Intermediate *HTT* CAG repeats worsen disease severity in amyotrophic lateral sclerosis

INTRODUCTION

Recent research has indicated a connection between amyotrophic lateral sclerosis (ALS) and Huntington's disease (HD), an inherited neurological condition caused by a trinucleotide CAG repeat expansion within exon 1 of the Huntingtin gene (*HTT*; MIM:613004).[1] The same pathogenic CAG repeat expansions (40 or more CAG repeats) observed in patients with HD have been identified in patients with frontotemporal dementia (FTD)/ALS.[1] Similarly, non-pathogenic intermediate-length CAG repeats in the *ATXN2* gene are a well-established factor associated with increased ALS risk and faster disease progression.[2] Given these observations, we investigated the impact of intermediate *HTT* alleles on survival in two cohorts of patients diagnosed with ALS.

METHODS

Discovery cohort

The study population consisted of 1181 patients with ALS identified through the Piemonte and Valle d'Aosta Register for ALS (PARALS) (online supplemental etable 1). Among these, 996 samples were part of the original report of *HTT* repeat expansions in individuals with FTD/ALS.[3] The characteristics of the PARALS register are described in online supplemental materials. These subjects were negative for pathogenic mutations in *C9orf72*, *SOD1*, *TARDBP* and *FUS*.

Replication cohort

As a replication cohort, we used clinical and genomic data from the Answer ALS database (https://dataportal.answerals.org/home). We retrieved whole-genome sequence data from 376 patients with ALS who did not have mutations in the four major ALS genes (online supplemental etable 1). The data processing pipeline is described in online supplemental materials.

HTT CAG repeat analysis

We used ExpansionHunter (V.5.0.0) to estimate repeat lengths of *HTT* expansions from the whole-genome sequence data. *HTT* CAG alleles were deemed 'fully penetrant' if they carried 40 or more repeats, 'incompletely penetrant' if they carried between 36 and 39 repeats and were designated as 'intermediate' if they carried between 27 and 35 repeats. Healthy subjects typically have 26 or fewer CAG repeats.

Statistical methods

Survival was calculated from symptom onset to death, tracheostomy or censoring date (31 December 2021 in the PARALS cohort; last available follow-up in the Answer ALS cohort). Survival times were calculated using the Kaplan-Meier method. Multivariate survival analysis was performed using the Cox proportional hazards model, modelling the presence of intermediate expansion as a binary variable. In a separate additional analysis, we included the number of *HTT* CAG repeats as a continuous variable. All models were adjusted for relevant clinical predictors; sensitivity analyses were conducted to evaluate the effect of population structure, haplogroups or genetic modifiers of HTT alleles. Further details on statistical methods and analysis are given in Supplementary Materials.

Data availability

The individual-level sequence data are available on the dbGaP web portal, accession number phs001963.[3] The clinical data are available on reasonable request by interested researchers. The programming code used to analyse the data is available at https://github.com/maurigrassano/CAG_HTT_in_ALS.

RESULTS

PARALS cohort

In the PARALS cohort, we discovered one patient with ALS carrying a pathogenic *HTT* allele (40 repeats) (online supplemental etable 2); this sample was not reported in the Dewan *et al* paper.[3] In addition, we identified four ALS cases carrying incomplete penetrance CAG alleles (36–39 repeats, representing 0.33% of the cohort). These five subjects were excluded from the analyses.

We identified 79 (6.67%) ALS cases in the discovery cohort who carried an intermediate-length *HTT* CAG allele (ie, 27–35 repeats). None of these patients manifested atypical symptoms or clinical features distinctive of HD; their clinical characteristics are reported in online supplemental etable 3. Patients carrying intermediate *HTT* alleles had a median survival time of 29.3 months compared with the 34.5 months of those without (HR = 1.37, 95% CI 1.03 to 1.82, p=0.0318, figure 1A). There was no relationship between the number of HTT CAG repeats (8–35 repeats as continuous variables) and survival time (p=0.2597). The effect of HTT alleles on survival was unrelated to population structure or haplogroups (online supplemental efigure 1).

Answer ALS cohort

Within the replication cohort, we detected 30 (7.98%) ALS cases carrying *HTT* expansion in the 27–35 range (online supplemental etable 2). Their clinical characteristics were comparable to the rest of the cohort (online supplemental etable 4). Survival analysis confirmed that these patients had a worse prognosis, with a median survival time of 29.5 months compared with the 56.4 months observed in patients without intermediate *HTT* alleles (HR=1.85, 95% CI 1.06 to 3.25, p=0.0310, figure 1B). No correlation was observed between the number of *HTT* CAG repeats and survival duration (p=0.1048).

DISCUSSION

This study found that patients with ALS carrying intermediate-length *HTT* expansions had a more severe disease course characterised by reduced survival rates. Our study underscores the importance of genetic factors in determining the natural history of ALS and supports the notion that additional studies in this area should be prioritised.

Our study has limitations, primarily the small number of cases carrying the intermediate *HTT* expansions available for phenotype analysis. Nevertheless, we replicated our findings in an independent cohort, suggesting that the *HTT* gene does influence ALS survival. Thus, this study is an additional example of the link connecting *HTT* with multiple neurodegenerative phenotypes and longevity (see online supplemental etable 5). It remains unclear whether the detrimental effects of CAG expansions in the *HTT* gene on ALS result from a direct interaction between poly-Q residues and TDP-43, or from other potential mechanisms.[3 4] Further research is needed to elucidate the precise pathways involved. Notably, an enhancement of TDP-43 aggregation and toxicity has been hypothesised for polyQ expansions within ataxin 2 (encoded by the *ATXN2* gene). Intriguingly, therapeutic lowering of ataxin 2 reduces TDP-43 aggregation, improving survival and motor function in a disease model.[5] This observation led to the development of gene-directed therapy for ALS targeting

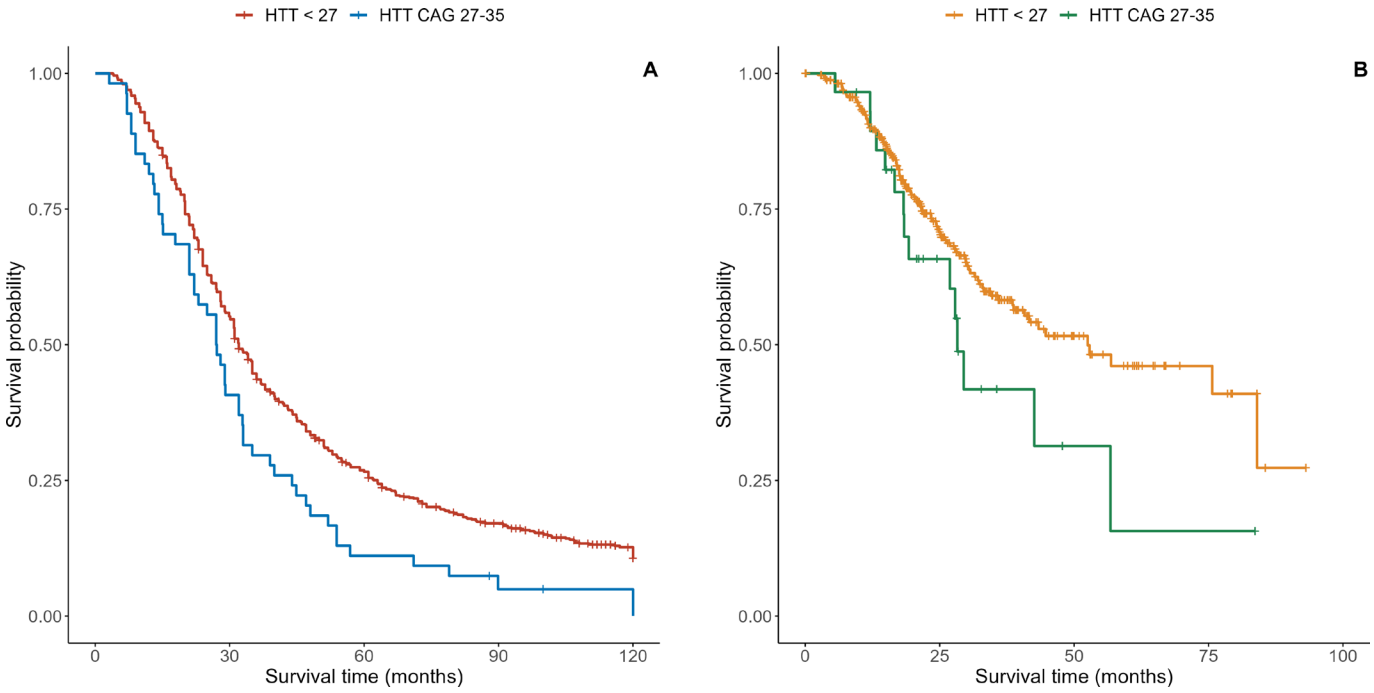


Figure 1 Survival from disease onset according to *HTT* CAG intermediate number of repeats in the Piemonte and Valle d'Aosta Register for ALS (PARALS) (A) and Answer amyotrophic lateral sclerosis (ALS) (B) cohort. (A) PARALS cohort (n = 1181). The blue line represents the survival of 79 patients with ALS carrying 27–35 CAG repeats (ie, intermediate length), and the red line represents the survival of 1102 patients with ALS carrying 26 or fewer CAG repeats (ie, normal length). (B) Answer ALS cohort (n = 476). The green line represents the survival of 30 patients with ALS carrying 27–35 CAG repeats (ie, intermediate length), and the orange line represents the survival of 346 patients with ALS carrying 26 or fewer CAG repeats (ie, normal length).

polyQ repeats (ClinicalTrials.gov Identifier: NCT04494256).

In conclusion, our study contributes to the growing evidence linking *HTT* expansions to ALS pathology. It emphasises the crucial role of genetics in shaping disease progression and opens new avenues for intervention and treatment strategies, offering hope for improved outcomes for patients with ALS.

Maurizio Grassano ,[1,2] **Antonio Canosa** ,[1] **Sandra D'Alfonso,**[3] **Lucia Corrado** ,[3] **Giorgia Brodini,**[1] **Emanuele Koumantakis,**[4] **Paolo Cugnasco,**[1] **Umberto Manera** ,[5] **Rosario Vasta** ,[1] **Francesca Palumbo** ,[1] **Letizia Mazzini** ,[6,7] **Salvatore Gallone** ,[8] **Cristina Moglia** ,[1] **Ramita Dewan,**[2] **Ruth Chia** ,[2] **Jinhui Ding,**[2] **Clifton Dalgard** ,[9,10] **Raphael J Gibbs,**[2] **Sonja Scholz** ,[11,12] **Andrea Calvo** ,[1] **Bryan Traynor** ,[2,13] **Adriano Chio** [1]

[1]Department of Neurosciences Rita Levi Montalcini, University of Turin, Torino, Piemonte, Italy
[2]National Institute on Aging Laboratory of Neurogenetics, Bethesda, Maryland, USA
[3]Department of Health Sciences Interdisciplinary Research Center of Autoimmune Diseases, University of Eastern Piedmont Amedeo Avogadro School of Medicine, Novara, Piemonte, Italy
[4]Department of Public Health and Pediatrics, University of Turin, Torino, Italy
[5]'Rita Levi Montalcini' Department of Neuroscience, University of Turin, Torino, Piemonte, Italy
[6]University Hospital Maggiore della Carità, Novara, Piemonte, Italy
[7]University of Eastern Piedmont Amedeo Avogadro School of Medicine, Novara, Piemonte, Italy
[8]Azienda Ospedaliero Universitaria Città della Salute e della Scienza di Torino, Torino, Piemonte, Italy
[9]Department of Anatomy, Physiology & Genetics, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA
[10]The American Genome Center, Collaborative Health Initiative Research Program, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA
[11]National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, USA
[12]Department of Neurology, Johns Hopkins University, Baltimore, Maryland, USA
[13]Johns Hopkins University, Baltimore, Maryland, USA

Correspondence to Dr Maurizio Grassano; maurizio.grassano@unito.it

Acknowledgements The authors thank the Laboratory of Neurogenetics (NIH) and ALS Centers staff for their collegial support and technical assistance.

Contributors Concept and design: MG, ACanosa, BT, and AChio. Acquisition, analysis, or interpretation of data: all authors. Drafting of the manuscript: MG, ACanosa, SS, ACalvo, BT, and AChio. Critical review of the manuscript for important intellectual content and administrative, technical, or material support: SD, LC, GB, EK, PC, UM, RV, FP, LM, SG, CM, RD, RC, JD, CD and RJG. Statistical analysis: MG and EK. Obtained funding: SS, BT and AChio. Supervision: BT and AChio.

Funding This work was in part supported by the Italian Ministry of Health (Ministero della Salute, Ricerca Sanitaria Finalizzata, grant RF-2016-02362405), the European Commission's Health Seventh Framework Programme (FP7/2007-2013 under grant agreement 259867), the Italian Ministry of Education, University and Research (Progetti di Ricerca di Rilevante Interesse Nazionale [PRIN] grant 2017SNW5MB), the Joint Programme–Neurodegenerative Disease Research (ALS-Care, Strength and Brain-Mend projects), granted by the Italian Ministry of Education, University and Research, the Horizon 2020 Programme (project Brainteaser under grant agreement 101017598), the Fondazione Mario e Anna Magnetto and the Thierry Latran Foundation (INSPIRED project). This work was also supported in part by the Intramural Research Programs of the National Institute on Aging (grant Z01-AG000933) and the National Institute of Neurological Disorders and Stroke (grant ZIANS003154). This study was performed under the Department of Excellence grant of the Italian Ministry of Education, University and Research to the 'Rita Levi Montalcini' Department of Neuroscience, University of Torino, Italy, and to the Department of Health Sciences, University of Eastern Piedmont, Novara, Italy. The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

Competing interests AChio serves on scientific advisory boards for Mitsubishi Tanabe, Roche, Biogen, Denali Pharma, AC Immune, Biogen, Lilly, and Cytokinetics and has received a research grant from Biogen. BT holds the US, Canadian and European patents on the clinical testing and therapeutic intervention for the hexanucleotide repeat expansion in C9orf72. BT and SS received research support from Cerevel Therapeutics. ACalvo has received a research grant from Cytokinetics. MG has received grants from the American Academy of Neurology, the American Brain Foundation and the ALS Association.

Patient consent for publication Not applicable.

Ethics approval This study was approved by Comitato Etico Azienda Ospedaliero-Universitaria Città della Salute e della Scienza, Torino (protocol number #0038876). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/jnnp-2024-333998).

MG and AC contributed equally.
BT and AC contributed equally.

To cite Grassano M, Canosa A, D'Alfonso S, *et al*. *J Neurol Neurosurg Psychiatry* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/jnnp-2024-333998

Received 8 April 2024
Accepted 23 July 2024

J Neurol Neurosurg Psychiatry 2024;**0**:1–3. doi:10.1136/jnnp-2024-333998

ORCID iDs
Maurizio Grassano http://orcid.org/0000-0001-6714-6897
Antonio Canosa http://orcid.org/0000-0001-5876-4079
Lucia Corrado http://orcid.org/0000-0002-1737-8267
Umberto Manera http://orcid.org/0000-0002-9995-8133
Rosario Vasta http://orcid.org/0000-0002-0393-4736
Francesca Palumbo http://orcid.org/0000-0002-7070-2409
Letizia Mazzini http://orcid.org/0000-0003-2479-9120
Salvatore Gallone http://orcid.org/0000-0002-3156-4613
Cristina Moglia http://orcid.org/0000-0001-7377-7222
Ruth Chia http://orcid.org/0000-0002-4709-7423
Clifton Dalgard http://orcid.org/0000-0003-2025-8239
Sonja Scholz http://orcid.org/0000-0002-6623-0429
Andrea Calvo http://orcid.org/0000-0002-5122-7243
Bryan Traynor http://orcid.org/0000-0003-0527-2446
Adriano Chio http://orcid.org/0000-0001-9579-5341

REFERENCES

1 Dewan R, Chia R, Ding J, *et al*. Pathogenic Huntingtin Repeat Expansions in Patients with Frontotemporal Dementia and Amyotrophic Lateral Sclerosis. *Neuron* 2021;109:448–60.
2 Chio A, Moglia C, Canosa A, *et al*. Exploring the phenotype of Italian patients with ALS with intermediate *ATXN2* polyQ repeats. *J Neurol Neurosurg Psychiatry* 2022;93:1216–20.
3 Fuentealba RA, Udan M, Bell S, *et al*. Interaction with polyglutamine aggregates reveals a Q/N-rich domain in TDP-43. *J Biol Chem* 2010;285:26304–14.
4 Coudert L, Nonaka T, Bernard E, *et al*. Phosphorylated and aggregated TDP-43 with seeding properties are induced upon mutant Huntingtin (mHtt) polyglutamine expression in human cellular models. *Cell Mol Life Sci* 2019;76:2615–32.
5 Becker LA, Huang B, Bieri G, *et al*. Therapeutic reduction of ataxin-2 extends lifespan and reduces pathology in TDP-43 mice. *Nat New Biol* 2017;544:367–71.

Supplementary Files

Table of Contents

Supplementary Methods.

PARALS Cohort

Subjects

The study population consisted of incident patients identified through the Piemonte and Valle d'Aosta Register for ALS (PARALS), diagnosed with definite, probable, and probable laboratory-supported ALS according to the El Escorial revised diagnostic criteria[1] between January 1st 2007 and December 31st 2019. The PARALS is a prospective epidemiologic registry established in 1995 in Piemonte and Valle d'Aosta regions in Northern Italy.[2] Patients were followed at the ALS Centers in Turin and Novara (Piemonte, Italy).

Neuropsychological evaluation

A comprehensive battery of neuropsychological tests, administered by experienced neuropsychologists with a deep understanding of ALS, was conducted on each patient.3 All study participants were rigorously tested at diagnosis or during the first follow-up visit (2 months later). Patients' cognitive status was meticulously classified according to the revised ALS-FTD Consensus Criteria[4] into five categories: ALS with normal cognition (ALS-CN); ALS with behavioral impairment (ALS-Bi); ALS with cognitive impairment (ALS-Ci); ALS with cognitive and behavioral impairment (ALS-Cbi) ALS with FTD (ALS-FTD).

Whole-genome sequencing.

Genomic DNA was extracted from whole blood using a Maxwell RSC Instrument (Promega Corp., Madison, WI, USA). Fluorometric quantitation of the genomic DNA samples was performed using the PicoGreen dsDNA assay (Thermo Fisher). PCR-free, paired-end libraries were constructed by automated liquid handlers using the Illumina TruSeq chemistry according to the manufacturer's protocol. DNA samples underwent sequencing on an Illumina HiSeq X Ten sequencer (v.2.5 chemistry, Illumina) using 150 bp, paired-end cycles. All samples were sequenced using the same platform.

Sequence alignment, variant calling.

Genome sequence data were processed using the pipeline standard developed by the Centers for Common Disease Genomics (CCDG; https://www.genome.gov/27563570/). The GRCh38 reference genome was used for alignment, as per the CCDG standard. The Broad Institute's implementation of the functional equivalence standardized pipeline was used for whole-genome sequence alignments and processing. This pipeline, which incorporates the GATK (2016) Best Practices, was implemented in the workflow description language (WDL) for deployment and execution on the Google Cloud Platform. Single-nucleotide variants (SNV) and insertion-deletions (InDels) variants were called from the processed whole-genome sequence data following the GATK Best Practices using another Broad Institute workflow for joint discovery and Variant Quality Score Recalibration (VQSR). Workflows for WGS sample processing and joint discovery are publicly available (https://github.com/gatk-workflows/broad-prod-wgs-germline-snps-indels). All whole-genome sequence data were processed using the same pipeline.

Quality control.

For sample-level quality control checks, genomes were excluded from the analysis for the following reasons: (1) a high contamination rate (>5%), (2) an excessive heterozygosity rate (exceeding +/− 0.15 F-statistic), (3) a low call rate (≤ 95%), (4) discordance between reported sex and genotypic sex, (5) duplicate samples (determined by pi-hat statistics > 0.8), (6) non-European ancestry based on principal

components analysis when compared to the HapMap 3 Genome Reference Panel, and (7) samples that were related (defined as having a pi-hat > 0.125).

For variant-level quality control, exclusions were made for: (1) variants with non-random missing data between cases and controls ($P \leq 1 \times 10{-}4$), (2) variants showing haplotype-based non-random missing data ($P \leq 1 \times 10{-}4$), (3) variants with a missingness rate of 5% or more, (4) non-autosomal variants (chromosomes X, Y, and mitochondrial), (5) variants significantly deviating from Hardy-Weinberg equilibrium in controls ($P \leq 1 \times 10{-}6$), (6) variants located in VDJ recombination sites and centromeric regions +/− 10 kb, (7) variants with allele frequencies in the aged control group (Wellderly cohort) that significantly differed from other controls (FDR-corrected chi-square test $P < 0.05$), (8) variants whose minor allele frequencies in our control groups significantly differed from those reported in the NHLBI Trans-Omics TOPMed database (freeze 5b; www.nhlbiwgs.org) or gnomAD (version 3.0) (FDR-corrected chi-square test $P < 0.05$), (9) variants that failed TOPMed's variant calling criteria, and (10) spanning deletions.

Repeat expansion analysis

ExpansionHunter - Targeted software (version 3.0.1) was used to estimate repeat lengths of known, disease-causing expansions from whole-genome sequencing data. This algorithm has been validated using experimentally confirmed samples carrying pathogenic expansions, including HTT[5]. ExpansionHunter explicitly models the *HTT* CAG and CCG repeats and the interrupting sequence region in the gene.

Principal component analysis

To examine population structure and account for underlying population stratification, we generated principal-component analysis (PCA) from common single nucleotide variants using flashPCA (github.com/gabraham/flashpca) as described elsewhere[6].

Definitions of HTT haplotypes

For haplotype definition, individuals with at least one repeat allele of 27–35 CAG were evaluated at SNPs across HTT as previously described[7,8,9]. The selection of variants, mainly single-nucleotide polymorphisms (SNPs), and samples initially used to characterize HTT haplotypes on Huntington's Disease (HD) expanded chromosomes and normal chromosomes were described elsewhere[10]

Genetic modifiers of *HTT* CAG phenotype

We evaluated variants that reportedly affect the age of onset of Huntington's Disease. We assessed the sequence downstream of the CAG repeat in HTT: the variant that results in complete loss of interrupting (LOI) adenine nucleotides in this region (associated with dramatically earlier HD age at onset) was evaluated assessed by evaluating perfect tag SNPs (rs193119731, rs145048189, rs73198489 and rs73200492); the duplicated (CAA-CAG)2 variant (associated with later age at onset) was instead assessed through a SNP in perfect linkage disequilibrium (rs10006977)[11]. Additionally, we evaluated common (MAF > 5%) genome-wide significant loci associated with HD age at onset after (rs34017474, rs1799977, rs35811129, rs701383, rs3791767, rs274883, rs74302792, and rs79136984)[12] (data not shown).

Answer ALS Cohort

Subjects

Participants were recruited from eight neuromuscular clinics distributed across the USA (Johns Hopkins University, Massachusetts General Hospital, Ohio State, Emory University, Washington University, Northwestern University, Cedars-Sinai and Texas Neurology). The study was approved by local

institutional review boards. All subjects provided written informed consent that was uniform across all sites and included an agreement for the broad sharing of data for medical research purposes. Subjects with sALS, fALS and related MNDs (referred to as non-ALS MNDs), including those with primary lateral sclerosis, progressive bulbar palsy and progressive muscular atrophy, along with asymptomatic ALS gene mutation carriers, were enrolled in AALS. Additional enrollment details are provided elsewhere[13].

Whole-genome sequencing and analysis.
Whole blood was also collected in EDTA tubes and sent to the New York Genome Center (NYGC) (https://www.nygenome.org/) for DNA extraction, whole genome sequencing and sample QC. Whole-genome sequencing libraries were prepared, and sequencing was performed on an Illumina NovaSeq 6000 sequencer using 2X150 bp cycles. Sequence data were processed on an NYGC automated pipeline. Sequence runs were assessed, and only high-quality FASTQ data (exhibiting a 99.9% base call accuracy) were processed. Paired-end reads were aligned to the GRCh38 human reference using the BurrowsWheeler Aligner (BWA-MEMv0.7.8) and processed using the GATK best-practices workflow, which includes marking of duplicate reads by the use of Picard tools (v1.83, http://picard.sourceforge.net), local realignment around indels, and base quality score recalibration (BQSR) via Genome Analysis Toolkit (GATK v3.4.0)

Data cleaning in the Answer ALS dataset
The data cleaning process was implemented as follows:
1. Data download. From the ANSWER ALS Data Portal (AALS-01184, https://dataportal.answerals.org/, accessed February 14, 2024)., we downloaded whole-genome sequencing data (in cram format) of ALS patients with: (a) definite, probable or possible ALS; (b) less than 25 hexanucleotide repeats in the *C9orf72* gene; (c) negative for pathogenic mutations in *SOD1*, *TARDBP* and *FUS* genes (starting patients=439).
2. Dataset Merging.: The datasets - subjects, AALSHXFX, ANSASFD, ALSFRS_R, Demographics, and Vital_Capacity - were merged using the "SubjectUID" to match baseline values across datasets. The baseline was defined as the point of patient entry into the study.
3. Survival status and survival time definition: We ensured the availability of sex and survival status for all participants. Survival time was calculated from the date of diagnosis to the occurrence of patient death, tracheostomy, or the last known follow-up. Patients alive at the last follow-up were considered censored. Survival time (converted to months) and status (coded as death=1, censored=0) were derived from the Mortality and Tracheostomy dataset.
4. Site of onset definition. Onset sites were then categorized using the values of hxblb, hxax, hxli, and hxot variables (onset site bulbar, axial, limb, or other, respectively). Cases with unavailable or ambiguous onset site information were excluded (remaining patients = 439).
4. Missing covariates. Participants lacking baseline ALSFRS-R scores, diagnostic delay (time between onset and diagnosis), time from diagnosis to study enrollment or baseline respiratory function (FVC% or SVC%) were removed (remaining patients=376).
5. Respiratory Function. To preserve cohort size and allow a robust analysis, we incorporated respiratory function into the model as a categorical variable defined as follows: (a) the highest FVC or SVC trial was used to assess the respiratory domain; (b) three categories were created based on the percentage of the predicted values: <50%, 50%-70%, >70%.

Statistical Analysis

In the PARALS cohort, survival was calculated from symptom onset to death, tracheostomy, or censoring date (December 31, 2021); in the Answer ALS cohort, survival was calculated from symptom onset to death, tracheostomy, or last available follow-up. Survival times were calculated using the Kaplan-Meier method. Survival analysis was performed using the Cox proportional hazards model (stepwise backwards) with a $P < 0.1$ retention criterion, modelling the presence of intermediate alleles (IA) as a binary variable. In a separate additional analysis, we included the number of *HTT* CAG repeats as a continuous variable. All models were adjusted for the following clinical predictors: age at symptom onset, diagnostic delay (defined as the time from symptom onset to diagnosis), bulbar onset, presence of frontotemporal dementia (for the PARALS cohort) and the ALS Functional Rating Scale-Revised (ALSFRS-R) score assessed either at the time of diagnosis (for the PARALS cohort) or at enrollment (for the Answer ALS cohort). Additionally, we adjusted for respiratory function, measured as the percentage of predicted forced vital capacity (FVC%) at diagnosis for the PARALS cohort and the percentage of slow vital capacity (SVC%) at enrollment for the Answer ALS cohort.

First, we performed a sensitivity analysis to exclude the possibility that the observed effect resulted from a genetic relationship between subjects or population substructures. Therefore, we repeated the analyses, incorporating as covariates in our model the first five principal components generated from common single nucleotide variants.

Next, we investigated whether *HTT* haplotypes or variants interrupting the CAG repeat tracts modified the clinical impact of *HTT* intermediate alleles on ALS survival. We fitted survival models incorporating interaction terms between *HTT* intermediate alleles and haplotypes and *HTT* intermediate alleles and downstream variants.

Finally, to evaluate the potential non-linear effects of *HTT* CAG repeat size variations, we used a model with *HTT* CAG repeat sizes (longer allele) and a quadratic term for the *HTT* allele as explanatory variables.

Results of sensitivity analysis are presented in eFigure 1.

Statistical analyses were carried out using R (version 4.0) using the survfit() or coxph() function of the survival package (version 3.5-5).

eTable 1. Demographic characteristics of the study cohorts.

	PARALS (n = 1,181)	**Answer ALS (n = 376)**
Mean age (SD), years	66.7 (10.6)	60.1 (11.00)
Sex (female, %)	551 (46.7%)	129 (34.3%)
Site of disease onset (bulbar, %)	393 (33.3%)	89 (23.7%)

SD = Standard Deviation.

eTable 2. Number of patients with *HTT* repeats ≥ 27 categorized according to longest allele.

Longest repeat	PARALS (n = 1,181)	Answer ALS (n = 376)
≤ 26	1,102	346
27	26	12
28	10	6
29	12	3
30	12	4
31	9	3
32	4	1
33	3	0
34	2	1
35	2	0
36	1	0
37	1	0
38	1	0
39	1	0
40	1	0

eTable 3. Demographic and clinical characteristics of patients according to *HTT* CAG intermediate number of repeats in the PARALS cohort. *HTT*+ refers to patients carrying 27–35 CAG repeats, and *HTT*- refers to patients carrying 26 or fewer CAG repeats.

	HTT+	***HTT-***	***P****
Age at disease onset (years, SD)	69.1 (10.2)	65.7 (10.8)	0.0154
Sex (female, %)	33 (41.8%)	566 (47.9%)	0.3910
Site of onset (bulbar, %)	26 (33.3%)	394 (33.4%)	0.6441
ALS-FTD (%)	12 (20.7%)	129 (15.8%)	0.4843
Diagnostic delay (months, mean, SD)	9.0 (7.0)	12.1 (10.9)	0.0415
ALSFRS-R (points, mean, SD)	40.7 (6.0)	40.8 (6.2)	0.9190
ATXN2 ≥31 polyQ repeats (%)	-	44 (3.7%)	

*Results of ANOVA (continuous variables) and chi-squared (categorical variables) statistics, significant P values are bolded; SD = standard deviation; ALS-FTD = ALS with comorbid FTD; ECAS = Edinburgh Cognitive and Behavioural ALS Screen; ALSFRS = score of the ALS Functional Rating Scale – Revised at diagnosis.

eTable 4. Clinical characteristics of patients according to *HTT* CAG intermediate number of repeats in the Answer ALS cohort. *HTT*+ refers to patients carrying 27–35 CAG repeats, and *HTT*- refers to patients carrying 26 or fewer CAG repeats.

	HTT+	***HTT-***	***P****
Age at disease onset (years, SD)	60.0 (10.6)	60.1 (11.0)	0.9701
Sex (female, %)	33 (41.8%)	566 (47.9%)	0.1440
Site of onset (bulbar, %)	10 (33.3%)	79 (22.9%)	0.2023
ALS-CBS total score (mean, SD)	16.5 (3.5)	16.2 (3.2)	0.6680
Diagnostic delay (months, median, IQR)	12.5 (11.1)	11.2 (10.9)	0.7470
ALSFRS-R (points, mean, SD)	35.0 (9.2)	35.2 (7.8)	0.9351
ATXN2 ≥31 polyQ repeats (%)	-	5 (1.3%)	

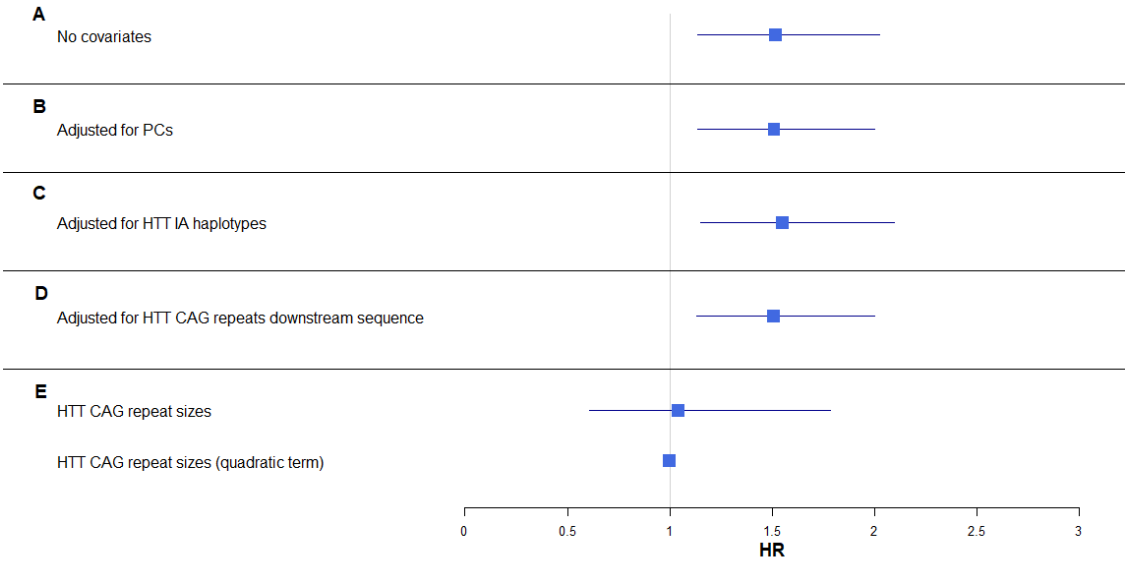
*Results of ANOVA (continuous variables) and chi-squared (categorical variables) statistics; SD = standard deviation; IQR = interquartile range; ALS-CBS= Amyotrophic Lateral Sclerosis Cognitive-Behavioral Screen; ALSFRS = score of the ALS Functional Rating Scale – Revised at the enrollment.

eTable 5. Brief description of previous studies reporting the effect of *HTT* intermediate alleles on non-HD phenotypes.

Study	Outcome	Study Cohort(s)	Main findings
Rosas et al, 2019[14]	Neurodegenerative Disease Risk	Alzheimer's disease (AD) (n = 1126), Parkinson's disease (PD) (n = 610), frontotemporal dementia (FTD) (n = 440) and 509 healthy controls (HCs)	Frequency of *HTT* IAs is higher in patients with FTD (6.9%) versus controls (2.9%)
Menéndez-González et al, 2019[15]	Neurodegenerative Disease Risk	AD (n = 1126), PD (n = 610), frontotemporal lobar degeneration (FTLD) (n = 225) and 509 healthy controls (HCs)	The frequency of HTT IA was significantly higher among patients with AD than among HCs (p = 0.011, OR = 2.11, 95% CI = 1.19-3.74).
Gardiner et al, 2017[16]	Depression	2981 participants aged 18–65 years (including 1973 subjects with a lifetime diagnosis of major depressive disorder); 510 participants aged 60–93 years (including 378 patients with depression)	Non-linear association between the risk of lifetime depression and *HTT* CAG repeat size (both relatively short and relatively large alleles are associated with an increased risk of depression (β−0.292 and β 0.006 for the linear and the quadratic term; both P ≤ 0.009)
Ingannato et al, 2022[17]	Longevity	Centenarians (n = 143) compared with pathological controls with cognitive decline (n = 232) and healthy controls (n = 104).	Higher frequency of IAs in Centenarians compared to pathological controls with cognitive decline (p = 0.031; OR = 2.3097 95% CI 1.0591 to 5.0371)
Schultz et al, 2021[18]	Cognitive performance	502 young adults (≤30 years old) without a history of depression, apathy, or cognitive deficits	Increased number of *HTT* CAG repeats associated with higher scores in multiple cognitive performance tests
Gardiner et al, 2019[19]	Cognitive function in older adults	5786 participants aged 70-83 years old	*HTT* CAG repeat number is significantly associated with the decline in cognitive function (p = 0.005)
Lee et al, 2021[20]	Intelligence	316 children aged 6–18 years old	There is a non-linear effect of the number of HTT CAG repeats (inverted U-shape pattern, p = 0.006) on the measure of general intelligence. Increasing repeat length was associated with higher scores up until 40–41; HTT CAG > 41 was associated with declining scores.
Faquih et al. 2023[21]	Metabolomic measurements	10257 individuals from three European cohort	The metabolomic profile associated with larger HTT CAG repeat sizes is unfavourable (elevated levels of low-density lipoprotein (LDL)-, very low-density lipoprotein- and intermediate density lipoprotein (IDL)-related metabolites; decreased levels of very large high-density lipoprotein (HDL)-related

			metabolites.

eFigure 1. Summary of the results of sensitivity analyses.
The effect of *HTT* Intermediate alleles (IA) on ALS survival (**A**) persisted when principal components (PCs) were included to account for population substructure (**B**). Neither *HTT* haplotypes (**C**) nor known *HTT* variants affecting age at onset in Huntington's Disease (**D**) substantially modified the effect of *HTT* intermediate alleles on ALS. Finally, we did not observe a linear or U-shaped correlation between the number of *HTT* CAG repeats (**E**).



References

1. Brooks BR, Miller RG, Swash M, Munsat TL. El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph Lateral Scler Other Motor Neuron Disord. 2000;1(5). doi:10.1080/146608200300079536
2. Chiò A, Mora G, Moglia C, et al. Secular Trends of Amyotrophic Lateral Sclerosis. JAMA Neurol. 2017;74(9). doi:10.1001/jamaneurol.2017.1387
3. Iazzolino B, Pain D, Peotta L, et al. Validation of the revised classification of cognitive and behavioural impairment in ALS. J Neurol Neurosurg Psychiatry. 2019;90(7):734-739. doi:10.1136/jnnp-2018-319696
4. Strong MJ, Abrahams S, Goldstein LH, et al. Amyotrophic lateral sclerosis - frontotemporal spectrum disorder (ALS-FTSD): Revised diagnostic criteria. Amyotroph Lateral Scler Front Degener. 2017;18(3-4). doi:10.1080/21678421.2016.1267768
5. Dolzhenko E, Deshpande V, Schlesinger F, et al. ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. *Bioinformatics*. 2019;35(22):4754-4756. doi:10.1093/bioinformatics/btz431
6. Abraham G, Qiu Y, and Inouye M. FlashPCA2: principal component analysis of biobank-scale genotype datasets. *Bioinformatics*. 2017;33(17): 2776-2778. doi:10.1093/bioinformatics/btx299
7. Kay C, Collins JA, Skotte NH, et al. Huntingtin Haplotypes Provide Prioritized Target Panels for Allele-specific Silencing in Huntington Disease Patients of European Ancestry. *Mol Ther*. 2015;23(11):1759-1771. doi:10.1038/mt.2015.128
8. Chao, M., Gillis, T., Atwal, R. *et al.* Haplotype-based stratification of Huntington's disease. *Eur J Hum Genet* **25**, 1202–1209 (2017). https://doi.org/10.1038/ejhg.2017.12
9. Kay C, Collins JA, Wright GEB, et al. The molecular epidemiology of Huntington disease is related to intermediate allele frequency and haplotype in the general population. *Am J Med Genet B Neuropsychiatr Genet*. 2018;177(3):346-357. doi:10.1002/ajmg.b.32618
10. Lee JM, Gillis T, Mysore JS *et al*: Common SNP-based haplotype analysis of the 4p16.3 Huntington disease gene region. *Am J Hum Genet* 2012; **90**: 434–444.
11. Wright GEB, Collins JA, Kay C, et al. Length of Uninterrupted CAG, Independent of Polyglutamine Size, Results in Increased Somatic Instability, Hastening Onset of Huntington Disease. *Am J Hum Genet*. 2019;104(6):1116-1126. doi:10.1016/j.ajhg.2019.04.007
12. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset. *Cell*. 2019 Aug 8;178(4):887-900.e14. doi: 10.1016/j.cell.2019.06.036
13. Baxi, E.G., Thompson, T., Li, J. et al. Answer ALS, a large-scale resource for sporadic and familial ALS combining clinical and multi-omics data from induced pluripotent cell lines. Nat Neurosci 25, 226–237 (2022). https://doi.org/10.1038/s41593-021-01006-0
14. Rosas I, Martínez C, Clarimón J, et al. Role for ATXN1, ATXN2, and HTT intermediate repeats in frontotemporal dementia and Alzheimer's disease. *Neurobiol Aging*. 2020;87:139.e1-139.e7. doi:10.1016/j.neurobiolaging.2019.10.017
15. Menéndez-González M, Clarimón J, Rosas-Allende I, et al. HTT gene intermediate alleles in neurodegeneration: evidence for association with Alzheimer's disease. *Neurobiol Aging*. 2019;76:215.e9-215.e14. doi:10.1016/j.neurobiolaging.2018.11.014

16. Gardiner SL, van Belzen MJ, Boogaard MW, et al. Huntingtin gene repeat size variations affect risk of lifetime depression. *Transl Psychiatry*. 2017;7(12):1277. Published 2017 Dec 11. doi:10.1038/s41398-017-0042-1
17. Ingannato A, Bagnoli S, Bessi V, et al. Intermediate alleles of HTT: A new pathway in longevity. *J Neurol Sci*. 2022;438:120274. doi:10.1016/j.jns.2022.120274
18. Schultz JL, Saft C, Nopoulos PC. Association of CAG Repeat Length in the Huntington Gene With Cognitive Performance in Young Adults. *Neurology*. 2021;96(19):e2407-e2413. doi:10.1212/WNL.0000000000011823
19. Gardiner SL, Trompet S, Sabayan B, et al. Repeat variations in polyglutamine disease-associated genes and cognitive function in old age. *Neurobiol Aging*. 2019;84:236.e17-236.e28. doi:10.1016/j.neurobiolaging.2019.08.002
20. Lee JK, Conrad A, Epping E, et al. Effect of Trinucleotide Repeats in the Huntington's Gene on Intelligence. *EBioMedicine*. 2018;31:47-53. doi:10.1016/j.ebiom.2018.03.031
21. Faquih TO, Aziz NA, Gardiner SL, et al. Normal range CAG repeat size variations in the HTT gene are associated with an adverse lipoprotein profile partially mediated by body mass index. *Hum Mol Genet*. 2023;32(10):1741-1752. doi:10.1093/hmg/ddad020