Mutations in the tail domain of the neurofilament heavy chain gene increase the risk of amyotrophic lateral sclerosis

Heather Marriott MSc[1,2], Thomas P. Spargo MSc[1,2], Ahmad Al Khleifat PhD[1], Isabella Fogh[1], Project MinE ALS Sequencing Consortium, Peter M Andersen MD, PhD[3], Nazli A. Başak PhD[4], Johnathan Cooper-Knock PhD[5], Philippe Corcia MD, PhD[6,7], Philippe Couratier[8,9], Mamede de Carvalho[10], Vivian Drory MD[11,12], Jonathan D. Glass MD[13], Marc Gotkine MD[14,15], Orla Hardiman PhD[16], John E. Landers PhD[17], Russell McLaughlin PhD[18] , Jesús S. Mora Pardina[19], Karen E. Morrison[20], Susana Pinto MD, PhD[10], Monica Povedano MD[21], Christopher E. Shaw MD[1], Pamela J. Shaw MD[5], Vincenzo Silani MD[22,23], Nicola Ticozzi MD[22,23], Philip van Damme MD, PhD[24,25], Leonard H. van den Berg MD, PhD[26], Patrick Vourc'h PhD[6,27], Markus Weber PhD[28], Jan H. Veldink PhD[26], Richard J. Dobson PhD[2,29,30,31], Patrick Schwab PhD[32], Ammar Al-Chalabi MD, PhD[1,33*], Alfredo Iacoangeli PhD[1,2,29*,**]

[1]Maurice Wohl Clinical Neuroscience Institute, Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, SE5 8AF, UK

[2]Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, SE5 8AF UK

[3]Department of Clinical Science, Umeå University, Umeå SE-901 85, Sweden

[4]Koc University, School of Medicine, Translational Medicine Research Center, NDAL, Istanbul, 34450, Turkey

[5]Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield S10 2HQ, UK

[6]UMR 1253, Université de Tours, Inserm, Tours 37044, France

[7]Centre de référence sur la SLA, CHU de Tours, Tours 37044, France

[8]Centre de référence sur la SLA, CHRU de Limoges, Limoges, France

[9]UMR 1094, Université de Limoges, Inserm, Limoges 87025, France

[10]Instituto de Fisiologia, Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Lisbon 1649-028, Portugal

[11]Department of Neurology, Tel-Aviv Sourasky Medical Centre, Tel-Aviv 64239, Israel

[12]Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv 6997801, Israel

[13]Department of Neurology, Emory University School of Medicine, Atlanta, Georgia, GA 30322, USA

[14]Faculty of Medicine, Hebrew University of Jerusalem, Jerusalem 91904, Israel

[15]Agnes Ginges Center for Human Neurogenetics, Department of Neurology, Hadassah Medical Center, Jerusalem 91120, Israel

[16]Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin D02 PN40, Ireland

[17]Department of Neurology, University of Massachusetts Medical School, Worcester, MA 01655, USA

[18]Complex Trait Genomics Laboratory, Smurfit Institute of Genetics, Trinity College Dublin, Dublin D02 PN40, Ireland

[19]ALS Unit, Hospital San Rafael, Madrid, Spain

[20]School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast BT9 7BL, UK

[21]Functional Unit of Amyotrophic Lateral Sclerosis (UFELA), Service of Neurology, Bellvitge University Hospital, L'Hospitalet de Llobregat, Barcelona 08907, Spain

[22]Department of Neurology-Stroke Unit and Laboratory of Neuroscience, Istituto Auxologico Italiano, IRCCS, Milan 20149, Italy

[23]Department of Pathophysiology and Transplantation, "Dino Ferrari" Center, Università degli Studi di Milano, Milan 20122

[24]University Hospitals Leuven, Department of Neurology, Leuven 3000, Belgium

[25]VIB, Center for Brain and Disease Research, Leuven, Belgium & Neuroscience Department, Leuven Brain Institute, KU Leuven, Belgium

[26]Department of Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht 3584 CX, Netherlands

[27]Service de Biochimie et Biologie molécularie, CHU de Tours, Tours 37044, France

[28]Neuromuscular Diseases Unit/ALS Clinic, Kantonsspital St. Gallen, 9007 St. Gallen, Switzerland

[29]NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, UK

[30]Institute of Health Informatics, University College London, London, NW1 2DA, UK

[31]NIHR Biomedical Research Centre at University College London Hospitals NHS Foundation Trust, London, UK

[32]GlaxoSmithKline, Artificial Intelligence and Machine Learning

[33]King's College Hospital, London, SE5 9RS, UK

* these authors contributed equally, ** corresponding author

ABSTRACT

Objective: Genetic variation in the neurofilament heavy chain gene (*NEFH*) has been convincingly linked to the pathogenesis of multiple neurodegenerative diseases, however, the relationship between *NEFH* mutations and ALS susceptibility has not been robustly explored. We therefore wanted to determine if genetic variants in *NEFH* modify ALS risk.

Methods: We performed fixed and random effects model meta-analysis of published case-control studies reporting *NEFH* variant frequencies using next-generation sequencing, microarray or PCR-based approaches. Comprehensive screening and rare variant burden analysis of *NEFH* variation in the Project MinE ALS whole-genome sequencing data set was also conducted.

Results: We identified 12 case-control studies that reported *NEFH* variant frequencies, for a total of 9,496 samples (4,527 ALS cases and 4,969 controls). Fixed effects meta-analysis found that rare (MAF<1%) missense variants in the tail domain of *NEFH* increase ALS risk (OR 4.56, 95% CI 2.13-9.72, p<0.0001). A total of 591 rare *NEFH* variants, mostly novel (78.2%), were found in the Project MinE dataset (8,903 samples: 6,469 cases and 2,434 controls). Burden analysis showed ultra-rare (MAF <0.1%) pathogenic missense variants in the tail domain are associated with ALS (OR 1.94, 95% CI 0.86-4.37, Madsen-Browning p=0.039), replicating and confirming the meta-analysis finding. High-frequency rare (MAF 0.1-1%) tail in-frame deletions also confer susceptibility to ALS (OR 1.18, 95% CI 0.67-2.07, SKAT-O p=0.03), which supports previous findings.

Interpretation: This study shows that *NEFH* tail domain variants are a risk factor of ALS and supports the inclusion of missense and in-frame deletion *NEFH* variants in ALS genetic screening panels.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a relentlessly progressive and fatal neurodegenerative disease resulting from upper and lower motor neuron loss [1]. As ALS displays considerable clinical and genetic heterogeneity, it is essential that its genetic mechanisms are defined appropriately in order to develop modifying therapies and enhance personalised medicine approaches [2]. Around 40-45 genes are implicated in ALS and are involved in cellular processes such as autophagy, DNA damage repair, protein degradation, mitochondrial function and cellular/axonal transport [3]. The neurofilament heavy chain gene (*NEFH*), encodes the neurofilament heavy subunit protein (NF-H), which regulates several of these activities in an effort to maintain neuronal homeostasis.

Neurofilament protein subunits preserve neuronal architecture by using their side-arms to construct cross-bridges with cytoskeletal components such as microtubules and actin filaments, forming a stable filament-centred matrix that allows intracellular signalling, mitochondrial localisation and ER transport to occur [4]. This is predominantly orchestrated by the phosphorylation of the head and tail domains of neurofilament genes. For instance, phosphorylation of the head domain acts as a primer for matrix formation, controlling polymerisation of the NF-H subunit in the cell body before the subunits move to the axon, where the lysine-serine-proline (KSP) repeat of the tail domain is phosphorylated to construct the matrix structure and stabilise the neurofilament side arms [5]. As a result, *NEFH* disruption could influence selective motor neuron degeneration in the brain and spinal cord of affected individuals with ALS via dysregulation of neuronal function [6].

Frameshift and missense mutations in *NEFH* have been convincingly linked to the pathogenesis of various neurological diseases, including Charcot-Marie-Tooth disease type 2CC [7–10], spinal muscular atrophy [11] and Alzheimer's disease [12].

Several lines of evidence suggest hyperphosphorylation of the KSP repeat causes axonal aggregation of phosphorylated NF-H (pNF-H), thereby compromising neuronal integrity and increasing circulating pNF-H levels in the serum and CSF [5]. Raised pNF-H levels have already been established as a robust biomarker for ALS progression, survival [13,14], patterns of motor neuron involvement, and can clinically distinguish ALS from mimics such as hereditary spastic paraplegia, spinal muscular atrophy and myasthenia gravis [15]. While pNF-H demonstrates prognostic value, there have not been robust studies examining the relationship between *NEFH* mutations and ALS susceptibility. The association between small insertions and deletions in the KSP repeat and ALS risk has been reported in a number of studies [16–18], however, this relationship has not been widely reproduced. Still, *NEFH* is included in the majority of genetic screening panels worldwide.

Therefore, a large scale, targeted and comprehensive investigation of the role of common and rare *NEFH* variants in ALS is greatly needed. This study aims to fill such a gap by first performing a meta-analysis of published ALS case-control

studies that reported *NEFH* variants and second conducting a large-scale investigation of *NEFH* variation using genetic data from the Project MinE international ALS whole-genome sequencing consortium.

SUBJECTS/MATERIALS AND METHODS

Systematic Review

This study was performed in accordance with the 2020 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [19].

Eligibility Criteria

Primary research articles published between January 1993 and October 2021 were included if they reported *NEFH* variant frequencies in ALS patients via a candidate or panel gene approach (targeted panel resequencing, variant screening), whole genome sequencing, whole exome sequencing, microarray or PCR-based approaches. Studies were excluded if they were clinical, functional or epidemiological, or if *NEFH* variants were not identified (in targeted gene panel studies) or were identified in non-ALS cases only.

Information Sources, Search Strategy and Screening Process

Relevant studies were identified by searching PubMed, Embase and Medline databases with the search terms ‘‘amyotrophic lateral sclerosis” OR “ALS’’ in combination with “neurofilament heavy chain gene”, NEFH,” “NFH” OR “NF-H.” After removing duplicate records, title and abstract screening was then performed against the eligibility criteria. Studies which advanced to full text screening were subject to backward citation screening using Web of Science to identify any suitable articles which may have been missed. Full text screening of database and citation identified records was then performed. The search strategy was independently performed, and the results were crosschecked by two members of the team.

Data Collection Process and Data Synthesis

Once all eligible records were identified, the following study characteristics were extracted; author, publication year, study design, screening method and genetic technology used to detect *NEFH* variants, population (country of origin), study groups, sex and age of ALS groups and diagnostic criteria applied for recruitment into the study. In addition, for each

variant identified, the following information was obtained: name according to HGVS protein nomenclature, mutation type, *NEFH* domain location, rsID, and pathogenicity status according to SIFT [20] and PolyPhen [21] prediction software. Study-specific variant information i.e. frequency in cases and/or controls, odds ratios (ORs) and 95% confidence intervals with p-values and other ALS-associated gene variants carried in *NEFH*-positive individuals, were also extracted. Population-specific *NEFH* variant frequencies were added to each variant record using the gnomAD v2.1.1 non-neuro database [22]. If the rsID was not supplied, dbSNP [23] and gnomAD were searched with the corresponding variant entry. Similarly, for variants without pathogenicity predictions, gnomAD and the hg19 Variant Effect Predictor (VEP) web tool [24] was used to obtain variant consequence status.

Meta-Analysis

Individual missense and exonic insertion and deletion variants which were supported by two or more case-control studies were eligible for variant-level meta-analysis. Subgroup meta-analysis was also performed according to combinations of population-specific gnomAD non-neuro frequency (ultra-rare: < 0.1%, rare: < 1%, or common: > 1%), domain (head, rod, or tail) and variant type. Studies which identified variants which were absent from gnomAD but present in more than one control were classified as common for the stratified analysis. Synonymous variants were excluded from the analysis. Inverse-variance weighted meta-analyses were conducted with both the fixed-effect (Cochran-Mantel-Haenszel) and random-effect (DerSimonian-Laird) models. Crude ORs were calculated from the extracted data. Between-study heterogeneity was assessed using the combination of the I^2 test and Cochran-Q statistic, with significant heterogeneity indicated when $I^2 > 50\%$ and $Q < 0.10$. Publication bias was assessed with both Egger's and Harbord's test, with p-values < 0.05 classed as displaying significant outcome heterogeneity and selective reporting. All statistical analyses were performed using the *metabin* and *metabias* functions of the meta R package [25].

Genetic Screening

Whole-genome sequencing samples collected as part of the Project MinE ALS sequencing consortium [26] were used to investigate the impact of *NEFH* variants in ALS and for replicating the literature based meta-analysis results. Sample overlap between Project MinE and the studies included in the meta-analysis was investigated by contacting the authors of the studies that included patients of the same nationality as the Project MinE participants. No overlap was found. No Briefly, genomic DNA from venous blood was isolated using standard methods and assessed with gel electrophoresis

before PCR-free 100bp paired-end sequencing was performed on the Illumina HiSeq2000 platform, which yielded ~40x coverage. The full dataset consists of 9,050 samples, which includes 6,603 individuals defined as having pure ALS as well as 2,447 age and sex matched controls. After standard quality control measures, the final filtered dataset comprised of 6,469 ALS cases and 2,434 controls from 13 countries (Supplementary Table 1) for which whole-genome SNV and small indel data were available [27]. Structural variant data generated with Manta [28] were available for approximately two thirds of samples (4,686 ALS cases and 1,859 controls). Variants were aligned to hg19. Variants were then annotated with VEP for both functional consequence/type (i.e. UTR, intronic, missense, indel, synonymous) and VEP-specific classification of the impact of the variant (i.e. high, moderate, low, modifier). Screening of structural variants called with Manta v0.28.0 [28] was also performed; SURVIVOR v1.0.9 [29] was used to create a union callset before variants greater than 100,000bp were excluded to reduce false positives. The remaining variants were then annotated with AnnotSV [30] and CADD-SV [31] to assess their potential pathogenicity. All results files were converted into a matrix with the VariantAnnotation R package [32], from which case-control frequencies were calculated. For the review-identified variants and structural variants present in Project MinE, Firth logistic regression was performed using RVTests [33] under default settings to assess potential associations between variant status and ALS susceptibility. Results were corrected for sex and the first 10 principal components.

Rare Variant Burden Analysis

Burden analysis of all *NEFH* variants identified in the Project MinE samples was performed with RVTests [33], using Madsen-Browning and SKAT-O methods under default settings. Results were corrected for sex and the first 10 principal components. Variants were initially grouped by frequency (ultra-rare: <0.1%, high-frequency rare: 0.1-1%), according to the highest value in control databases (gnomAD non-neuro non-Finnish European and Project MinE controls), before being grouped by functional domain (whole gene, head, rod, tail) for which the genomic coordinates were obtained with the *ensembldb* R package. For each functional domain, variant burden was calculated for several variant types (missense, synonymous, insertion, deletion, 3' UTR, 5' UTR, intronic) and VEP impact classes (high, moderate, low, modifier). Variant burden was also repeated, sub-setting missense variants into predicted pathogenicity classes, according to SIFT and/or Polyphen scores ("deleterious," "deleterious low confidence," "possibly damaging" or "probably damaging").

RESULTS

Study Selection

The literature systematic review process flowchart is presented in Figure 1. The initial search identified 29 articles which were eligible for title and abstract screening, of which 16 articles were the wrong study type, disease, or instances where genetic screening did not include *NEFH* or identify *NEFH* variants even if *NEFH* was present in the targeted sequencing panel. Backward citation searching of the remaining 13 articles found an additional 251 records for screening. Manual full text inspection removed a further 242 records (2 from database search and 240 from citation search) as the inclusion criteria were not met. In total, 22 studies involving a total of 10,959 individuals (6,090 ALS cases and 4,869 controls) from 14 countries were included in the present study.

Study Characteristics

An overview of the characteristics of all included studies is given in Table 1. The people were most frequently sampled from Asian (N Studies = 7) and European (N Studies = 7) populations, with a family disease history reported in 77% of studies. Diagnostic criteria were applied to support inclusion in 15 studies (68%), with varying definitions of El Escorial criteria employed in 93% of those. A combination of El Escorial and Awaji-Shima criteria was used in one study. The average age of recruitment of the ALS patients ranged from 30.7 to 62.1 (median 58.1), with a male: female ratio ranging between 0.60 and 1.78 (median 1.38) across studies. When separating by country, Asian populations had a younger median age at recruitment and a higher median male: female ratio than European populations (Asian: age 52.01, sex ratio 1.52; European: age 60.1, sex ratio 1.22). A case-control design was adopted in 12 studies (55%), with 6 investigating *NEFH* variation in ALS via candidate gene-based methods. Gene panels which included *NEFH* were used in 13 studies, with a further 2 opting for custom variant panel screening. The most popular genetic technology used to identify *NEFH* variants was whole-exome sequencing (N Studies = 6) and the combination of whole-exome sequencing with validation approaches such as PCR and Sanger sequencing (N Studies = 6).

Variant Characteristics

A total of 59 *NEFH* variants were identified from the included studies. Full details of each variant are documented in Supplementary Table 2, with their genomic coordinates and base pair substitutions (for hg19) available in Supplementary Table 3. Missense variants were the most represented (67.8%), followed by inframe deletions (13.6%), synonymous

variants (13.6%), inframe insertions (1.7%), frameshift deletions (1.7%) and stop-gained SNVs (1.7%). Insertion/deletion variants (indels) ranged from 3bp to 48bp in length and exclusively occupied the tail (Figure 2). Seven times as many in-frame deletions were found in the KSP repeat than the lysine-glutamic acid-proline (KEP) segment. Only 2 variants were found in the head domain (Figure 2). Only 17 variants (28.8%) were reported in more than one study. Ten people with *NEFH* variants also harboured variants in other ALS-associated genes, including *SOD1*, *FUS*, *OPTN*, *SETX*, *ALS2* and *CHMP2B* (Supplementary Table 2).

Meta-Analysis of Previously Published Studies

The twelve case-control studies we identified were selected for meta-analysis. A total of 34 deletion, insertion and missense variants were reported across these studies (displayed in the top panel of Figure 3) in a total of 9,496 individuals (4,527 cases; 4,969 controls). Of these, 8 variants (3 in-frame deletions and 5 missense) were identified in two or more case-control studies and therefore were included in the variant-level meta-analysis. No singular variant was shown to significantly confer or reduce risk for ALS (Supplementary Table 4). One of the deletion variants, K790del, displayed a significantly high level of heterogeneity according to both Cochran's Q and I^2 metrics (Supplementary Table 4).

We performed meta-analyses of *NEFH* variants based on the aggregation of variants stratified by frequency, domain and variant type (see methods). We found that rare missense variants in the tail domain increase the risk of ALS (Table 2, Figure 4), with an OR of 4.56 (95% CI 2.13-9.72, $p<0.0001$) under the fixed-effects model. There was no evidence of inter-study heterogeneity (Cochran's $Q = 2.30$, $p=0.51$, $I^2 = 0\%$) or publication bias (Egger $p=2.11$, Harbord $p=1.85$). We also found that rare missense and rare tail variants were also associated with an increased risk of ALS (Table 2), with ORs of 2.37 (95% CI = 1.39-4.04, $p=0.0015$) and 2.42 (95% CI = 1.28-4.58, $p=0.0066$), although we determined that the rare missense tail variants were driving this result, as removing these from the meta-analyses caused both associations to be lost. Across all categories, deletion variants did not significantly increase or reduce susceptibility for ALS (Table 2).

Screening of SNV/indel *NEFH* variants in the Project MinE cohort

We next screened the whole *NEFH* gene in the Project MinE dataset (6,469 ALS cases and 2,434 controls) to obtain a complete landscape of *NEFH* variation in ALS. A total of 591 SNV and indel variants were identified (Figure 5a). Additional information on all variants are available in Supplementary Table 5. The KSP repeat harboured the highest number of variants (61; 10.32%). Interestingly, intronic regions contained 65% of all variants found in the cohort, with

220 (57.29%) existing as singletons (in either one case or one control). In fact, there were 351 singletons totalling 59.39% of all variants identified in the cohort (Figure 5b), with 220 (62.68%) being in intronic regions of *NEFH*; 3.4x that of the tail domain (65 variants; 18.52%). When accounting for variants from different sources, 462 (78.17%) were identified only in cases and/or controls and not the review or gnomAD non-neuro non-Finnish database (Figure 5c), and are therefore classified as 'novel' for the purposes of this study.

For the *NEFH* variants identified from the systematic review, 16 out of 59 (27.1%) were found in Project MinE, with 11 of these occurring in Project MinE cases and controls, review and gnomAD (Figure 5c). Examination of case-control frequencies of review-identified variants present in Project MinE (Table 3) suggested that K790del could be protective against ALS (0.139% cases, 0.04% controls; Beta (SE) = -1.03 (0.47), p=0.03). Using the Project MinE cohort as an additional study for meta-analysis of individual variants did not offer any other insights into their contribution to ALS risk (Supplementary Table 6).

Screening of *NEFH* structural variants in the Project MinE cohort

Only 4 structural variants were identified in a subset of the Project MinE cohort (Table 3). All were located in the KSP and KEP segments of the tail domain, with none being likely pathogenic according to the CADD-SV pathogenicity prediction tool (≥ 15). When comparing case-control frequencies to elucidate the association of the structural variants with ALS, the 113bp KEP segment deletion was found to be protective against ALS (17.95% cases vs 23.91% controls; Beta (SE) = -0.34 (0.061), p=2.60E-08).

Rare variant burden analysis in the Project MinE cohort

All of the SNV/indel variants from Project MinE were subject to burden analysis at two frequency levels (Ultra-Rare (<0.1%) and High-frequency rare (0.1-1%)) to assess the contribution of different variant classes to ALS risk (Supplementary Tables 7 and 8) stratified by domain. We found that ultra-rare pathogenic missense tail variants (PolyPhen and/or SIFT) were associated with an increased risk of developing ALS (OR 1.94, 95% CI 0.86-4.37; Madsen Browning p=0.039), which replicated and confirmed the result of the meta-analysis. Stratifying this by subdomain revealed that the KEP repeat drove this result (OR 5.65, 95% CI 0.75-42.83, Madsen Browning p=0.02), and that other domains mildly increased ALS risk albeit insignificantly (Supplementary Table 7). In line with previous reports, ultra-rare tail domain in-frame deletions had a large impact on ALS risk, but this finding is at the border of the significance testing threshold (OR

3.01, 95% CI 0.69-13.12, Madsen-Browning p=0.052). A similar but significant effect was observed for low-frequency rare in-frame deletions (Supplementary Table 8), with an OR of 1.18 (95% CI 0.67-2.07, SKAT-O p=0.03). Ultra-rare pathogenic missense variants and high-frequency rare in-frame deletions identified and assessed in these burden analyses are detailed in Figure 3 (bottom panel), with their genomic coordinates, variant frequencies and pathogenicity status detailed in Supplementary Table 9.

When assessing the role of moderate impact variants (missense and indel variants, as defined by VEP), both ultra and high-frequency rare tail-domain variants increased ALS risk (Ultra-Rare OR 1.69, 95% CI 1.00-2.87, Madsen-Browning p=0.024; High-Frequency Rare OR 1.13, 95% CI 0.70-1.84, SKAT-O p=0.04), as did moderate impact variants throughout the whole gene (OR 1.47, 95% CI 0.98-2.22, Madsen-Browning p=0.032). Interestingly, ultra-rare intronic, 5'UTR and modifying impact variants (intronic, 5'UTR and 3'UTR combined) also significantly increased ALS risk (Supplementary Table 7).

DISCUSSION

In this study, we found that missense tail variants in *NEFH* are associated with an increased risk of ALS. The meta-analysis of 3 previous case-control reports which documented rare (MAF<1%) missense tail domain *NEFH* variant frequencies in a total of 1164 ALS patients and 2,177 controls yielded an OR of 4.56 (p<0.0001). This association was replicated, although with a lower OR, when performing an ultra-rare variant burden analysis of pathogenic missense tail variants in the Project MinE dataset (OR 1.94, Madsen-Browning p=0.039). This is likely due to the discrepancy in sample size between the two cohorts, as Project MinE contains more than 5.5 times the number of cases used in the meta-analysis, with smaller sample sizes often reporting a larger effect size (OR) for significant relationships in either direction [53] and also the 'winner's curse' effect commonly observed in genetic association discovery studies [54].These findings hold high validity as the vast majority of variants in the meta-analysis were deleterious and possibly/probably damaging according to SIFT and PolyPhen pathogenic prediction tools (Supplementary Table 2), which were the same criteria used for missense variants in the burden analysis to be considered pathogenic. Furthermore, removing the tail domain variants from the rare missense variant meta-analysis, which does not take domain-specific effects into account, as well as performing ultra-rare burden analysis of pathogenic missense variants in the whole gene and head and rod domain, nullified the association with ALS, thus proving that missense tail variants were drivers of ALS risk. Additionally, we found that pathogenic ultra-rare missense variants in the KEP repeat were the main drivers of the association with ALS risk in Project MinE (OR 5.65, Madsen Browning

p=0.02). However, pathogenic ultra-rare missense variants in the KSP repeat still showed a the same direction of effect in the Project MinE burden analysis.

Unlike its *NEFM* and *NEFL* counterparts, the mechanism by which missense tail variants alter the functionality of the *NEFH* gene has not been fully elucidated [6]. Despite this, it is plausible to suggest that these mutations, especially those in the KSP repeat segment, could modify the effects of phosphorylation, thereby changing the conformation of the NF-H subunit in such a way that simultaneously increases the propensity of pNF-H aggregate formation in the axon and disrupts energy metabolism and protein transport. Indeed, *NEFH* also has a complex mRNA and protein-linked stoichiometry of which has not been studied here and could provide us with additional insights into the genetic basis of NF-H inclusion formation. For instance, downregulation of two exclusively spinal cord and CSF-expressed miRNAs, miR-92a-3p, miR-9-5p, bind to recognition elements in the 3'UTR of *NEFH* to increase the expression of NEFH mRNA transcripts and NF-H protein levels in people with ALS [55,56]. Moreover, there is emerging evidence which suggests that microglial-secreted protein factors can influence NEFH transcript expression and contribute to NF-H inclusion pathology, albeit in the absence of *NEFH* mutations [57]. Therefore, future studies should build on what we have reported here by incorporating genetic evidence of missense tail mutations with proteomic and transcriptomic data to determine if the aberrant stoichiometry of NF-H is due solely to the action of the mutation on phosphorylation sites within the tail or is a product of a larger interaction between miRNA, protein and glial targets.

We also found that high-frequency rare (MAF 0.1-1%) small in-frame deletions in the tail domain confer susceptibility to ALS within Project MinE (OR 1.18, SKAT-O p=0.03), which agrees with previous findings in the literature [16,17]. However, the literature based meta-analysis did not find a significant association with ALS risk for either rare (MAF<1%, OR 0.90, p=0.85) or ultra-rare (MAF<0.1%, OR 0.94, p=0.93) tail deletions (Table 2). This discrepancy could again be due to the relatively small sample sizes used in the meta-analysis compared to in Project MinE, or that there may be subdomain-specific effects occurring in the tail that the meta-analysis design could not account for. Potentially, deletions in the KSP repeat could be associated with an increased risk for ALS and that perhaps deletions in the KEP segment may dilute this association having a protective effect. This is plausible given that we found a novel protective 113bp deletion in the KEP region (p=2.60E-08), present in 18% of cases and 24% controls. Another possibility is that the rarity of these deletions coupled with the small sample sizes and genetic technologies of studies used in the meta-analysis resulted in decreased power to detect any significant associations or increases in ALS risk. This was the case for several ultra-rare variant

categories which were examined using both meta-analysis and variant burden methods, where there were either higher ORs i.e. in-frame deletions, or lower ORs i.e. missense head, missense rod, present in the Project MinE burden analysis. An unexpected finding was that the presence of ultra-rare (MAF<0.1%) intronic rod variants were significantly associated with increased ALS risk (Supplementary Table 7, OR 1.40, Madsen Browning p=0.004), as the rod domain is highly conserved across all of the neurofilament subunit family [4], and there has been no evidence to date to support the role of non-coding DNA in neurofilaments to neurodegeneration. Further research on this is needed to establish the role of non-coding neurofilament variants in ALS.

In conclusion, we showed that missense mutations and in-frame deletions in the tail domain of *NEFH* are associated with the increase of ALS risk, using a two-tiered meta-analysis and variant burden approach which leveraged *NEFH* variant information of 11,130 ALS patients and 7,416 controls from both the literature and the largest whole-genome sequencing consortium for ALS, Project MinE. Our results support the inclusion of missense variants and in-frame deletions in the tail of *NEFH* in ALS sequencing panels.

ACKNOWLEDGEMENTS

The authors are supported by South London and Maudsley NHS Foundation Trust; MND Scotland; Motor Neurone Disease Association; National Institute for Health Research; Darby Rimmer MND Foundation; Spastic Paraplegia Foundation and Rosetrees Trust. H.M is supported by GlaxoSmithKline and the KCL funded centre for Doctoral Training (CDT) in Data-Driven Health. A.A.K is funded by ALS Association Milton Safenowitz Research Fellowship (grant number22-PDF-609.DOI :10.52546/pc.gr.150909.), The Motor Neurone Disease Association (MNDA) Fellowship (Al Khleifat/Oct21/975-799), The Darby Rimmer Foundation, and The NIHR Maudsley Biomedical Research Centre. A.I is funded by the Motor Neurone Disease Association and The NIHR Maudsley Biomedical Research Centre. A.A-C is an NIHR Senior Investigator (NIHR202421) and has received support from an EU Joint Programme - Neurodegenerative Disease Research (JPND) project. The work is supported through the following funding organisations under the aegis of JPND - www.jpnd.eu *(United Kingdom, Medical Research Council* (MR/L501529/1; MR/R024804/1) *and Economic and Social Research Council* (ES/L008238/1)*)* and through the Motor Neurone Disease Association, My Name'5 Doddie Foundation, and Alan Davidson Foundation. This study represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust

and King's College London. The authors acknowledge use of the research computing facility at King's College London, Rosalind (https://rosalind.kcl.ac.uk). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. Project MinE Belgium was supported by a grant from IWT (n° 140935), the ALS Liga België, the National Lottery of Belgium and the KU Leuven Opening the Future Fund. PVD holds a senior clinical investigatorship of FWO-Vlaanderen (G077121N) and is supported by the E. von Behring Chair for Neuromuscular and Neurodegenerative Disorders, the ALS Liga België and the KU Leuven funds "Een Hart voor ALS", "Laeversfonds voor ALS Onderzoek" and the "Valéry Perrier Race against ALS Fund". Several authors of this publication are member of the European Reference Network for Rare Neuromuscular Diseases (ERN-NMD)

AUTHOR CONTRIBUTIONS

H.M, A.A-C and A.I contributed to conception and design of the study; H.M, T.P.S, A.A.K and A.I contributed to the acquisition and analysis of data; all authors contributed to drafting the text or preparing the figures. Details about the contributing members of the Project MinE ALS Sequencing Consortium and their affiliated institutions are available in Supplementary Table 10.

POTENTIAL CONFLICTS OF INTEREST

JVH reports to have sponsored research agreements with Biogen and Astra Zeneca. AAC reports consultancies or advisory boards for Amylyx, Apellis, Biogen, Brainstorm, Cytokinetics, GenieUs, GSK, Lilly, Mitsubishi Tanabe Pharma, Novartis, OrionPharma, Quralis, Sano, Sanofi, and Wave Pharmaceuticals. The other authors have nothing to report.

REFERENCES

1. Brown RH, Al-Chalabi A. Amyotrophic Lateral Sclerosis. N Engl J Med 2017;377(2):162–172.
2. Zou Z-Y, Liu C-Y, Che C-H, Huang H-P. Toward precision medicine in amyotrophic lateral sclerosis. Ann Transl Med 2016;4(2):27.
3. Mejzini R, Flynn LL, Pitout IL, et al. ALS Genetics, Mechanisms, and Therapeutics: Where Are We Now? Front Neurosci 2019;13:1310.
4. Yuan A, Rao MV, Veeranna, Nixon RA. Neurofilaments and Neurofilament Proteins in Health and Disease. Cold Spring Harb Perspect Biol 2017;9(4):a018309.
5. Didonna A, Opal P. The role of neurofilament aggregation in neurodegeneration: lessons from rare inherited neurological disorders. Mol Neurodegener 2019;14(1):19.
6. Theunissen F, West PK, Brennan S, et al. New perspectives on cytoskeletal dysregulation and mitochondrial mislocalization in amyotrophic lateral sclerosis. Transl Neurodegener 2021;10(1):46.
7. Ikenberg E, Reilich P, Abicht A, et al. Charcot-Marie-Tooth disease type 2CC due to a frameshift mutation of the neurofilament heavy polypeptide gene in an Austrian family. Neuromuscul Disord 2019;29(5):392–397.
8. Jacquier A, Delorme C, Belotti E, et al. Cryptic amyloidogenic elements in mutant NEFH causing Charcot-Marie-Tooth 2 trigger aggresome formation and neuronal death. Acta Neuropathol Commun 2017;5(1):55.
9. Pipis M, Cortese A, Polke JM, et al. Charcot-Marie-Tooth disease type 2CC due to NEFH variants causes a progressive, non-length-dependent, motor-predominant phenotype. J Neurol Neurosurg Psychiatry 2022;93(1):48–56.
10. Yan J, Qiao L, Peng H, et al. A novel missense pathogenic variant in NEFH causing rare Charcot-Marie-Tooth neuropathy type 2CC. Neurol Sci 2021;42(2):757–763.
11. Ando M, Higuchi Y, Okamoto Y, et al. An NEFH founder mutation causes broad phenotypic spectrum in multiple Japanese families. J Hum Genet 2022;67(7):399-403.
12. Yemni EA, Monies D, Alkhairallah T, et al. Integrated Analysis of Whole Exome Sequencing and Copy Number Evaluation in Parkinson's Disease. Sci Rep 2019;9(1):3344.
13. Puentes F, Lombardi V, Lu C-H, et al. Humoral response to neurofilaments and dipeptide repeats in ALS progression. Ann Clin Transl Neurol 2021;8(9):1831–1844.

14. Xu Z, Henderson RD, David M, McCombe PA. Neurofilaments as Biomarkers for Amyotrophic Lateral Sclerosis: A Systematic Review and Meta-Analysis. PLoS One 2016;11(10):e0164625.
15. Poesen K, De Schaepdryver M, Stubendorff B, et al. Neurofilament markers for ALS correlate with extent of upper and lower motor neuron disease. Neurology 2017;88(24):2302–2309.
16. Al-Chalabi A, Andersen P, Nilsson P, et al. Deletions of the heavy neurofilament subunit tail in amyotrophic lateral sclerosis. Hum Mol Genet 1999;8(2):157–164.
17. Figlewicz DA, Krizus A, Martinoli MG, et al. Variants of the heavy neurofilament subunit are associated with the development of amyotrophic lateral sclerosis. Hum Mol Genet 1994;3(10):1757–1761.
18. Tomkins J, Usher P, Slade JY, et al. Novel insertion in the KSP region of the neurofilament heavy gene in amyotrophic lateral sclerosis (ALS). Neuroreport 1998;9(17):3967–3970.
19. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
20. Ng P, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31(13):3812–3814.
21. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7(4):248–249.
22. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581(7809):434–443.
23. Sherry S, Ward M, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29(1):308–311.
24. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biol 2016;17(1):122.
25. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health 2019;22(4):153–160.
26. Project MinE ALS Sequencing Consortium. Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. Eur J Hum Genet 2018;26(10):1537–1546.
27. Raczy C, Petrovski R, Saunders CT, et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. Bioinformatics 2013;29(16):2041–2043.

28. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 2016;32(8):1220–1222.
29. Jeffares DC, Jolly C, Hoti M, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. Nat Commun 2017;8:14061.
30. Geoffroy V, Herenger Y, Kress A, et al. AnnotSV: an integrated tool for structural variations annotation. Bioinformatics 2018;34(20):3572–3574.
31. Kleinert P, Kircher M. A framework to score the effects of structural variants in health and disease. Genome Res 2022;32(4):766–777.
32. Obenchain V, Lawrence M, Carey V, et al. VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants. Bioinformatics 2014;30(14):2076–2078.
33. Zhan X, Hu Y, Li B, et al. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics 2016;32(9):1423–1426.
34. Vechio JD, Bruijn LI, Xu Z, et al. Sequence variants in human neurofilament proteins: absence of linkage to familial amyotrophic lateral sclerosis. Ann Neurol 1996;40(4):603–610.
35. Garcia ML, Singleton AB, Hernandez D, et al. Mutations in neurofilament genes are not a significant primary cause of non-SOD1-mediated amyotrophic lateral sclerosis. Neurobiol Dis 2006;21(1):102–109.
36. Daoud H, Valdmanis PN, Gros-Louis F, et al. Resequencing of 29 candidate genes in patients with familial and sporadic amyotrophic lateral sclerosis. Arch Neurol 2011;68(5):587–593.
37. Couthouis J, Raphael AR, Daneshjou R, Gitler AD. Targeted exon capture and sequencing in sporadic amyotrophic lateral sclerosis. PLoS Genet 2014;10(10):e1004704.
38. Nakamura R, Sone J, Atsuta N, et al. Next-generation sequencing of 28 ALS-related genes in a Japanese ALS cohort. Neurobiol Aging 2016;39:219.e1–8.
39. Kruger S, Battke F, Sprecher A, et al. Rare Variants in Neurodegeneration Associated Genes Revealed by Targeted Panel Sequencing in a German ALS Cohort. Front Mol Neurosci 2016;9:92.
40. Pang SY-Y, Hsu JS, Teo K-C, et al. Burden of rare variants in ALS genes influences survival in familial and sporadic ALS. Neurobiol Aging 2017;58:238.e9-238.e15.
41. Morgan S, Shatunov A, Sproviero W, et al. A comprehensive analysis of rare genetic variation in amyotrophic lateral sclerosis in the UK. Brain 2017;140(6):1611–1618.

42. Nishiyama A, Niihori T, Warita H, et al. Comprehensive targeted next-generation sequencing in Japanese familial amyotrophic lateral sclerosis. Neurobiol Aging 2017;53:194.e1-194.e8.
43. Garton FC, Benyamin B, Zhao Q, et al. Whole exome sequencing and DNA methylation analysis in a clinical amyotrophic lateral sclerosis cohort. Mol Genet Genomic Med 2017;5(4):418–428.
44. Müller K, Brenner D, Weydt P, et al. Comprehensive analysis of the mutation spectrum in 301 German ALS families. J Neurol Neurosurg Psychiatry 2018;89(8):817–827.
45. Zhang H, Cai W, Chen S, et al. Screening for possible oligogenic pathogenesis in Chinese sporadic ALS patients. Amyotroph Lateral Scler Frontotemporal Degener 2018;19(5–6):419–425.
46. Liu Z-J, Lin H-X, Wei Q, et al. Genetic Spectrum and Variability in Chinese Patients with Amyotrophic Lateral Sclerosis. Aging Dis 2019;10(6):1199–1206.
47. Tripolszki K, Gampawar P, Schmidt H, et al. Comprehensive Genetic Analysis of a Hungarian Amyotrophic Lateral Sclerosis Cohort. Front Genet 2019;10:732.
48. Chen W, Xie Y, Zheng M, et al. Clinical and genetic features of patients with amyotrophic lateral sclerosis in southern China. Eur J Neurol 2020;27(6):1017–1022.
49. Lin F, Lin W, Zhu C, et al. Sequencing of neurofilament genes identified NEFH Ser787Arg as a novel risk variant of sporadic amyotrophic lateral sclerosis in Chinese subjects. BMC Med Genomics 2021;14(1):222.
50. Giguet-Valard A-G, Bellance R, Jeannin S, et al. SOD1-related ALS with anticipation in a large family from Martinique. Amyotroph Lateral Scler Frontotemporal Degener 2021;22(7-8):545-551.
51. Shepheard SR, Parker MD, Cooper-Knock J, et al. Value of systematic genetic screening of patients with amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry 2021;92(5):510–518.
52. McCann E, Henden L, Fifita J, et al. Evidence for polygenic and oligogenic basis of Australian sporadic amyotrophic lateral sclerosis. J Med Genet 2021;58(2):87–95.
53. Lin L. Bias caused by sampling error in meta-analysis with small sample sizes. PLoS One 2018;13(9):e0204056.
54. Xiao R, Boehnke M. Quantifying and correcting for the winner's curse in genetic association studies. Genet Epidemiol 2009;33(5):453–462.
55. Campos-Melo D, Hawley ZCE, Strong MJ. Dysregulation of human NEFM and NEFH mRNA stability by ALS-linked miRNAs. Mol Brain 2018;11(1):43.

56. Foggin S, Mesquita-Ribeiro R, Dajas-Bailador F, Layfield R. Biological Significance of microRNA Biomarkers in ALS—Innocent Bystanders or Disease Culprits? Front Neurol 2019;10:578.
57. Allison RL, Adelman JW, Abrudan J, et al. Microglia Influence Neurofilament Deposition in ALS iPSC-Derived Motor Neurons. Genes (Basel) 2022;13(2):241

FIGURE LEGENDS

Figure 1. PRISMA flowchart of the study systematic review process. The left of the figure outlines screening for articles identified via PubMed, Embase and Medline databases, whilst the right outlines the process for articles found via backwards citation screening of articles undergoing full-text screening.

Figure 2. *NEFH* domain distribution of the 59 variants identified from the systematic review. Colours characterise the different variant types. KEP = lysine-glutamic acid-proline; KSP = lysine-serine-proline.

Figure 3. Schematic depicting the locations of the gene variants included in the meta-analysis and in both meta-analysis and Project MinE (top), as well as the variants that were found to increase the risk for ALS with burden analysis (bottom). Green = only present in cases. Amber = present in cases and controls. Red = only present in controls.

Figure 4. Forest plot demonstrating that rare (MAF<1%) missense variants in the tail domain increase the risk of ALS.

Figure 5. Results of the SNV/indel screening analysis in the Project MinE cohort. Additional information on all 591 variants identified are available in the Supplementary Information. a) Proportion of variants found in various gene domains and untranslated regions (top), and in exons and introns (bottom). b) Breakdown of the 351 *NEFH* singletons by domain. c) A Venn diagram illustrating the overlap of the *NEFH* variants in Project MinE cases and controls, the systematic review and the gnomAD v2.1.1 database. The value for variants that are only in gnomAD (933) refers to the remaining *NEFH* variants in the catalogue after accounting for variants shared with Project MinE or the review.

TABLES

Table 1. Summary characteristics of all included studies identified from the systematic review.

Reference	Study Type	Discovery Method	Genetic Technology	Population (Country)	Sample Groups	Sex (M:F (Ratio))	Age (years (SD or range))	Diagnostic Criteria
17	Case-Control	Candidate Gene/Region	PCR	France and America	Case: 356 SALS Control: 306 neurologically healthy unrelated individuals	-	-	Definite ALS (El Escorial)
34	Case-Control	Candidate Gene/Region	PCR-SSCP	-	Case: 100 FALS + 75 SALS Control: 100 unrelated individuals	-	-	-
18	Case-Control	Candidate Gene/Region	PCR	UK	Case: 164 ALS Control: 207 age-matched unrelated individuals	-	-	-
16	Case-Control	Candidate Gene/Region	PCR-SSCP	UK and Scandinavia (Denmark, Norway, Sweden and Finland)	Case: UK: 19 FALS + 188 SALS Scandinavia: 59 FALS + 264 SALS Control: UK: 219 age, sex and	UK SALS: 109:79 (1.38) Scandinavian Cases: 194:129 (1.50)	UK SALS: 55.2 Scandinavian Cases: 61.5	-

					ethnicity-matched individuals Scandinavia: 228 age, sex and ethnicity-matched individuals			
35	Case-Control	Candidate Gene/Region	PCR	America	Case: 100 FALS + 100 SALS Control: 100 neurologically healthy individuals	-	-	-
36	Case-Control	Gene Panel	PCR	France and Canada	Case: 80 FALS + 110 SALS Control: 190 neurologically healthy individuals	104:86 (1.21)	55.4 ± 13.1	Definite or probable ALS (El Escorial)
37	Case-Control	Gene Panel	WES	America	Case: 242 SALS Control: 29 age-matched individuals	131:111 (1.18)	60 (44-82)	Definite, probable or possible ALS (El Escorial)
38	Case-Control	Gene Panel	WES + PCR	Japan	Case: 39 FALS + 469 SALS Control: 191 neurologically healthy individuals	298:210 (1.42)	62.1 (53.5-68.4)	Definite, probable, probable laboratory-supported or possible ALS (El Escorial)

39	Case	Gene Panel	NGS	Germany	Case: 80 ALS	44:36 (1.22)	60.1 (29-88)	-
40	Case-Control	Gene Panel	WGS + WES	China (Hong-Kong)	Case: 8 FALS + 46 SALS Control: 699 volunteer individuals	FALS: 3:5 (0.60) SALS: 28:18 (1.56)	FALS: 41.4 ± 8.71 SALS: 58.1 ± 13.45	Definite, probable or probable laboratory-supported ALS (El Escorial)
41	Case-Control	Gene Panel	WES + PCR	UK	Case: 131 FALS + 995 SALS Control: 613 neurologically healthy age and ethnicity-matched individuals	FALS: 67:64 (1.05) SALS: 567:428 (1.32)	FALS: 56 (24-85) SALS: 61 (25-88)	-
42	Case	Gene Panel	WES + Sanger	Japan	51 FALS	-	-	El Escorial
43	Case	Variant Panel	WES	Australia	120 ALS	75:45 (1.67)	61 ± 10.1	Definite or probable ALS (revised El Escorial)
44	Case	Gene Panel	WES + PCR	Germany	171 FALS	-	-	El Escorial
45	Case-Control	Gene Panel	WES + Sanger	China	Case: 311 SALS Control: 200 neurologically healthy individuals	199:112 (1.78)	51.92 ± 10.8	Definite, probable, probable laboratory-supported or possible ALS (El Escorial)

46	Case	Gene Panel	WES	China	24 FALS + 21 early-onset SALS	FALS: 13:11 (1.18) SALS: 13:8 (1.63)	FALS: 40.3 ± 14.8 SALS: 30.7 ± 11.5	El Escorial
47	Case	Gene Panel	WES	Hungary	107 ALS	45:62 (0.73)	60 (30-79)	El Escorial + Awaji-Shima
48	Case	Gene Panel	WES + Sanger	China	268 ALS	160:108 (1.48)	52.1 ± 10.4	Definite or probable ALS (El Escorial)
49	Case-Control	Candidate Gene/Region	PCR	China	Case: 671 SALS Control: 1787 neurologically healthy individuals	410:261 (1.57)	53.45 ± 9.96	El Escorial
50	Case Report	Whole Genome	WES	Martinique	5 related FALS	-	-	-
51	Case	Gene Panel	WES	UK	7 FALS + 93 SALS	54:46 (1.17)	60.4 (22-87)	Definite, probable, probable laboratory-supported or possible ALS (revised El Escorial)
52	Case	Variant Panel	WGS	Australia	616 SALS	346:221 (1.57)	60 ± 12	Definite or probable ALS (El Escorial)

All case-control studies (12) proceeded to the meta-analysis stage. FALS = familial ALS, SALS = sporadic ALS, WES = whole exome sequencing, WGS = whole genome sequencing, PCR = polymerase chain reaction, NGS = next-generation sequencing, PCR-SSCP = polymerase chain reaction-single-strand conformation polymorphism.

Table 2. Results of the subgroup meta-analysis.

MAF Category	Subgroup	No. Studies	N Cases	N Controls	Fixed OR (95% CI; p-value)	Random OR (95% CI; p-value)	Heterogeneity (Q statistic; p-value)	Heterogeneity (I-squared)	Publication Bias (Egger t statistic; p-value)	Publication Bias (Harbord t statistic; p-value)
Ultra-Rare (< 0.1%)										
	Tail	7	1914	2063	1.82 (0.68-4.88; 0.23)	1.82 (0.68-4.88; 0.23)	3.08 (0.80)	0%	4.36; **7.0E-03**	3.00; **0.03**
	Missense	6	1675	1521	2.65 (0.95-7.40; 0.06)	2.65 (0.95-7.40; 0.06)	3.28 (0.66)	0%	1.27; 0.27	0.58; 0.59
	Head Missense	2	285	290	1.63 (0.20-13.35; 0.65)	1.63 (0.20-13.35; 0.65)	0.25 (0.62)	0%	N/A	N/A
	Rod Missense	5	1368	1321	3.46 (0.86-13.97; 0.08)	3.46 (0.86-13.97; 0.08)	1.16 (0.89)	0%	-1.60; 0.21	-1.66; 0.20
	Tail Missense	3	864	1101	5.14 (0.88-30.21; 0.07)	5.14 (0.88-30.21; 0.07)	0.01 (0.99)	0%	2.67; 0.23	11.21; 0.06
	Tail Deletion	3	886	753	0.94 (0.26-3.38; 0.93)	0.94 (0.26-3.38; 0.93)	0.52 (0.77)	0%	2.25; 0.27	2.45; 0.25

Rare (< 1%)										
	Tail	9	3711	4463	2.42 (1.28-4.58; **6.6E-03**)	2.53 (1.01-6.33; **0.05**)	12.51 (0.13)	36%	-0.38; 0.72	0.35; 0.74
	Tail (exc. Missense)	5	2176	1575	1.05 (0.37-3.01; 0.93)	1.05 (0.37-3.01; 0.93)	4.32 (0.36)	7%	1.22; 0.31	0.16; 0.88
	Missense	8	2546	3408	2.37 (1.39-4.04; **1.5E-03**)	2.67 (1.26-5.65; **0.01**)	8.60 (0.28)	19%	0.60; 0.57	0.57; 0.59
	Missense (exc. Tail)	7	1875	1621	1.34 (0.66-2.71; 0.42)	1.34 (0.66-2.71; 0.42)	3.02 (0.81)	0%	0.43; 0.69	0.67; 0.53
	Rod Missense	7	1879	1621	1.53 (0.73-3.25; 0.26)	1.53 (0.73-3.25; 0.26)	2.31 (0.89)	0%	0.93; 0.39	1.63; 0.16
	Tail Missense	4	1164	2177	4.56 (2.13-9.72; **< 1.0E-04**)	4.67 (1.79-12.19; **1.6E-03**)	2.30 (0.51)	0%	2.11; 0.17	1.85; 0.21

	Tail Deletion	4	2012	1366	0.90 (0.30-2.74; 0.85)	0.90 (0.30-2.74; 0.85)	3.62 (0.31)	17%	0.75; 0.53	-0.50; 0.66
Common (> 1%)										
	Tail	6	2025	2442	1.03 (0.89-1.19; 0.72)	1.07 (0.86-1.32; 0.57)	7.62 (0.18)	34%	1.50; 0.21	1.37; 0.24
	Missense	5	1821	2223	1.03 (0.88-1.21; 0.73)	1.11 (0.78-1.58; 0.57)	8.11 (0.09)	**51%**	0.65; 0.56	0.66; 0.56
	Tail Missense	5	1818	2223	1.03 (0.87 – 1.21; 0.75)	1.11 (0.76-1.63; 0.58)	8.53 (0.07)	**53%**	0.61; 0.58	0.62; 0.58
	Tail Deletion	2	578	930	1.03 (0.78-1.35; 0.86)	1.03 (0.78-1.35; 0.86)	0.48 (0.49)	0%	N/A	N/A
All	Rod Missense	7	1879	1621	1.26 (0.66-2.42; 0.48)	1.26 (0.66-2.42; 0.48)	2.53 (0.87)	0%	1.50; 0.19	2.18; 0.08
	Tail Missense	8	2982	4400	1.08 (0.92-1.27; 0.34)	1.76 (0.92-3.37; 0.09)	20.64 (**4.3E-03**)	**66%**	2.39; **0.05**	2.49; **0.05**
	Tail Deletion	5	2383	2077	1.02 (0.78-1.33; 0.89)	1.02 (0.78-1.33; 0.89)	4.02 (0.40)	1%	0.29; 0.79	0.35; 0.75

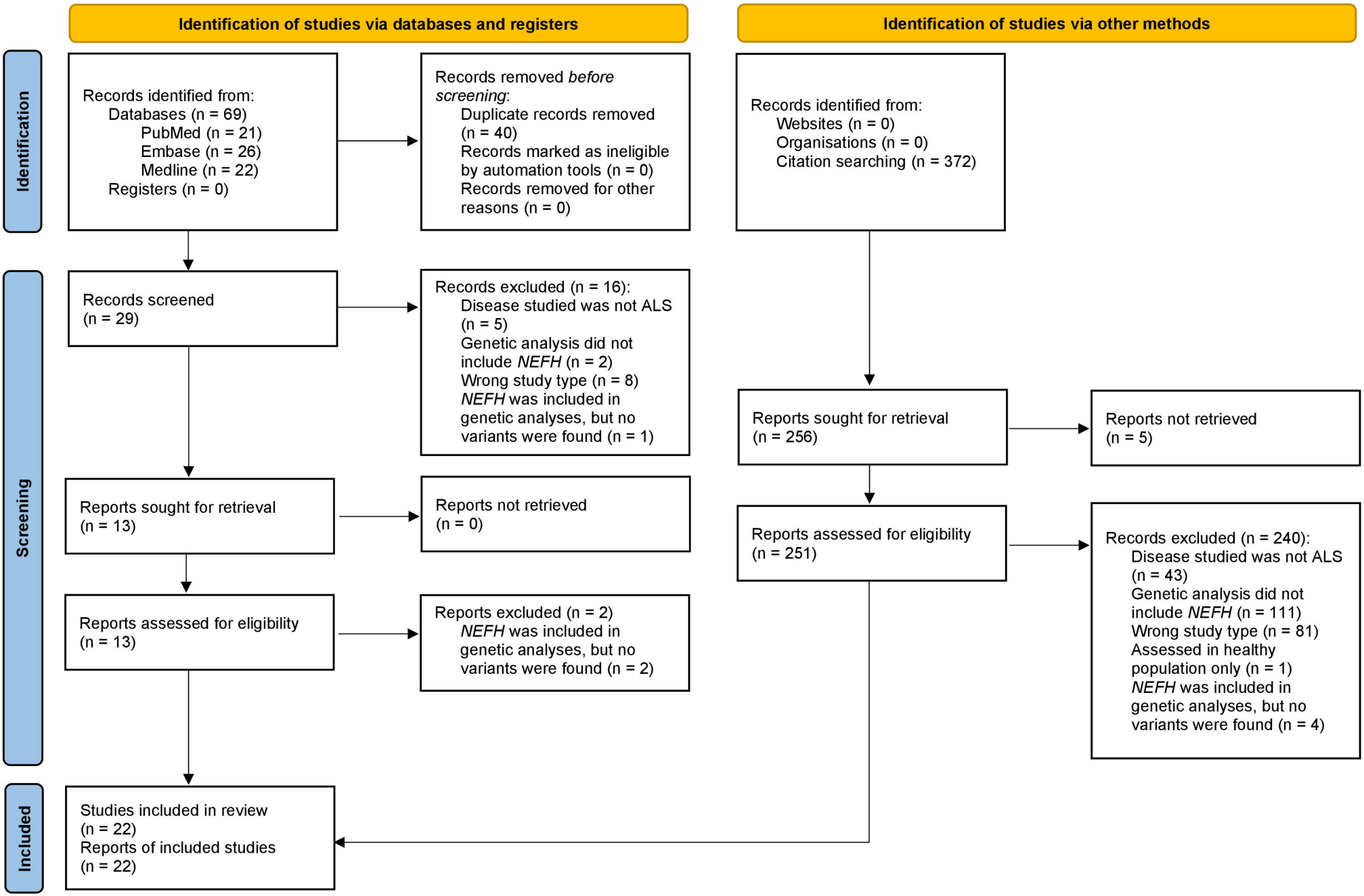
N/A values represent instances where publication bias could not be calculated as the minimum number of studies required for calculation was not reached

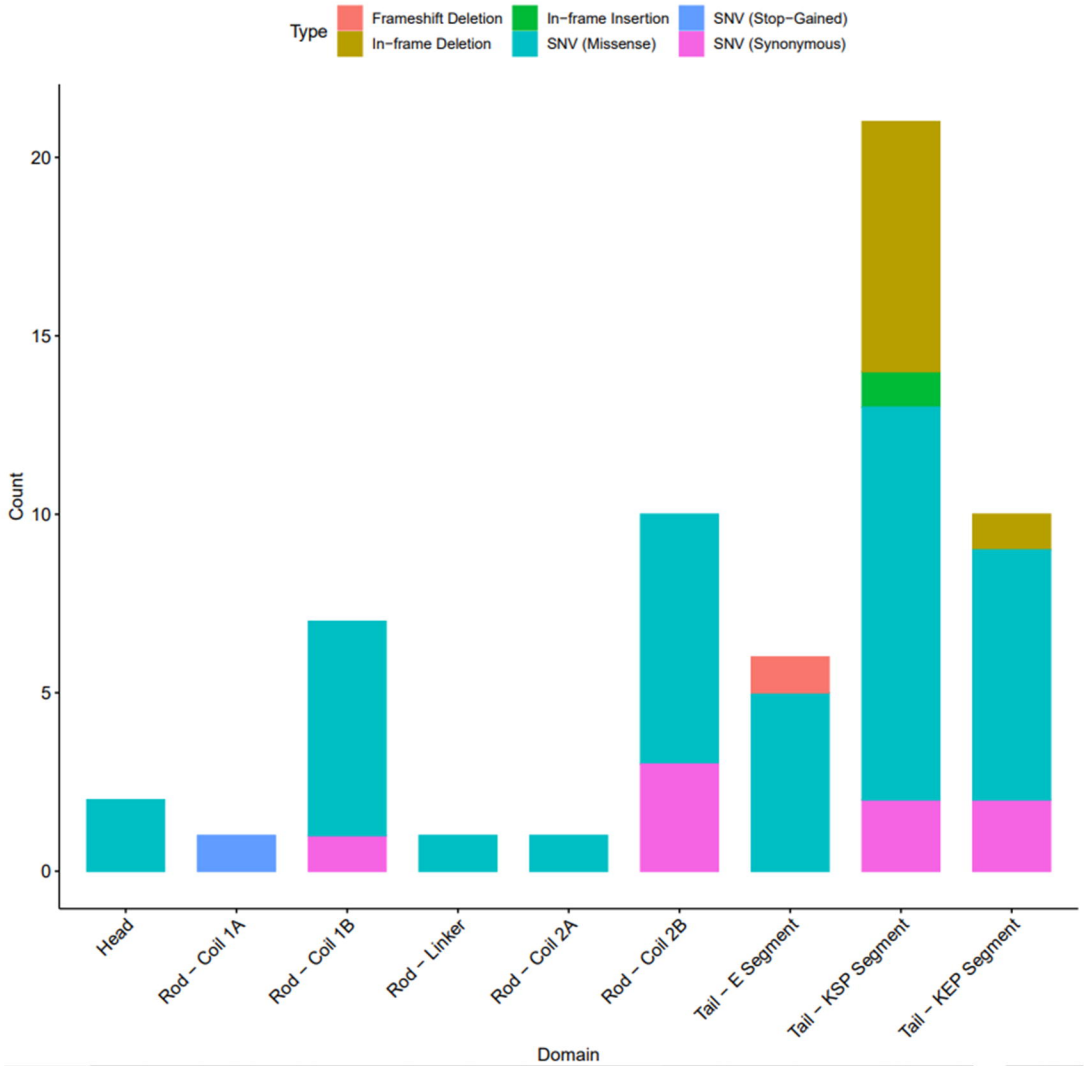
Table 3. Case-control frequencies of the 16 SNV/indel variants found in both the systematic review and in the Project MinE cohort, and the 4 structural variants identified in a subset of Project MinE.

Variant	Position	Type	Pathogenicity Prediction	CADD -SV Score	Total Frequency (Case/Control)	Beta (SE)	P-Value	OR (95% CI)
A90V	29876520	Missense	Tolerated/Benign	N/A	7 (6/1)	0.49 (0.95)	0.61	2.26 (0.27-18.77)
G249S	29876996	Missense	Tolerated/Benign	N/A	92 (69/23)	0.03 (0.23)	0.91	1.13 (0.70-1.82)
A314V	29879421	Missense	Deleterious/ Probably Damaging	N/A	1 (1/0)	0.22 (2.11)	0.92	1.13 (0.05-27.73)
E463K	29885016	Missense	Deleterious/ Probably Damaging	N/A	1527 (1103/424)	-0.0065 (0.06)	0.91	0.97 (0.86-1.10)
T642M	29885554	Missense	Tolerated/Benign	N/A	1 (0/1)	-2.47 (2.11)	0.24	0.13 (0.005-3.08)
K647N	29885570	Missense	Deleterious/ Possibly Damaging	N/A	1 (1/0)	0.40 (2.12)	0.85	1.13 (0.05-27.73)
P777L	29885959	Missense	Deleterious/ Probably Damaging	N/A	1 (1/0)	0.14 (2.12)	0.95	1.13 (0.05-27.73)
E805A	29886043	Missense	Deleterious/ Possibly Damaging	N/A	2578 (1873/705)	-0.028 (0.05)	0.54	1.00 (0.90-1.11)
K867N	29886230	Missense	Deleterious/ Possibly Damaging	N/A	1 (1/0)	0.40 (2.12)	0.85	1.13 (0.05-27.73)
A400A	29881828	Synonymous	N/A	N/A	2569 (1867/702)	0.03 (0.05)	0.57	1.00 (0.90-1.11)
S580S	29885369	Synonymous	N/A	N/A	199 (147/52)	0.10 (0.16)	0.56	1.07 (0.77-1.47)

D919D	29886386	Synonymous	N/A	N/A	148 (105/43)	-0.05 (0.18)	0.77	0.92 (0.64-1.31)
V928V	29886413	Synonymous	N/A	N/A	8492 (6198/2324)	-0.007 (0.04)	0.85	1.08 (0.86-1.36)
K790del	29885996	Small Deletion	N/A	N/A	18 (9/9)	-1.03 (0.47)	**0.03**	0.38 (0.15-0.95)
E658_K665del	29885604	Small Deletion	N/A	N/A	29 (18/11)	0.02 (0.64)	0.98	0.61 (0.29-1.30)
E664_P669del	29885622	Small Deletion	N/A	N/A	13 (10/3)	-2.20 (2.11)	0.30	1.25 (0.35-4.56)
INS_61	29877834-29877895	Large Deletion	N/A	14.27	1 (1/0)	0.47 (2.11)	0.82	1.19 (0.049-29.24)
INS_56	29879816-29879872	Large Deletion	N/A	0.75	1 (0/1)	-2.38 (2.11)	0.26	0.13 (0.0054-3.25)
DEL_1169	29880841-29882010	Large Deletion	Likely Pathogenic	9.83	1 (1/0)	0.65 (2.11)	0.76	1.19 (0.049-29.24)
DEL_113	29885279-29885870	Large Deletion	VUS	4.85	1743 (1161/582)	-0.34 (0.061)	**2.60E-08**	0.72 (0.64-0.81)

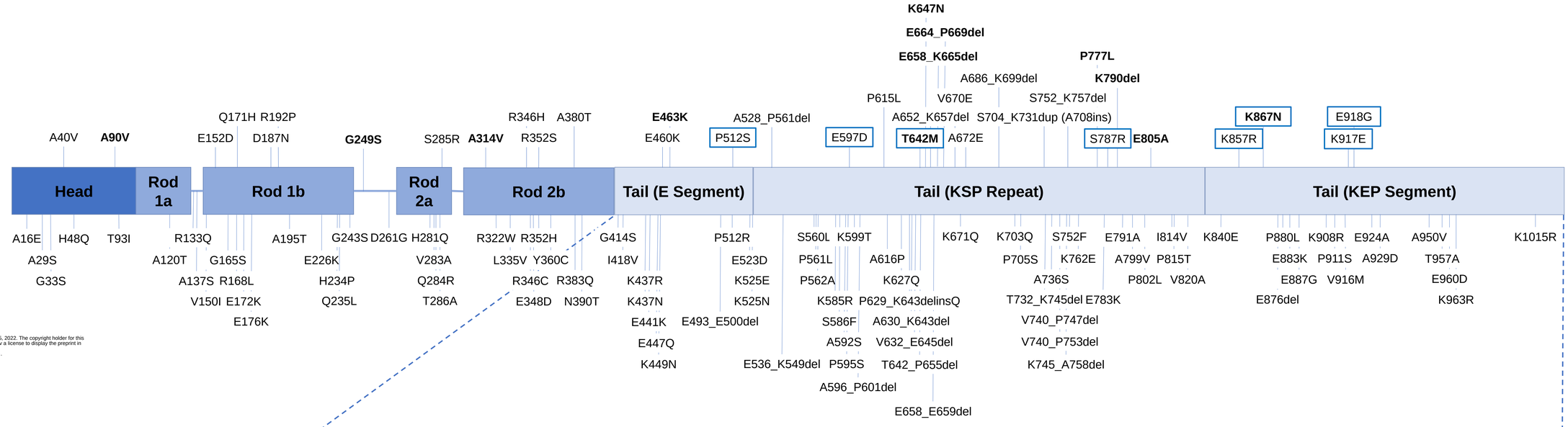
Significance is denoted by Firth logistic regression p-value of ≤ 0.05 (corrected for sex and 10 principal components). N/A refers to a pathogenicity prediction that could not be reached as the variant type was inappropriate for the tool (SIFT/PolyPhen for SNVs and indels; ACMG and CADD-SV for SVs).


Identification of studies via databases and registers
Identification of studies via other methods
Identification
Screening
Included
Records identified from:
Databases (n = 69)
PubMed (n = 21)
Embase (n = 26)
Medline (n = 22)
Registers (n = 0)
Records removed *before screening*:
Duplicate records removed (n = 40)
Records marked as ineligible by automation tools (n = 0)
Records removed for other reasons (n = 0)
Records identified from:
Websites (n = 0)
Organisations (n = 0)
Citation searching (n = 372)
Records screened (n = 29)
Records excluded (n = 16):
Disease studied was not ALS (n = 5)
Genetic analysis did not include *NEFH* (n = 2)
Wrong study type (n = 8)
NEFH was included in genetic analyses, but no variants were found (n = 1)
Reports sought for retrieval (n = 256)
Reports not retrieved (n = 5)
Reports sought for retrieval (n = 13)
Reports not retrieved (n = 0)
Reports assessed for eligibility (n = 251)
Records excluded (n = 240):
Disease studied was not ALS (n = 43)
Genetic analysis did not include *NEFH* (n = 111)
Wrong study type (n = 81)
Assessed in healthy population only (n = 1)
NEFH was included in genetic analyses, but no variants were found (n = 4)
Reports assessed for eligibility (n = 13)
Reports excluded (n = 2)
NEFH was included in genetic analyses, but no variants were found (n = 2)
Studies included in review (n = 22)
Reports of included studies (n = 22)

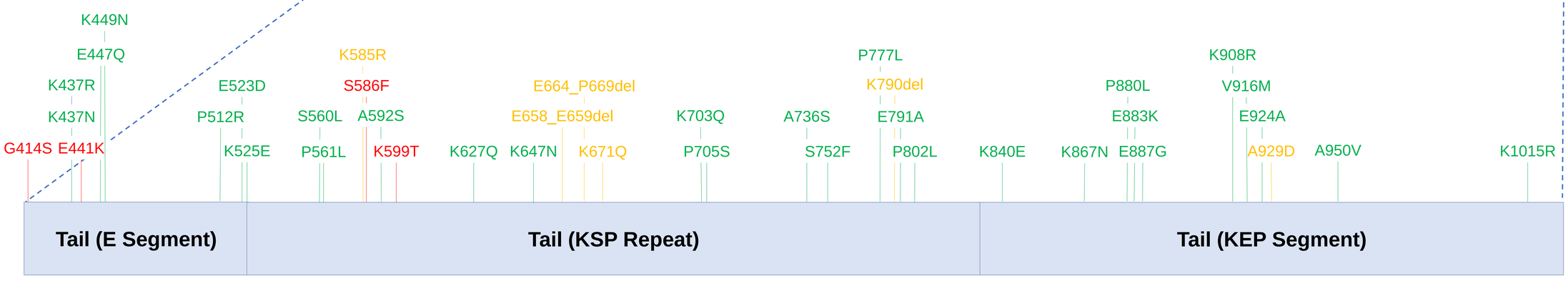

Type
Frameshift Deletion
In−frame Deletion
In−frame Insertion
SNV (Missense)
SNV (Stop−Gained)
SNV (Synonymous)
Count
0
5
10
15
20
Head
Rod − Coil 1A
Rod − Coil 1B
Rod − Linker
Rod − Coil 2A
Rod − Coil 2B
Tail − E Segment
Tail − KSP Segment
Tail − KEP Segment
Domain

Variants included in literature-based meta-analysis

Missense and in-frame deletions identified in Project MinE



Ultra-rare (MAF <0.1%) pathogenic missense and high-frequency rare (MAF 0.1-1%) in-frame deletion mutations identified from burden analysis as increasing ALS risk

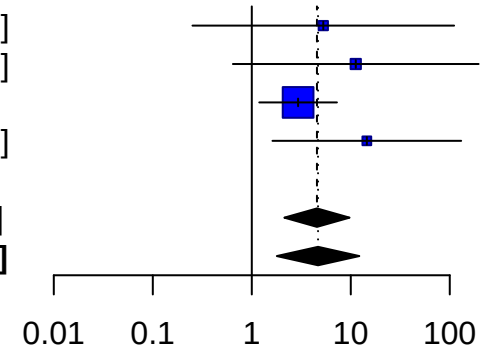


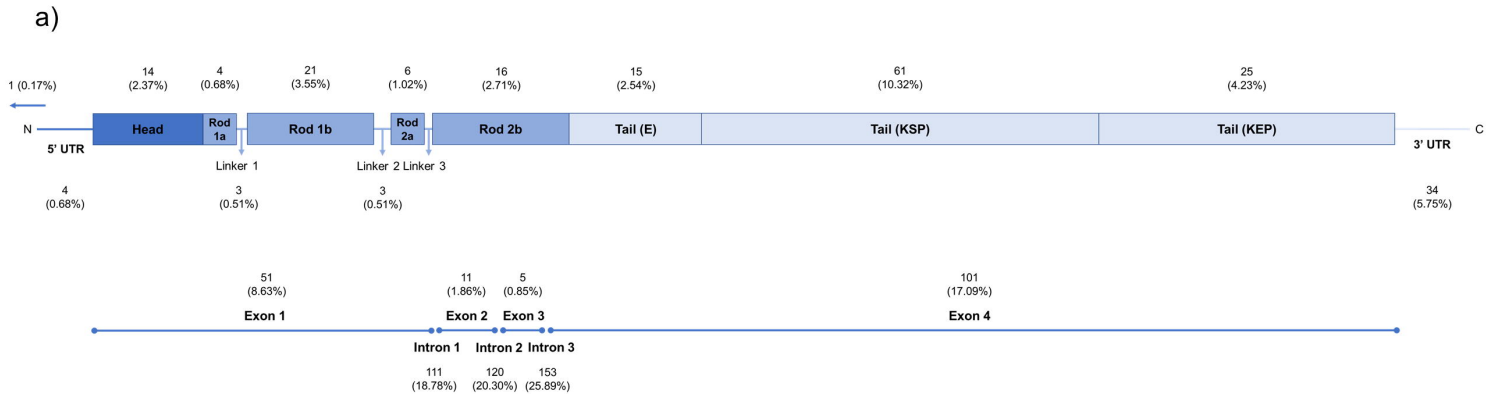
Study	ALS Events	ALS Total	Control Events	Control Total	Weight (Fixed)	Weight (random)	OR	95% CI
Daoud et al., 2011	2	182	0	190	7.1%	9.3%	5.28	[0.25; 110.67]
Zhang et al., 2018	8	311	0	200	8.7%	10.4%	11.23	[0.64; 195.66]
Lin et al., 2021	12	371	8	711	77.9%	63.4%	2.94	[1.19; 7.25]
Lin et al., 2021 (Replication)	4	300	1	1076	6.3%	16.9%	14.53	[1.62; 130.46]
Total (fixed effects, 95% CI)		**1164**		**2177**	**100.0%**	**--**	**4.56**	**[2.13; 9.72]**
Total (random effects, 95% CI)					**--**	**100.0%**	**4.67**	**[1.79; 12.19]**

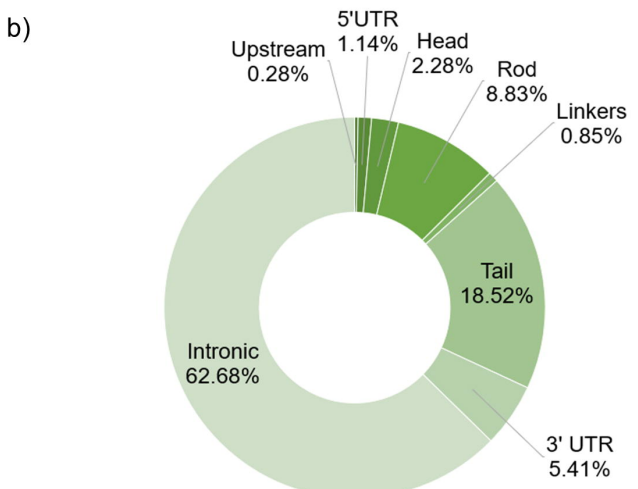
Heterogeneity:Tau^2 = 0.1651; Chi^2 = 2.30, df = 3 (P = 0.51); I^2 = 0%

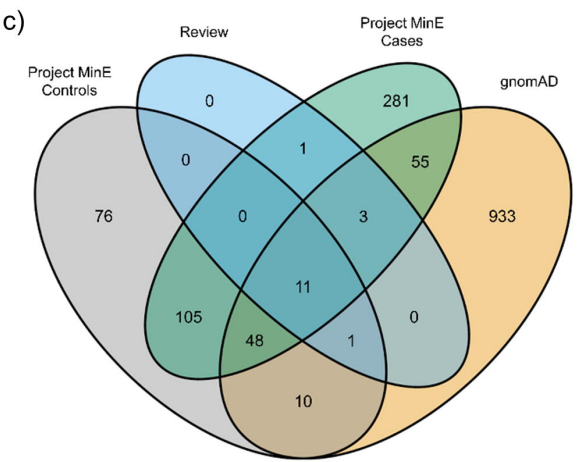
Total (Fixed Effects) P < 0.0001

Total (Random Effects) P = 0.0016




a)
1 (0.17%)
14 (2.37%)
4 (0.68%)
21 (3.55%)
6 (1.02%)
16 (2.71%)
15 (2.54%)
61 (10.32%)
25 (4.23%)
N
Head
Rod 1a
Rod 1b
Rod 2a
Rod 2b
Tail (E)
Tail (KSP)
Tail (KEP)
C
5' UTR
3' UTR
Linker 1
Linker 2
Linker 3
4 (0.68%)
3 (0.51%)
3 (0.51%)
34 (5.75%)
51 (8.63%)
11 (1.86%)
5 (0.85%)
101 (17.09%)
Exon 1
Exon 2
Exon 3
Exon 4
Intron 1
Intron 2
Intron 3
111 (18.78%)
120 (20.30%)
153 (25.89%)


b)
Upstream 0.28%
5'UTR 1.14%
Head 2.28%
Rod 8.83%
Linkers 0.85%
Tail 18.52%
Intronic 62.68%
3' UTR 5.41%


c)
Review
Project MinE Cases
Project MinE Controls
gnomAD
0
281
1
0
55
76
0
3
933
11
105
0
48
1
10