

OPINION

Open Access



Short structural variants as informative genetic markers for ALS disease risk and progression

Frances Theunissen^{1,2*} , Loren L. Flynn^{1,2,3,4}, Ryan S. Anderton^{2,5} and P. Anthony Akkari^{1,2,3,4,6}

Abstract

There is considerable variability in disease progression for patients with amyotrophic lateral sclerosis (ALS) including the age of disease onset, site of disease onset, and survival time. There is growing evidence that short structural variations (SSVs) residing in frequently overlooked genomic regions can contribute to complex disease mechanisms and can explain, in part, the phenotypic variability in ALS patients. Here, we discuss SSVs recently characterized by our laboratory and how these discoveries integrate into the current literature on ALS, particularly in the context of application to future clinical trials. These markers may help to identify and differentiate patients for clinical trials that have a similar ALS disease mechanism(s), thereby reducing the impact of participant heterogeneity. As evidence accumulates for the genetic markers discovered in *SQSTM1*, *SCAF4*, and *STMN2*, we hope to improve the outcomes of future ALS clinical trials.

Keywords: Genetic marker, Structural variant, Amyotrophic lateral sclerosis, Clinical trials, Participant selection, Enrichment tool, Responder sub-population

Background

Many key cellular processes are known to be disrupted in amyotrophic lateral sclerosis (ALS) such as RNA metabolism, mitochondrial dysfunction, oxidative stress, protein aggregation, impaired axonal transport, and cytoskeletal dysfunction, reviewed in [1]. Variation in the expression of genes involved in these processes may increase disease risk and/or influence the rate of disease progression [2–4]. At present, there is a lack of genetic markers for the different ALS disease subtypes, as well as a lack of genetic indicators of disease risk and/or trajectory. The heterogeneous clinical presentation and diverse rates of progression makes identifying ALS patients with similar disease mechanisms extremely

challenging, and undoubtedly contributes to clinical trial failures [5].

Establishing molecular targets and genetic markers for ALS can lead to improved patient stratification for clinical trials and might enable positive treatments to be identified for specific patient subgroups. Van Eijk and Eijkemans [6] recently demonstrated that genotypic data for *unc-13* homolog a (*UNC13A*), myelin-associated oligodendrocyte basic protein (*MOBP*), and the repeat expansion in *c9orf72*-*SMCR8* complex subunit (*C9orf72*) could influence both primary and secondary outcomes including survival, ALS functional rating scale (ALFSRS) and forced vital capacity (FVC) measures. Additionally, a retrospective meta-analysis of three lithium carbonate clinical trials revealed that contrary to the reported negative outcomes, patients with the *UNC13A* (C/C) genotype had actually responded to the lithium carbonate [7]. This study provides evidence that genetic markers can inform clinical trial outcomes and should

* Correspondence: Frances.theunissen@perron.uwa.edu.au

¹Perron Institute for Neurological and Translational Science, First floor, RR block, QEII Medical Centre, 8 Verdun St, Nedlands, WA 6009, Australia

²Centre for Neuromuscular and Neurological Disorders, University of Western Australia, Nedlands, WA, Australia

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

be incorporated into clinical trial design. Intuitively, genetic regions that are highly variable, known as structural variants (SVs), will likely be more informative as genetic markers than single nucleotide polymorphisms (SNP), due to a larger number of potential genotypes [8]. These variable regions of the genome have been predominantly under-characterized [9, 10]; however, the scientific community is beginning to appreciate the need to investigate polymorphic loci as potential disease modifying regions.

Structural variants (SVs) have typically been defined as insertions, deletions, inversions, and microsatellites that can be repeated hundreds of times within the genome that are greater than 1 kb in length [11]. Short structural variants (SSVs) have been predominately overlooked in the context of ALS and encompass that same class of variants as the classical definition (e.g., short tandem repeat, microsatellite, insertion/deletion, inversion, polynucleotide repeat) but are much shorter in size, typically < 50 bp in length [12]. Approximately 4 million SSVs exist within the human genome and have been previously described [13, 14]. It is possible that some of the “missing heritability” in ALS could be explained by more common SSVs with small effect sizes that have not yet been identified [13, 15]. Importantly, changes in the size and composition of both SV/SSVs can have a significant impact on the binding of regulatory elements that modulate RNA processing and gene expression [16]. SSVs have been implicated in many complex diseases, including ALS and other neurodegenerative diseases [17], such as Parkinson’s and Alzheimer’s disease [18–20]. The ability of SSVs to alter gene expression is dependent on their location within and around the gene or intergenic region, with their effects occurring via several mechanisms including the following: influencing the binding of regulatory elements, mRNA splicing and processing, genome folding and higher order structure, and translation. This may differentiate mechanisms of disease pathogenesis, including risk of disease, risk for a specific phenotype, symptom presentation, disease course, and response to treatment, between individuals [21].

Recent in silico mapping of known ALS-linked genes has predicted a number of as yet unresolved short tandem repeats within each of these genes, that are likely polymorphic, and could influence gene expression and contribute to disease risk for ALS [15, 22]. Importantly, in the most recent ALS genome wide association study (GWAS) conducted by Van Rheenen and Van der Spek [23], 15 risk loci were identified, with 8 loci previously reported in GWAS studies [24–26]. This GWAS study was unique; in addition to screening pathogenic rare burden variants, it also incorporated short tandem repeats, RNA-sequencing, and methylation datasets to prioritize causal genes within identified ALS risk loci.

Of note, a polymorphic tetra-nucleotide repeat downstream of ALS-linked gene NIMA-related kinase 1 (*NEK1*) was reported to be associated with increased ALS disease risk. The *NEK1* SSV was in linkage disequilibrium with the top hit *NEK1* SNP reported in this study and may help to further explain phenotypic variability and disease penetrance between patients carrying different *NEK1* mutations, particularly since the SNP alone could not reliably determine its contribution to ALS risk [23]. Furthermore, another recent study has investigated SVs in 6580 whole genome sequences (4315 ALS and 1880 controls) from the Project MinE cohort to determine genotype-phenotype correlations in known ALS genes. Al Khleifat, Iacoangeli [27] reported that structural variants in *C9orf72* (repeat expansion), Valosin containing protein *VCP* (inversion) and erb-b2 receptor tyrosine kinase 4, *ERBB4* (deletion) are variously associated with ALS disease risk and phenotype. The variant caller used in this study detected multiple classes of SVs, however, small inversions and insertion/deletions < 200 bp could not be analyzed by the Manta platform [27]. Further investigation into variants outside of the known ALS genes in addition to wet lab validation is required to gain a holistic understanding of the contribution of genetic variation to ALS risk and phenotype [27].

In addition to contributing to disease risk, SSVs can also be an informative tool for clinical trial participant selection, as demonstrated in Alzheimer’s disease, in the case of translocase of outer mitochondrial membrane 40 (*TOMM40*) [28]. Although the age-of-onset distributions for Alzheimer’s disease have been well-established since 1993 [29], the basis of this distribution was only partially explained by the apolipoprotein E (*APOE*) genotype, suggesting that other genetic factors must contribute to age-of-onset for the disease [30]. In fact, the combination of the *APOE* genotype alongside the SSV genotype, a poly-T repeat in *TOMM40*, was subsequently shown to account for > 98% of the clinical age-of-onset distributions in the Caucasian population [20, 28]. Combining these genotypes allowed the generation of the clinical age-of-onset risk algorithm [31] to be developed to inform participant selection for the TOMMORROW clinical trial [32], with selection based on the genotype of individuals and their corresponding risk of disease/age-of-onset prediction. Evidently, this finding demonstrates that informative SSVs can be used as an enrichment tool for clinical trials, thus informing participant selection.

Innovative approach for SSV discovery

In ALS, the clinical presentation can manifest differently in family members sharing the same ALS-linked variant (incomplete penetrance), suggesting multiple factors, including both genetic and non-genetic factors, can

contribute to the progression of the disease [15, 21, 33]. With this in mind, we have used an established structural variant evaluation system (dbSSV) [22] to identify candidate SSVs within and around confirmed ALS loci [15]. Using a systematic approach to candidate GWAS loci, analysis of SSVs in these genomic regions will likely identify common variants that subtly influence gene function, in some cases leading to ALS. The dbSSV software [22] focuses specifically on identifying SSVs and scores them against 24 different properties, including data describing the location and definition of the variation, variability indicators, repeat context, gene context, transcription factor and microRNA binding sites, other regulatory markers, conservation, position within a linkage disequilibrium block, GWAS signals, and tissue-specific regulatory signals. Based on these scores, a short list of SSVs is generated with each total score suggesting the likelihood of the variant having significant biological effects and contributing to disease risk. Using this method, we have now identified and published several novel genetic markers, discussed below. Investigating SSVs in ALS-linked genes will help to better understand differences between individual patient phenotypes and disease progression.

Insertion/deletion in *SQSTM1* is associated with disease in familial *SOD1* patients

Sequestosome 1 (*SQSTM1*) encodes P62, an adaptor protein that binds ubiquitylated protein aggregates and delivers them to the autophagosome for degradation. With an essential role in protein clearance, it is not surprising that P62 dysfunction is implicated in neurodegenerative diseases that are governed by abnormal protein inclusions. Thus, SSVs within *SQSTM1* may contribute to the diverse presentations observed between ALS patients by influencing P62 expression and autophagic clearance. In our recently published association study of a North American cohort of familial ALS (fALS) and sporadic (sALS) patients ($n = 403$) and age matched controls ($n = 562$), a small cytosine/adenine (CAAA) insertion deletion (I/D) was associated with fALS, particularly in familial superoxide dismutase 1 (*SOD1*) mutation positive patients, but not with sALS patients [34]. Furthermore, the presence of the insertion variant appears to translate to a stepwise decrease in *SQSTM1* expression in healthy olfactory neurosphere-derived cells, with the I/I genotype resulting in a 2.5-fold reduction in transcript levels [34]. The observations of the insertion/deletion influencing *SQSTM1* expression has previously been reported by an independent study in a screen of 17 different healthy tissue types [35]. By this weight of evidence, the CAAA variant needs to be further examined as a contributor to ALS disease mechanisms, particularly since *SQSTM1* plays a critical role in

autophagy, and mutations in this gene can be a direct cause of ALS and other neurodegenerative diseases [36, 37]. Future therapeutic studies to modulate autophagy should take into consideration the potential impact of this SSV on P62 expression.

Poly-T repeat in *SCAF4* is a genetic marker for disease risk and survival in familial ALS

Studies of common *SOD1*-ALS mutations have suggested that a disease modifying factor located nearby *SOD1* may influence the penetrance of *SOD1* p.D91A (D90A) and p.A5V (AV4) mutations (historical names noted in brackets are not reflective of the amino acid position), thus contributing broadly to ALS risk [38, 39]. Our laboratory reported an 11-18 thymine repeat, located within the 3'UTR of the neighboring gene, SR-related CTD associated factor 4 (*SCAF4*) [40]. The function of *SCAF4* has recently been elucidated, with an essential role in RNA processing through regulating transcript elongation [41]. Variants in *SCAF4* have been reported to cause impaired RNA processing and can result in neurodevelopmental disorders [42]. Interestingly, the poly-T variant in *SCAF4* is flanked by two binding sites for transcription factor RNA polymerase II subunit A (POLR2A); therefore, changes in the length of this variant may influence the binding of POLR2A, thus affecting transcription of nearby genes (*SOD1*) and may therefore play a role in ALS related neurodegeneration. A case control association study in fALS patients ($n = 190$) and healthy age matched controls ($n = 560$) revealed that the 18 T repeat is associated with ALS risk for the entire cohort, including those without *SOD1*-linked mutations ($n = 27$). The 18 T allele was also associated with a 26 month reduction in survival time but was not associated with age at disease onset [40]. Future studies should investigate the functional impact of this variant and determine whether thymine repeat length influences transcript elongation of genes that are regulated by *SCAF4*. Additionally, this SSV may also help to identify patients that could benefit from a *SOD1* targeted therapy, such as the *SOD1* suppressing antisense therapy Tofersen (currently in phase III extension study) [43].

Short tandem repeat in *STMN2* is associated with sporadic ALS disease risk and clinical phenotype

The microtubule regulator stathmin-2 (*STMN2*) was recently identified as a gene with strong therapeutic potential for ALS [44, 45]. *STMN2* is involved in axonal growth and repair and is directly regulated by the ALS-linked gene TAR DNA-binding protein 43 (*TARDBP*). When *TARDBP* protein TDP-43 is depleted or mislocalized, as occurs in ALS, *STMN2* protein is also depleted [44, 45]. To understand the implications of this TDP-43:STMN2 relationship and ALS phenotype, it is

essential to characterize the natural variation in *STMN2* expression.

We recently examined the *STMN2* gene for candidate SSVs that may affect the binding of regulatory elements and, therefore, influence gene expression. We identified a variable length cytosine/adenine (CA) short tandem repeat within *STMN2* in a cohort of Caucasian sALS patients ($n = 321$) and healthy age matched controls ($n = 332$). We have identified the first genetic association with *STMN2* and both sALS disease risk and age-at-disease onset [46]. In 143 patients where end-point survival data was available, when categorized according to site of disease onset (bulbar vs spinal), the bulbar cases carrying the risk genotype had significantly shorter survival times than other bulbar cases. Moreover, this effect on survival was not abrogated when controlling for sex or age. Furthermore, in an Australian longitudinal sALS cohort ($n = 67$), ALSFRS scores were significantly lower in patients carrying the risk genotype. Following these clinical associations, stathmin-2 mRNA expression was shown to be reduced in sALS patient olfactory neurosphere-derived cells. When accounting for CA genotype, a trend for reduced expression of stathmin-2 mRNA was also observed in sALS patients and control laser-captured spinal motor neurons [46].

This work points to a novel mechanism by which this SSV may regulate *STMN2* gene expression and could further explain the recently elucidated *STMN2* cryptic exon mechanism [44, 45] and its influence on neurodegenerative disease phenotype [47, 48]. The genetic validation of *STMN2* significantly adds to the weight of evidence that this gene is important in ALS [46]. The *STMN2* genetic marker may therefore be a useful tool for cohort selection in clinical trials or to stratify patient response. Further, this discovery has broad implications for clinical assessment and therapeutic development and should be incorporated in future clinical trials targeting this gene.

Application of SSVs to clinical trials

Recently, the Treatment Research Initiative to Cure ALS (TRICALS) has brought the urgent need to reform ALS clinical trial design to the forefront of the literature [5, 49]. In particular, the major concerns raised are related to the stringent patient selection criteria and analytical strategy of phase 3 clinical trials. Van Eijk and Nikolakopoulos [49] highlight the use of patient risk profiles as a strategy to provide more informative selection criteria that will help to improved randomization, enable risk-based subgroup analyses, and increase the statistical power of clinical trials. Patient risk profiles are based on a multivariate analysis of several patient characteristics (i.e., age of onset, site of onset, vital capacity, diagnostic delay, ALSFRS etc.), creating a “prognostic summary” for

each patient [49]. This is likely to help increase the number of eligible patients for trials, while reducing patient drop out and exclusion rates, thus increasing the generalizability of the trial [49]. The authors noted that they did not include *C9orf72* repeat expansion as a factor in their predictive model, as it did not significantly improve the predictive power of the model and may pose additional logistical constraints (i.e., genotyping and counseling patients) prior to clinical trial randomization [49]. However, one must consider that although the *C9orf72* repeat expansion is applicable to 40% of fALS patients and 6% sALS patients [50], incomplete penetrance is likely to be the complicating factor that influences the predictive capacity of this variant. Incorporating genetic variants into the prognostic summary may not improve prognostic predictions, since the clinical characteristics that the variant may be associated with are already individually incorporated into the model. The true value of patient genotype information is in balancing the clinical trial treatment arms, allowing sub-group analyses for patients that may have similar underlying disease mechanisms or ALS risk factors. It is likely that the eligibility window for patient risk profiles will need evaluation on a study-by-study basis, particularly since this will be influenced by the type of trial, i.e., broad vs genotype targeted treatment [49]. However, the use of patient genotype information will be pertinent to maximizing the information gained from a prospective clinical trial on potential responders. Across the most frequently studied indications, it has been shown that genetically validated targets are twice as likely to succeed in clinical development, transitioning from phase 1 to approval [51]. Therefore, the use of association studies to identify ALS genetic markers will help to guide selection of appropriate drug targets for sub-populations and help lower the rate of failure in clinical development programs.

Tofersen is currently the most promising disease modifying therapeutic for ALS and recently completed phase III (NCT02623699) with its long-term extension study currently postponed (NCT03070119) [43]. Tofersen was initially designed as a genetically targeted therapeutic for *SOD1* mutation positive patients [43]; however, evidence is accumulating that this therapy may also benefit other forms of sALS, where *SOD1* misfolding and aggregation is also a pathological feature [52, 53]. The use of genetic information in identifying the target population (mutation positive patients) in its initial early trial stage was integral to the phase I/II trial success [43]. Incorporation of genetic information in early-stage clinical development allowed initial efficacy to be shown and can be built upon in the hope of treating a broader patient population. Genetic variants that are associated with *SOD1* aggregation within sALS

cohorts will be necessary to identify those sALS patients who are likely to respond to a SOD1 targeting therapy. With the evolution of personalized treatment approaches using antisense oligonucleotides for neurodegenerative disorders [54], there is an absolute need to incorporate informative SSV genetic markers into early stage drug development. This will help to identify patients that have similar ALS disease mechanisms that would benefit from a targeted antisense therapy.

Conclusion

Currently, there is no effective treatment for ALS, and over the past 20 years, more than 60 controlled trials of putative ALS therapeutics have failed to demonstrate clinical efficacy [55, 56]. Current treatment options are non-specific and only extend survival by ~3 months in some patients, and furthermore, at present, there is no method to determine which patients are more likely to respond to a particular ALS therapeutic.

The continued discovery and evaluation of novel SSVs will undoubtedly shed light on the pathogenic mechanisms of ALS neurodegeneration. At present, there are few biomarkers/genetic markers that allow patient stratification according to disease mechanism [57–60] and treatment efficacy can only be evaluated by clinical measures during current clinical trials [34, 38]. With increasing evidence from our laboratory that SSVs do contribute to ALS risk and have disease-modifying effects [15, 34, 40, 46], investigations need to incorporate SSVs into genetic studies and clinical trial design [5]. There is an urgent need to establish well-characterized genetic markers that can be used to inform on the validity of certain treatment approaches. As ALS is a complex and heterogeneous disorder, with a varied clinical phenotype and disease trajectory, personalized medicine approaches will be more likely to result in successful treatments because they can directly target the underlying disease mechanism [54]. Therefore, it is crucial to be able to identify patient subgroups and develop compounds that are more likely to be effective in genetically defined subgroups of patients, thus reducing the impact of participant heterogeneity.

The authors recognize the challenges we face in current ALS clinical trials. Moving forward, we must examine the potential of these SSVs as a tool for patient stratification in retrospective clinical trials cohorts. This will result in the accelerated development of these genetic markers, fast tracking them into current and future clinical trials. Undoubtedly, as the data accumulates for these genetic markers, we are hopeful this will translate into identifying responder populations of ALS patients, allowing drug development to continue for specific subsets of patients. This is likely to significantly change the way clinical trials are conducted in ALS moving forward.

Abbreviations

ALS: Amyotrophic lateral sclerosis; SVs: Structural variants; SSVs: Short structural variants; *UNC13A*: Unc-13 homolog a; *MOBP*: Myelin-associated oligodendrocyte basic protein; *C9orf72*: C9orf72-SMCR8 complex subunit; ALFSRS: ALS functional rating scale; FVC: Forced vital capacity; SNP: Single nucleotide polymorphism; GWAS: Genome wide association study; *NEK1*: NIMA-related kinase 1; *TOMM40*: Translocase of outer mitochondrial membrane 40; *APOE*: Apolipoprotein E; dbSSV: Structural variant evaluation system; *SQSTM1*: Sequestosome 1; fALS: Familial ALS; sALS: Sporadic ALS; *SOD1*: Superoxide dismutase 1; I/D: Insertion deletion; *SCAF4*: SR-related CTD associated factor 4; *POLR2A*: RNA polymerase II subunit A; *STMN2*: Stathmin-2; *TARDBP*: TAR DNA-binding protein 43; TRICALS: Treatment Research Initiative to Cure ALS

Acknowledgements

FT and PAA would like to thank Professor Frank Mastaglia, Professor Ammar Al-Chalabi, Dr Alfredo Iacoangeli, and Dr Ahmad Al Khleifat for their thoughtful discussions regarding the content of this manuscript.

Authors' contributions

FT: conception and initial draft. FT, LLF, RSA, and PAA: critical editing of the manuscript. All authors have read and approved the final manuscript.

Funding

No funding to report.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author details

¹Perron Institute for Neurological and Translational Science, First floor, RR block, QEII Medical Centre, 8 Verdun St, Nedlands, WA 6009, Australia. ²Centre for Neuromuscular and Neurological Disorders, University of Western Australia, Nedlands, WA, Australia. ³Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University, Perth, WA, Australia. ⁴Black Swan Pharmaceuticals, Wake Forrest, NC, USA. ⁵Faculty of Medicine, Nursing, Midwifery and Health Sciences, University of Notre Dame Australia, Fremantle, WA 6160, Australia. ⁶Division of Neurology, Duke University Medical Centre, Duke University, Durham, NC, USA.

Received: 29 September 2021 Accepted: 6 December 2021

Published online: 17 January 2022

References

- Mejzini R, Flynn LL, Pitout IL, Fletcher S, Wilton SD, Akkari PA. ALS genetics, mechanisms, and therapeutics: where are we now? *Front Neurosci.* 2019;13:1310.
- D'Erchia AM, Gallo A, Manzari C, Raho S, Horner DS, Chiara M, et al. Massive transcriptome sequencing of human spinal cord tissues provides new insights into motor neuron degeneration in ALS. *Sci Rep.* 2017;7(1):10046. <https://doi.org/10.1038/s41598-017-10488-7>.
- Krokidis MG, Vlamos P. Transcriptomics in amyotrophic lateral sclerosis. *Front Biosci (Elite Ed).* 2018;10:103–21.
- Recabarren-Leiva D, Alarcón M. New insights into the gene expression associated to amyotrophic lateral sclerosis. *Life Sci.* 2018;193:110–23. <https://doi.org/10.1016/j.lfs.2017.12.016>.
- Van Eijk RPA, Kliest T, McDermott CJ, Roes KCB, Van Damme P, Chio A, et al. TRICALS: creating a highway toward a cure. *Amyotroph Lateral Scler and*

- Frontotemporal Degener. 2020;21(7-8):1–6. <https://doi.org/10.1080/21678421.2020.1788092>.
6. Van Eijk RP, Eijkemans MJ, Nikolakopoulos S, Jansen MD, Westeneng H-J, Van Eijk KR, et al. Pharmacogenetic interactions in amyotrophic lateral sclerosis: a step closer to a cure? *Pharmacogenomics J.* 2019;1–7.
 7. Van Eijk RP, Jones AR, Sproviero W, Shatunov A, Shaw PJ, Leigh PN, et al. Meta-analysis of pharmacogenetic interactions in amyotrophic lateral sclerosis clinical trials. *Neurol.* 2017;89(18):1915–22. <https://doi.org/10.1212/WNL.0000000000004606>.
 8. Roses A. Polyallelic structural variants can provide accurate, highly informative genetic markers focused on diagnosis and therapeutic targets: Accuracy vs. Precision *Clin Pharmacol Ther.* 2016;99(2):169–71. <https://doi.org/10.1002/cpt.288>.
 9. Cameron DL, Di Stefano L, Papenfuss AT. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nat Commun.* 2019;10(1):3240. <https://doi.org/10.1038/s41467-019-11146-4>.
 10. Ebbert MTW, Jensen TD, Jansen-West K, Sens JP, Reddy JS, Ridge PG, et al. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biol.* 2019;20(1):97. <https://doi.org/10.1186/s13059-019-1828-7>.
 11. Conrad DF, Hurler ME. The population genetics of structural variation. *Nat Genet.* 2007;39(7):S30–S6. <https://doi.org/10.1038/ng2042>.
 12. Mahmoud M, Gobet N, Cruz-Dávalos DI, Mounier N, Dessimoz C, Sedlazeck FJ. Structural variant calling: the long and the short of it. *Genome Biol.* 2019;20(1):246. <https://doi.org/10.1186/s13059-019-1828-7>.
 13. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nat Rev Genet.* 2009;10(4):241–51. <https://doi.org/10.1038/nrg2554>.
 14. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68–74. <https://doi.org/10.1038/nature15393>.
 15. Theunissen F, Flynn LL, Anderton RS, Mastaglia F, Pytte J, Jiang L, et al. Structural variants may be a source of missing heritability in sALS. *Front Neurosci.* 2020;14:47.
 16. Chiang C, Scott AJ, Davis JR, Tsang EK, Li X, Kim Y, et al. The impact of structural variation on human gene expression. *Nat Genet.* 2017;49(5):692–9. <https://doi.org/10.1038/ng.3834>.
 17. Beck J, Poulter M, Hensman D, Rohrer JD, Mahoney CJ, Adamson G, et al. Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. *Am J Hum Genet.* 2013;92(3):345–53. <https://doi.org/10.1016/j.ajhg.2013.01.011>.
 18. Bakeberg MC, Hoes ME, Gorecki AM, Theunissen F, Pfaff AL, Kenna JE, et al. The TOMM40 '523' polymorphism in disease risk and age of symptom onset in two independent cohorts of Parkinson's disease. *Sci Rep.* 2021;11(1):6363. <https://doi.org/10.1038/s41598-021-85510-0>.
 19. Bakeberg MC, Gorecki AM, Pfaff AL, Hoes ME, Kócs S, Akkari PA, et al. TOMM40 '523' poly-T repeat length is a determinant of longitudinal cognitive decline in Parkinson's disease. *NPJ Parkinson's Dis.* 2021;7(1):56. <https://doi.org/10.1038/s41531-021-00200-y>.
 20. Roses A, Lutz M, Amrine-Madsen H, Saunders A, Crenshaw D, Sundseth S, et al. A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease. *Pharmacogenomics J.* 2010;10(5):375–84. <https://doi.org/10.1038/tpj.2009.69>.
 21. Shatunov A, Al-Chalabi A. The genetic architecture of ALS. *Neurobiol Dis.* 2021;147:105156. <https://doi.org/10.1016/j.nbd.2020.105156>.
 22. Saul R, Lutz MW, Burns DK, Roses AD, Chiba-Falek O. The SSV evaluation system: a tool to prioritize short structural variants for studies of possible regulatory and causal variants. *Hum Mutat.* 2016;37(9):877–83. <https://doi.org/10.1002/humu.23023>.
 23. Van Rheenen W, Van der Spek RAA, Bakker MK, JFav V, Hop PJ, RAJ Z, et al. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat Genet.* 2021;53(12):1636–48.
 24. 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. <https://doi.org/10.1038/ng.3622>.
 25. Benyamin B, He J, Zhao Q, Gratten J, Garton F, Leo PJ, et al. Cross-ethnic meta-analysis identifies association of the GPX3-TNIP1 locus with amyotrophic lateral sclerosis. *Nat Commun.* 2017;8(1):611. <https://doi.org/10.1038/s41467-017-00471-1>.
 26. 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;97(6):1268–83. e6.
 27. Al Khleifat A, Iacoangeli A, Jfja VV, Bowles H, RAJ Z, Moisse M, et al. Structural variation analysis of 6,500 whole genome sequences in amyotrophic lateral sclerosis. *NPJ Genom Med.* 2021. [https://kclpure.kcl.ac.uk/portal/en/publications/structural-variation-analysis-of-6500-whole-genome-sequences-inamyotrophic-lateral-sclerosis\(c59ada0f-35ff-4d57-a26c-16ac3e6720f4\).html](https://kclpure.kcl.ac.uk/portal/en/publications/structural-variation-analysis-of-6500-whole-genome-sequences-inamyotrophic-lateral-sclerosis(c59ada0f-35ff-4d57-a26c-16ac3e6720f4).html).
 28. Roses AD, Akkari PA, Chiba-Falek O, Lutz MW, Gottschalk WK, Saunders AM, et al. Structural variants can be more informative for disease diagnostics, prognostics and translation than current SNP mapping and exon sequencing. *Expert Opin Drug Metab Toxicol.* 2016;12(2):135–47. <https://doi.org/10.1517/17425255.2016.1133586>.
 29. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small G, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science.* 1993;261(5123):921–3. <https://doi.org/10.1126/science.8346443>.
 30. Raybould R, Sims R. Searching the dark genome for Alzheimer's disease risk variants. *Brain Sci.* 2021;11(3):332. <https://doi.org/10.3390/brainsci11030332>.
 31. Crenshaw DG, Gottschalk WK, Lutz MW, Grossman I, Saunders AM, Burke JR, et al. Using genetics to enable studies on the prevention of Alzheimer's disease. *Clin Pharmacol Ther.* 2013;93(2):177–85. <https://doi.org/10.1038/clpt.2012.222>.
 32. Burns DK, Chiang C, Welsh-Bohmer KA, Brannan SK, Culp M, O'Neil J, et al. The TOMMORROW study: design of an Alzheimer's disease delay-of-onset clinical trial. *Alzheimer's Dement: Transl Res Clin Interv.* 2019;5(1):661–70. <https://doi.org/10.1016/j.trci.2019.09.010>.
 33. 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. [https://doi.org/10.1016/S1474-4422\(14\)70219-4](https://doi.org/10.1016/S1474-4422(14)70219-4).
 34. Pytte J, Anderton RS, Flynn LL, Theunissen F, Jiang L, Pitout I, et al. Association of a structural variant within the SQSTM1 gene with amyotrophic lateral sclerosis. *Neurol Genet.* 2020;6:e406. <https://doi.org/10.1212/NXG.0000000000000406>.
 35. Fotsing SF, Margoliash J, Wang C, Saini S, Yanicky R, Shleizer-Burko S, et al. The impact of short tandem repeat variation on gene expression. *Nat Genet.* 2019;51(11):1652–9. <https://doi.org/10.1038/s41588-019-0521-9>.
 36. Fecto F, Yan J, Vemula SP, Liu E, Yang Y, Chen W, et al. SQSTM1 Mutations in familial and sporadic amyotrophic lateral sclerosis. *JAMA Neurol.* 2011;68(11):1440–6.
 37. Rea SL, Majcher V, Searle MS, Layfield R. SQSTM1 mutations—bridging Paget disease of bone and ALS/FTLD. *Exp Cell Res.* 2014;325(1):27–37. <https://doi.org/10.1016/j.yexcr.2014.01.020>.
 38. Al-Chalabi A, Andersen PM, Chioza B, Shaw C, Sham PC, Robberecht W, et al. Recessive amyotrophic lateral sclerosis families with the D90A SOD1 mutation share a common founder: evidence for a linked protective factor. *Hum Mol Genet.* 1998;7(13):2045–50. <https://doi.org/10.1093/hmg/7.13.2045>.
 39. Saeed M, Yang Y, Deng HX, Hung WY, Siddique N, Dellefave L, et al. Age and founder effect of SOD1 A4V mutation causing ALS. *Neurology.* 2009;72(19):1634–9. <https://doi.org/10.1212/01.wnl.0000343509.76828.2a>.
 40. Pytte J, Flynn LL, Anderton RS, Mastaglia FL, Theunissen F, James I, et al. Disease-modifying effects of an SCAF4 structural variant in a predominantly SOD1 ALS cohort. *Neurol Genet.* 2020;6:e470. <https://doi.org/10.1212/NXG.0000000000000470>.
 41. Gregersen LH, Mitter R, Ugalde AP, Nojima T, Proudfoot NJ, Agami R, et al. SCAF4 and SCAF8, mRNA anti-terminator proteins. *Cell.* 2019;177(7):1797–813. e18.
 42. Fliedner A, Kirchner P, Wiesener A, van de Beek I, Waisfisz Q, van Haelst M, et al. Variants in SCAF4 cause a neurodevelopmental disorder and are associated with impaired mRNA processing. *Am J Hum Genet.* 2020;107(3):544–54. <https://doi.org/10.1016/j.ajhg.2020.06.019>.
 43. 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;383(2):109–19. <https://doi.org/10.1056/NEJMoa2003715>.
 44. Ze M, Lopez-Erasquin 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.
 45. Klim JR, Williams LA, Limone F, San Juan IG, Davis-Dusenbery BN, Mordes DA, et al. ALS-implicated protein TDP-43 sustains levels of STMN2, a mediator of motor neuron growth and repair. *Nat Neurosci.* 2019;1.

46. 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;13:127.
47. Prudencio M, Humphrey J, Pickles S, Brown A-L, Hill SE, Kachergus J, et al. Truncated stathmin-2 is a marker of TDP-43 pathology in frontotemporal dementia. *J Clin Investig.* 2020;130(11):6080–92. <https://doi.org/10.1172/JCI139741>.
48. Wang Q, Zhang Y, Wang M, Song W-M, Shen Q, McKenzie A, et al. The landscape of multiscale transcriptomic networks and key regulators in Parkinson's disease. *Nat Commun.* 2019;10(1):1–15. <https://doi.org/10.1038/s41467-019-13144-y>.
49. Van Eijk RPA, Nikolakopoulos S, Roes KCB, Kendall L, Han SS, Lavrov A, et al. Challenging the established order: innovating clinical trials for amyotrophic lateral sclerosis. *Neurology.* 2021. <https://doi.org/10.1212/WNL.0000000000012545>.
50. Roggenbuck J. C9orf72 and the care of the patient with ALS or FTD: Progress and recommendations after 10 years. *Neurol Genet.* 2021;7(1).
51. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. *Nat Genet.* 2015;47(8):856–60. <https://doi.org/10.1038/ng.3314>.
52. Rotunno MS, Bosco DA. An emerging role for misfolded wild-type SOD1 in sporadic ALS pathogenesis. *Front Cell Neurosci.* 2013;7:253.
53. Hayashi Y, Homma K, Ichijo H. SOD1 in neurotoxicity and its controversial roles in SOD1 mutation-negative ALS. *Adv Biol Regul.* 2016;60:95–104. <https://doi.org/10.1016/j.jbbior.2015.10.006>.
54. Bennett CF, Kordasiewicz HB, Cleveland DW. Antisense drugs make sense for neurological diseases. *Ann Rev Pharmacol Toxicol.* 2020;61:2021.
55. Petrov D, Mansfield C, Moussy A, Hermine O. ALS clinical trials review: 20 years of failure. Are we any closer to registering a new treatment? *Front Aging Neurosci.* 2017;9:68.
56. Wobst HJ, Mack KL, Brown DG, Brandon NJ, Shorter J. The clinical trial landscape in amyotrophic lateral sclerosis—past, present, and future. *Med Res Rev.* 2020;40(4):1352–84. <https://doi.org/10.1002/med.21661>.
57. Vejux A, Namsi A, Nury T, Moreau T, Lizard G. Biomarkers of amyotrophic lateral sclerosis: current status and interest of oxysterols and phytosterols. *Front Mol Neurosci.* 2018;11:12.
58. Khalil M, Teunissen CE, Otto M, Piehl F, Sormani MP, Gatteringer T, et al. Neurofilaments as biomarkers in neurological disorders. *Nat Rev Neurol.* 2018;14(10):577–89. <https://doi.org/10.1038/s41582-018-0058-z>.
59. Agah E, Saleh F, Moghaddam HS, Saghadzadeh A, Tafakhori A, Rezaei N. CSF and blood biomarkers in amyotrophic lateral sclerosis: protocol for a systematic review and meta-analysis. *Syst Rev.* 2018;7(1):237. <https://doi.org/10.1186/s13643-018-0913-4>.
60. Mitsumoto H, Saito T. A prognostic biomarker in amyotrophic lateral sclerosis. *Clin Neurol.* 2018;58(12):729–36. <https://doi.org/10.5692/clinica/lnurol.cn-001220>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

