

Intronic NEFH variant is associated with reduced risk for sporadic ALS and later age of disease onset

Frances Theunissen (✉ Frances.theunissen@perron.uwa.edu.au)

Perron Institute for Neurological and Translational Science

Ryan S Anderton

University of Notre Dame Australia

Frank L Mastaglia

Perron Institute for Neurological and Translational Science

Ian James

Murdoch University

Richard Bedlack

Duke University

P. Anthony Akkari

Perron Institute for Neurological and Translational Science

Article

Keywords:

Posted Date: March 30th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1456790/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Neurofilament heavy (NEFH) is one of the critical proteins required for the formation of the neuronal cytoskeleton and polymorphisms in *NEFH* are reported as a rare cause of sporadic ALS (sALS). In the current study, a candidate tetranucleotide (TTTA) repeat variant in *NEFH* was selected using an *in-silico* SSV evaluation algorithm and investigated in two cohorts of North American sALS patients, both separately and combined (Duke cohort n = 138, Coriell cohort n = 333; combined cohort n = 471), compared to a group of healthy controls from the Coriell Institute biobank (n = 496). Stratification according to site of disease onset revealed that the 9 TTTA allele was associated with a 2x reduced risk specifically confined to spinal-onset sALS patients in the Duke cohort ($p = 0.001$). Furthermore, carriage of the 10 TTTA allele was associated with a 2.7 year later age of disease onset in the larger combined sALS cohort ($p = 0.02$). These results suggest that the 9 and 10 TTTA motif length may have a protective advantage for potentially lowering the risk of sALS and delaying the age of disease onset, however these results need to be replicated in larger multicenter and multi-ethnic cohorts.

Introduction

Neurofilament heavy protein (NEFH) is a type IV intermediate filament. In neurons, assembly of the light, medium and heavy neurofilament proteins is required in a specific stoichiometry (4:2:1 respectively) to form the foundations of the neuronal cytoskeleton^{1,2}, to facilitate axonal transport by maintaining axonal diameter, and to direct and position mitochondria along the axon^{3,4}. In ALS, one of the histopathological features of disease is the presence of cytoplasmic inclusions containing intermediate filament proteins (light, medium, and heavy) in axonal spheroids and within the cell body of motor neurons⁵⁻⁸. The mechanisms behind this increased propensity for aggregation has not yet been fully elucidated, however, there is evidence for hyper-phosphorylation^{5,9-11}, dysregulated mRNA expression¹²⁻¹⁵ and altered microRNA levels in ALS patients that can influence the stability of neurofilament transcripts^{16,17} and could contribute to this altered stoichiometry. Typically, the mRNA levels of the smaller neurofilaments (*NEFL* and *INA*) are reported as dysregulated in ALS, whilst *NEFH* mRNA is reported as showing no significant differences between ALS patients and controls^{15,18,19}. Alternatively, another report suggests that expression may shift away from *NEFH* to favor the smaller neurofilaments, as an adaptive strategy to reduce energy requirements under conditions of cellular stress during neurodegeneration²⁰.

On a genetic level, mutations and polymorphisms in all three NEFH protein domains (head, rod and tail) have been reported to be associated with ALS²¹⁻²⁷, although studies have typically focused on exonic variants. In particular, large insertions or deletions in the tail region of NEFH have been shown to impact phosphorylation sites critical for its interaction and assembly with other smaller neurofilament proteins^{21-23,28}. However, many studies report variants in NEFH as a rare cause of ALS, occurring in only ~1% of sporadic cases²¹⁻²⁷. Despite huge advances in the field of ALS genetics, there is still an urgent need for the characterization of genetic markers that may help to underpin disease heterogeneity and phenotypic variation, particularly in sALS patients^{3,29,30}. Therefore, exploring more common short structural variation residing in non-coding regions may help to tag or directly identify genetic factors that may contribute to the role of neurofilaments in ALS disease modification.

In the current study, a candidate intronic TTTA repeat variant in *NEFH* was selected using an *in-silico* short structural variant evaluation algorithm³¹ and investigated using PCR, polyacrylamide gel fractionation and Sanger sequencing. A case-control study based on TTTA allele frequencies was performed in two North American cohorts of sporadic ALS (sALS) patients (separately and combined) to determine if TTTA alleles/genotypes are associated with sALS disease risk or have disease-modifying effects.

Methods

Study participants and ethics approval from case control cohorts

A combined cohort of 471 Caucasian North American sALS patients and 496 Caucasian North American healthy age-matched controls were used in this study. Written informed consent was obtained from all participants in accordance with Health Insurance Portability and Accountability Act. Participants were diagnosed by board-certified neurologists and met the revised El Escorial World Federation of Neurology criteria for diagnosis of ALS³². DNA and clinical information including age of onset, site of disease onset and survival duration, were collected from 138 sALS patients at Duke ALS clinics in accordance with the Health Insurance Portability and Accountability Act (Pro00040665/323682) and the Declarations of Helsinki. In addition, DNA samples from 333 sALS patients and 496 population control samples and accompanying clinical data were obtained from the NINDS Repository, Coriell Institute for Medical Research (available from <https://www.coriell.org/1/NINDS>). Retrospective genotyping of DNA samples was approved by the Human Research Ethics committee of the University of Western Australia (RA/4/20/5308).

Polymerase chain reaction

Endpoint PCR reactions were prepared to a final volume of 10 µl, containing; 7.2 µl dH₂O (Baxter Healthcare, NSW, Australia), 2 µl MyFi reaction buffer (Bioline, NSW, Australia), 0.05 µl MyFi DNA polymerase (Bioline, NSW, Australia), 0.375 µl forward and reverse primer 5'-GCCTGCTTTGCAGAGCTG-3' and 5'-GGCTCAAGAGATTCTCGTGCC-3', respectively. (Integrated DNA Technologies, Iowa, USA) at 200 ng/µl, and 10 ng DNA. The amplification protocol followed, an initial hold temperature 95°C for 4 min 30 sec and 35 cycles of denaturation at 95°C for 30 sec, annealing 59°C for 30 sec and extension at 72°C for 1 min. PCR products were then fractionated on a 2% agarose (w/v) gel (Scientifix Pty Ltd, VIC, Australia) in 1 x TAE and stained with red safe nucleic acid stain (iNtRON Biotechnology, Scientifix Pty Ltd, VIC, Australia) prior to imaging using the BioRad Chemidoc™ MP Imaging System.

Polyacrylamide fractionation and Sanger sequencing

PCR products were fractionated on 8% (w/v) 29:1 polyacrylamide gel (BioRad, CA, USA) in 1 x TBE as previously described³³. Electrophoresis fragment separation was performed at 100 V for 8 hr on the DCode™ Universal Mutation Detection System (BioRad, CA, USA). Gels were stained in 1 x TBE containing SYBR® Gold nucleic acid gel stain (Thermo Fisher Scientific, MA, USA) for 4 min before visualization using a BioRad Chemidoc™ MP Imaging System. PCR products from homozygous samples were purified and Sanger sequenced by the Australian Genome Research Facility (AGRF, Perth Australia). Analysis was conducted using Finch-TV software (version 1.5.0; Geospiza Inc). For high throughput genotyping, a FAM-labelled forward primer was included and the methodology was performed as stated above. PCR products were sent to for capillary separation at the Australian Genome Research Facility (AGRF, Perth Australia). Results were analyzed on peak scanner software (version 1.0; Thermo Fisher Scientific).

Statistical analysis

Differences in distributions or allelic frequencies were assessed using independent samples Mann-Whitney U or chi-squared tests as appropriate. Case-control genotype/allele associations were assessed using binary logistic regression controlling for patient sex in the covariate adjusted model, with data coded as either the absence (0) or presence (1) of each *NEFH* allele or genotype. Analyses were stratified according to cohort and site of onset, and also combined. Significant associations withstanding Bonferroni correction for multiple comparisons are indicated. Joint significance of genotype/allele associations were assessed via case-control logistic regression. General linear models were used for the association of age of onset with genotype/allele, accounting for patient sex. Kaplan-Meier curves for survival duration were estimated taking into account site of disease onset. Genotypes with low frequency (n < 3) were excluded from statistical analyses. A *p*-value below 0.05 was considered statistically significant. Analyses were carried out in IBM SPSS Statistics version 25.0 (IBM Co., Armonk, NY, USA).

Results

Identification of polymorphic TTTA short structural variant in *NEFH*

A SV evaluation algorithm was used to identify potential polymorphic variants within *NEFH*³¹, with candidate variants scored according to 24 different properties, previously described³⁴. The selected TTTA variant (rs140814097) was subsequently investigated on public genomic databases NCBI and ensemble genome browser NC_000022.11 and 22: 29483828–29483863 respectively. The recorded entries for this region show a multitude of “rs numbers” logged for this variant, indicating this genetic locus is likely polymorphic and warranted further investigation (allele frequency information available from <https://www.ncbi.nlm.nih.gov/snp/rs140814097>). The candidate TTTA variant resides 255bp past exon 2 in the *NEFH* primary transcript encoding 1020 amino acids (Fig. 1a). Polyacrylamide electrophoresis fractionation of *NEFH* PCR products revealed polymorphic alleles with varying numbers of TTTA repeat motifs, and this was confirmed via Sanger sequencing (Fig. 1b). Additionally, capillary fragment separation confirmed variable length TTTA repeat genotypes, with alleles ranging from 6 to 15 TTTA repeats. Representative plots are shown (Fig. 1c), with a single fluorescent signal peak indicating homozygous samples and two signal peaks representing heterozygous samples. As the signal peaks move further apart this indicates a greater base pair difference between the individual *NEFH* alleles.

Distribution of TTTA alleles

We next investigated the distribution of the TTTA variant in the two sALS cohorts (n = 138 and n = 333) and 496 healthy control cases (cohort demographics can be seen in Table 1). The distribution of allele lengths ranged from 6 to 15 TTTA repeats in both cohorts of sALS patients and in the controls. No significant differences were found in the allele frequency distributions between the Duke and Coriell sALS samples (Fig. 2a). Therefore, both sALS cohorts were analyzed individually and also as a combined cohort (combined allele distribution, Fig. 2b). Along with self-reported ethnicity and country of origin data for each participant, each *NEFH* allele distribution was also compared to Webstr database (<http://webstr.ucsd.edu/>) and was similar to both GTE_x (predominantly European self-reported ancestry) and 1000 genomes European allele distributions, providing confidence that the population in this study is reflective of a Caucasian population of European

descent. The results are presented as each sALS cohort analyzed individually, in addition to all sALS patients combined, unless stated otherwise.

Table 1
Demographics of study participants used for case-control association study

	Duke cases (n = 138)	Coriell cases (n = 333)	Combined (n = 471)	Controls (n = 496)
Males	76 (55.07)	188 (56.46)	264 (56.05)	256 (51.61)
Age (years)	57.36 (10.91)	54.12 (13.55)	55.05 (12.91)	55.03 (16.76)
Females	62 (44.93)	145 (43.54)	207 (43.94)	240 (48.39)
Age (years)	59.84 (9.46)	58.54 (12.82)	58.93 (11.90)	57.26 (13.47)
Site Spinal	86 (63.24)	249 (76.14)	335 (72.35)	-
Bulbar	50 (36.76)	78 (23.85)	128 (27.65)	-
Age of onset	58.47 (10.32)	56.04 (13.40)	56.75 (12.61)	-
Disease duration (months)	52.7 (39.66)	-	-	-
Note: Displayed as numbers of participants (percentage) or means (standard deviations).				

TTTA variant and sALS phenotype

We first investigated the proportion of ALS subtypes (spinal vs bulbar) within the Duke and Coriell cohorts. The Duke and Coriell cohorts differed in the proportion of spinal and bulbar onset patients ($p = 0.0066$, Table 1). Within the Duke cohort, carriage of the 9 TTTA allele was significantly higher in the bulbar onset cases ($p = 0.01$) however, this effect was not observed in the Coriell cohort ($p = 0.38$). With the significant enrichment of the 9 TTTA allele in bulbar cases within the Duke cohort we wanted to investigate whether the *NEFH* TTTA variant influenced site-specific risk for sALS.

TTTA variant and sALS disease risk

To explore the potential association between *NEFH* TTTA repeat length and risk of sALS, a case-control association study was conducted. Binary logistic regression models were used to analyze the carriage of each allele compared to all other alleles in the Duke and Coriell sALS cohorts both separately and combined (Table 2). Comparison of the 9 TTTA allele between Duke sALS cases and controls, stratified by site of disease onset, revealed reduced risk specifically for spinal-onset sALS ($p = 0.001$), whilst no effect was seen for bulbar-onset sALS ($p = 0.60$). This effect was not replicated within the Coriell ($p = 0.38$) or the combined cohort ($p = 0.03$). To determine if specific genotypes were contributing to this allelic association we followed up by investigating the genotype carriage in each individual cohort and the combined cohort (Table 3). The 9,10 genotype was associated with a 2x reduced risk for sALS in the Duke cohort, in both the naïve and the covariate-adjusted model for patient sex. However, this was not statistically significant in the Coriell or combined sALS cohort. With both the 9 TTTA allele and 9,10 genotype showing statistical significance in the Duke cohort we wanted to further investigate this association. Within the spinal cases of the Duke cohort, when considering carriage of the 9 TTTA allele, 10 TTTA allele and the 9,10 TTTA genotype; carriage of the 9 TTTA allele remained significant (corrected $p = 0.001$) and was not abrogated by carriage of the 10 TTTA allele or 9,10 TTTA genotype which were both non-significant in the joint model.

Table 2
Corrected regression models evaluating the association between *NEFH* TTTA alleles and disease risk in two sporadic ALS cohorts stratified by site of disease onset

	Duke cases			Coriell cases			Combined		
	n	OR (95% CI)	p	n	OR (95% CI)	p	n	OR (95% CI)	p
7 Spinal	34	1.46 (0.91–2.34)	0.12	82	1.09 (0.79–1.52)	0.59	116	1.18 (0.88–1.58)	0.28
Bulbar	15	0.96 (0.51–1.81)	0.91	20	0.77 (0.45–1.33)	0.35	35	0.84 (0.55–1.30)	0.44
8 Spinal	31	1.12 (0.69–1.8)	0.66	89	1.11 (0.80–1.52)	0.53	120	1.11 (0.83–1.49)	0.47
Bulbar	20	1.33 (0.73–2.42)	0.34	22	0.78 (0.46–1.32)	0.46	42	0.97 (0.64–1.47)	0.90
9 Spinal	28	0.45 (0.28–0.73)	0.001*	121	0.87 (0.64–1.18)	0.38	149	0.74 (0.56–0.97)	0.03
Bulbar	28	1.17 (0.65–2.10)	0.60	43	1.13 (0.71–1.83)	0.61	71	1.15 (0.78–1.70)	0.49
10 Spinal	49	1.52 (0.96–2.42)	0.08	119	1.06 (0.78–1.44)	0.71	168	1.16 (0.88–1.53)	0.30
Bulbar	19	0.71 (0.39–1.29)	0.26	38	1.10 (0.68–1.77)	0.70	57	0.93 (0.63–1.37)	0.71
11 Spinal	10	0.95 (0.47–1.94)	0.89	26	0.85 (0.52–1.38)	0.51	36	0.87 (0.56–1.36)	0.55
Bulbar	8	1.39 (0.62–3.11)	0.42	8	0.83 (0.38–1.81)	0.64	16	1.04 (0.58–1.88)	0.89
Data taken from binary logistic regression models with correction for patient sex									

P and OR (95% CI) values are calculated for the comparison to all other alleles.

OR, odds ratio; CI, confidence interval, *p*, statistical significance

Values in bold with * denote statistical significance after Bonferroni correction $p < 0.005$

Table 3

Corrected regression models evaluating the association between *NEFH*TTTA genotypes and disease risk in two sporadic ALS cohorts

NEFH TTTA	Duke cohort (n = 138)					Coriell cohort (n = 333)					Combined cohort (n = 471)				
	n	Naïve+		Corrected#		n	Naïve+		Corrected#		n	Naïve+		Corrected#	
		OR (95% CI)	p	OR (95% CI)	p		OR (95% CI)	p	OR (95% CI)	p		OR (95% CI)	p	OR (95% CI)	p
7,7	1	-	-	-	-	11	1.03 (0.47– 2.24)	0.95	1.01 (0.46– 2.21)	0.97	12	0.78 (0.37– 1.68)	0.53	0.78 (0.36–1.67)	0.52
7,8	12	1.48 (0.74– 2.97)	0.27	1.48 (0.73– 2.96)	0.28	21	1.04 (0.59– 1.86)	0.88	1.04 (0.59– 1.86)	0.89	33	1.17 (0.70– 1.95)	0.55	1.17 (0.70– 1.95)	0.55
7,9	15	1.17 (0.63– 2.15)	0.63	1.18 (0.64– 2.19)	0.60	25	0.78 (0.47– 1.29)	0.33	0.78 (0.47– 1.30)	0.34	40	0.88 (0.57– 1.38)	0.59	0.89 (0.57– 1.38)	0.61
7,10	20	1.62 (0.92– 2.84)	0.09	1.60 (0.91– 2.81)	0.10	43	1.42 (0.91– 2.20)	0.12	1.40 (0.90– 2.18)	0.13	63	1.48 (0.99– 2.20)	0.06	1.46 (0.98– 2.18)	0.06
7,11	3	0.90 (0.25– 3.22)	0.87	0.89 (0.25– 3.19)	0.85	5	0.61 (0.22– 1.76)	0.37	0.62 (0.22– 1.79)	0.38	8	0.70 (0.28– 1.72)	0.43	0.70 (0.28– 1.73)	0.44
8,8	4	1.03 (0.33– 3.17)	0.96	1.02 (0.33– 3.14)	0.98	13	1.40 (0.65– 3.02)	0.39	1.39 (0.65– 3.01)	0.40	17	1.29 (0.63– 2.65)	0.49	1.28 (0.62– 2.63)	0.50
8,9	14	0.94 (0.51– 1.76)	0.86	0.94 (0.50– 1.75)	0.84	37	1.05 (0.67– 1.63)	0.85	1.03 (0.66– 1.61)	0.89	51	1.02 (0.68– 1.53)	0.94	1.01 (0.67– 1.52)	0.97
8,10	18	1.23 (0.69– 2.17)	0.48	1.24 (0.70– 2.19)	0.47	27	0.72 (0.45– 1.17)	0.19	0.74 (0.45– 1.20)	0.22	45	0.87 (0.57– 1.31)	0.50	0.88 (0.58– 1.33)	0.54
8,11	4	1.03 (0.33– 3.13)	0.96	1.02 (0.33– 3.16)	0.97	15	1.62 (0.77– 3.41)	0.20	1.62 (0.77– 3.41)	0.20	19	1.45 (0.72– 2.92)	0.30	1.45 (0.72– 2.92)	0.30
9,9	10	0.59 (0.29– 1.19)	0.14	0.60 (0.30– 1.20)	0.15	38	0.97 (0.63– 1.50)	0.90	0.98 (0.63– 1.51)	0.91	48	0.86 (0.57– 1.29)	0.46	0.86 (0.57– 1.92)	0.47
9,10	13	0.53 (0.29– 0.99)	0.05*	0.52 (0.28– 0.98)	0.04*	59	1.10 (0.76– 1.60)	0.60	1.10 (0.76– 1.59)	0.62	72	0.93 (0.65– 1.32)	0.66	0.92(0.65– 1.30)	0.62
9,11	5	0.94 (0.35– 2.58)	0.91	0.95 (0.35– 2.60)	0.93	7	0.54 (0.22– 1.30)	0.17	0.54 (0.22– 1.30)	0.17	12	0.67 (0.32– 1.37)	0.26	0.66 (0.31– 1.37)	0.26
10,10	13	1.46 (0.75– 2.86)	0.27	1.49 (0.76– 2.92)	0.25	22	0.99 (0.57– 1.74)	0.98	1.01 (0.58– 1.77)	0.96	35	1.13 (0.69– 1.84)	0.64	1.15 (0.70– 1.88)	0.58
10,11	4	1.03 (0.33– 3.17)	0.96	1.01 (0.33– 3.12)	0.99	8	0.85 (0.35– 2.04)	0.71	0.82 (0.34– 1.99)	0.66	12	0.90 (0.41– 1.97)	0.79	0.87 (0.40– 1.91)	0.74

+ Data taken from binary logistic regression models without correction for covariates

Data taken from binary logistic regression models with correction for patient sex

P and OR (95% CI) values are calculated for the comparison to all other genotypes. The following genotypes were excluded from analysis due to low frequency: 6,7; 8,12; 9,15; 10,13; 11,11.

OR, odds ratio; CI, confidence interval, p, statistical significance

Values in bold with * denote statistical significance $p < 0.05$

TTTA variant and age of disease onset

To examine the effect of TTTA repeat length on age of disease onset, a general linear model was used. Due to the variability in age of onset data from both Duke and Coriell cohorts, this was investigated in the larger combined sALS cohort to help negate the potential of detecting spurious associations. Within the combined cohort, males had on average a 3.7 year earlier age of onset compared to females ($p = 0.02$). Compared to patients with spinal onset, those with the bulbar onset ALS had a 4.3 year later age of disease onset ($p = 0.001$). In the combined sALS cohort, when analyzed allelically (Table 4), carriage of the 10 TTTA allele was associated with a 2.5 year later age of disease onset in the naïve model ($p = 0.03$). This effect was not abrogated when taking into account the significant covariates including patient sex and site of disease onset, with an estimated mean difference of 2.7 years later age of disease onset following Bonferroni correction ($p = 0.02$). No *NEFH* genotypes passed the statistical significance threshold for being associated with age of disease onset (Table 4).

Table 4
Naïve generalized linear model evaluating the association between *NEFH* TTTA alleles/genotypes and age of disease onset in the combined sALS cohort

NEFH TTTA repeat length	Combined cohort (n = 471)						
	n	Naïve		95% confidence interval		Estimated means	
		β -CoE	p	Lower	Upper	Present	Absent
7	157	0.93	0.45	-1.48	3.34	56.13	57.06
8	165	0.82	0.50	-1.57	3.20	56.22	57.04
9	224	1.51	0.20	-0.77	3.78	55.96	57.47
10	227	-2.50	0.03*	-4.77	-0.23	58.05	55.55
11	53	0.06	0.97	-3.54	3.66	56.70	56.76
NEFH genotype							
7,7	12	1.29	0.73	-5.93	8.51	55.50	56.79
7,8	33	1.92	0.40	-2.54	6.37	54.97	56.89
7,9	40	2.60	0.21	-1.48	6.67	54.37	56.97
7,10	63	-1.90	0.27	-5.24	1.44	58.40	56.50
7,11	8	3.69	0.41	-5.11	12.49	53.12	56.82
8,8	17	-1.42	0.65	-7.51	4.68	58.12	56.70
8,9	51	3.22	0.08	-0.43	6.87	53.88	57.10
8,10	45	-2.80	0.16	-6.67	1.06	59.29	56.49
8,11	19	1.06	0.72	-4.72	6.84	55.74	56.80
9,9	48	-0.48	0.80	-4.24	3.28	57.19	56.70
9,10	72	-0.64	0.69	-3.80	2.53	57.29	56.66
9,11	12	-0.42	0.91	-7.64	6.80	57.17	56.74
10,10	35	-0.64	0.77	-4.97	3.70	57.34	56.71
10,11	12	-1.45	0.69	-8.67	5.77	58.17	56.72

The following genotype were excluded from analysis due to low frequency: 6,7; 8,12; 9,15; 10,13; 11,11. Values in bold with * denote statistical significance $p < 0.05$

TTTA repeat alleles and survival duration in the Duke cohort

End-point survival data were available only for the 138 sALS cases from the Duke cohort. Within this cohort there was a significantly reduced survival time with increasing age at disease onset ($p = 0.009$). Survival duration was analyzed according to TTTA allele carriage, with site of disease onset taken into consideration. There was no significant difference in end-point survival when stratified by initial site of disease

onset (Fig. 3a), nor was there any significant difference in end-point survival for carriage of the 9 or 10 TTTA alleles, or other alleles investigated (Fig. 3b-f).

Discussion

This study investigated a multi-allelic intronic TTTA structural variant in *NEFH* as a candidate risk factor for sALS, and modifier of age and site of disease onset and survival duration. Within the Duke cohort carriage of the 9 TTTA allele was significantly enriched in bulbar onset cases. Interestingly, in the Duke cohort the 9 TTTA allele was associated with a 2x reduction in sALS risk, but only in spinal onset patients. Within the spinal onset cases from the Duke cohort, the reduction in risk remained significant after considering carriage of the 10 TTTA allele or the 9,10 genotype, confirming this effect was in fact driven by the 9 TTTA allele. Despite the 9 TTTA association not replicating in the Coriell or combined cohort, it is important to note that the trend for reduced risk in the spinal onset patients was consistent in the Coriell and combined cohorts. On the other hand, there was no association for increased disease risk for any of the *NEFH* TTTA alleles/genotypes in either the Duke or Coriell cohorts, or in the combined sALS cohort.

Following the association with disease risk, *NEFH* alleles/genotypes were investigated as potential modifiers of sALS age of disease onset using general linear modelling. The association with age at disease onset was investigated in the combined sALS cohort to minimize the effect of variability across each separate sALS cohort. Carriage of the 10 TTTA allele was found to be associated with a 2.7 year later mean age at disease onset when accounting for allele carriage, sex and site of disease onset, further supporting the notion of the 9 and 10 TTTA alleles being associated with protective traits in sALS.

Lastly, *NEFH* alleles were investigated as a potential modifier of survival in the Duke cohort. Due to the limited number of patients with end-point survival data, we were only able to analyze the effect of *NEFH* alleles and not genotypes, which we acknowledge as a limitation of this study. We analyzed the carriage of *NEFH* alleles, taking into account the site of disease onset. Although the survival trajectories for spinal and bulbar onset patients did not differ within this cohort, it is important to include this covariate since short structural variants have previously been shown to stratify sALS patient sub-phenotype³³⁻³⁵. In the present study, no significant associations were detected between *NEFH* alleles and survival.

A number of possible explanations were considered for the differences in the 9 and 10 TTTA allele associations found in the two sALS cohorts. Firstly, in view of the smaller size of the Duke cohort, the possibility of a spurious association was considered, but in view of the magnitude of the associations and the follow up analyses accounting for the carriage of the 9 TTTA allele, 10 TTTA allele or the 9,10 TTTA genotype, this is considered unlikely. Differences in ethnicity were excluded on the basis of a comparison of self-reported data on ethnicity and country-of-origin from the two cohorts. However, consideration still needs to be given to differences in genetic diversity between the two cohorts, based on the sources of patient samples and geographic patterns of patient recruitment. It is pertinent in this regard that the Duke cohort were from one geographical location (Duke ALS Clinic, Durham North Carolina), whereas cases from the Coriell biobank are multicenter in origin and were recruited from geographically more diverse locations.

Currently, both light and heavy neurofilaments (pNFL and pNFH) are considered among the most promising disease biomarkers for ALS diagnosis and progression, and have been extensively reviewed^{5,36,37}. Levels of both CSF and serum pNFL and pNFH have been shown to distinguish patients with ALS from controls and other neurodegenerative diseases^{38,39}. Various studies have also reported pNFL and pNFH to be correlated with pheno-conversion, clinical measures of disease progression or the clinical subtype of ALS³⁹⁻⁴². Typically pNFL levels are considered a more sensitive prognostic marker, displaying a significant correlation between pNFL in the CSF and in serum³⁸. On the other hand, pNFH in the CSF does not correlate well with serum pNFH levels but this may be due to the masking of epitopes or post-translational modifications influencing antibody detection^{38,42}. Despite this, a study has reported significantly higher serum pNFH concentrations in pyramidal, bulbar and classic ALS phenotypes compared to flail arm ALS, primary muscular atrophy and primary lateral sclerosis subtypes, with a positive correlation between pNFH and progression rate, suggesting that faster neuronal degeneration may be a determinant of serum pNFH levels⁴³.

With evidence of *NEFH* playing a role at both the genetic level and as a prognostic marker for ALS, one must consider the potential impact of previously overlooked short structural variants within the *NEFH* gene in helping to distinguish clinical variability, particularly in sporadic ALS. A recent report has stated that 21% of sALS patients carry a genetic variant in an ALS associated gene and 13% of patients carry more than one genetic variant (including variants of unknown significance)⁴⁴. Of note, patients carrying two variants developed disease at a significantly earlier age, with variants in known ALS genes being of potential clinical importance in as many as 42% of sALS patients⁴⁴. With *NEFH* having a clear link to ALS disease and progression, future work should explore the impact of short structural variants in *NEFH* and other ALS associated genes as additional markers of disease risk and prognosis.

Conclusion

Neurofilament heavy plays a critical role in maintaining the structural integrity of the cytoskeleton within neurons, and when damaged leaks out of the axon into the CSF. Previous genetic studies of *NEFH* have primarily focused on exonic variants which are reported as a rare cause of sALS. The aim of this study was to investigate possible risk and disease-modifying effects of a candidate intronic TTTA variant in *NEFH* within two independent sALS cohorts. Our findings in the Duke cohort point to a substantial reduction in disease risk in carriers of the 9 TTTA allele, specifically in spinal-onset sALS. Collectively, the data suggests that the 9 and 10 TTTA motif lengths may have a protective advantage, potentially lowering both the risk of sALS and promoting a 2.7 year later age at disease onset, as seen in the combined sALS cohort. These results need to be replicated in larger multicenter cohorts with future studies considering more ethnically diverse populations to determine if this risk/disease modifying locus is specific to only Caucasian patients, and to validate this variant as a genetic marker for sALS.

Declarations

Author contributions

FT, RSA PAA: conception and study design. FT: data collection, statistical analysis and interpretation. RSA and IJ: statistical support. FT: initial drafting of manuscript. RB: contribution of patient DNA samples and collection of clinical information. FT, RSA, FLM, IJ, RB and PAA: critical review and editing of the manuscript. All authors read and approved the final manuscript.

Data availability

Clinical information for sALS patients and population controls from the Coriell biobank are available from <https://www.coriell.org/1/NINDS>. Any data pertaining to this article and not published within this article may be requested through collaboration.

Competing interests

The authors declare that they have no known competing interests.

Ethics approval and consent to participate

Samples from Duke ALS clinic were collected in accordance with the Health Insurance Portability and Accountability Act (Pro00040665/323682) and the Declarations of Helsinki. Retrospective genotyping of DNA samples was approved by the Human Research Ethics committee of the University of Western Australia (RA/4/20/5308).

Consent for Publication

Not applicable

Funding

This work was funded by the Perron Institute for Neurological and Translational Science and the Giumelli Foundation. The funders had no role in the preparation of the manuscript and the decision to publish.

References

1. Scott, D., Smith, K.E., O'Brien, B.J. & Angelides, K.J. Characterization of mammalian neurofilament triplet proteins. Subunit stoichiometry and morphology of native and reconstituted filaments. *J. Biol. Chem.* **260**, 10736–10747 (1985).
2. Straube-West, K., Loomis, P.A., Opal, P. & Goldman, R.D. Alterations in neural intermediate filament organization: functional implications and the induction of pathological changes related to motor neuron disease. *J. Cell Sci.* **109**, 2319–2329 (1996).
3. Theunissen, F. *et al.* New perspectives on cytoskeletal dysregulation and mitochondrial mislocalization in amyotrophic lateral sclerosis. *Transl. Neurodegener.* **10**, 1–16 (2021).
4. Wagner, O.I. *et al.* Mechanisms of mitochondria-neurofilament interactions. *J. Neurosci.* **23**, 9046–9058 (2003).
5. Zucchi, E. *et al.* Neurofilaments in motor neuron disorders: towards promising diagnostic and prognostic biomarkers. *Mol. Neurodegener.* **15**, 58 (2020).
6. Hirano, A., Donnenfeld, H., Sasaki, S. & Nakano, I. Fine structural observations of neurofilamentous changes in amyotrophic lateral sclerosis. *J. Neuropathol. Exp. Neurol.* **43**, 461–470 (1984).

7. Sobue, G. *et al.* Phosphorylated high molecular weight neurofilament protein in lower motor neurons in amyotrophic lateral sclerosis and other neurodegenerative diseases involving ventral horn cells. *Acta Neuropathol.* **79**, 402–408 (1990).
8. Itoh, T. *et al.* Phosphorylated high molecular weight neurofilament protein in the peripheral motor, sensory and sympathetic neuronal perikarya: system-dependent normal variations and changes in amyotrophic lateral sclerosis and multiple system atrophy. *Acta Neuropathol.* **83**, 240–245 (1992).
9. Sihag, R.K., Inagaki, M., Yamaguchi, T., Shea, T.B. & Pant, H.C. Role of phosphorylation on the structural dynamics and function of types III and IV intermediate filaments. *Exp. Cell Res.* **313**, 2098–2109 (2007).
10. Didonna, A. & Opal, P. The role of neurofilament aggregation in neurodegeneration: lessons from rare inherited neurological disorders. *Mol. Neurodegener.* **14**, 19 (2019).
11. Pant, H.C. Dephosphorylation of neurofilament proteins enhances their susceptibility to degradation by calpain. *Biochem. J.* **256**, 665–668 (1988).
12. Ge, W.-W., Leystra-Lantz, C., Wen, W. & Strong, M.J. Selective loss of trans-acting instability determinants of neurofilament mRNA in amyotrophic lateral sclerosis spinal cord. *J. Biol. Chem.* **278**, 26558–26563 (2003).
13. Bergeron, C. *et al.* Neurofilament light and polyadenylated mRNA levels are decreased in amyotrophic lateral sclerosis motor neurons. *J. Neuropathol. Exp. Neurol.* **53**, 221–230 (1994).
14. Menzies, F.M. *et al.* Selective loss of neurofilament expression in Cu/Zn superoxide dismutase (SOD1) linked amyotrophic lateral sclerosis. *J. Neurochem.* **82**, 1118–1128 (2002).
15. Strong, M.J., Leystra-Lantz, C. & Ge, W.-W. Intermediate filament steady-state mRNA levels in amyotrophic lateral sclerosis. *Biochem. Biophys. Res. Commun.* **316**, 317–322 (2004).
16. Hawley, Z.C.E., Campos-Melo, D. & Strong, M.J. MiR-105 and miR-9 regulate the mRNA stability of neuronal intermediate filaments. Implications for the pathogenesis of amyotrophic lateral sclerosis (ALS). *Brain Res.* **1706**, 93–100 (2019).
17. Campos-Melo, D., Hawley, Z.C.E. & Strong, M.J. Dysregulation of human NEFM and NEFH mRNA stability by ALS-linked miRNAs. *Mol. Brain.* **11**, 43 (2018).
18. Strong, M.J. Neurofilament metabolism in sporadic amyotrophic lateral sclerosis. *J. Neurol Sci.* **169**, 170–177 (1999).
19. Wong, N.K.Y., He, B.P. & Strong, M.J. Characterization of neuronal intermediate filament protein expression in cervical spinal motor neurons in sporadic amyotrophic lateral sclerosis (ALS). *J. Neuropathol. Exp. Neurol.* **59**, 972–982 (2000).
20. Zucchi, E. *et al.* A motor neuron strategy to save time and energy in neurodegeneration: adaptive protein stoichiometry. *J. Neurochem.* **146**, 631–641 (2018).
21. Figlewicz, D.A. *et al.* Variants of the heavy neurofilament subunit are associated with the development of amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **3**, 1757–1761 (1994).
22. Figlewicz, D.A., Rouleau, G.A., Krizus, A. & Julien, J.-P. Polymorphism in the multi-phosphorylation domain of the human neurofilament heavy-subunit-encoding gene. *Gene* **132**, 297–300 (1993).
23. Al-Chalabi, A. *et al.* Deletions of the heavy neurofilament subunit tail in amyotrophic lateral sclerosis. *Human Mol. Genet.* **8**, 157–164 (1999).
24. Skvortsova, V. *et al.* Analysis of heavy neurofilament subunit gene polymorphism in Russian patients with sporadic motor neuron disease (MND). *Eur. J. Hum. Genet.* **12**, 241–244 (2004).
25. Tripolszki, K. *et al.* Comprehensive genetic analysis of a Hungarian amyotrophic lateral sclerosis cohort. *Front. Genet.* **10**, 732 (2019).
26. Tomkins, J. *et al.* Novel insertion in the KSP region of the neurofilament heavy gene in amyotrophic lateral sclerosis (ALS). *Neuroreport* **9**, 3967–3970 (1998).
27. Lin, F. *et al.* Sequencing of neurofilament genes identified NEFH Ser787Arg as a novel risk variant of sporadic amyotrophic lateral sclerosis in Chinese subjects. *BMC Medical Genom.* **14**, 222 (2021).
28. Al-Chalabi, A. & Miller, C.C.J. Neurofilaments and neurological disease. *Bioessays* **25**, 346–355 (2003).
29. Theunissen, F. *et al.* Structural Variants May Be a Source of Missing Heritability in sALS. *Front. Neurosci.* **14**(2020).
30. Theunissen, F., Flynn, L.L., Anderton, R.S. & Akkari, P.A. Short structural variants as informative genetic markers for ALS disease risk and progression. *BMC Med.* **20**, 11 (2022).
31. Saul, R., Lutz, M.W., Burns, D.K., Roses, A.D. & Chiba-Falek, O. The SSV evaluation system: a tool to prioritize short structural variants for studies of possible regulatory and causal variants. *Hum. Mutat.* **37**, 877–883 (2016).
32. Brooks, B.R. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. *J. Neurol Sci.* **124**, 96–107 (1994).

33. Theunissen, F. *et al.* Novel STMN2 variant linked to amyotrophic lateral sclerosis risk and clinical phenotype. *Front. Aging Neurosci.* **13**, 127 (2021).
34. Pytte, J. *et al.* Association of a structural variant within the SQSTM1 gene with amyotrophic lateral sclerosis. *Neurol. Genet.* **6**(2020).
35. Al Khleifat, A. *et al.* Structural variation analysis of 6,500 whole genome sequences in amyotrophic lateral sclerosis. *NPJ Genom. Med.* (2021).
36. Gagliardi, D. *et al.* Diagnostic and prognostic role of blood and cerebrospinal fluid and blood neurofilaments in amyotrophic lateral sclerosis: a review of the literature. *Int. J. Mol. Sci.* **20**, 4152 (2019).
37. Falzone, Y.M. *et al.* Current application of neurofilaments in amyotrophic lateral sclerosis and future perspectives. *Neural Regen. Res.* **16**, 1985 (2021).
38. Halbgebauer, S. *et al.* Comparison of CSF and serum neurofilament light and heavy chain as differential diagnostic biomarkers for ALS. *J Neurol. Neurosurg. Psychiatry.* (2021).
39. Verde, F. *et al.* Neurofilament light chain in serum for the diagnosis of amyotrophic lateral sclerosis. *J Neurol. Neurosurg. Psychiatry.* **90**, 157–164 (2019).
40. De Schaepdryver, M. *et al.* Comparison of elevated phosphorylated neurofilament heavy chains in serum and cerebrospinal fluid of patients with amyotrophic lateral sclerosis. *J Neurol. Neurosurg. Psychiatry.* **89**, 367–373 (2018).
41. Poesen, K. *et al.* Neurofilament markers for ALS correlate with extent of upper and lower motor neuron disease. *Neurology* **88**, 2302 (2017).
42. Benatar, M. *et al.* Neurofilaments in pre-symptomatic ALS and the impact of genotype. *Amyotroph. Lateral Scler. Frontotemporal Degener.* **20**, 538–548 (2019).
43. Falzone, Y.M. *et al.* Serum phosphorylated neurofilament heavy-chain levels reflect phenotypic heterogeneity and are an independent predictor of survival in motor neuron disease. *J. Neurol.* **267**, 2272–2280 (2020).
44. Shephard, S.R. *et al.* Value of systematic genetic screening of patients with amyotrophic lateral sclerosis. *J Neurol. Neurosurg. Psychiatry.* **92**, 510–518 (2021).

Figures

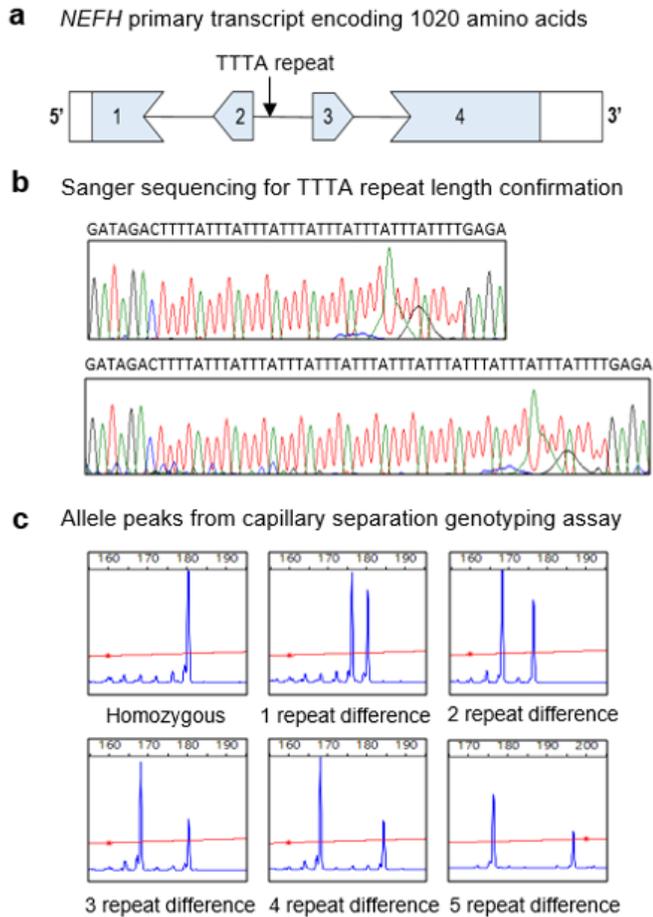


Figure 1

Identification and wet lab validation of polymorphic TTTA structural variant in *NEFH*. (a) The primary *NEFH* transcript and location of TTTA repeat 255bp past exon 2. (b) Sanger sequencing demonstrating variable length of the TTTA repeat motif. (c) Representation on *NEFH* alleles from capillary separation genotyping assay. The blue peaks depict the fluorescent signal intensity with each peak corresponding to the size of the PCR product capturing the *NEFH* TTTA variant.

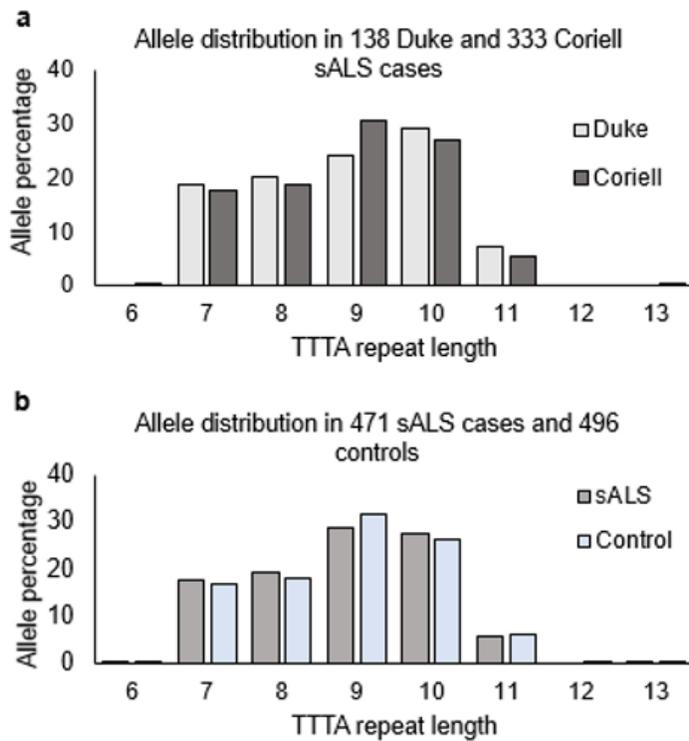


Figure 2

Allele distributions for the *NEFH* TTTA structural variant in two North American Caucasian cohorts (a) Comparison of TTTA allele distribution of 138 Duke sALS cases and 333 Coriell sALS cases. (b) TTTA allele distribution in the 471 combined sALS cohort compared to 496 controls. Only one sALS patient in the study had a TTTA repeat length of 15 (not shown in figure).

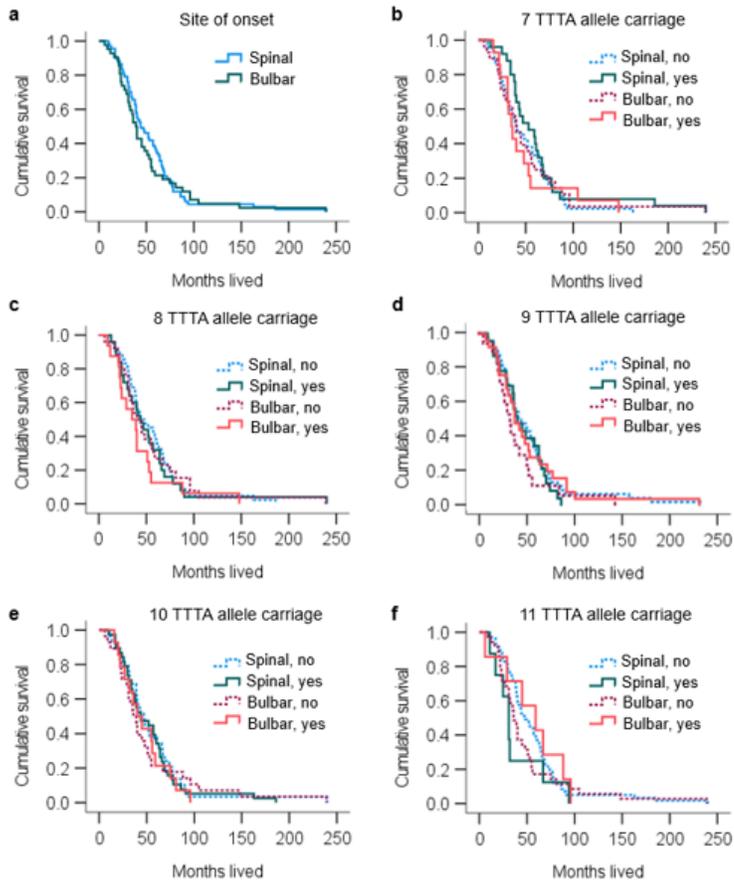


Figure 3

Cumulative survival curves for 138 sALS patients from Duke University (a) Survival according to initial site of disease onset. (b-f) Survival by presence (yes) or absence (no) of *NEFH* TTTA alleles and the initial site of sALS disease onset.