NC and FTLD-TDP are known to overlap and the similarity of their TDP progression patterns has not yet been fully determined, although some studies suggest that they can be differentiated.<sup>16</sup> Whilst SuStain could differentiate LATE-NC and FTLD-TDP with high accuracy, it is important to note that there was a systematic under-sampling of individuals with early stages of FTLD-TDP and late stages of LATE-NC in our dataset. We did observe subtle differences for the few individuals at overlapping middle stages of FTLD-TDP or LATE-NC. However, we also observed a clear decrease in the confidence of classifying these individuals, suggesting that these subtle differences may not be strong enough to differentiate one condition from the other at an individual level. Similarly, there was considerable uncertainty in the early stages of the FTLD-TDP progression pattern and the late stages of the LATE-NC progression pattern. There is some controversy as to whether FTLD-TDP and LATE-NC share similar origins, where FTLD-TDP may be an earlier-onset and far more aggressive variant of LATE-NC. Based on our observations, we can neither confirm nor reject this hypothesis, especially since early FTLD-TDP is undersampled in our cohort. Closer study of TDP-43 morphology in LATE-NC and FTLD-TDP may help to resolve this controversy.

Our study has a number of limitations that warrant further consideration in future work. The neuropathological sample was obtained from a tertiary academic centre and may have a referral bias. FTLD-TDP is also known to be highly heterogeneous, potentially with distinct progression patterns in carriers of genetic mutations for FTLD-TDP, however our study was underpowered to resolve mutation-specific patterns. The SuStain algorithm infers subtypes with distinct progression patterns from cross-sectional data, relying on the assumption that TDP-43 proteinopathies progress sequentially from region to region, as is the case in all neuropathological studies. The SuStain algorithm also assumes that each individual has a single TDP-43 proteinopathy (FTLD-TDP, ALS or LATE-NC), an assumption violated by several individuals in this study, leading to misclassification. Our study determined progression patterns using neuropathological samples from a large number of regions. Whilst ideally suited to determining the progression pattern of each TDP-43 proteinopathy, this means that the derived staging systems may require sampling of a larger range of regions than is necessary to differentiate one proteinopathy from another. There are also several regions that were not sampled in the present study that merit further investigation in future studies. Our study only considered the total amount of TDP-43 pathology in each region and did not consider the specific TDP-43 inclusion morphologies, which may provide further stratification if considered in future work, e.g. TDP type may enhance the differentiation of FTLD-TDP and LATE-NC<sup>46</sup> but was only available for a subset of individuals in our study. In this study, we focused on comparing SuStain to staging schema of predefined neuropathological groups, running SuStain separately in each group. While consistent with current neuropathological practice, this may bias our results towards only reproducing these preconceived diagnostic categories. For example, diagnosis of LATE-NC in this study required amygdalar TDP-43 pathology, which may bias the amygdala as the earliest region to show TDP-43 burden. A similar issue may be present with the spinal cord and ALS, though other criteria are used in the neuropathological definition of ALS as well.

Overall, we develop an empirical pathological staging system for ALS, FTLD-TDP and LATE-NC, using data-driven disease progression modelling. We demonstrate that this staging system is sufficient for staging and accurate classification and can provide

probabilistic information that aids classification of individuals that deviate from the stereotypical progression pattern for each TDP-43 proteinopathy. Our results demonstrate that there is substantial heterogeneity amongst ALS and FTLD-TDP progression patterns, whilst we identify that LATE-NC is relatively homogeneous in this cohort. Data-driven disease progression modelling is a useful tool for neuropathological staging and management and can aid in unravelling neuropathological heterogeneity.

## Acknowledgements

We would like to acknowledge the late John Q. Trojanowski, MD, PhD, who also participated in study design and interpretation of results. We thank Manuela Neumann and Elisabeth Kremmer for providing the phosphorylation-specific TDP-43 antibody 1D3.

## Funding

A.L.Y. is supported by a Skills Development Fellowship from the Medical Research Council (MRC) [MR/T027800/1]. J.W.V. acknowledges funding from the National Institutes of Health (NIH) (T32MH019112) and the SciLifeLab & Wallenberg Data Driven Life Science Program (grant: KAW 2020.0239). R.O. has received research funding from European Research Council, ZonMw, NWO, National Institute of Health, Alzheimer Association, Alzheimer Nederland, Stichting Dioraphte, Cure Alzheimer's fund, Health Holland, ERA PerMed, Alzheimerfonden, Hjärtfonden. D.J.I. was supported by NIH grant R01-NS109260. This work was supported by the Penn Udall Center of Excellence in Parkinson's Disease Research, the Penn Center for Neurodegenerative Disease Research, and the Department of Pathology and Laboratory Medicine of the Penn Alzheimer's Disease Research Center. The current work and other research at these centres were supported by the NIH (P30 AG072979, P01AG066597, U19AG062418). O.H. was supported by the Swedish Research Council (2016-00906), the Knut and Alice Wallenberg foundation (2017-0383), the Marianne and Marcus Wallenberg foundation (2015.0125), the Strategic Research Area MultiPark (Multidisciplinary Research in Parkinson's disease) at Lund University, the Swedish Alzheimer Foundation (AF-939932), the Swedish Brain Foundation (FO2021-0293), The Parkinson foundation of Sweden (1280/20), the Cure Alzheimer's fund, the Konung Gustaf V:s och Drottning Victorias Frimurarestiftelse, the Skåne University Hospital Foundation (2020-O000028), Regionalt Forskningsstöd (2020-0314) and the Swedish federal government under the ALF agreement (2018-Projekt0279).

## Supplementary material

Supplementary material is available at *Brain* online.

## Competing interests

O.H. has acquired research support (for the institution) from ADx, AVID Radiopharmaceuticals, Biogen, Eli Lilly, Eisai, Fujirebio, GE Healthcare, Pfizer and Roche. In the past 2 years, he has received consultancy/speaker fees from AC Immune, Amylyx, Alzpath, BioArctic, Biogen, Cerveau, Fujirebio, Genentech, Novartis, Roche and Siemens. L.E. receives consulting fees from Biogen, PTC Therapeutics, Apellis Pharmaceuticals and Edgewise Therapeutics. D.W. has served as a paid consultant to Eli Lilly, GE

Healthcare and Qynapse. He serves on a DSMB for Functional Neuromodulation. He is a site investigator for a clinical trial sponsored by Biogen. R.O. has received research support from Avid Radiopharmaceuticals, Janssen Research & Development, Roche, Quanterix and Optina Diagnostics, and has given lectures in symposia sponsored by GE Healthcare. He is an editorial board member of *Alzheimer's Research & Therapy* and the *European Journal of Nuclear Medicine and Molecular Imaging*. All other authors report no competing interests.

## References

- Nelson PT, Dickson DW, Trojanowski JQ, et al. Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. *Brain*. 2019;142:1503-1527.
- Brettschneider J, Del Tredici K, Toledo JB, et al. Stages of pTDP-43 pathology in amyotrophic lateral sclerosis. *Ann Neurol*. 2013;74:20-38.
- Brettschneider J, Del Tredici K, Irwin DJ, et al. Sequential distribution of pTDP-43 pathology in behavioral variant frontotemporal dementia (bvFTD). *Acta Neuropathol*. 2014;127:423-439.
- Josephs KA, Murray ME, Whitwell JL, et al. Staging TDP-43 pathology in Alzheimer's disease. *Acta Neuropathol*. 2014;127:441-450.
- Nag S, Yu L, Wilson RS, Chen EY, Bennett DA, Schneider JA. TDP-43 pathology and memory impairment in elders without pathologic diagnoses of AD or FTLD. *Neurology*. 2017;88:653-660.
- Kim EJ, Vatsavayai S, Seeley WW. Neuropathology of dementia. In: Miller BL, Boeve BF, eds. *The behavioral neurology of dementia*. Cambridge University Press; 2016:94-122.
- Grad LI, Rouleau GA, Ravits J, Cashman NR. Clinical Spectrum of Amyotrophic Lateral Sclerosis (ALS). *Cold Spring Harb Perspect Med*. 2017;7:a024117.
- Warren JD, Rohrer JD, Rossor MN. Clinical review. Frontotemporal dementia. *BMJ*. 2013;347:f4827.
- Lee EB, Porta S, Michael Baer G, et al. Expansion of the classification of FTLD-TDP: distinct pathology associated with rapidly progressive frontotemporal degeneration. *Acta Neuropathol*. 2017;134:65-78.
- van Es MA, Hardiman O, Chio A, et al. Amyotrophic lateral sclerosis. *Lancet*. 2017;390:2084-2098.
- McMillan CT, Wu J, Rascofsky K, et al. Defining cognitive impairment in amyotrophic lateral sclerosis: an evaluation of empirical approaches. *Amyotroph Lateral Scler Frontotemporal Degener*. 2022;23:517-526.
- Strong MJ, Abrahams S, Goldstein LH, et al. Amyotrophic lateral sclerosis—frontotemporal spectrum disorder (ALS-FTSD): revised diagnostic criteria. *Amyotroph Lateral Scler Frontotemporal Degener*. 2017;18:153-174.
- DeJesus-Hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*. 2011;72:245-256.
- Renton AE, Majounie E, Waite A, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*. 2011;72:257-268.
- Josephs KA, Mackenzie I, Frosch MP, et al. LATE To the PART-y. *Brain*. 2019;142:e47.
- Robinson JL, Porta S, Garrett FG, et al. Limbic-predominant age-related TDP-43 encephalopathy differs from frontotemporal lobar degeneration. *Brain*. 2020;143:2844-2857.

17. Nana AL, Sidhu M, Gaus SE, et al. Neurons selectively targeted in frontotemporal dementia reveal early stage TDP-43 pathobiology. *Acta Neuropathol.* 2019;137:27-46.
18. Pascuzzo R, Oxtoby NP, Young AL, et al. Prion propagation estimated from brain diffusion MRI is subtype dependent in sporadic Creutzfeldt-Jakob disease. *Acta Neuropathol.* 2020;140:169-181.
19. Young AL, Marinescu RV, Oxtoby NP, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. *Nat Commun.* 2018;9:4273.
20. Vogel JW, Young AL, Oxtoby NP, et al. Four distinct trajectories of tau deposition identified in Alzheimer's disease. *Nat Med.* 2021;27:871-881.
21. Young AL, Bocchetta M, Russell LL, et al. Characterizing the clinical features and atrophy patterns of MAPT-related frontotemporal dementia with disease progression modeling. *Neurology.* 2021;97:e941-e952.
22. Eshaghi A, Young AL, Wijeratne PA, et al. Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. *Nat Commun.* 2021;12:2078.
23. Young AL, Vogel JW, Aksman LM, et al. Ordinal SuStaIn: subtype and stage inference for clinical scores, visual ratings, and other ordinal data. *Front Artif Intell.* 2021;4:613261.
24. Toledo JB, Van Deerlin VM, Lee EB, et al. A platform for discovery: the University of Pennsylvania integrated neurodegenerative disease biobank. *Alzheimers Dement.* 2014;10:477-484.e1.
25. Mackenzie IRA, Neumann M, Baborie A, et al. A harmonized classification system for FTLTDP pathology. *Acta Neuropathol.* 2011;122:111-113.
26. 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;123:1-11.
27. Aksman LM, Wijeratne PA, Oxtoby NP, et al. Pysustain: a python implementation of the subtype and stage inference algorithm. *SoftwareX.* 2021;16:100811.
28. Chiò A, Calvo A, Moglia C, Mazzini L, Mora G, PARALS study group. Phenotypic heterogeneity of amyotrophic lateral sclerosis: a population based study. *J Neurol Neurosurg Psychiatry.* 2011;82:740-746.
29. Ravits JM, La Spada AR. ALS Motor phenotype heterogeneity, focality, and spread: Deconstructing motor neuron degeneration. *Neurology.* 2009;73:805-811.
30. Sabatelli M, Conte A, Zollino M. Clinical and genetic heterogeneity of amyotrophic lateral sclerosis. *Clin Genet.* 2013;83:408-416.
31. Irwin DJ, Cairns NJ, Grossman M, et al. Frontotemporal lobar degeneration: defining phenotypic diversity through personalized medicine. *Acta Neuropathol.* 2015;129:469-491.
32. Seelaar H, Rohrer JD, Pijnenburg YAL, Fox NC, van Swieten JC. Clinical, genetic and pathological heterogeneity of frontotemporal dementia: a review. *J Neurol Neurosurg Psychiatry.* 2011;82:476-486.
33. Nishihira Y, Tan CF, Onodera O, et al. Sporadic amyotrophic lateral sclerosis: two pathological patterns shown by analysis of distribution of TDP-43-immunoreactive neuronal and glial cytoplasmic inclusions. *Acta Neuropathol.* 2008;116:169-182.
34. Takeuchi R, Tada M, Shiga A, et al. Heterogeneity of cerebral TDP-43 pathology in sporadic amyotrophic lateral sclerosis: evidence for clinico-pathologic subtypes. *Acta Neuropathol Commun.* 2016;4:61.
35. Mackenzie IR, Neumann M. Subcortical TDP-43 pathology patterns validate cortical FTLTDP subtypes and demonstrate unique aspects of C9orf72 mutation cases. *Acta Neuropathol.* 2020;139:83-98.
36. Josephs KA, Stroh A, Dugger B, Dickson DW. Evaluation of subcortical pathology and clinical correlations in FTLTDP subtypes. *Acta Neuropathol.* 2009;118:349-358.
37. Rohrer JD, Geser F, Zhou J, et al. TDP-43 subtypes are associated with distinct atrophy patterns in frontotemporal dementia. *Neurology.* 2010;75:2204-2211.
38. Sampathu DM, Neumann M, Kwong LK, et al. Pathological heterogeneity of frontotemporal lobar degeneration with ubiquitin-positive inclusions delineated by ubiquitin immunohistochemistry and novel monoclonal antibodies. *Am J Pathol.* 2006;169:1343-1352.
39. Mackenzie IRA, Baborie A, Pickering-Brown S, et al. Heterogeneity of ubiquitin pathology in frontotemporal lobar degeneration: classification and relation to clinical phenotype. *Acta Neuropathol.* 2006;112:539-549.
40. Jack CR, Wiste HJ, Schwarz CG, et al. Longitudinal tau PET in ageing and Alzheimer's disease. *Brain.* 2018;141:1517-1528.
41. Marshall GA, Fairbanks LA, Tekin S, Vinters HV, Cummings JL. Early-onset Alzheimer's disease is associated with greater pathologic burden. *J Geriatr Psychiatry Neurol.* 2007;20:29-33.
42. Nishihira Y, Tan CF, Hoshi Y, et al. Sporadic amyotrophic lateral sclerosis of long duration is associated with relatively mild TDP-43 pathology. *Acta Neuropathol.* 2009;117:45-53.
43. Spencer KR, Foster ZW, Rauf NA, et al. Neuropathological profile of long-duration amyotrophic lateral sclerosis in military veterans. *Brain Pathol.* 2020;30:1028-1040.
44. Stern Y. Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol.* 2012;11:1006-1012.
45. Cykowski MD, Arumanayagam AS, Powell SZ, et al. Patterns of amygdala region pathology in LATE-NC: subtypes that differ with regard to TDP-43 histopathology, genetic risk factors, and comorbid pathologies. *Acta Neuropathol.* 2022;143:531-545.
46. Nelson PT, Lee EB, Cykowski MD, et al. LATE-NC staging in routine neuropathologic diagnosis: an update. *Acta Neuropathol.* 2023;145:159-173.