Advances in the Genetic Classification of ALS

Johnathan Cooper-Knock[1], **Calum Harvey**[1], **Sai Zhang**[2,3], **Tobias Moll**[1], **Ilia Sarah Timpanaro**[4], **Kevin P. Kenna**[4], **Alfredo Iacoangeli**[5,6,7], **Jan H Veldink**[4]

[1]Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield, UK

[2]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

[3]Center for Genomics and Personalized Medicine, Stanford University School of Medicine, Stanford, CA, USA

[4]Department of Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherland

[5]Maurice Wohl Clinical Neuroscience Institute, Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

[6]Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

[7]National Institute for Health Research Biomedical Research Centre and Dementia Unit, South London and Maudsley NHS Foundation Trust and King's College London, London, UK

Abstract

Purpose of review—Amyotrophic lateral sclerosis (ALS) is an archetypal complex disease where disease risk and severity are, for the majority of patients, the product of interaction between multiple genetic and environmental factors. We are in a period of unprecedented discovery with new large-scale genome-wide association study (GWAS) and accelerating discovery of risk genes. However, much of the observed heritability of ALS is undiscovered and we are not yet approaching elucidation of the total genetic architecture which will be necessary for comprehensive disease subclassification.

Recent findings—We summarise recent developments and discuss the future. New machine learning models will help to address nonlinear genetic interactions. Statistical power for genetic discovery may be boosted by reducing the search-space using cell-specific epigenetic profiles and expanding our scope to include genetically correlated phenotypes. Structural variation, somatic heterogeneity and consideration of environmental modifiers represent significant challenges which will require integration of multiple technologies and a multidisciplinary approach including clinicians, geneticists and pathologists.

Summary—The move away from fully penetrant Mendelian risk genes necessitates new experimental designs and new standards for validation. The challenges are significant but the potential reward for successful disease subclassification is large-scale and effective personalized medicine.

Keywords

Amyotrophic lateral sclerosis; genetics; personalised medicine

Introduction

Amyotrophic lateral sclerosis (ALS) is a rapidly progressive and universally fatal late age of onset neurodegenerative disease involving loss of motor neurons. ALS is relatively common with a lifetime risk of ~1/400 (1). ALS is also an archetypal complex disease where 10% of patients suffer monogenic disease but the majority of disease, known as sporadic, is determined by an interaction of multiple environmental and genetic risk factors. More than 30 ALS genes have been previously identified; in monogenic disease the most frequent mutations are found within *C9ORF72*, *SOD1*, *TARDBP* and *FUS*. Changes in known ALS genes are found in ~21% of ALS patients, and the presence of more than one variant is associated with lower age of onset (2). However, there is still substantial missing heritability which is not explained by known ALS genes. It seems likely that the majority of ALS patients carry multiple risk variants which have not yet been identified. Our ability to profile and subdivide sporadic ALS is likely to determine the future of personalised medicine interventions for the majority of ALS patients.

The world is in the midst of the global COVID-19 pandemic which has significantly impeded research progress. Despite this there has been significant progress in the field of ALS genetics. A new ALS genome-wide association study (GWAS) has recently been reported (3). The large sample size of this new study is a significant step forwards and the authors have reported 8 new genome-wide significant loci (Figure 1). In addition several new ALS risk genes have been reported in recent years including *ATXN1 (4)*, *CAV1 (5)*, *SPTLC1 (6)*, *ACSL5 (7,8)*, *DNAJC7* (9), ANXA11 (10) and *GLT8D1* (11). There is a move away from fully penetrant genetic changes, and a move towards identification of genetic risk factors which are probably pathogenic only in conjunction with other genetic and environmental risk factors. Indeed the majority of monogenic disease has now been explained (12) and remaining undiscovered genetic changes with complete penetrance are likely to affect relatively small numbers of pedigrees.

Broad sense heritability for sporadic ALS is ~50% (13,14). This contrasts sharply with the estimates of SNP-based heritability from GWAS which are as low as 3.5% (3). More research is needed to identify the source of this missing heritability. GWAS estimates of heritability typically utilise linear methods which likely underestimate the effect of nonlinear interactions, which may be important in a complex biological system. In contrast, broad sense heritability is often calculated using relatives where nonlinear effects can be profiled more accurately. Using a hypothetical model we have previously shown that linear methods can dramatically fail to identify nonlinear heritability (15). Alternatively, modelling by

others has suggested that the impact of nonlinear effects may be negligible (16). Currently, genetic association studies are still dominated by linear methods which, like the heritability estimates, may be underpowered to identify certain genetic risk factors. New, nonlinear methodology is required to answer this question definitively.

A limiting factor in the discovery of ALS genetic risk loci is poor statistical power despite increasing sample sizes. This is an important reason why <10% of SNP-based heritability has been assigned to a specific locus. We have recently demonstrated that statistical power can be improved by reducing the search space to areas of the genome which are functional in a cell type of interest, such as motor neurons (17). Recent advances in single cell profiling of relevant CNS tissues promise much in this regard (18,19). Moreover, integration of other biological data such as protein-protein interaction data and known gene functions can further aid gene prioritization (20,21)

The challenge of genetic classification is to group patients according to management strategy. It is noteworthy that risk genes increasingly converge on certain biological pathways (Figure 2) which may represent distinct therapeutic targets. It has been known for some time that approximately 98% of ALS patients share TDP-43 mislocalization (22) and new research linking this change to molecular and clinical phenotypes suggests that this may represent a final common pathway of ALS amenable to therapeutic intervention (23–25). Cell and animal models are important for drug development, but are usually produced via genetic manipulation and therefore rely on accurate genetic classification. Mistakes in this area can be significant, for example therapies developed for the SOD1-mouse model of ALS have largely failed to translate because SOD1-ALS is not representative of the majority of ALS patients having a distinct molecular basis and neuropathology (26).

Impact of new GWAS

The new GWAS from van Rheenen et al (3) includes a meta-analysis of 117 cohorts, 29,612 ALS patients and 122,656 controls, which represents a significant step change in the number of cases compared to previous studies (27,28). The study has reported 15 genome-wide significant loci (Figure 1) of which 8 have previously been identified but 7 are new. The study also included whole genome sequencing of 6,538 ALS patients and 2,415 controls which aided comparison of rare and common variant signals. With imputation (29), the GWAS included variants with minor allele frequency (MAF) >0.1% whereas whole genome sequencing profiles all variants. In brief, all identified loci fall within four categories: 1) Rare variants within coding sequence which likely to be directly causal, e.g. within *SOD1*, *KIF5A* and *CFAP410*; 2) Single nucleotide polymorphisms (SNPs) that are in linkage with a pathogenic repeat expansion such as the G4C2-repeat expansion within *C9ORF72*; 3) Loci that consist of both common regulatory variants and rare coding variants that are distinct (not in significant linkage) but have overlapping functional consequences, e.g. variants associated with *NEK1* and *TBK1; 4)* Lastly, remaining loci are those where there is no direct link to a causal gene through coding variants or repeat expansions; these are likely to have relatively subtle effects mediated through as yet unknown changes in gene and protein expression.

Both regulatory and coding variants within the *NEK1* and *TBK1* loci are thought to cause loss of function (LoF) of the target gene in ALS patients. Similarly we have profiled ALS-associated regulatory variation and identified other non-GWAS hits where there is a convergence of common and rare variant signal in a LoF mechanism (5,17). Moreover, a similar phenomenon has been reported in other disease areas (30). Regulatory variants are thought to have lower effect size and more redundancy (31) whereas coding variants typically show higher penetrance. In reality this is likely to be more of a spectrum where population frequencies may be a measure of selection pressure and indirectly effect size.

The new GWAS also took advantage of growing genetic profiles of other diseases to measure genetic correlation between neurodegenerative diseases including progressive supranuclear palsy (PSP), Alzherimer's disease (AD), Parkinson's disease (PD) and frontotemporal dementia (FTD). This is an area of increasing understanding but it is clear that ALS-associated risk variants can be associated with multiple phenotypes, most notably FTD. This has been observed previously in the context of populations (32) and pedigrees (33,34). Shared genetic risk amongst different phenotypes may also go some way to explain why ALS is strongly associated with rare variants (28) suggesting a role for selection pressure, despite a late age of onset that should have a minimal effect on reproductive fitness. It is potentially significant that ALS-associated mutations have been associated with molecular mechanisms underlying fertilisation (35).

To date genetic association studies have largely relied on short sequencing reads of up to 150bp. Accurate assessment of SNPs is achieved, but there is an inherent uncertainty in assessment of variants which are longer than the basic read length. This is compounded by a positive bias towards sequence of intermediate GC-content (36) which can reduce coverage of repetitive genomic regions. A key example is the failure for many years to identify the intronic G4C2-repeat expansion within *C9ORF72* which is the most common genetic cause of ALS and FTD (37,38). A number of other repeat-expansions have been associated with ALS including short-tandem repeats (STR) (4,39) and variable-number tandem repeats (40). The new GWAS (3) includes a genome-wide screen for ALS-associated STR using short read sequencing data. This is achieved by taking advantage of previously annotated STRs (41) and identification of anchored read pairs containing repetitive sequence and adjacent sequence which can be confidently mapped to the genome (42). Expansion of a STR downstream of NEK1, a known ALS gene (43), was thus significantly associated with ALS risk.

Genetic classification of a polygenic disease

The promise of genetic classification is nothing short of personalised medicine. For monogenic disease this is already becoming a reality (44). However, a prerequisite for widely applicable subclassification is an approximation of the total genetic architecture of ALS. Current efforts to capture genetic architecture are dominated by polygenic risk scores (PRS) developed from GWAS data (45) but this is a linear method and success has been limited (46). A recent effort to produce a PRS for ALS (47) reported limited ability to provide individualised predictions with a maximum AUC of only 0.57. Advances in machine learning have led to new tools such as deep learning, which are the state of the art in many

classification problems (48) and hold promise for analysis of complex biological systems (49). Machine learning approaches offer an opportunity to integrate multiple data-types; a model based on a graph representation of protein-protein interactions, known gene functions and gene-disease associations has been proposed for the prediction of novel ALS genes (20). A retrospective validation study of predictions made using this model in early 2019, revealed that those predictions were enriched (p=0.012) for subsequently discovered ALS genes (21).

It is suggested that ALS is a multistep process involving both genetic and environmental factors (50,51). As a result it may be impossible to effectively classify ALS based on genetic variation in isolation. Progress has been made to identify environmental modifiers of ALS risk using Mendelian randomisation (MR), including strenuous exercise (52) and blood lipids (3,53). MR relies on measured genetic liability to a particular environmental exposure and therefore it is amenable to study of the gene-environment interaction (54). However, efforts to perform this at scale will likely require either clear hypotheses, or large sample sizes. The gut microbiome is an effective integrator of environmental effects (55) and therefore microbiome profiling alongside host genetics may be an effective method of measuring gene-environment interactions. By matching multiple profiles within the same patients, including for example metabolome and microbiome, it may be possible to boost power to identify new genetic risk while simultaneously determining the target group. Such multi-omics data is increasingly available from consortia such as AnswerALS (https://www.answerals.org/).

In a fascinating development, the new ALS GWAS (3) has shown that severity and risk of ALS are genetically independent (3). The suggestion is that genetic classification, and therefore identification of effective therapeutic targets, may require integration of separate risk and severity signals.

ALS risk genes converge in pathways

The new ALS GWAS (3) took advantage of a new RNA-seq dataset including cortical tissues from 31,684 samples known as Metabrain (56). Using brain specific co-expression networks the authors localised ALS genetic risk to Gene Ontology terms related to vesicle mediated transport and autophagy. Vesicle mediated transport is involved in synaptic neurotransmission but also in vesicle fusion, which is an essential component of autophagy (57). Our study of the genetic architecture of ALS in the context of motor neuron function highlighted function within the distal axon (17) and demonstrated that axonal dysfunction associated with LoF of *KANK1*, a putative new ALS gene, may be upstream of TDP-43 mislocalization.

Various databases have complied lists of ALS risk genes and variants including ALSoD (https://alsod.ac.uk/), GCEP (https://clinicalgenome.org/tools/educational-resources/gene-disease-validity-topics/gcep-presentations/) and ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/?term=amyotrophic+lateral+sclerosis). Overall there is an observed convergence of risk genes within a relatively small number of pathways (58) including cytoskeletal dynamics, protein homeostasis including autophagy, RNA processing and immune cell function (Figure 2). The observed convergence may be subject to some publication bias

which is difficult to combat. However, overall there is good evidence that these pathways should be subject to an intense search for effective therapeutic targets, and future genetic subclassification may be focused on these pathways which have already been identified. However, if severity and risk of ALS are genetically independent then perhaps the field needs to shift efforts to discovery of biological pathways important for disease outcome measures such as age at onset, survival and cognitive impairment.

In the past six months the ALS field is evaluating an evolving story which comes close to a universal pathogenic mechanism. TDP-43 is said to be the central protein for ALS because mislocalization of nuclear TDP-43 within motor neurons and the formation of cytoplasmic neuronal TDP-43-positive inclusions links >98% of ALS patients (22) and predicts the severity of neuronal loss (59). TDP-43 has long been shown to regulate RNA splicing but a specific role in control of splicing of cryptic exons was only discovered in 2015 (60). It has been recently shown that TDP-43 controls the expression of cryptic exons within *UNC13A* mRNA. *UNC13A* is the subject of known ALS GWAS locus where ALS-associated SNPs, are linked to risk of disease (61) and patient survival (25,62). When TDP-43 is lost from the nucleus then cryptic exons are included within the mature *UNC13A* mRNA which are otherwise excluded; in some cases this leads to nonsense mediated decay (NMD) and loss of protein function. Crucially, ALS-associated GWAS SNPs within *UNC13A* exacerbate the effect of TDP-43 dysfunction leading to LoF of UNC13A protein (23,24) suggesting that this mechanism is upstream in disease pathogenesis. Going forwards, cryptic exon inclusion may become a unifying theme in ALS pathology; new discoveries may need to be evaluated for an interaction with TDP-43 mislocalization and/or UNC13A. For example, it has been shown that TDP-43 normally suppresses a cryptic polyadenylation site within the axonal protein STMN2; loss of nuclear TDP-43 leads to NMD of the *STMN2* mRNA and near-total knock-down of the protein (63,64). LoF of STMN2 has been associated with motor neuron toxicity (65), genetic variants within *STMN2* have been associated with ALS severity (66) and the interaction between TDP-43 and *STMN2* mRNA is the focus of active translational research.

Future Directions for discovery of ALS genetic risk

We have already stated that the prospect for discovery of new monogenic changes in ALS is now limited to what are likely very rare mutations with undetermined relevance for the majority of patients. We have argued that new approaches are required to address complex inheritance patterns in sporadic ALS including nonlinear machine learning models, integration of epigenetic profiling in vulnerable cell types, recognition of associated pleiotropic phenotypes (particularly FTD) and correlation with measures of environmental risk factors to draw out gene-environment interactions. The case-control design that informs many GWAS and even rare variant studies, such as Project MinE (67), is effective but increasing sample sizes are unlikely to be sufficient alone to delineate a total genetic architecture of ALS, given the inherent lack of phenotypic detail. On the other hand, classical, family based studies will also be of limited value, given the large degree of incomplete penetrance and partly recognized pleiotropy.

Fundamentally genetic risk is inherited and therefore in complex diseases, family members carry higher than background genetic risk even if they are unlikely to develop disease (68). As a result, exclusion of unaffected family members is likely to lead to a loss of substantial statistical power. To take advantage of this we purpose a 'super pedigree' approach utilising non-Mendelian pedigrees with variable penetrance and pleiotropy. The focus would be on identification of drivers of genetic risk of ALS and related disorders. Candidate related disorders are dyslipidemia, body mass index (BMI), cognitive disorders and psychiatric disease.

We have described the recent discovery of several structural variants associated with ALS including STRs and VNTRs. Similar technology should allow more accurate profiling of copy number variants (CNV) a number of which have been associated with rare instances of ALS (69). Interestingly TDP-43 mislocalization has been associated with aberrant expression of LINE1 retrotransposons (70) which are a frequent source of structural variation within the genome. It is likely that a significant number of as yet undiscovered structural variants are associated with ALS. Long-read sequencing technology is leading to advances in this area (71) but cost, throughput and accuracy may prohibit large-scale genome wide long read sequencing studies for some time (72). The near future is likely to be dominated by improvements in imputing structural variants from short read sequencing data, perhaps using models trained using long read sequencing; indeed several exciting new genome builds are being released (73) which will be essential in this area.

Finally, there is a case for somatic mosaicism in ALS. Somatic mosaicism could in theory explain a number of observed phenomena in ALS: 1) The adult onset of the disease which could result from accumulation of mutations within transcriptionally active brain regions (74); 2) The multifocal and heterogeneous onset of disease; 3) The failure to delineate either a total genetic architecture or establish broadly relevant environmental risk factors for ALS (75); 4) The selective involvement of particular neuronal populations, particularly within the motor system. Notably, the most common genetic risk factor for ALS, G4C2-repeat expansion of *C9ORF72*, shows marked somatic mosaicism (76). Technical challenges have limited discovery of other somatic mutations, for example it is difficult to distinguish a rare change affected in a limited number of cells from technical variation arising during sequencing (77). Also, by definition mutations which have occurred in dying cells are hard to capture at post-mortem. Somatic heterogeneity has long been studied in the context of cancer biology and perhaps methodology can be adapted from this field as described in (78) and (79). ALS is characterised by selective vulnerability of specific cell populations within the CNS and so non-CNS tissues may be an effective control for technical variation within an individual. However, many existing ALS-associated mutations are notable for their global expression and it may be that non-genetic factors, or even cell-specific functions (17–19), are a more important determinant of selective vulnerability. Addressing this problem will require a multidisciplinary approach involving clinicians, pathologists and geneticists.

Importance of validation for incompletely penetrant changes

There is already a case to be made that certain ALS-associated mutations are false positives (80). Whilst a period of evaluation is necessary for any new discovery, the ALS field needs

to agree consensus on genetic risk factors if we are ultimately to reach a clinical grade test for disease subclassification.

A common cause of false positives is population variability and it is therefore important that case-control designs include matched control populations but also that biological signals replicate in different populations as has been observed in the most recent GWAS (2) where 8 loci were genome-wide significant in European and Asian ancestry groups (Figure 1).

Experimental validation is a helpful tool, particularly when classifying risk genes by pathway and ultimately therapeutic target, as has already been discussed. However, some models are likely to be more representative than others. The field is currently favouring patient-derived models (81) which can recapitulate key molecular pathologies including TDP-43 mislocalization (82) and astrocyte toxicity towards motor neurons (83). These models also have the advantage that they can be adapted for drug screening but still require robust readouts that are relevant to clinical phenotypes. Biobanks such as AnswerALS are producing iPSC-derived motor neurons from patients at scale alongside other omics and clinical data (https://www.answerals.org/) which will facilitate identification of such readouts.

Conclusion

In conclusion, we are experiencing a rapid expansion in the genetic profiling of ALS headlined by new GWAS (3), the largest disease-specific whole genome sequencing consortium for any disease in Project MinE (https://www.projectmine.com/) and large multi-omics datasets such as AnswerALS (https://www.answerals.org/). This is leading to unprecedented discovery of risk genes and loci and is moving the field away from monogenic fully penetrant changes towards risk variants with variable penetrance which interact with other genetic and environmental factors. This new era brings new challenges with a need for new methods. Future studies will need to account for reduced penetrance, genetic pleiotropy and nonlinear inheritance. Both structural variation within the genome and somatic heterogeneity are promising areas of research. Lastly, we predict that successful genetic subclassification of ALS will have a dramatic positive influence on therapeutic developments, both by directly targeting mutations with genetic therapy approaches but also by enabling the generation of new, more valid ALS models for drug screening to identify neuroprotective compounds.

Acknowledgements

References

1. Kiernan MC, Vucic S, Cheah BC, Turner MR, Eisen A, Hardiman O, et al. Amyotrophic lateral sclerosis. Lancet. 2011; Mar 12; 377 (9769) 942–55. [PubMed: 21296405]

2. Shepheard SR, Parker MD, Cooper-Knock J, Verber NS, Tuddenham L, Heath P, et al. Value of systematic genetic screening of patients with amyotrophic lateral sclerosis. JNNP. 2021; 92: 510–518.

3. van Rheenen W, van der Spek RAA, Bakker MK, van Vugt JJFA, Hop PJ, Zwamborn RAJ, et al. Common and rare variant association analyses in Amyotrophic Lateral Sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. bioRxiv medRxiv. 2021.

4. Tazelaar GHP, Boeynaems S, De Decker M, van Vugt JJFA, Kool L, Goedee HS, et al. ATXN1 repeat expansions confer risk for amyotrophic lateral sclerosis and contribute to TDP-43 mislocalization. Brain Commun. 2020; May 19. 2 (2) fcaa064 [PubMed: 32954321]

5. Cooper-Knock J, Zhang S, Kenna KP, Moll T, Franklin J, Allen S, et al. Rare Variant Burden Analysis within Enhancers Identifies CAV1 as a New ALS Risk Gene. Cell Rep. 2020; Dec 1. 33 (9) 108456 [PubMed: 33264630]

6. Mohassel P, Donkervoort S, Lone MA, Nalls M, Gable K, Gupta SD, et al. Childhood amyotrophic lateral sclerosis caused by excess sphingolipid synthesis. Nat Med. 2021; May 31. doi: 10.1038/s41591-021-01346-1

7. Iacoangeli A, Lin T, Al Khleifat A, Jones AR, Opie-Martin S, Coleman JRI, et al. Genome-wide Meta-analysis Finds the ACSL5-ZDHHC6 Locus Is Associated with ALS and Links Weight Loss to the Disease Genetics. Cell Rep. 2020; Oct 27. 33 (4) 108323 [PubMed: 33113361]

8. Nakamura R, Misawa K, Tohnai G, Nakatochi M, Furuhashi S, Atsuta N, et al. A multi-ethnic meta-analysis identifies novel genes, including ACSL5, associated with amyotrophic lateral sclerosis. Commun Biol. 2020; Sep 23. 3 (1) 526. [PubMed: 32968195]

9. Farhan SMK, Howrigan DP, Abbott LE, Klim JR, Topp SD, Byrnes AE, et al. Exome sequencing in amyotrophic lateral sclerosis implicates a novel gene, DNAJC7, encoding a heat-shock protein. Nat Neurosci. 2019; Nov 25; 22 (12) 1966–74. [PubMed: 31768050]

10. Smith BN, Topp SD, Fallini C, Shibata H, Chen HJ, Troakes C, et al. Mutations in the vesicular trafficking protein annexin A11 are associated with amyotrophic lateral sclerosis. Sci Transl Med. 2017; 9 (388) [PubMed: 28469040]

11. Cooper-Knock J, Moll T, Ramesh T, Castelli L, Beer A, Robins H, et al. Mutations in the Glycosyltransferase Domain of GLT8D1 Are Associated with Familial Amyotrophic Lateral Sclerosis. Cell Rep. 2019; 26 (9) 2298–306. e5 [PubMed: 30811981]

12. Renton AE, Chiò A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. Nat Neurosci. 2014; Jan; 17 (1) 17–23. [PubMed: 24369373]

13. Ryan M, Heverin M, McLaughlin RL, Hardiman O. Lifetime Risk and Heritability of Amyotrophic Lateral Sclerosis. JAMA Neurol. 2019; Jul 22. doi: 10.1001/jamaneurol.2019.2044

14. Trabjerg BB, Garton FC, van Rheenen W, Fang F, Henderson RD, Mortensen PB, et al. ALS in Danish Registries: Heritability and links to psychiatric and cardiovascular disorders. Neurol Genet. 2020; Apr. 6 (2) e398 [PubMed: 32211514]

15. Li J, Li X, Zhang S, Snyder M. Gene-Environment Interaction in the Era of Precision Medicine. Cell. 2019; 177 (1) 38–44. [PubMed: 30901546]

16. Hivert V, Sidorenko J, Rohart F, Goddard ME, Yang J, Wray NR, et al. Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals. Am J Hum Genet. 2021; May 6. 108 (5) 962. [PubMed: 33961780]

17. Zhang S, Cooper-Knock J, Weimer AK, Shi M, Moll T, Harvey C, et al. Genome-wide Identification of the Genetic Basis of Amyotrophic Lateral Sclerosis [Internet]. bioRxiv. 2021. 2020.11.14.382606 cited 2021 May 28

18. Bakken T, Jorstad N, Hu Q, Lake B, Tian W, Kalmbach B, et al. Evolution of cellular diversity in primary motor cortex of human, marmoset monkey, and mouse [Internet]. BioRxiv. 2020. cited 2021 Jul 13

19. Pineda SS, Lee H, Fitzwalter BE, Mohammadi S. Single-cell profiling of the human primary motor cortex in ALS and FTLD. bioRxiv. 2021.

20. Bean DM, Al-Chalabi A, Dobson RJB, Iacoangeli A. A Knowledge-Based Machine Learning Approach to Gene Prioritisation in Amyotrophic Lateral Sclerosis. Genes. 2020; Jun 19. 11 (6) doi: 10.3390/genes11060668

21. Hu J, Lepore R, Dobson RJB, Al-Chalabi A, Bean DM, Iacoangeli A. DGLinker: flexible knowledge-graph prediction of disease-gene associations. Nucleic Acids Res. 2021; Jul 2. 49 (W1) W153–61 [PubMed: 34125897]

22. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science. 2006; 314 (5796) 130–3. [PubMed: 17023659]

23. Green EM, Seeley WW, Petrucelli L, Gitler AD. TDP-43 represses cryptic exon inclusion in FTD/ALS gene UNC13A. bioRxiv. 2021.

24. Ule J, Buratti E, Humphrey J, Ward ME, Fratta P. Common ALS/FTD risk variants in UNC13A exacerbate its cryptic splicing and loss upon TDP-43 mislocalization. bioRxiv. 2021.

25. Diekstra FP, van Vught PWJ, van Rheenen W, Koppers M, Pasterkamp RJ, van Es MA, et al. UNC13A is a modifier of survival in amyotrophic lateral sclerosis. Neurobiol Aging. 2012; Mar. 33 (3) 630. e3–8

26. Stephenson J, Amor S. Modelling amyotrophic lateral sclerosis in mice. Drug Discov Today Dis Models. 2017; Dec 1. 25–26: 35–44.

27. Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, Chia R, et al. Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. Neuron. 2018; Mar 21; 97 (6) 1268–83. e6 [PubMed: 29566793]

28. van Rheenen W, Shatunov A, Dekker AM, McLaughlin RL, Diekstra FP, Pulit SL, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nat Genet. 2016; 48 (9) 1043–8. [PubMed: 27455348]

29. Halperin E, Stephan DA. SNP imputation in association studies. Nat Biotechnol. 2009; Apr; 27 (4) 349–51. [PubMed: 19352374]

30. Castel SE, Cervera A, Mohammadi P, Aguet F, Reverter F, Wolman A, et al. Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. Nat Genet. 2018; Sep; 50 (9) 1327–34. [PubMed: 30127527]

31. Kvon EZ, Waymack R, Gad M, Wunderlich Z. Enhancer redundancy in development and disease. Nat Rev Genet. 2021; May; 22 (5) 324–36. [PubMed: 33442000]

32. McLaughlin RL, Schijven D, van Rheenen W, van Eijk KR, O'Brien M, Kahn RS, et al. Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. Nat Commun. 2017; Mar 21. 8 14774 [PubMed: 28322246]

33. Kim HJ, Kim NC, Wang Y-D, Scarborough EA, Moore J, Diaz Z, et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature. 2013; Mar 28; 495 (7442) 467–73. [PubMed: 23455423]

34. Johnson JO, Mandrioli J, Benatar M, Abramzon Y, Van Deerlin VM, Trojanowski JQ, et al. Exome sequencing reveals VCP mutations as a cause of familial ALS. Neuron. 2010; Dec 9; 68 (5) 857–64. [PubMed: 21145000]

35. Han SM, Cottee PA, Miller MA. Sperm and oocyte communication mechanisms controlling C. elegans fertility. Dev Dyn. 2010; May; 239 (5) 1265–81. [PubMed: 20034089]

36. Benita Y, Oosting RS, Lok MC, Wise MJ, Humphery-Smith I. Regionalized GC content of template DNA as a predictor of PCR success. Nucleic Acids Res. 2003; Aug 15. 31 (16) e99 [PubMed: 12907751]

37. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. 2011; Oct 20; 72 (2) 245–56. [PubMed: 21944778]

38. Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron. 2011; Oct 20; 72 (2) 257–68. [PubMed: 21944779]

39. Elden AC, Kim H-J, Hart MP, Chen-Plotkin AS, Johnson BS, Fang X, et al. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature. 2010; Aug 26; 466 (7310) 1069–75. [PubMed: 20740007]

40. Course MM, Gudsnuk K, Smukowski SN, Winston K, Desai N, Ross JP, et al. Evolution of a Human-Specific Tandem Repeat Associated with ALS. Am J Hum Genet. 2020; Sep 3; 107 (3) 445–60. [PubMed: 32750315]

41. Mousavi N, Shleizer-Burko S, Yanicky R, Gymrek M. Profiling the genome-wide landscape of tandem repeat expansions. Nucleic Acids Res. 2019; Sep 5. 47 (15) e90 [PubMed: 31194863]

42. Dolzhenko E, Bennett MF, Richmond PA, Trost B, Chen S, van Vugt JJFA, et al. ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data [Internet]. Genome Biology. 2020; 21 doi: 10.1186/s13059-020-02017-z

43. Kenna KP, van Doormaal PT, Dekker AM, Ticozzi N, Kenna BJ, Diekstra FP, et al. NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. Nat Genet. 2016; 48 (9) 1037–42. [PubMed: 27455347]

44. Miller T, Cudkowicz M, Shaw PJ, Andersen PM, Atassi N, Bucelli RC, et al. Phase 1–2 Trial of Antisense Oligonucleotide Tofersen for SOD1 ALS. N Engl J Med. 2020; Jul 9; 383 (2) 109–19. [PubMed: 32640130]

45. Consortium TIS, The International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder [Internet]. Nature. 2009; 460: 748–52. DOI: 10.1038/nature08185 [PubMed: 19571811]

46. Zheutlin AB, Ross DA. Polygenic Risk Scores: What Are They Good For? Biol Psychiatry. 2018; Jun 1. 83 (11) e51–3 [PubMed: 29759133]

47. Restuadi R, Garton FC, Benyamin B, Lin T, Williams KL, Vinkhuyzen A, et al. Polygenic risk score analysis for amyotrophic lateral sclerosis leveraging cognitive performance, educational attainment and schizophrenia. Eur J Hum Genet. 2021; Apr 27. doi: 10.1038/s41431-021-00885-y

48. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015; 521: 436–44. doi: 10.1038/ nature14539 [PubMed: 26017442]

49. Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. Mol Syst Biol. 2016; Jul 29. 12 (7) 878. [PubMed: 27474269]

50. Chiò A, Mazzini L, D'Alfonso S, Corrado L, Canosa A, Moglia C, et al. The multistep hypothesis of ALS revisited. 91 Neurology. 2018; e635–42 doi: 10.1212/wnl.0000000000005996

51. Al-Chalabi A, Calvo A, Chio A, Colville S, Ellis CM, Hardiman O, et al. Analysis of amyotrophic lateral sclerosis as a multistep process: a population-based modelling study. Lancet Neurol. 2014; 13 (11) 1108–13. [PubMed: 25300936]

52. Julian TH, Glascow N, Dylan Fisher Barry A, Moll T, Harvey C, Klimentidis YC, et al. Physical exercise is a risk factor for amyotrophic lateral sclerosis: Convergent evidence from mendelian randomisation, transcriptomics and risk genotypes. EBioMed. 2021; 68 103397 doi: 10.1101/2020.11.24.20238063

53. Bandres-Ciga S, Noyce AJ, Hemani G, Nicolas A, Calvo A, Mora G, et al. Shared polygenic risk and causal inferences in amyotrophic lateral sclerosis. Ann Neurol. 2019; 85 (4) 470–81. [PubMed: 30723964]

54. Spiller W, Slichter D, Bowden J, Davey Smith G. Detecting and correcting for bias in Mendelian randomization analyses using Gene-by-Environment interactions. Int J Epidemiol. 2019; Jun 1; 48 (3) 702–12. [PubMed: 30462199]

55. Boddy SL, Giovannelli I, Sassani M, Cooper-Knock J, Snyder MP, Segal E, et al. The gut microbiome: a key player in the complexity of amyotrophic lateral sclerosis (ALS). BMC Med. 2021; Jan 20. 19 (1) 13. [PubMed: 33468103]

56. de Klein N, Tsai EA, Vochteloo M, Baird D, Huang Y, Chen C-Y, et al. Brain expression quantitative trait locus and network analysis reveals downstream effects and putative drivers for brain-related diseases [Internet]. bioRxiv. 2021. 2021.03.01.433439 cited 2021 Jul 12

57. Amaya C, Fader CM, Colombo MI. Autophagy and proteins involved in vesicular trafficking. FEBS Lett. 2015; Nov 14; 589 (22) 3343–53. [PubMed: 26450776]

58. Ghasemi M, Brown RH Jr. Genetics of Amyotrophic Lateral Sclerosis. Cold Spring Harb Perspect Med. 2018; May 1. 8 (5) doi: 10.1101/cshperspect.a024125

59. Hergesheimer RC, Chami AA, de Assis DR, Vourc'h P, Andres CR, Corcia P, et al. The debated toxic role of aggregated TDP-43 in amyotrophic lateral sclerosis: a resolution in sight? Brain. 2019; May 1; 142 (5) 1176–94. [PubMed: 30938443]

60. Ling JP, Pletnikova O, Troncoso JC, Wong PC. TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. Science. 2015; Aug 7; 349 (6248) 650–5. [PubMed: 26250685]

61. van Es MA, Veldink JH, Saris CGJ, Blauw HM, van Vught PWJ, Birve A, et al. Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. Nat Genet. 2009; Oct; 41 (10) 1083–7. [PubMed: 19734901]

62. Chiò A, Mora G, Restagno G, Brunetti M, Ossola I, Barberis M, et al. UNC13A influences survival in Italian amyotrophic lateral sclerosis patients: a population-based study. Neurobiol Aging. 2013; Jan. 34 (1) 357. e1–5

63. Melamed Z, López-Erauskin J, Baughn MW, Zhang O, Drenner K, Sun Y, et al. Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. Nat Neurosci. 2019; 22 (2) 180–90. [PubMed: 30643298]

64. Klim JR, Williams LA, Limone F, Juan IGS, Davis-Dusenbery BN, Mordes DA, et al. ALS-implicated protein TDP-43 sustains levels of STMN2, a mediator of motor neuron growth and repair [Internet]. Nature Neuroscience. 2019; 22: 167–79. doi: 10.1038/s41593-018-0300-4 [PubMed: 30643292]

65. Klim JR, Pintacuda G, Nash LA, Guerra San Juan I, Eggan K. Connecting TDP-43 Pathology with Neuropathy. Trends Neurosci. 2021; Jun; 44 (6) 424–40. [PubMed: 33832769]

66. Theunissen F, Anderton RS, Mastaglia FL, Flynn LL, Winter SJ, James I, et al. Novel STMN2 Variant Linked to Amyotrophic Lateral Sclerosis Risk and Clinical Phenotype. Front Aging Neurosci. 2021; Mar 26. 13 658226 [PubMed: 33841129]

67. 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; Oct; 26 (10) 1537–46. [PubMed: 29955173]

68. Hanby MF, Scott KM, Scotton W, Wijesekera L, Mole T, Ellis CE, et al. The risk to relatives of patients with sporadic amyotrophic lateral sclerosis. Brain. 2011; Dec; 134 (Pt 12) 3454–7. [PubMed: 21933809]

69. Morello G, Guarnaccia M, Spampinato AG, La Cognata V, D'Agata V, Cavallaro S. Copy Number Variations in Amyotrophic Lateral Sclerosis: Piecing the Mosaic Tiles Together through a Systems Biology Approach. Mol Neurobiol. 2018; Feb; 55 (2) 1299–322. [PubMed: 28120152]

70. Liu EY, Russ J, Cali CP, Phan JM, Amlie-Wolf A, Lee EB. Loss of Nuclear TDP-43 Is Associated with Decondensation of LINE Retrotransposons. Cell Rep. 2019; Apr 30; 27 (5) 1409–21. e6 [PubMed: 31042469]

71. Loomis EW, Eid JS, Peluso P, Yin J, Hickey L, Rank D, et al. Sequencing the unsequenceable: Expanded CGG-repeat alleles of the fragile X gene [Internet]. Genome Research. 2013; 23: 121–8. doi: 10.1101/gr.141705.112 [PubMed: 23064752]

72. Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS. Long reads: their purpose and place. Hum Mol Genet. 2018; Aug 1. 27 (R2) R234–41 [PubMed: 29767702]

73. Logsdon GA, Vollger MR, Hsieh P, Mao Y, Liskovykh MA, Koren S, et al. The structure, function and evolution of a complete human chromosome 8. Nature. 2021; May; 593 (7857) 101–7. [PubMed: 33828295]

74. Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. Science. 2015; Oct 2; 350 (6256) 94–8. [PubMed: 26430121]

75. Al-Chalabi A, Hardiman O. The epidemiology of ALS: a conspiracy of genes, environment and time. Nat Rev Neurol. 2013; Nov; 9 (11) 617–28. [PubMed: 24126629]

76. Buchman VL, Cooper-Knock J, Connor-Robson N, Higginbottom A, Kirby J, Razinskaya OD, et al. Simultaneous and independent detection of C9ORF72 alleles with low and high number of GGGGCC repeats using an optimised protocol of Southern blot hybridisation. Mol Neurodegener. 2013; 8: 12. [PubMed: 23566336]

77. Shi W, Ng CKY, Lim RS, Jiang T, Kumar S, Li X, et al. Reliability of Whole-Exome Sequencing for Assessing Intratumor Genetic Heterogeneity. Cell Rep. 2018; Nov 6; 25 (6) 1446–57. [PubMed: 30404001]

78. Spence JM, Spence JP, Abumoussa A, Burack WR. Ultradeep analysis of tumor heterogeneity in regions of somatic hypermutation. Genome Med. 2015; Mar 12. 7 (1) 24. [PubMed: 25874000]

79. Kim J, Kim D, Lim JS, Maeng JH, Son H, Kang H-C, et al. The use of technical replication for detection of low-level somatic mutations in next-generation sequencing. Nat Commun. 2019; Mar 5. 10 (1) 1047. [PubMed: 30837471]

80. Project MinE ALS Sequencing Consortium. CHCHD10 variants in amyotrophic lateral sclerosis: Where is the evidence? Ann Neurol. 2018; Jul; 84 (1) 110–6. [PubMed: 30014597]

81. Fujimori K, Ishikawa M, Otomo A, Atsuta N, Nakamura R, Akiyama T, et al. Modeling sporadic ALS in iPSC-derived motor neurons identifies a potential therapeutic agent. Nat Med. 2018; Oct; 24 (10) 1579–89. [PubMed: 30127392]

82. Sun X, Song J, Huang H, Chen H, Qian K. Correction to: Modeling hallmark pathology using motor neurons derived from the family and sporadic amyotrophic lateral sclerosis patient-specific iPS cells. Stem Cell Res Ther. 2019; Mar 15. 10 (1) 97. [PubMed: 30876443]

83. Meyer K, Ferraiuolo L, Miranda CJ, Likhite S, McElroy S, Renusch S, et al. Direct conversion of patient fibroblasts demonstrates non-cell autonomous toxicity of astrocytes to motor neurons in familial and sporadic ALS. Proc Natl Acad Sci U S A. 2014; Jan 14; 111 (2) 829–32. [PubMed: 24379375]

Ket Points

- Rapid progress in ALS genetics has led to the discovery of new loci and new genes but does not completely explain the observed heritability.

- The focus of research has moved away from fully penetrant monogenic variants towards risk variants with incomplete penetrance.

- Specific challenges and opportunities for gene discovery include genetic pleiotropy, gene-environment interactions, structural variation and nonlinear genetic interactions.

- A comprehensive genetic architecture for ALS could facilitate disease subclassification and lead to personalised medicine.
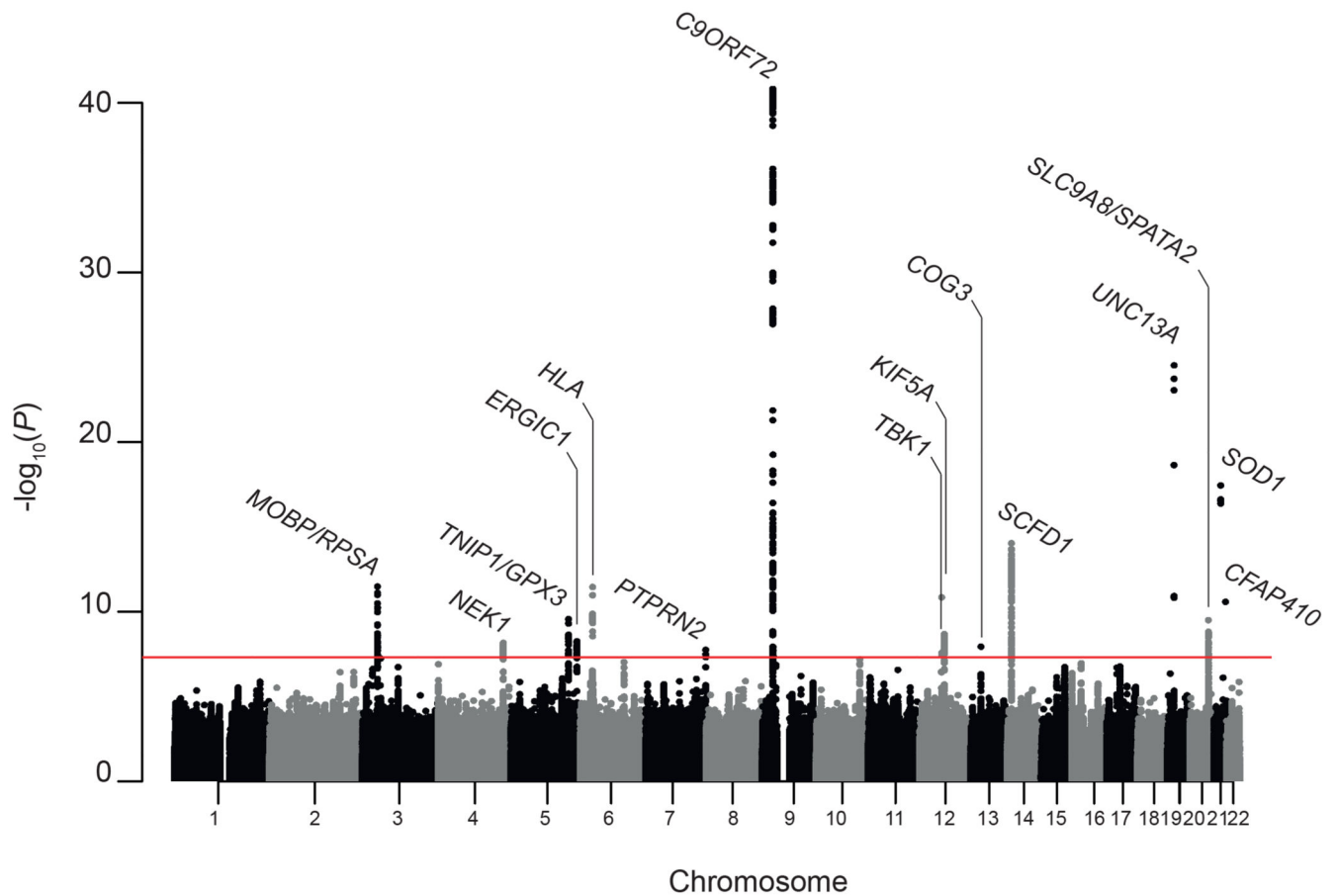
Figure 1. Manhattan plot for new ALS genome-wide association study (GWAS) including cross-ancestry meta-analysis.

Red line represents the p-value threshold for genome-wide significance based on Bonferroni multiple testing correction (p=5e-08). Genes linked to loci by prioritization analysis are labelled.
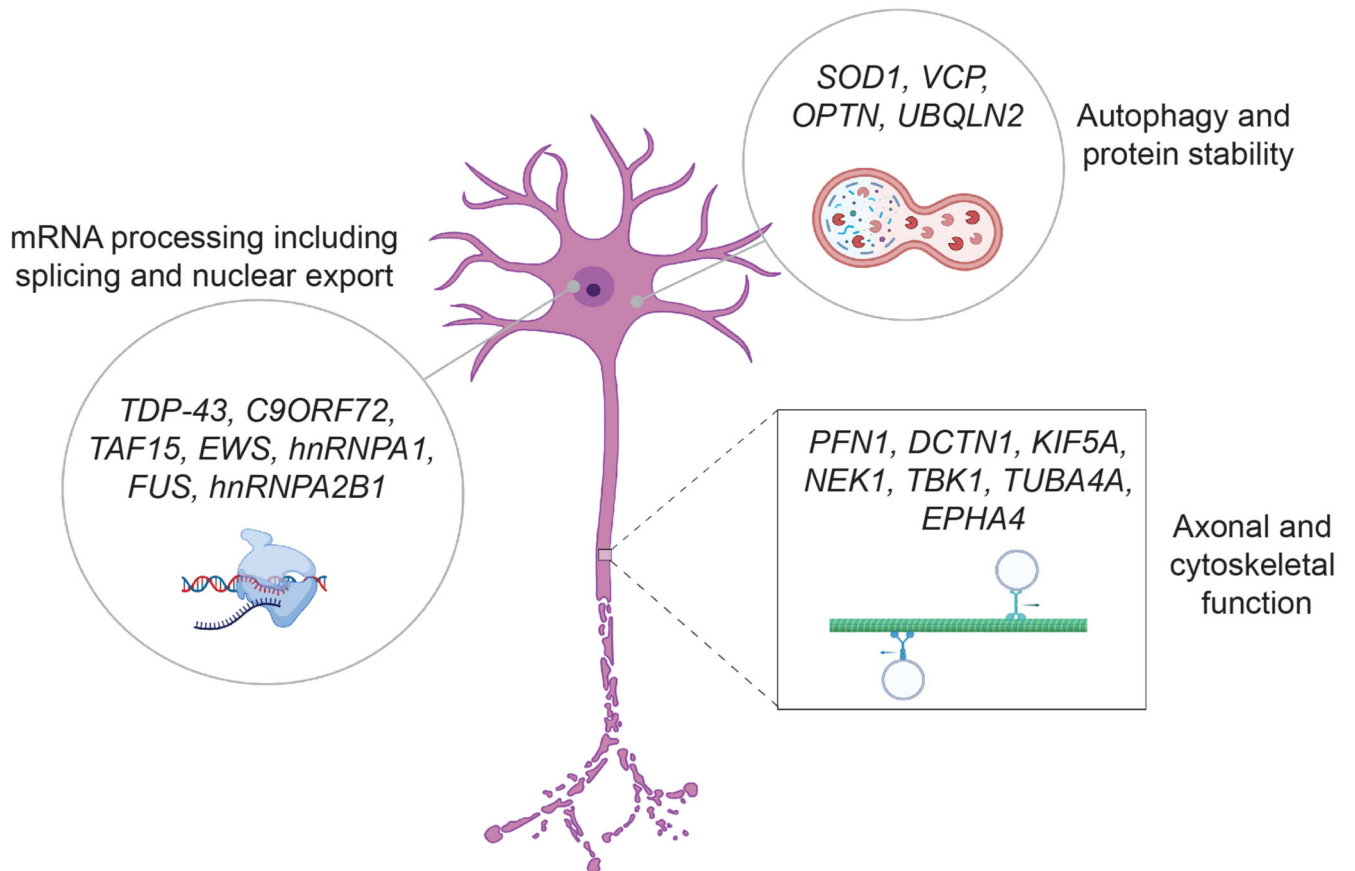
Figure 2. Existing ALS risk genes converge within biological pathways including mRNA processing, autophagy and axonal function.
Figure indicates example risk genes and relevant subcellular localisation.