Random forest modelling of neuropathological features identifies microglial activation as an accurate pathological classifier of C9orf72-related amyotrophic lateral sclerosis

Olivia M. Rifai[1-5], James Longden[2], Judi O'Shaughnessy[2-4], Michael D.E. Sewell[1,3], Karina McDade[2-4], Michael J.D. Daniels[3], Sharon Abrahams[4,6], Siddharthan Chandran[2-4], Barry McColl[3], Christopher R. Sibley[4,5,7,8], Jenna M. Gregory[2-4]

[1]Translational Neuroscience PhD Programme, Centre for Clinical Brain Sciences, University of Edinburgh, 49 Little France Crescent, Edinburgh EH16 4SB

[2]Centre for Clinical Brain Sciences, University of Edinburgh, 49 Little France Crescent, Edinburgh EH16 4SB

[3]UK Dementia Research Institute, University of Edinburgh, 49 Little France Crescent, Edinburgh EH16 4SB

[4]Euan MacDonald Centre for Motor Neurone Disease Research, University of Edinburgh, 49 Little France Crescent, Edinburgh EH16 4SB

[5]Centre for Discovery Brain Sciences, University of Edinburgh, 15 George Square, Edinburgh EH8 9XD

[6]Human Cognitive Neuroscience-Psychology, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ

[7]Institute of Quantitative Biology, Biochemistry and Biotechnology, School of Biological Sciences, University of Edinburgh, The King's Buildings, Edinburgh EH9 3FF

[8]Simons Initiative for the Developing Brain, University of Edinburgh, 15 George Square, Edinburgh EH8 9XD

Corresponding authors:

Olivia M. Rifai

Email: olivia.rifai@ed.ac.uk

Jenna M. Gregory

Email: jenna.gregory@ed.ac.uk

Telephone: 0131 242 9494

Fax: No fax available

Acknowledgments

Abstract

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are regarded as two ends of a pathogenetic spectrum, termed ALS-frontotemporal spectrum disorder (ALS-FTSD). However, it is currently difficult to predict where on the spectrum an individual will lie, especially for patients with *C9orf72* hexanucleotide repeat expansions (HRE), a mutation associated with both ALS and FTD. It has been shown that both inflammation and protein misfolding influence aspects of ALS and ALS-FTSD disease pathogenesis, such as the manifestation or severity of motor or cognitive symptoms. Previous studies have highlighted markers which may influence *C9orf72*-associated disease presentation in a targeted fashion, though there has yet to be a systematic and quantitative assessment of common immunohistochemical markers to investigate the significance of these pathways in an unbiased manner. Here we report the first extensive digital pathological assessment with random forest modelling of pathological markers often used in neuropathology practice. This study profiles glial activation and protein misfolding in a cohort of deeply clinically profiled *post-mortem* tissue from patients with a *C9orf72* HRE, who either met the criteria for a diagnosis of ALS or ALS-FTSD. We show that microglial immunohistochemical staining features, both morphological and spatial, are the best independent classifiers of disease status and that clinicopathological associations exist between microglial activation status and cognitive dysfunction in ALS-FTSD patients with *C9orf72* HRE. Furthermore, we show that spatially resolved changes in FUS staining are also an accurate predictor of disease status, implying that liquid-liquid phase shift of this aggregation-prone RNA-binding protein may be important in ALS caused by a *C9orf72* HRE. Our findings provide further support to the hypothesis of dysfunctional immune regulation and proteostasis in the pathogenesis of *C9orf72* ALS and provide a framework for digital analysis of commonly used neuropathological stains as a tool to enrich our understanding of clinicopathological associations between cohorts.

Keywords

Introduction

Amyotrophic lateral sclerosis (ALS) is a clinically heterogeneous neurodegenerative disease, with substantial variation in prognosis between patients and disease duration ranging from months to decades[61]. The most common known monogenetic cause of ALS is a hexanucleotide (G_4C_2) repeat expansion (HRE) in the first intron of the *C9orf72* gene, accounting for approximately 25% of familial and 10% of sporadic cases (C9-ALS)[20, 36, 54]. *C9orf72* HREs are associated with diverse clinical presentations, ranging from ALS and ALS-frontotemporal spectrum disorder (ALS-FTSD)[51], to parkinsonism and psychosis[2, 9, 21, 40, 49, 63]. Clinical trials in people with ALS are often based on phenotypic outcome measures like survival, functional rating scores (FRS) such as the ALS-FRS, and cognitive scores such as the Edinburgh Cognitive and Behavioural ALS Screen (ECAS). Thus, the clinical heterogeneity seen in C9-ALS, even with the same underlying monogenetic cause, is a hugely complicating factor in identifying potential therapeutics. Understanding the molecular drivers of clinical heterogeneity will enable more targeted and personalised therapies to be explored and improve stratification for clinical trials aimed at reducing specific symptom burden.

Thus far, pathological examination of *post-mortem* tissue from over 200 cases with a *C9orf72* HRE exists in the literature, primarily in ALS and FTD cases[12–17]. This collection of research suggests that, despite differences in clinical presentation, *C9orf72* mutations are very similar pathologically, exhibiting unifying features such as a p62-positive granule cell layer[4, 25, 57], dipeptide repeat (DPR) protein deposits[6, 19, 35], and cortical aggregates of the RNA-binding protein (RBP) TDP-43[25, 40, 49, 57]. They are further differentiated from sporadic ALS pathology by TDP-43-negative, p62-positive staining[25]. Indeed, other RBPs are seen to misfold certain ALS subtypes, but whether they are involved in C9-ALS remains to be determined. Interestingly, it has recently been shown that the wildtype form the RBP fused in sarcoma (FUS) mislocalises in sporadic ALS *post-mortem tissue* and vasolisin-containing protein (VCP)-related ALS induced pluripotent stem cell-derived motor neurons; previously this was only documented a property of only the ALS-associated mutant form [59]. Wildtype FUS dysregulation in a *post-mortem* C9-ALS context has yet to be observed, though research demonstrating the association of C9orf72 protein with FUS in stress granules, as well as the negative effects of *C9orf72*-related DPRs on the phase transition regulation of RBPs such as FUS and TDP-43 suggests it is possible[14, 41]. Few studies have shown markers that may be associated with the clinical heterogeneity observed. For example, greater TDP-43 burden in the spinal cord has been previously associated with a shorter disease duration[13]. Additionally,

we have previously shown that cognitive impairment is linked with the presence of TDP-43 pathology in extra-motor brain regions, though a subgroup of cases was shown to also have extra-motor TDP-43 pathology despite being cognitively unaffected. Thus, extra-motor TDP-43 pathology is a specific, but not sensitive, marker of cognitive impairment in ALS[26]. While informative, these findings do not account for all clinical heterogeneity observed in C9-ALS. Further investigation into sensitive and specific markers of C9-ALS and associated symptoms using clinically annotated tissue is thus warranted.

Another pathway of interest related to *C9orf72,* and which may influence phenotypic heterogeneity, is immune signalling. Studies examining both *C9orf72* knockdown and haploinsufficiency resulting from HREs have demonstrated immune dysregulation [5, 56]. Additionally, it has recently been suggested that differential inflammatory profiles may influence the manifestation or severity of motor and cognitive symptoms across the spectrum[23]. Indeed, a recent study observed differences in neuroinflammatory profiles in the cerebrospinal fluid of participants with ALS, FTD, or ALS-FTD, though each group possessed similar mutations[45]. Furthermore, levels of inflammatory factors, such as dysfunctional regulatory T cells and levels of interleukins (ILs) in plasma, as well as microglial activation in the cervical corticospinal tract, have been reported to correlate with disease progression[8, 10, 34]. Despite these associations, clinical trials exploring the use of anti-inflammatory drugs for ALS therapy have thus far been unsuccessful[47]. This may be attributed to non-specific mechanisms of action, such as a general blocking of inflammation, when a more nuanced, pathway-specific manipulation may be necessary. Additionally, the bias of preclinical studies toward *SOD1* models must also be considered. Indeed, *SOD1* mutations only account for 2% of ALS cases[53, 54]. As for FTD, there is little focus on anti-inflammatory therapies at present[58].

Previous studies have highlighted some markers which may influence *C9orf72*-associated disease, though there has yet to be a systematic and quantitative assessment of immunohistochemical markers to investigate the significance of these pathways in an unbiased manner. Here we present the first extensive digital pathological assessment of morphological and pathological markers commonly used in neuropathology practice to profile glia-related inflammation and protein misfolding in a cohort of deeply clinically profiled *post-mortem* tissue from patients with a *C9orf72* HRE, who either met the criteria for a diagnosis of ALS or ALS-FTSD. Furthermore, we used machine learning to train a random forest disease classifier using tens of millions of features derived from our digital analysis, to identify relevant markers of C9-ALS status.

Methods

Case identification and cognitive profiling

Tissue from 10 ALS *post-mortem* cases with a *C9orf72* HRE was obtained from the Medical Research Council (MRC) Edinburgh Brain Bank (Table 1). All cases had corresponding whole genome sequencing and pathogenic *C9orf72* HRE were confirmed by repeat-primed polymerase chain reaction (PCR). Tissue from 10 age- and sex-matched control cases was obtained from the Edinburgh Sudden Death Brain Bank. Control cases had no history of neurological conditions or evidence of neurodegenerative disease pathology. All *post-mortem* tissue was collected with ethics approval from East of Scotland Research Ethics Service (16/ES/0084) in line with the Human Tissue (Scotland) Act (2006). Use of *post-mortem* tissue for studies was reviewed and approved by the Edinburgh Brain Bank ethics committee and the Academic and Clinical Central Office for Research and Development (ACCORD) medical research ethics committee (AMREC). Clinical data were collected as part of the Scottish Motor Neurone Disease Register (SMNDR) and Care Audit Research and Evaluation for Motor Neurone Disease (CARE-MND)[33] platform, with ethics approval from Scotland A Research Ethics Committee (10/MRE00/78 and 15/SS/0216). Donor patients underwent neuropsychological testing with the Edinburgh Cognitive and Behavioural ALS Screen (ECAS)[1] and all patients consented to the use of their data. Clinical correlates of motor dysfunction/disease progression include disease duration (months) and sequential ALS functional rating score (ALSFRS) data points. Clinical correlates of cognition include Edinburgh Cognitive and Behavioural ALS Screen (ECAS) scores for ALS-specific and ALS non-specific subdomain scores.

Immunohistochemistry (IHC)

Brain tissue was taken *post-mortem* from standardised Brodmann areas (BA) BA4, BA39, BA44, BA46 and fixed in 10% formalin for a minimum of 72 h. Tissue was dehydrated in an ascending alcohol series (70–100%), followed by three 4 h washes in xylene. Three successive 5 h paraffin wax embedding stages were performed followed by cooling and sectioning of formalin-fixed, paraffin-embedded (FFPE) tissue on a Leica microtome into 4 μm thick serial sections that were collected on Superfrost (ThermoFisher Scientific) microscope slides. Sections were dried overnight at 40°C before immunostaining. Sections were dewaxed using successive xylene washes, followed by alcohol hydration and treatment with picric acid for quenching of lipofuscin. For pTDP-43, CD68, Iba1 and FUS staining, antigen retrieval was carried out in citric acid buffer (pH 6) in a pressure cooker

for 30 min, after which immunostaining was performed using the Novolink Polymer detection system with a 2B Scientific (Oxfordshire, UK) anti-phospho(409–410)-TDP-43 antibody at a 1 in 4000 dilution, Agilent (Cheadle, UK) anti-CD68 antibody at a 1 in 100 dilution, Abcam (Cambridge, UK) anti-Iba1 antibody at a 1 in 3000 dilution, and Bio-Techne (Minneapolis, USA) anti-FUS antibody at a 1 in 500 dilution. For GFAP staining, no antigen retrieval step was carried out, and immunostaining was performed using the Novolink Polymer detection system with a Dako anti-GFAP antibody at a 1 in 800 dilution. Counterstaining was performed using 3,3'-diaminobenzidine (DAB) chromogen counterstained with haematoxylin, according to standard operating procedures.

Imaging and digital pathology analysis

Whole tissue sections were scanned with Brightfield at 40x magnification using a Hamamtsu NanoZoomer XR. Using NDP.view2 viewing software (Hamamatsu), regions of interest (ROIs) were taken from key regions for quantification. For both grey and white matter, 3 vascular-adjacent and 3 non-vascular-adjacent regions of interest (ROIs) (500 μm x 500 μm) were taken per tissue section for each stain. ROIs were defined as vascular-adjacent when centered on a vessel larger than a small capillary, with larger, cross-sectional vessels captured where possible. ROIs were considered nonvascular adjacent if they were at least ~250 μm away from any vessel larger than a small capillary (Figure 1a, Figure S1a). ROIs were analysed using QuPath software[7] superpixel segmentation in a manner previously described[39], with the modification of using difference of gaussians (DoG), rather than simple linear iterative clustering (SLIC) segmentation, to obtain shape features. For vascular-adjacent ROIs, annotations were drawn manually around the vessel (endothelial cell border) and around the area of tissue retraction if present; vessel annotations were subtracted from the image for subsequent automated analyses. Images were segmented into superpixels with stain-specific segmentation parameters selected to best identify the staining of interest (downsample = 1, gaussian sigma = 8.0 for Iba1 and GFAP, gaussian sigma = 5.0 for CD68, FUS and TDP-43). Superpixels were classified using three thresholds for DAB Mean Intensity, according to the pathological precedence of Allred scoring of oestrogen receptor immunohistochemistry in breast cancer[22], and intensity and shape features were calculated for each superpixel. For vascular-adjacent ROIs, distance to vessel (normalised to the square root of the non-vessel-containing area of the image to account for differences in vessel calibre) and distance to superpixels by threshold class were also calculated. Measurements were exported at the image (number of positive detections per threshold class) and superpixel level (threshold class, intensity, shape, and distance features).

Nuclear and cytoplasmic FUS localisation were quantified using cell segmentation in QuPath. For neuronal quantification, only grey matter ROIs were included, cells were segmented (pixel size = 0.5 µm, background radius = 0, median filter radius = 0, sigma = 4, minimum nuclear area = 150 µm^2, maximum nuclear area = 5000 µm^2, intensity threshold = 0.1, max background intensity = 0, cell expansion = 10 µm) and nuclear, cytoplasmic, and whole cell intensity and shape features were calculated and exported. For glial quantification, both grey and white matter ROIs were included, cells were segmented (pixel size = 0.5 µm, background radius = 0, median filter radius = 0, sigma = 4, minimum nuclear area = 10 µm^2, maximum nuclear area = 100 µm^2, intensity threshold = 0.1, max background intensity = 0, cell expansion = 5 µm) and nuclear, cytoplasmic, and whole cell intensity and shape features were calculated and exported. Scripts used in these analyses can be found in the Supplementary Information. Manual TDP-43 grading was performed as described previously[26] and was visualised using GraphPad Prism (version 9.2.0).

Data were visualised in RStudio (R version 4.1.1) with "ggplot2" (version 3.3.5) and "Ggally" (version 2.2.1) packages. Datasets that were found to be normally distributed via Shapiro-Wilk's test were subjected to parametric tests; unpaired t-test for two-group comparisons, and ANOVA with Tukey's multiple comparisons correction for three-group comparisons. Datasets that were found to be non-normal via Shapiro-Wilk's test were subsequently subjected to non-parametric tests; Mann-Whitney U for two-group comparisons, pairwise Wilcoxon with Holm-Sidak multiple comparisons correction for three-group comparisons and Spearman's test for correlations. Statistical comparisons were only conducted between groups with n ≥ 3. Results were averaged over 3 ROIs capturing the same region and presented as ungrouped or grouped by brain region, grey or white matter, and vascular or non-vascular adjacent. For box plots, data were further averaged when certain stratifications were not included to avoid pseudoreplication (e.g., when comparisons were made between grey and white matter within a brain region without grey matter vascular-adjacent and grey matter non-vascular-adjacent subregions, vascular-adjacent and non-vascular adjacent results were averaged for grey matter and white matter). For correlations, all stratified data were included even when data were ungrouped to explore relationships between two stains across all images. Multiple linear regression models were fitted where significant correlations were found, and coefficients were compared by t-tests on individual regression coefficients.

Machine learning

The automated analysis of IHC images in QuPath generated approximately 80.2 million superpixel features and 5.2 million cell segmentation features, 56.4 million of which (66%) were used to train random forest classifiers using disease versus control datasets. Classifiers were trained with 1000 trees per classifier and Information Gain as the split criterion. Classifiers were then tested on the remaining 33% of the dataset. Cross-tables were generated for regions and X^2 tests were performed to determine which regions allowed the random forest to correctly predicted disease significantly better than chance. Sensitivity (100*true positives/total disease) and specificity (100*true negatives/total control) scores were also calculated (Supplementary Table 1). Finally, leave-one-out analyses were performed to determine the predictive power of each feature (Supplementary Table 2).

RT-qPCR

Reverse transcription quantitative polymerase chain reaction (RT-qPCR) was conducted by extracting RNA from human FFPE using the RNAstorm FFPE RNA extraction kit (Cell Data Sciences) on two 10 µm curls per sample cut from BA4. RNA was eluted in 50 µL nuclease-free water, after which sample concentrations were measured using a NanoDrop 1000 (ThermoFisher) and concentrated further for 10-15 minutes at 45°C using an Eppendorf Concentrator Plus if needed. 125 ng of RNA was reverse transcribed using Superscript IV (Life Technologies) as per the manufacturer's instructions, after which 10 ng cDNA was amplified with PowerUp SYBR Green Master Mix (ThermoFisher) and forward and reverse primers to *FUS* and housekeeping 18S ribosomal RNA gene *18Sr* (0.5 µM), selected to control for age- and degradation-related artefact. FUS primers were specifically designed to target two non-contiguous central portions of the sequence so as to better detect product that may be partially degraded due to the nature of FFPE tissue preparation and storage[3]. Primer sequences are listed below, with alternate *FUS* primer sequences that failed to amplify cDNA listed in Supplementary Information.

18Sr Forward: 5'-GGCCCTGTAATTGGAATGAGTC-3'

18Sr Reverse: 5'-CCAAGATCCAACTACGAGCTT-3'

FUS Forward: 5'-ACTTCCTTTCTTTTGCTCTCAC-3'

FUS Reverse: 5'-ATTAAGCCCCACTAAGCCC-3'

Quantitative PCR was run on eight control and eight C9-ALS samples in 10 ng triplicates using qTOWER^3 84 (Analytik Jena GMBH) (50°C for 2 min, 95°C for 2 min, 40 x 95°C for 5s, 60°C for 30 s). One control was removed due to off-target amplification as determined by a melting curve. Cycle threshold (Ct) values for *FUS* were normalised (ΔCt) to *18Sr* Ct values, and fold changes ($2^{-\Delta\Delta Ct}$) were calculated relative to the average ΔCt of control samples. Six controls and seven cases that did not exhibit off-target amplification and were consistent among triplicates were included in the final analysis. Conditions were visually inspected for homogeneity of variance and an unpaired t-test was performed to test significance between conditions.

Results

Microglial immunohistochemical staining features are the best independent classifiers of C9-ALS

To investigate pathological differences between C9-ALS and controls, *post-mortem* tissue from a deeply clinically phenotyped cohort (Table 1) was stained with antibodies commonly used in the neuropathological work-up of ALS cases in all brain banks. Specifically, we examined antibody immunoreactivity to microglia-enriched marker Iba1, microglia/macrophage lysosomal activation marker CD68, astrocyte activation marker GFAP, and RNA-binding proteins FUS and TDP-43, which are known to mislocalise and aggregate in ALS[44, 59]. Threshold-based stain intensities, as well as morphology and spatial features, were calculated using superpixel and cell segmentation methods in QuPath (Figure 1a-b) for the ALS-vulnerable motor cortex (BA4) and for three vulnerable cognitive regions related to the domains assessed in the ECAS[1]; BA39 (language), BA44 (fluency) and BA46 (executive). These were further stratified by non-vascular-adjacent and vascular-adjacent regions in grey and white matter. Collectively this equated to 4,800 images for digital analysis, with 240 images per case. To determine the relationship between the total number of stain-positive superpixels between stains and identify differences between C9-ALS and control cases, correlation analyses were performed. Significant positive correlations were found between the number of Iba1+ and CD68+ superpixels in disease (R = 0.427 p < 0.001) and controls (R = 0.453, p < 0.001) (Figure 1c). Additionally, significant positive correlations were found between the number of CD68+ and pTDP-43+ superpixels in disease (R = 0.360, p < 0.001) and controls (R = 0.250, p < 0.01) (Figure 1c). The number of GFAP+ superpixels was found to positively correlate with CD68+ superpixels in disease (R = 0.466, p < 0.001) and controls (R = 0.457, p < 0.001), demonstrating consistency between glial

activation markers (Figure 1c). Notably, this did not simply represent an increase in cell number; see digital cell recognition analysis (Figure S1b-c). Finally, total GFAP+ superpixels were also found to positively correlate with pTDP-43+ superpixels in C9-ALS (R = 0.224, p < 0.01) and controls (R = 0.216, p < 0.01), suggesting a relationship between astrogliosis and TDP-43 aggregation (Figure 1c). When visually inspected, the data spread was most distinct between C9-ALS and controls for Iba1+ versus CD68+ and CD68+ versus pTDP43+ correlations, the details of which are further characterised in Figure 3.

Classification potential of immunohistochemical staining features was interrogated using a random forest machine learning model to determine whether the model could be trained to predict C9-ALS versus control status based on phenotypic tissue features. Overall, digital analysis of immunohistochemical staining conducted in this way was found to be an accurate disease classifier with 69% sensitivity (X^2 p < 0.0001) and 66% specificity (X^2 p < 0.0001) overall (Figure 1d; Supplementary Table 1). Interestingly, all regions were independently predictive of C9-ALS status, irrespective of whether the regions were clinically involved (i.e., motor versus cognitive involvement (Supplementary Table 1). When broken down by stain, CD68, FUS and Iba1 stains were both sensitive (i.e., 67%, 64% and 78%, X^2 p < 0.0001, p < 0.001, p < 0.0001, respectively) and specific (i.e., 60%, 69% and 84%, X^2 p < 0.05, p < 0.0001, p < 0.0001, respectively) when predicting disease status, while GFAP staining was sensitive (79%, X^2 p < 0.0001) but not specific (53%, X^2 p = 0.527), and pTDP-43 staining was specific (64%, X^2 p < 0.001) but not sensitive (51%, X^2 p = 0.8744) to disease status, consistent with previous findings (Supplementary Table 1) [27]. This meant that GFAP and pTDP-43 staining yielded more false positive and false negative results, respectively. This may imply that the astrocyte activation marker used in this study is not an adequate predictor of C9-ALS status, though this does not rule out the potential usefulness of other astrocyte-related markers that were not tested, such as ALDH1L1 and AQP4. FUS and CD68 staining were accurate classifiers of disease overall, though this was not maintained within subregions. Iba1 was the most predictive stain, with accurate classification in all four brain regions, in most grey and white matter regions within those and even at the vascular level in BA39 non-vascular-adjacent grey and white matter. The predictive power of microglia-related stains underscores disease-related changes to microglia in C9-ALS and reveals microglia-related stains to be sensitive and specific classifiers of disease status. Finally, almost all included morphological and spatial features had some predictive power for classifying disease. Notably, the features with the highest predictive power for disease included neuronal and glial Haematoxylin and DAB parameters from FUS cell

segmentation, distance of negative superpixels to vessel, percentage of class 1+ superpixel detections, and solidity and circularity of class 2+ superpixel detections (Supplementary Table 2).

Interestingly, FUS localisation features from our cell segmentation analysis appeared to have more predictive power of disease status than superpixel count features (Supplementary Table 2) which may also act as a proxy measure for neuronal loss, as FUS is a predominantly nuclear protein. However, no significant reductions in the number of neurons, glia, or superpixels detected in FUS analyses were found between C9-ALS and controls (Figure S1b-d); in fact, a significant increase in glia in BA44 white matter and superpixels detected in BA44 and BA46 white matter were detected in C9-ALS (Figure S1b-d). Given previous reports of FUS mislocalisation and nuclear and cytoplasmic condensates in sporadic ALS[59], and the involvement of C9orf72 protein in FUS-related stress granule dynamics[14], we next sought to assess for digitally apparent differences in localisation and aggregation of FUS in our cohort. Further investigation of nuclear and cytoplasmic FUS staining differences revealed significant increases in nuclear/cytoplasmic mean intensity ratio in BA4 and BA46 glia, but not in neurons (Figure 2a-c). We next interrogated whether this change in intensity was associated with increased FUS expression, or rather nuclear FUS aggregation, using reverse transcription quantitative polymerase chain reaction (RT-qPCR). *FUS* primers were specifically designed to target a central region of the sequence so as to better detect product that may be partially degraded due to the nature of FFPE tissue preparation and storage[3], and revealed no significant difference in *FUS* expression between C9-ALS and controls (Figure 2d). Further support to the notion that this effect could be related to failed degradation and aggregation of FUS, rather than expression differences, is that cytoplasmic condensates in both neurons and glia are readily identifiable in these cases (Figure 2e).

C9-ALS post-mortem tissue exhibits microgliosis-related changes

As features of the microglia-related stains Iba1 and CD68 were the most accurate predictors of C9-ALS status, and distribution of total positive superpixels for these stains between C9-ALS and controls appeared to be the most distinct, we further explored these stains and their relationships at a regional level. Investigation of the aforementioned microglia-related correlations with multiple linear regression revealed that y-intercepts (disease: y = 10.5437, control: y = 1.9610), but not slopes (disease: a = 0.3733, control: a = 0.4162), of the linear models of Iba1- versus CD68-positive superpixels were significantly different between C9-ALS and control cases (p < 0.001), indicating elevated microgliosis in disease (Figure 3a). This general pattern is consistent across most

regions, and although power is reduced when the data are stratified, significant correlations persist in BA4 white matter regions (Figure 3b). For the relationship between CD68- and pTDP-43-positive superpixel number, multiple linear regression revealed that both slopes (disease: a = 0.0528, control: a = 0.001497; $p < 0.001$) and y-intercepts (disease: y = -0.24604, control: y = 0.065258; $p < 0.01$) of linear models were significantly different between C9-ALS and controls (Figure 3c), suggesting a relationship between microglial activation and TDP-43 aggregation in disease. A general relationship between CD68 and pTDP-43 is preserved particularly across grey matter regions, as well as BA4 and BA39 white matter (Figure 3d). However, the only significant correlation was found in BA4 vascular-adjacent grey matter, which may be due to the low number of pTDP-43+ detections combined with low power resulting from regional stratification.

Analysis of specific features of the strongest classifier of C9-ALS status, Iba1, revealed significant increases in Iba1-positive superpixel area in all included white matter regions as well as in extramotor grey matter (Figure 4a). Interestingly, there were also significant decreases in superpixel circularity of class 1+ and class 3+ Iba1 superpixels (digital mask demonstrated in Figure 1a) between disease and control cases, predominantly in white matter regions (Figure 4b). While this might be unexpected, as microglial activation morphology is more circular and ameboid compared to the ramified homeostatic morphology[55], it must be noted that these findings reveal changes to area and circularity of superpixels containing Iba1 stain, thus indicating disease-related morphological changes of a non-physiological nature and are not necessarily suggestive of activation status. Finally, significant decreases in the distance of class 1+ and class 2+ Iba1 superpixels from the central vessel were observed in C9-ALS, predominantly in BA4 grey matter (Figure 4c), which may reflect migration to vessels as a response to systemic inflammation or compromised blood-brain barrier integrity[29]. It is also possible that the combined result of morphological changes and increased vessel-adjacent Iba1+ staining is indicative of an increased presence of non-microglial, Iba1+ monocyte-derived cells that have migrated from vessels in disease (Figure 4c).

Clinicopathological associations exist between microglial activation status and cognitive dysfunction

Findings from immunohistochemistry data were next analysed for relationships with clinical features by comparing cognitively unimpaired C9-ALS cases to those with C9-ALS-FTSD. The amount of CD68+ and pTDP-43+ staining in relevant regions was assessed between language-impaired cases as determined by ECAS. Language-impaired ALS cases (n = 4) were found to have significantly higher CD68+ staining in BA39 grey matter compared to controls (n = 10) in both grey ($p < 0.05$) and white ($p < 0.01$) matter, while this was not the

case for language-unimpaired cases (n = 3) (Figure 5a-c). Both C9-ALS subgroups (language-impaired and language-unimpaired) had more CD68+ superpixels per Iba1+ superpixel number in BA39, suggestive of elevated microglial lysosomal activation in ALS cases as mentioned previously (Figure 3), with significant differences in y-intercepts of linear models (language-impaired: y = 7.482, p < 0.001; language-unimpaired: y. = 4.1073, p < 0.05) compared to controls (y = 3.878) (Figure 5c). Interestingly, the slope of the linear model for language-impaired cases was significantly increased compared to controls (impaired: a = 1.037, control: a = 0.172, p < 0.05), while this was not the case for language-unimpaired cases (a = 0.7689), potentially suggesting a more reactive CD68+ microglial phenotype in language-impaired cases (Figure 5c). Two cases exhibited executive impairment and on average displayed higher CD68+ staining in BA46 grey and white matter, though no statistical analysis was conducted due to low power (Figure S2a). The sole case that did exhibit fluency impairment (C9-ALS-FTD) displayed increased CD68+ staining in BA44 grey and white matter compared to controls and other cases, though, again, no statistical analysis was performed due to absence of replicate cases for verification (Figure S2b). Furthermore, there were no significant differences in pTDP43+ staining in BA39 detected between language-impaired, language-unimpaired and control conditions using the digital analysis method (Figure 5d). When investigating disease duration-related differences, no significant differences in CD68+ staining were found between long survivors (i.e., disease duration of over 48 months[61]), short survivors and controls (Figure 5e). Total pTDP-43+ staining was significantly increased in long survivors in BA4 grey matter (p < 0.05), which may indicate a build-up of aggregates over time (Figure 5f). Finally, no significant differences in total GFAP+ staining were detected in BA39 between language-impaired, language-unimpaired and control cases, or in BA4 between long survivors, short survivors and controls (Figure S2c-d).

Due to the nature of the automated image analysis and ROI selection, we postulated that milder and sparser TDP-43 pathology could be underrepresented in our analysis. To test this, TDP-43 aggregate burden was manually graded as described previously[26] (Table 2). As suspected, in BA4, which contained the highest TDP-43 burden across cases, significant differences between neuronal and glial TDP-43 burden in C9-ALS versus control cases were picked up both by the automated and manual methods, though neither method resulted in a significant difference between long and short survivors (Figure S3a). In BA39, which had milder TDP-43 burden overall, the manual method picked up on significant differences between disease and controls that were not detected using the automated method (Figure S3b). However, both methods did not reveal a significant difference between language-impaired and language-unimpaired cases. Thus, while our automated method may be susceptible to

underrepresentation of milder pTDP-43 pathology, both methods were similar in their ability to discern differences, or lack thereof, between pTDP-43 pathology in C9-ALS clinical phenotypes. When comparing the scale of differences between language-impaired and language-unimpaired cases for both manually graded TDP-43 burden and digitally quantified CD68+ staining, there appeared to be a more distinct difference in the latter (Figure S3c). This is likely due to the existence of "mismatch" cases, or ALS cases that exhibit TDP-43 burden in cognitive regions but did not experience related cognitive impairment[27]; while cognitive impairment may predict the presence of TDP-43 burden in the related cognitive region, this is not the case vice versa. Indeed, all language-unimpaired cases in the studied cohort exhibited pTDP-43 pathology to some extent in BA39, often with comparable scoring to language-impaired cases (Table 2).

Discussion and Future Work

Few studies have previously investigated markers of neuroinflammation and their relationship with pathological proteins in ALS *post-mortem* tissue[10, 17, 50, 52]. This study is the first to conduct immunohistochemical staining analysis at this scale with unbiased, high-throughput quantification at the image, cell and superpixel level, with diverse regional coverage in deeply clinically phenotyped *post-mortem* tissue. Here we outlined a potential standard for the quantitative assessment of important neuropathological elements in ALS, building upon previous methods analysing FUS- and splicing factor proline- and glutamine-rich (SFPQ)-related differences at the single-cell level in sporadic ALS spinal cord tissue[28], and that described for Alzheimer's disease[43, 60]. Additionally, this study allowed for *post-mortem* tissue staining data to be directly compared with both motor and cognitive clinical features. Importantly, our random forest model revealed that, from a standard panel of neuropathological markers, the microglia-related stains Iba1 and CD68 were more predictive of C9-ALS status compared to controls than pTDP-43 or GFAP staining. The relevance of microglia in C9-ALS may pertain to the proposed pathogenic mechanism of *C9orf72* HREs being haploinsufficiency. Microglia have previously been shown to strongly express and require C9orf72 protein for normal function, and may therefore be particularly susceptible to the negative effects of this haploinsufficiency[38, 46, 48]. Indeed, *C9orf72* -/- peripheral myeloid cells have been demonstrated to have hyperactive type I interferon responses, mediated by impaired lysosomal degradation of the stimulator of interferon genes (STING) protein[38], thus suggesting that loss of normal C9orf72 function may lead to a proinflammatory state, as we observed in our data. Similarly, mouse *C9orf72* -/- microglia have also been shown to have higher levels of STING protein and increased expression of interferon response genes as well as other

activation response genes[32]. Qualitative pathological assessment has shown signs of increased microglial activation in *C9orf72* HRE *post mortem* tissue[18], and microglial activation in the cervical corticospinal tract has been shown to correlate with disease progression[11]. Our study supports and builds upon these findings, highlighting morphological and spatial features of microglia that may be at play, and suggesting an influence of microglial activation in extra-motor regions on cognitive impairment in C9-ALS-FTSD.

There are a few possible reasons as to why pTDP-43 staining was not found to be predictive of disease status in this study. For instance, the way ROIs were selected may lead images from disease cases to appear free of pTDP-43. This was most evident when comparing automated and manual grading methods, and quantification of pTDP-43 in brain regions with less aggregate burden (i.e., extra-motor regions) (Figure S3). In line with this, it is recommended that future analyses of pTDP-43 staining combine manual grading of burden with intensity, morphological and spatial features afforded by digital analysis. Alternatively, as has previously been described regarding cognitive impairment[26], it could be that while affected cases all displayed TDP-43 pathology, some unaffected cases, termed "mismatch cases", could be resilient to TDP-43 burden. Indeed, while TDP-43 pathology has been associated with more rapid cognitive decline[62], it has also been observed in healthy ageing[24, 42]. While it may also be the case that the presence of pathology may precede clinical manifestations, such as in Alzheimer's disease[31], and that healthy control or "mismatch" cases could have had the potential to develop symptoms later on, this would still indicate that while TDP-43 aggregation may have a central role in disease pathogenesis, its presence or burden may not serve as well as a classifier of C9-ALS in a given time point as other markers such as microglial activation. It may be that different conformations of TDP-43 aggregates not discerned in this study exist that lead to differential behaviour of microglia, similarly to what has been observed with microglial infiltration of amyloid plaques in Alzheimer's disease[12, 15]. This could explain our findings that language-impaired, BA39 TDP43+ cases have significantly more CD68+ staining than controls, while language-unimpaired cases, which are also BA39 TDP43+, do not (Table 2; Figure 5a-b). Furthermore, there appears to be a difference in the Iba1+ CD68+ relationship between language-impaired and language-unimpaired cases (Figure 5c). Recent studies have shown the immunogenic capacity of TDP-43 and its ability to activate microglia specifically[64, 65], though further research is needed to determine whether certain aggregate subtypes induce a differential response in microglia.

Finally, our random forest model revealed FUS staining to be an accurate classifier of C9-ALS status, and we also found differences in the nuclear/cytoplasmic FUS ratio in glia between control and C9-ALS cases. FUS is an RNA-binding protein that is known to undergo liquid-liquid phase separation (LLPS) and associate with membraneless organelles, some of which are localised to the nucleus such as paraspeckles and nuclear gems[37]. It is also known to autoregulate and increase its own expression as a result of impaired nuclear import[30]. We did not find significant differences in *FUS* expression between control and disease using RT-qPCR, which suggests possible LLPS or aggregation rather than compensatory overexpression. It has been shown that functional C9orf72 protein is required for removal of aggregation-prone FUS protein[14], as well as that C9orf72-related DPRs have been shown to induce aggregation of TDP-43, which similarly to FUS has an intrinsically disordered C-terminal region[16]. Thus, C9orf72 haploinsufficiency due to mutation or the interaction of FUS with DPRs may result in its impaired degradation; our observations of cytoplasmic condensate and paraspeckle formation in C9-ALS in the absence of increased *FUS* expression further support this hypothesis. Whilst this is the first comprehensive, unbiased identification of FUS dysregulation in C9-ALS *post-mortem* tissue, several caveats should be considered. For example, future studies with a higher sample size and cell-sorting methods may be needed to investigate expression changes further, as any cell-type-specific effects may be masked during measurement of bulk RNA expression. Additionally, the cell segmentation algorithm employed in this study was less robust than the superpixel segmentation algorithm; to avoid frequent detection of nonexistent cells over all images, haematoxylin thresholds for the identification of nuclei were set relatively high, resulting in some cells being excluded from the analysis (Figure 1b). Thus, our cell segmentation findings may not be as representative of all cells in the ROIs compared to our superpixel segmentation findings. As such, the question of how FUS staining is predictive of C9-ALS status and subsequent cytotoxic mechanisms requires further investigation.

There are limitations that should be noted regarding all digital analyses. For example, pre-set staining thresholds in automated digital analysis circumvents the possibility of bias that arises with manual grading, though it is not as accommodating of any natural variation in staining. Additionally, this study utilised manually selected ROIs based on predetermined criteria rather than entire tissue scans, which were designed to equally represent different regions of interest as part of an exploratory analysis (i.e., grey and white matter, vascular-adjacent and non-vascular-adjacent regions), though as a result, the analysis risked being less sensitive to sparser pathology. Thus, it may be worth assessing larger regions, such as entire grey and white matter regions within each tissue section, although this would require more sophisticated cortical and cellular annotation algorithms than are currently

available. Additionally, while our cohort included deeply clinically phenotyped *post-mortem* tissue and was sufficiently powered to detect differences in staining with regional resolution, our sample size did not include multiple fluency-impaired cases, nor sufficient cases with executive impairment to conduct meaningful comparisons regarding these types of cognitive impairment. Finally, while the random forest model provides important information about which stains are accurate classifiers of C9-ALS, it is currently unclear whether the predictive features are specific to C9-ALS, or whether they are common to other ALS subtypes or to neurodegeneration in general. To determine this, immunohistochemistry features from other neurodegenerative diseases would need to be added to the random forest model.

These limitations notwithstanding, our study demonstrates the importance of a standardised digital pathological assessment of commonly used immunohistochemical markers in ALS and other conditions, the results of which can be shared and meta-analysed to provide a comprehensive and highly powered model revealing pathological predictors of disease and clinical phenotypes. The identification of these predictors may hold promise for peripheral biomarkers and positron emission tomography (PET) ligand creation in the future and contribute to the essential understanding of disease-related specific pathways for the development of targeted therapeutics. Together, our high-throughput digital pathology and random forest modelling approach enabled detailed mapping and analysis of tens of millions of features derived from common neuropathological stains to characterise neuroinflammation and protein dynamics across a clinically heterogeneous C9-ALS cohort. The generation of large datasets using this method allows for the investigation of numerous neuropathological questions, many of which are outside the scope of this study. For this reason, all the scripts and data featured in this study are included in Supplementary Materials.

References

1.   Abrahams S, Newton J, Niven E, Foley J, Bak TH (2014) Screening for cognition and behaviour changes in ALS. null 15:9–14. doi: 10.3109/21678421.2013.805784

2.   Abramzon YA, Fratta P, Traynor BJ, Chia R (2020) The Overlapping Genetics of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. Front Neurosci 14:42–42. doi: 10.3389/fnins.2020.00042

3.   von Ahlfen S, Missel A, Bendrat K, Schlumpberger M (2007) Determinants of RNA Quality from FFPE Samples. PLOS ONE 2:e1261. doi: 10.1371/journal.pone.0001261

4.   Al-Sarraj S, King A, Troakes C, Smith B, Maekawa S, Bodi I, Rogelj B, Al-Chalabi A, Hortobágyi T, Shaw CE (2011) p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLD and MND/ALS. Acta Neuropathologica 122:691–702. doi: 10.1007/s00401-011-0911-2

5.   Atanasio A, Decman V, White D, Ramos M, Ikiz B, Lee H-C, Siao C-J, Brydges S, LaRosa E, Bai Y, Fury W, Burfeind P, Zamfirova R, Warshaw G, Orengo J, Oyejide A, Fralish M, Auerbach W, Poueymirou W, Freudenberg J, Gong G, Zambrowicz B, Valenzuela D, Yancopoulos G, Murphy A, Thurston G, Lai K-MV (2016) C9orf72 ablation causes immune dysregulation characterized by leukocyte expansion, autoantibody production and glomerulonephropathy in mice. Scientific Reports 6:23204. doi: 10.1038/srep23204

6.   Baborie A, Griffiths TD, Jaros E, Perry R, McKeith IG, Burn DJ, Masuda-Suzukake M, Hasegawa M, Rollinson S, Pickering-Brown S, Robinson AC, Davidson YS, Mann DMA (2015) Accumulation of dipeptide repeat proteins predates that of TDP-43 in frontotemporal lobar degeneration associated with hexanucleotide repeat expansions in C9ORF72 gene. Neuropathol Appl Neurobiol 41:601–612. doi: 10.1111/nan.12178

7.   Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG, James JA, Salto-Tellez M, Hamilton PW (2017) QuPath: Open source software for digital pathology image analysis. Scientific Reports 7:16878. doi: 10.1038/s41598-017-17204-5

8.   Beers DR, Zhao W, Wang J, Zhang X, Wen S, Neal D, Thonhoff JR, Alsuliman AS, Shpall EJ, Rezvani K, Appel SH (2017) ALS patients' regulatory T lymphocytes are dysfunctional, and correlate with disease progression rate and severity. JCI Insight 2:e89530–e89530. doi: 10.1172/jci.insight.89530

9.   Bourinaris T, Houlden H (2018) C9orf72 and its Relevance in Parkinsonism and Movement Disorders: A Comprehensive Review of the Literature. Movement Disorders Clinical Practice 5:575–585. doi: 10.1002/mdc3.12677

10. Brettschneider J, Toledo JB, Van Deerlin VM, Elman L, McCluskey L, Lee VM-Y, Trojanowski JQ (2012) Microglial Activation Correlates with Disease Progression and Upper Motor Neuron Clinical Symptoms in Amyotrophic Lateral Sclerosis. PLOS ONE 7:e39216. doi: 10.1371/journal.pone.0039216

11. Brettschneider J, Toledo JB, Van Deerlin VM, Elman L, McCluskey L, Lee VM-Y, Trojanowski JQ (2012) Microglial activation correlates with disease progression and upper motor neuron clinical symptoms in amyotrophic lateral sclerosis. PLoS One 7:e39216–e39216. doi: 10.1371/journal.pone.0039216

12. Brown MR, Radford SE, Hewitt EW (2020) Modulation of β-Amyloid Fibril Formation in Alzheimer's Disease by Microglia and Infection. Frontiers in Molecular Neuroscience 13:228. doi: 10.3389/fnmol.2020.609073

13. Cathcart SJ, Appel SH, Peterson LE, Greene EP, Powell SZ, Arumanayagam AS, Rivera AL, Cykowski MD (2021) Fast Progression in Amyotrophic Lateral Sclerosis Is Associated With Greater TDP-43 Burden in Spinal Cord. Journal of Neuropathology & Experimental Neurology. doi: 10.1093/jnen/nlab061

14. Chitiprolu M, Jagow C, Tremblay V, Bondy-Chorney E, Paris G, Savard A, Palidwor G, Barry FA, Zinman L, Keith J, Rogaeva E, Robertson J, Lavallée-Adam M, Woulfe J, Couture J-F, Côté J, Gibbings D (2018) A complex of C9ORF72 and p62 uses arginine methylation to eliminate stress granules by autophagy. Nature Communications 9:2794. doi: 10.1038/s41467-018-05273-7

15. Condello C, Yuan P, Schain A, Grutzendler J (2015) Microglia constitute a barrier that prevents neurotoxic protofibrillar Aβ42 hotspots around plaques. Nature Communications 6:6176. doi: 10.1038/ncomms7176

16. Cook CN, Wu Y, Odeh HM, Gendron TF, Jansen-West K, Del Rosso G, Yue M, Jiang P, Gomes E, Tong J, Daughrity LM, Avendano NM, Castanedes-Casey M, Shao W, Oskarsson B, Tomassy GS, McCampbell A, Rigo F, Dickson DW, Shorter J, Zhang Y-J, Petrucelli L (2020) C9orf72 poly(GR) aggregation induces TDP-43 proteinopathy. Sci Transl Med 12:eabb3774. doi: 10.1126/scitranslmed.abb3774

17. Cooper-Knock J, Green C, Altschuler G, Wei W, Bury JJ, Heath PR, Wyles M, Gelsthorpe C, Highley JR, Lorente-Pons A, Beck T, Doyle K, Otero K, Traynor B, Kirby J, Shaw PJ, Hide W (2017) A data-driven approach links microglia to pathology and prognosis in amyotrophic lateral sclerosis. Acta Neuropathologica Communications 5:23. doi: 10.1186/s40478-017-0424-x

18. Cooper-Knock J, Hewitt C, Highley JR, Brockington A, Milano A, Man S, Martindale J, Hartley J, Walsh T, Gelsthorpe C, Baxter L, Forster G, Fox M, Bury J, Mok K, McDermott CJ, Traynor BJ, Kirby J, Wharton SB, Ince PG, Hardy J, Shaw PJ (2012) Clinico-pathological features in amyotrophic lateral sclerosis with expansions in C9ORF72. Brain 135:751–764. doi: 10.1093/brain/awr365

19. Davidson Y, Robinson AC, Liu X, Wu D, Troakes C, Rollinson S, Masuda-Suzukake M, Suzuki G, Nonaka T, Shi J, Tian J, Hamdalla H, Ealing J, Richardson A, Jones M, Pickering-Brown S, Snowden JS, Hasegawa M, Mann DMA (2016) Neurodegeneration in frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9orf72 is linked to TDP-43 pathology and not associated with aggregated forms of dipeptide repeat proteins. Neuropathol Appl Neurobiol 42:242–254. doi: 10.1111/nan.12292

20. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, Nicholson AM, Finch NA, Flynn H, Adamson J, Kouri N, Wojtas A, Sengdy P, Hsiung G-YR, Karydas A, Seeley WW, Josephs KA, Coppola G, Geschwind DH, Wszolek ZK, Feldman H, Knopman DS, Petersen RC, Miller BL, Dickson DW, Boylan KB, Graff-Radford NR, Rademakers R (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron 72:245–256. doi: 10.1016/j.neuron.2011.09.011

21. Devenney EM, Ahmed RM, Halliday G, Piguet O, Kiernan MC, Hodges JR (2018) Psychiatric disorders in C9orf72 kindreds. Neurology 91:e1498. doi: 10.1212/WNL.0000000000006344

22. Fitzgibbons PL, Dillon DA, Alsabeh R, Berman MA, Hayes DF, Hicks DG, Hughes KS, Nofech-Mozes S (2013) Template for Reporting Results of Biomarker Testing of Specimens From Patients With Carcinoma of the Breast. Archives of Pathology & Laboratory Medicine 138:595–601. doi: 10.5858/arpa.2013-0566-CP

23. Fyfe I (2018) Do inflammatory profiles explain ALS–FTD spectrum? Nature Reviews Neurology 14:634–634. doi: 10.1038/s41582-018-0088-6

24. Geser F, Robinson JL, Malunda JA, Xie SX, Clark CM, Kwong LK, Moberg PJ, Moore EM, Van Deerlin VM, Lee VM-Y, Arnold SE, Trojanowski JQ (2010) Pathological 43-kDa transactivation response DNA-binding protein in older adults with and without severe mental illness. Arch Neurol 67:1238–1250. doi: 10.1001/archneurol.2010.254

25. Gregory JM, Livesey MR, McDade K, Selvaraj BT, Barton SK, Chandran S, Smith C (2020) Dysregulation of AMPA receptor subunit expression in sporadic ALS post-mortem brain. J Pathol 250:67–78. doi: 10.1002/path.5351

26. Gregory JM, McDade K, Bak TH, Pal S, Chandran S, Smith C, Abrahams S (2020) Executive, language and fluency dysfunction are markers of localised TDP-43 cerebral pathology in non-demented ALS. J Neurol Neurosurg Psychiatry 91:149. doi: 10.1136/jnnp-2019-320807

27. Gregory JM, McDade K, Bak TH, Pal S, Chandran S, Smith C, Abrahams S (2020) Executive, language and fluency dysfunction are markers of localised TDP-43 cerebral pathology in non-demented ALS. J Neurol Neurosurg Psychiatry 91:149. doi: 10.1136/jnnp-2019-320807

28. Hagemann C, Tyzack GE, Taha DM, Devine H, Greensmith L, Newcombe J, Patani R, Serio A, Luisier R (2021) Automated and unbiased discrimination of ALS from control tissue at single cell resolution. Brain Pathology 31:e12937. doi: 10.1111/bpa.12937

29. Haruwaka K, Ikegami A, Tachibana Y, Ohno N, Konishi H, Hashimoto A, Matsumoto M, Kato D, Ono R, Kiyama H, Moorhouse AJ, Nabekura J, Wake H (2019) Dual microglia effects on blood brain barrier permeability induced by systemic inflammation. Nature Communications 10:5816. doi: 10.1038/s41467-019-13812-z

30. Humphrey J, Birsa N, Milioto C, McLaughlin M, Ule AM, Robaldo D, Eberle AB, Kräuchi R, Bentham M, Brown A-L, Jarvis S, Bodo C, Garone MG, Devoy A, Soraru G, Rosa A, Bozzoni I, Fisher EMC, Mühlemann O, Schiavo G, Ruepp M-D, Isaacs AM, Plagnol V, Fratta P (2020) FUS ALS-causative mutations impair FUS autoregulation and splicing factor networks through intron retention. Nucleic Acids Research 48:6889–6905. doi: 10.1093/nar/gkaa410

31. Jack CR Jr, Holtzman DM (2013) Biomarker modeling of Alzheimer's disease. Neuron 80:1347–1358. doi: 10.1016/j.neuron.2013.12.003

32. Lall D, Lorenzini I, Mota TA, Bell S, Mahan TE, Ulrich JD, Davtyan H, Rexach JE, Muhammad AKMG, Shelest O, Landeros J, Vazquez M, Kim J, Ghaffari L, O'Rourke JG, Geschwind DH, Blurton-Jones M, Holtzman DM, Sattler R, Baloh RH (2021) C9orf72 deficiency promotes microglial-mediated synaptic loss in aging and amyloid accumulation. Neuron 109:2275-2291.e8. doi: 10.1016/j.neuron.2021.05.020

33. Leighton D, Newton J, Colville S, Bethell A, Craig G, Cunningham L, Flett M, Fraser D, Hatrick J, Lennox H, Marshall L, McAleer D, McEleney A, Millar K, Silver A, Stephenson L, Stewart S, Storey D, Stott G, Thornton C, Webber C, Gordon H, Melchiorre G, Sherlock L, Beswick E, Buchanan D, Abrahams S, Bateman A, Preston J, Duncan C, Davenport R, Gorrie G, Morrison I, Swingler R, Chandran S, Pal S (2019) Clinical audit research and evaluation of motor neuron disease (CARE-MND): a national electronic platform for prospective, longitudinal monitoring of MND in Scotland. null 20:242–250. doi: 10.1080/21678421.2019.1582673

34. Lu C-H, Allen K, Oei F, Leoni E, Kuhle J, Tree T, Fratta P, Sharma N, Sidle K, Howard R, Orrell R, Fish M, Greensmith L, Pearce N, Gallo V, Malaspina A (2016) Systemic inflammatory response and neuromuscular involvement in amyotrophic lateral sclerosis. Neurol Neuroimmunol Neuroinflamm 3:e244–e244. doi: 10.1212/NXI.0000000000000244

35. Mackenzie IRA, Frick P, Grässer FA, Gendron TF, Petrucelli L, Cashman NR, Edbauer D, Kremmer E, Prudlo J, Troost D, Neumann M (2015) Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. Acta Neuropathologica 130:845–861. doi: 10.1007/s00401-015-1476-2

36. Majounie E, Renton AE, Mok K, Dopper EGP, Waite A, Rollinson S, Chiò A, Restagno G, Nicolaou N, Simon-Sanchez J, van Swieten JC, Abramzon Y, Johnson JO, Sendtner M, Pamphlett R, Orrell RW, Mead S, Sidle KC, Houlden H, Rohrer JD, Morrison KE, Pall H, Talbot K, Ansorge O, Chromosome 9-ALS/FTD Consortium, French research network on FTLD/FTLD/ALS, ITALSGEN Consortium, Hernandez DG, Arepalli S, Sabatelli M, Mora G, Corbo M, Giannini F, Calvo A, Englund E, Borghero G, Floris GL, Remes AM, Laaksovirta H, McCluskey L, Trojanowski JQ, Van Deerlin VM, Schellenberg GD, Nalls MA, Drory VE, Lu C-S, Yeh T-H, Ishiura H, Takahashi Y, Tsuji S, Le Ber I, Brice A, Drepper C, Williams N, Kirby J, Shaw P, Hardy J, Tienari PJ, Heutink P, Morris HR, Pickering-Brown S, Traynor BJ (2012) Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. Lancet Neurol 11:323–330. doi: 10.1016/S1474-4422(12)70043-1

37. Mann JR, Donnelly CJ (2021) RNA modulates physiological and neuropathological protein phase transitions. Neuron 109:2663–2681. doi: 10.1016/j.neuron.2021.06.023

38. McCauley ME, O'Rourke JG, Yáñez A, Markman JL, Ho R, Wang X, Chen S, Lall D, Jin M, Muhammad AKMG, Bell S, Landeros J, Valencia V, Harms M, Arditi M, Jefferies C, Baloh RH (2020) C9orf72 in myeloid cells suppresses STING-induced inflammation. Nature 585:96–101. doi: 10.1038/s41586-020-2625-x

39. Morriss NJ, Conley GM, Ospina SM, Meehan III WP, Qiu J, Mannix R (2020) Automated Quantification of Immunohistochemical Staining of Large Animal Brain Tissue Using QuPath Software. Neuroscience 429:235–244. doi: 10.1016/j.neuroscience.2020.01.006

40. Murray ME, DeJesus-Hernandez M, Rutherford NJ, Baker M, Duara R, Graff-Radford NR, Wszolek ZK, Ferman TJ, Josephs KA, Boylan KB, Rademakers R, Dickson DW (2011) Clinical and neuropathologic heterogeneity of c9FTD/ALS associated with hexanucleotide repeat expansion in C9ORF72. Acta Neuropathol 122:673–690. doi: 10.1007/s00401-011-0907-y

41. Nanaura H, Kawamukai H, Fujiwara A, Uehara T, Aiba Y, Nakanishi M, Shiota T, Hibino M, Wiriyasermkul P, Kikuchi S, Nagata R, Matsubayashi M, Shinkai Y, Niwa T, Mannen T, Morikawa N, Iguchi N, Kiriyama T, Morishima K, Inoue R, Sugiyama M, Oda T, Kodera N, Toma-Fukai S, Sato M, Taguchi H, Nagamori S, Shoji O, Ishimori K, Matsumura H, Sugie K, Saio T, Yoshizawa T, Mori E (2021) C9orf72-derived arginine-rich poly-dipeptides impede phase modifiers. Nature Communications 12:5301. doi: 10.1038/s41467-021-25560-0

42. Nascimento C, Di Lorenzo Alho AT, Bazan Conceição Amaral C, Leite REP, Nitrini R, Jacob-Filho W, Pasqualucci CA, Hokkanen SRK, Hunter S, Keage H, Kovacs GG, Grinberg LT, Suemoto CK (2018) Prevalence of transactive response DNA-binding protein 43 (TDP-43) proteinopathy in cognitively normal older adults: systematic review and meta-analysis. Neuropathol Appl Neurobiol 44:286–297. doi: 10.1111/nan.12430

43. Neltner JH, Abner EL, Schmitt FA, Denison SK, Anderson S, Patel E, Nelson PT (2012) Digital pathology and image analysis for robust high-throughput quantitative assessment of Alzheimer disease neuropathologic changes. J Neuropathol Exp Neurol 71:1075–1085. doi: 10.1097/NEN.0b013e3182768de4

44. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, Bruce J, Schuck T, Grossman M, Clark CM, McCluskey LF, Miller BL, Masliah E, Mackenzie IR, Feldman H, Feiden W, Kretzschmar HA, Trojanowski JQ, Lee VM-Y (2006) Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. Science 314:130. doi: 10.1126/science.1134108

45. Oeckl P, Weydt P, Steinacker P, Anderl-Straub S, Nordin F, Volk AE, Diehl-Schmid J, Andersen PM, Kornhuber J, Danek A, Fassbender K, Fliessbach K, Jahn H, Lauer M, Müller K, Knehr A, Prudlo J, Schneider A, Thal DR, Yilmazer-Hanke D, Weishaupt JH, Ludolph AC, Otto M (2019) Different neuroinflammatory profile in amyotrophic lateral sclerosis and frontotemporal dementia is linked to the clinical phase. J Neurol Neurosurg Psychiatry 90:4. doi: 10.1136/jnnp-2018-318868

46. O'Rourke JG, Bogdanik L, Yáñez A, Lall D, Wolf AJ, Muhammad AKMG, Ho R, Carmona S, Vit JP, Zarrow J, Kim KJ, Bell S, Harms MB, Miller TM, Dangler CA, Underhill DM, Goodridge HS, Lutz CM, Baloh RH (2016) C9orf72 is required for proper macrophage and microglial function in mice. Science 351:1324–1329. doi: 10.1126/science.aaf1064

47. Petrov D, Mansfield C, Moussy A, Hermine O (2017) ALS Clinical Trials Review: 20 Years of Failure. Are We Any Closer to Registering a New Treatment? Front Aging Neurosci 9:68–68. doi: 10.3389/fnagi.2017.00068

48. Rizzu P, Blauwendraat C, Heetveld S, Lynes EM, Castillo-Lizardo M, Dhingra A, Pyz E, Hobert M, Synofzik M, Simón-Sánchez J, Francescatto M, Heutink P (2016) C9orf72 is differentially expressed in the central nervous system and myeloid cells and consistently reduced in C9orf72, MAPT and GRN mutation carriers. Acta Neuropathologica Communications 4:37. doi: 10.1186/s40478-016-0306-7

49. Shinagawa S, Naasan G, Karydas AM, Coppola G, Pribadi M, Seeley WW, Trojanowski JQ, Miller BL, Grinberg LT (2015) Clinicopathological Study of Patients With C9ORF72-Associated Frontotemporal Dementia Presenting With Delusions. J Geriatr Psychiatry Neurol 28:99–107. doi: 10.1177/0891988714554710

50. Spencer KR, Foster ZW, Rauf NA, Guilderson L, Collins D, Averill JG, Walker SE, Robey I, Cherry JD, Alvarez VE, Huber BR, McKee AC, Kowall NW, Brady CB, Stein TD (2020) Neuropathological profile of long-duration amyotrophic lateral sclerosis in military Veterans. Brain Pathology 30:1028–1040. doi: 10.1111/bpa.12876

51. Strong MJ, Abrahams S, Goldstein LH, Woolley S, Mclaughlin P, Snowden J, Mioshi E, Roberts-South A, Benatar M, HortobáGyi T, Rosenfeld J, Silani V, Ince PG, Turner MR (2017) Amyotrophic lateral sclerosis - frontotemporal spectrum disorder (ALS-FTSD): Revised diagnostic criteria. Amyotroph Lateral Scler Frontotemporal Degener 18:153–174. doi: 10.1080/21678421.2016.1267768

52. Tam OH, Rozhkov NV, Shaw R, Kim D, Hubbard I, Fennessey S, Propp N, NYGC ALS Consortium, Fagegaltier D, Harris BT, Ostrow LW, Phatnani H, Ravits J, Dubnau J, Gale Hammell M (2019) Postmortem Cortex Samples Identify Distinct Molecular Subtypes of ALS: Retrotransposon Activation, Oxidative Stress, and Activated Glia. Cell Rep 29:1164-1177.e5. doi: 10.1016/j.celrep.2019.09.066

53. Tang BL (2017) Amyotrophic lateral sclerosis disease modifying therapeutics: a cell biological perspective. Neural Regen Res 12:407–408. doi: 10.4103/1673-5374.202943

54. Taylor JP, Brown RH Jr, Cleveland DW (2016) Decoding ALS: from genes to mechanism. Nature 539:197–206. doi: 10.1038/nature20413

55. Torres-Platas SG, Comeau S, Rachalski A, Bo GD, Cruceanu C, Turecki G, Giros B, Mechawar N (2014) Morphometric characterization of microglial phenotypes in human cerebral cortex. Journal of Neuroinflammation 11:12. doi: 10.1186/1742-2094-11-12

56. Trageser KJ, Smith C, Herman FJ, Ono K, Pasinetti GM (2019) Mechanisms of Immune Activation by c9orf72-Expansions in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. Frontiers in Neuroscience 13:1298. doi: 10.3389/fnins.2019.01298

57. Troakes C, Maekawa S, Wijesekera L, Rogelj B, Siklós L, Bell C, Smith B, Newhouse S, Vance C, Johnson L, Hortobágyi T, Shatunov A, Al-Chalabi A, Leigh N, Shaw CE, King A, Al-Sarraj S (2012) An MND/ALS phenotype associated with C9orf72 repeat expansion: Abundant p62-positive, TDP-43-negative inclusions in cerebral cortex, hippocampus and cerebellum but without associated cognitive decline. Neuropathology 32:505–514. doi: 10.1111/j.1440-1789.2011.01286.x

58. Tsai RM, Boxer AL (2016) Therapy and clinical trials in frontotemporal dementia: past, present, and future. J Neurochem 138 Suppl 1:211–221. doi: 10.1111/jnc.13640

59. Tyzack GE, Luisier R, Taha DM, Neeves J, Modic M, Mitchell JS, Meyer I, Greensmith L, Newcombe J, Ule J, Luscombe NM, Patani R (2019) Widespread FUS mislocalization is a molecular hallmark of amyotrophic lateral sclerosis. Brain 142:2572–2580. doi: 10.1093/brain/awz217

60. Vega AR, Chkheidze R, Jarmale V, Shang P, Foong C, Diamond MI, White CL, Rajaram S (2021) Deep learning reveals disease-specific signatures of white matter pathology in tauopathies. Acta Neuropathologica Communications 9:170. doi: 10.1186/s40478-021-01271-x

61. Westeneng H-J, Debray TPA, Visser AE, van Eijk RPA, Rooney JPK, Calvo A, Martin S, McDermott CJ, Thompson AG, Pinto S, Kobeleva X, Rosenbohm A, Stubendorff B, Sommer H, Middelkoop BM, Dekker AM, van Vugt JJFA, van Rheenen W, Vajda A, Heverin M, Kazoka M, Hollinger H, Gromicho M, Körner S, Ringer TM, Rödiger A, Gunkel A, Shaw CE, Bredenoord AL, van Es MA, Corcia P, Couratier P, Weber M, Grosskreutz J, Ludolph AC, Petri S, de Carvalho M, Van Damme P, Talbot K, Turner MR, Shaw PJ, Al-Chalabi A, Chiò A, Hardiman O, Moons KGM, Veldink JH, van den Berg LH (2018) Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. The Lancet Neurology 17:423–433. doi: 10.1016/S1474-4422(18)30089-9

62. Wilson RS, Yu L, Trojanowski JQ, Chen E-Y, Boyle PA, Bennett DA, Schneider JA (2013) TDP-43 pathology, cognitive decline, and dementia in old age. JAMA Neurol 70:1418–1424. doi: 10.1001/jamaneurol.2013.3961

63. Yokoyama JS, Sirkis DW, Miller BL (2014) C9ORF72 hexanucleotide repeats in behavioral and motor neuron disease: clinical heterogeneity and pathological diversity. Am J Neurodegener Dis 3:1–18

64. Yu C-H, Davidson S, Harapas CR, Hilton JB, Mlodzianoski MJ, Laohamonthonkul P, Louis C, Low RRJ, Moecking J, De Nardo D, Balka KR, Calleja DJ, Moghaddas F, Ni E, McLean CA, Samson AL, Tyebji S, Tonkin CJ, Bye CR, Turner BJ, Pepin G, Gantier MP, Rogers KL, McArthur K, Crouch PJ, Masters SL (2020) TDP-43 Triggers Mitochondrial DNA Release via mPTP to Activate cGAS/STING in ALS. Cell 183:636-649.e18. doi: 10.1016/j.cell.2020.09.020

65. Zhao W, Beers DR, Bell S, Wang J, Wen S, Baloh RH, Appel SH (2015) TDP-43 activates microglia through NF-κB and NLRP3 inflammasome. Experimental Neurology 273:24–35. doi: 10.1016/j.expneurol.2015.07.019
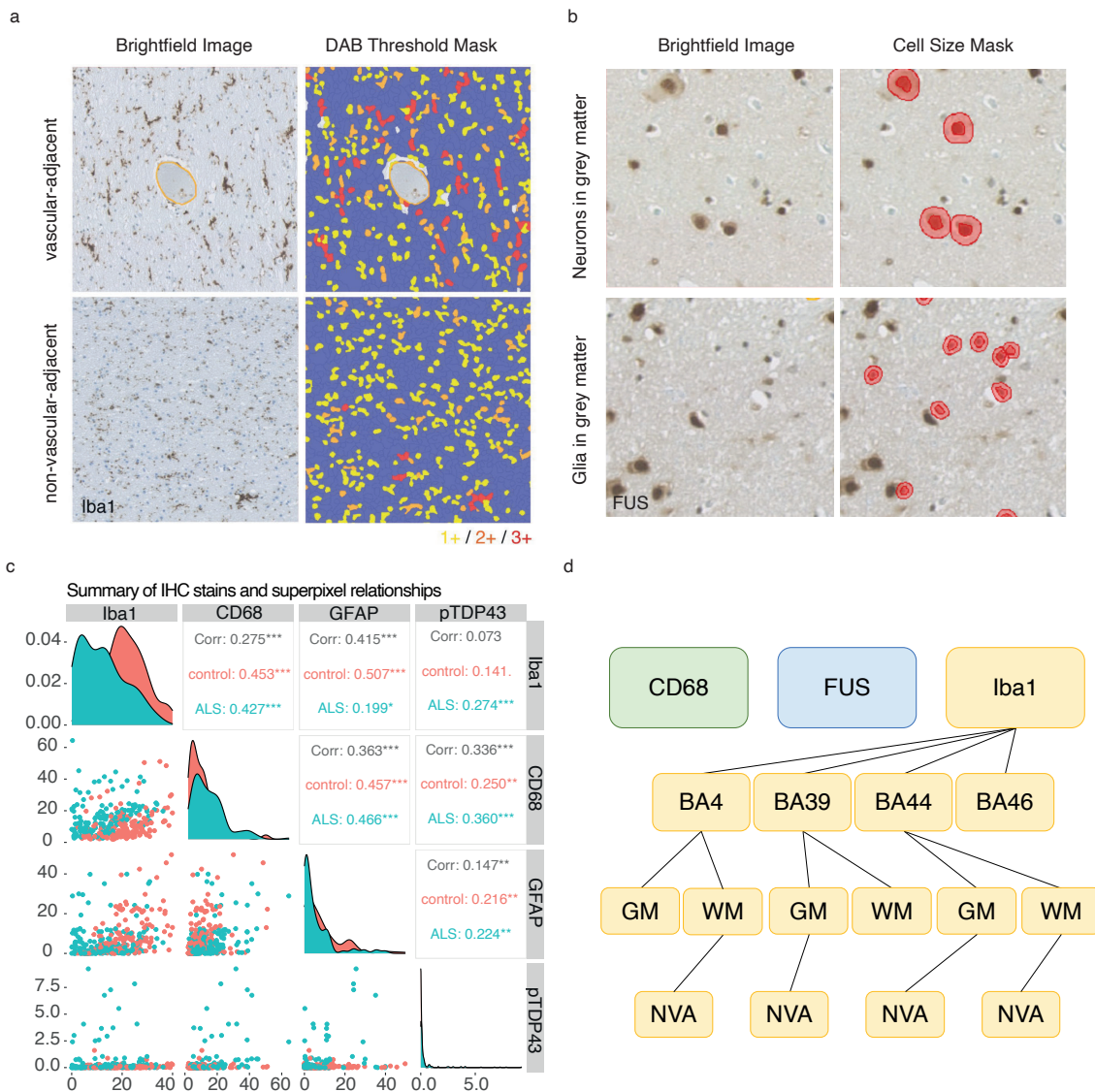
Figure 1. *Microglial immunohistochemical staining features are the best independent classifiers of disease status.* (a) Example region of interest comprising brightfield images of Iba1-stained tissue in vascular adjacent and non-vascular adjacent regions (left). Iba1 = brown DAB staining with a haematoxylin counterstain. Images on the right show a segmentation-analysis mask overlay created using QuPath digital pathology software coding DAB staining intensity from 1+ to 3+. (b) Example neuronal and glial cell masks created by nuclear area thresholding created using QuPath. (c) Correlation matrix demonstrating inter-relationship between superpixel analysis of histological stains. (d) Hierarchical organisation of results from Random Forest analysis demonstrating accuracy of each stain as a disease classifier. Note Iba1 demonstrates multi-level accuracy (i) stain, (ii) brain region, (iii) sub-region 1 (white matter vs grey matter) and (iv) sub-region 2 (vascular-adjacent and non-adjacent). BA, Brodmann area; GM, grey matter; WM, white matter; NVA, non-vascular-adjacent; VA, vascular-adjacent. Spearman, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.
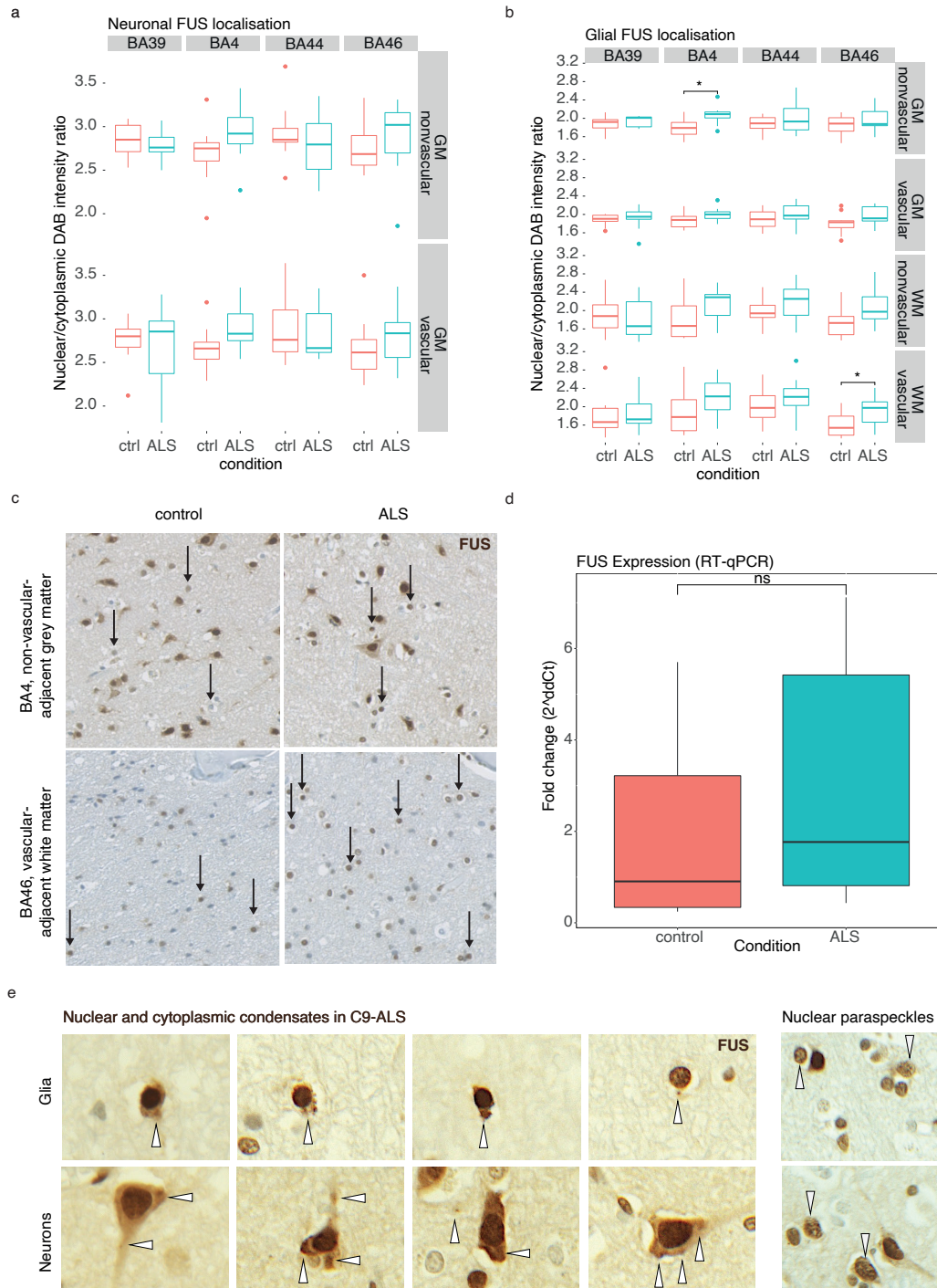
Figure 2. *Glia exhibit changes in FUS localisation.* Box and whisker plots demonstrating (a) neuronal (unpaired t-test) and (b) glial (Mann-Whitney U test) nuclear/cytoplasmic ratio of FUS immunostaining intensity by brain region, demonstrating increased nuclear/cytoplasmic ratio in disease. (c) Representative images of glial FUS staining with significant differences between control and ALS cases. (d) RT-qPCR of FUS expression demonstrating no significant difference between control and disease (unpaired t-test). (e) Example FUS nuclear and cytoplasmic condensates and nuclear paraspeckles in C9-ALS cases. BA, Brodmann area; GM, grey matter; WM, white matter; NVA, non-vascular-adjacent; VA, vascular-adjacent. * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001.
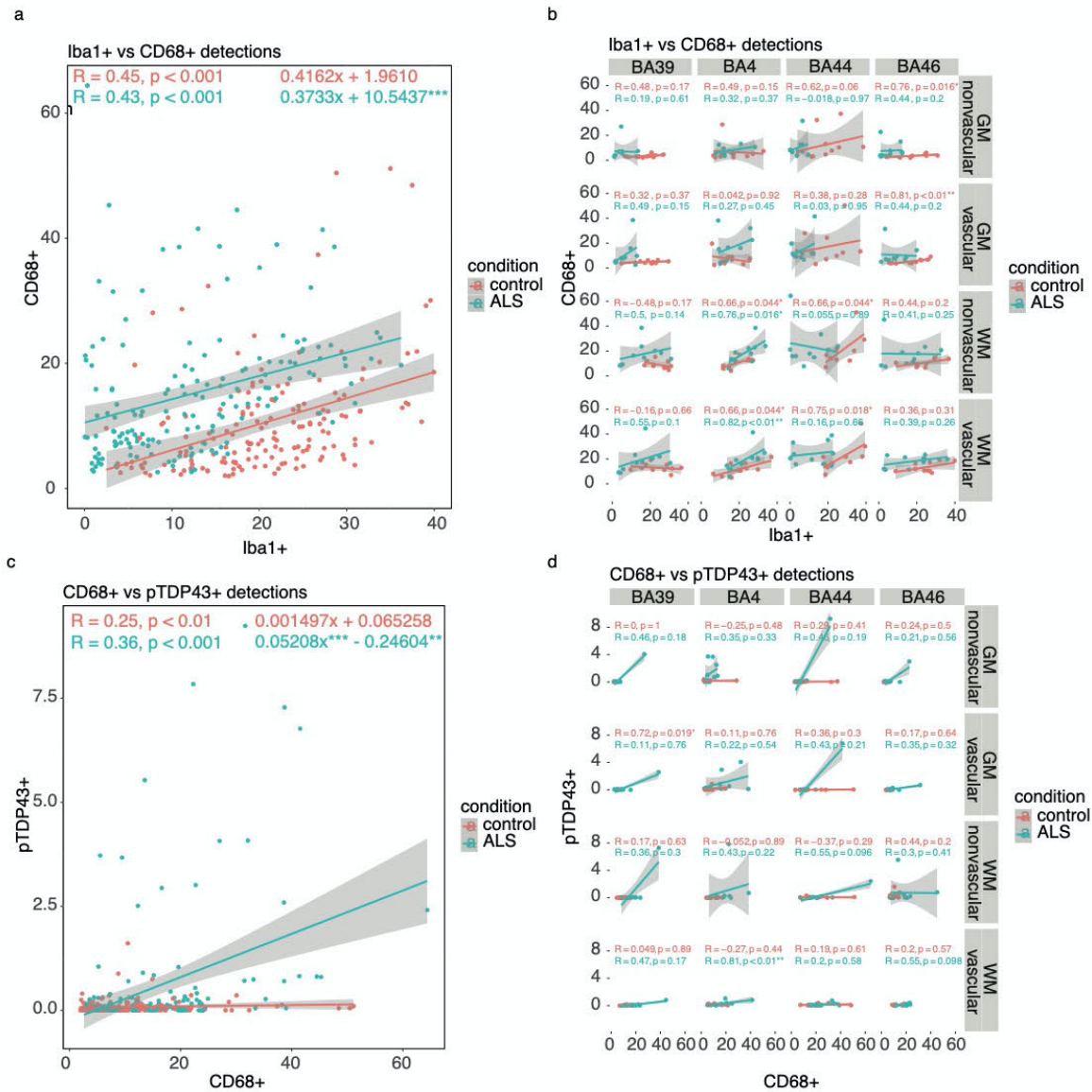
Figure 3. *C9-ALS post-mortem tissue exhibits microgliosis-related changes.* (a) Scatter plot demonstrating the relationship between CD68 (activated microglial stain) and Iba1 (total microglial stain) demonstrating an overall increase in activated microglia in disease vs control. (b) Correlation matrix plotting data from (a) by brain region subdivided in to grey and white matter and vascular adjacent and non-adjacent regions. (c) Scatter plot demonstrating the relationship between CD68 (activated microglial stain) and pTDP-43 staining demonstrating an increase in activated microglia with increasing TDP-43 staining. (d) Correlation matrix plotting data from (c) by brain region subdivided in to grey and white matter and vascular adjacent and non-adjacent regions. BA, Brodmann area; GM, grey matter; WM, white matter. Spearman, * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001.
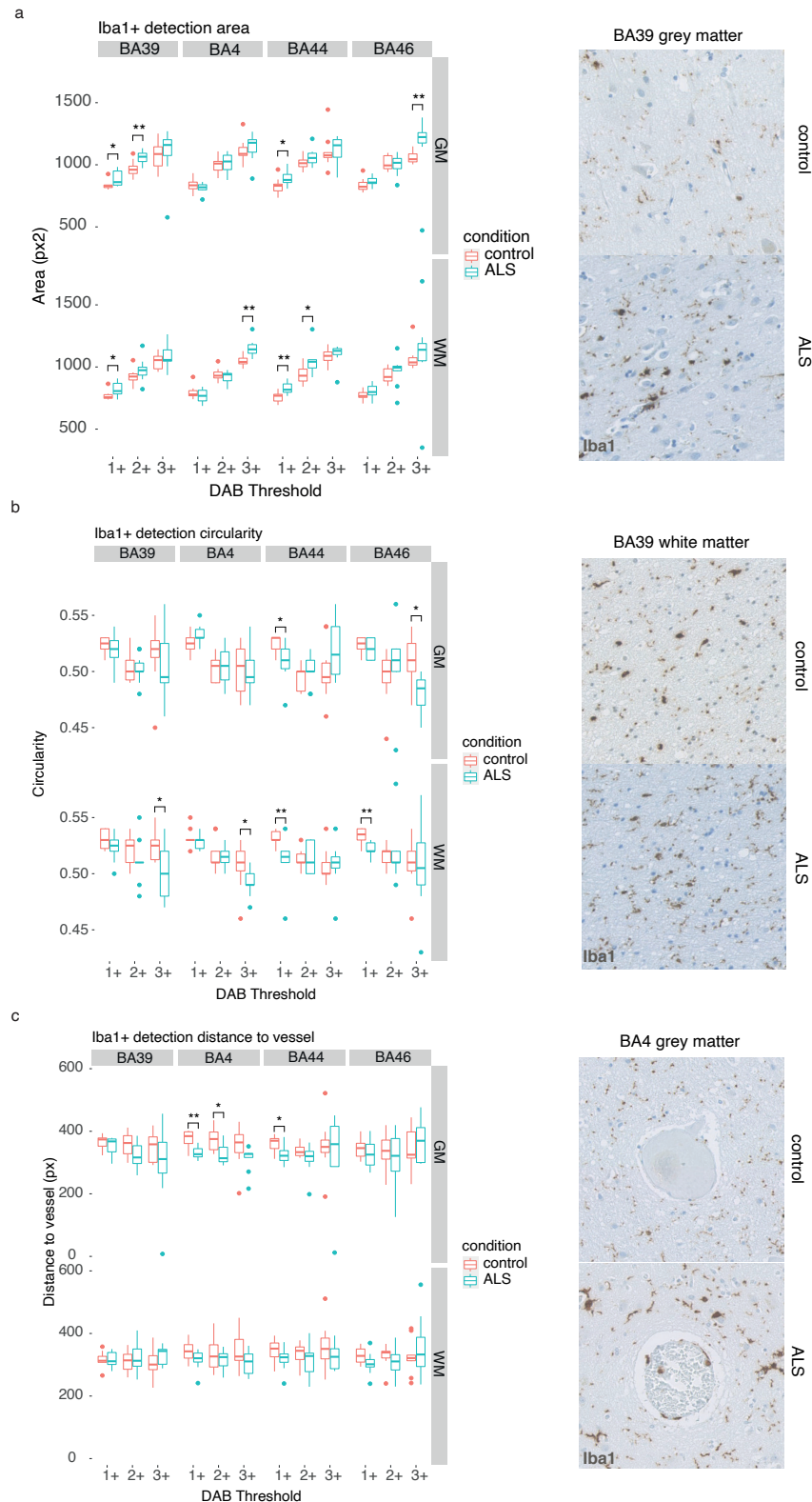
Figure 4. Analysis of microglial staining features reveals morphological and spatial differences in C9-ALS-FTD. (a-c) Examples of feature analysis for Iba1 staining, demonstrating differences in staining intensity (a), morphology (b) and proximity to blood vessels (c) in disease vs control cases. Example photomicrographs on right show control (above) and disease (below). Intensity classes 1-3 represent increasing DAB intensity, with 1 = mild, 2 = moderate, 3 = high. BA, Brodmann area; GM, grey matter; WM, white matter. Mann-Whitney U test, * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001.
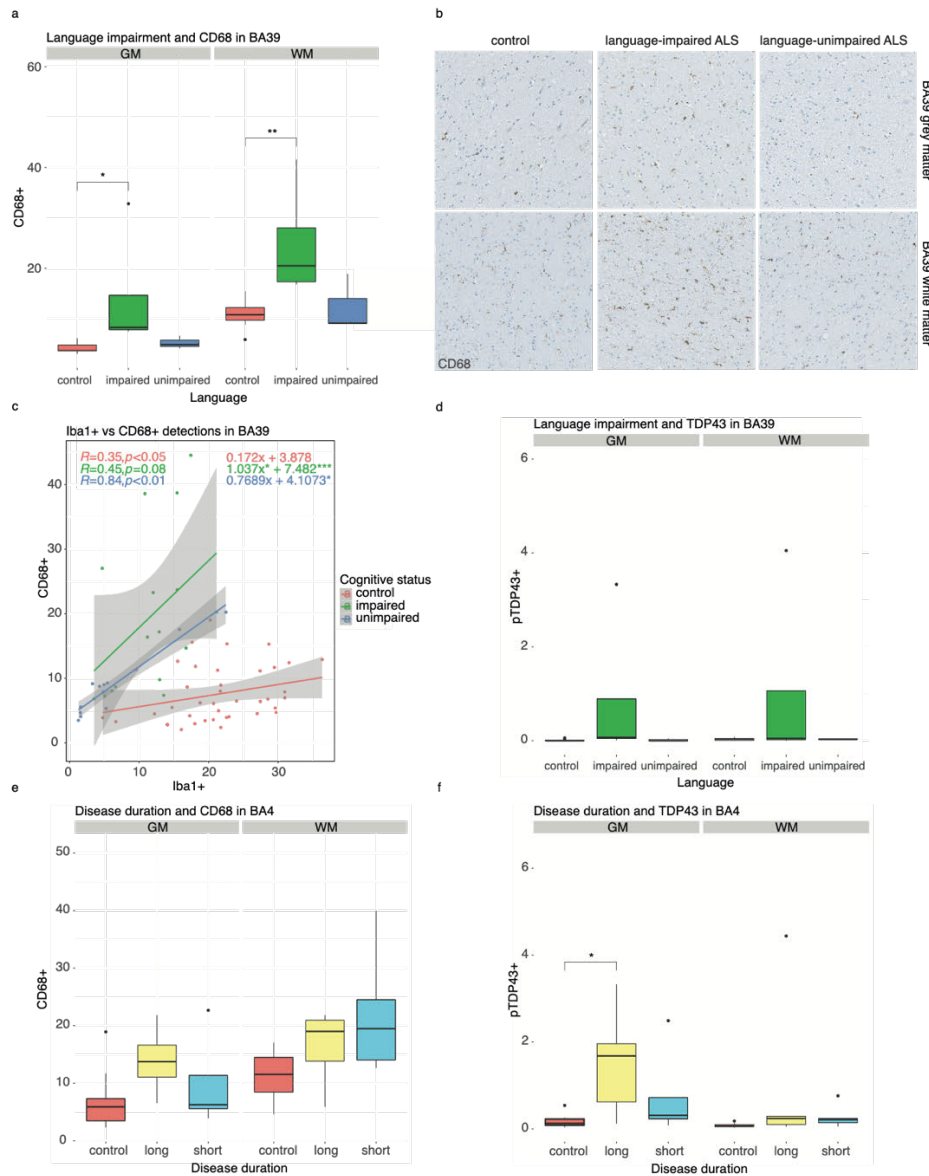
Figure 5. *Clinicopathological associations exist between microglial activation status and cognitive dysfunction.* (a) Box and whisker plot showing CD68 immunoreactivity in cases grouped by language impairment (controls, language impaired on Edinburgh Cognitive and Behavioural ALS Screen (ECAS) during life and language unimpaired), demonstrating increased immunoreactivity for CD68 in language impaired cases in BA39 – a brain region responsible for language function. (b) Example images (grey matter above and white matter below) with CD68 immunostaining (brown/DAB) and haematoxylin counterstain showing increased immunostaining in language impaired cases in BA39. (c) Scatter plot demonstrating higher CD68 activation status in BA39 of cases vs controls irrespective of cognitive status. (d) Box and whisker plot showing pTDP-43 immunoreactivity in cases grouped by language impairment (controls, language impaired on ECAS during life and language unimpaired), demonstrating increased immunoreactivity for pTDP-43 in language impaired cases in the grey matter of BA39 – a brain region responsible for language function. (e) Box and whisker plot showing CD68 immunoreactivity in BA4 (motor cortex) of cases grouped by disease severity (disease duration from symptom onset; controls, long disease duration and short disease duration). (f) Box and whisker plot showing pTDP-43 immunoreactivity in BA4 (motor cortex) of cases grouped by disease severity (disease duration from symptom onset; controls, long disease duration and short disease duration), demonstrating increased immunoreactivity for pTDP-43 in the grey matter of long survivors in the grey matter. Data in box and whisker plots are averaged across quantified regions not included in the presented stratification, such that each category contains n = number of cases, to avoid pseudoreplication. BA, Brodmann area; GM, grey matter; WM, white matter. Spearman, Wilcoxon with Holm Sidak, * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001.

	Age at death (y)	Disease duration (m)	Region of onset	Cognitive data	ALSFRS
Case 1	63	25	Lower Limb	Yes (unimpaired)	No
Case 2	50	29	bulbar	Yes (language dysfunction)	Yes (4 data points)
Case 3	63	30	Upper limb	No	Yes (3 data points)
Case 4	62	37	Mixed	Yes (unimpaired)	Yes (6 data points)
Case 5	65	44	Lower limb	No	Yes (2 data points)
Case 6	65	50	Upper limb	No	Yes (1 data point)
Case 7	62	50	Lower limb	Yes (executive dysfunction)	Yes (2 data points)
Case 8	43	57	Bulbar	Yes (language dysfunction)	No
Case 9	58	87	Lower limb	Yes (language dysfunction)	Yes (3 data points)
Case 10	63	119	Upper limb	Yes (FTD – 3 domains)	No

Table 1. C9orf72 cohort demographics.

Case	Neuronal (N) Glial (G)	TDP-43 score				ECAS
		BA4	BA39	BA44	BA46	
1	N	2	1	1	1	unimpaired (mismatch)
	G	2	1	1	1	
2	N	2	1	1	1	language
	G	3	1	1	1	
3	N	2	1	1	2	N/A
	G	2	1	1	1	
4	N	1	1	1	1	unimpaired (mismatch)
	G	2	1	1	2	
5	N	1	0	0	0	N/A
	G	2	1	0	0	
6	N	1	1	1	1	N/A
	G	1	0	0	0	
7	N	2	1	1	1	executive (language mismatch)
	G	0	0	0	0	
8	N	1	1	0	0	language
	G	2	0	0	0	
9	N	3	1	1	1	language
	G	3	1	1	1	
10	N	2	3	3	3	all (FTD)
	G	3	3	3	3	

Table 2. Manual TDP43 grading of ALS cases.