Journal of
Medical Genetics

Reduced penetrance of gene variants causing amyotrophic lateral sclerosis

Journal:	*Journal of Medical Genetics*
Manuscript ID	jmg-2023-109580.R1
Article Type:	Short report
Date Submitted by the Author:	23-Oct-2023
Complete List of Authors:	Douglas, Andrew; Oxford University Hospitals NHS Foundation Trust, Oxford Centre for Genomic Medicine; University of Southampton Faculty of Medicine, Human Development and Health Baralle, Diana; University of Southampton, Human Genetic and Genomics
Keywords:	Motor Neuron Disease, Genetic Predisposition to Disease, Dementia, Genetics, Population

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Reduced penetrance of gene variants causing amyotrophic lateral sclerosis

Authors:

Andrew Graham Lim Douglas[1,2]

Diana Baralle[2,3]

Affiliations:

1. Oxford Centre for Genomic Medicine, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

2. Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK

3. Wessex Clinical Genetics Service, University Hospital Southampton NHS Foundation Trust, Southampton, UK

Corresponding author:

Andrew Douglas

Tel: +44(0)1865226024

E-mail: andrew.douglas@ndcn.ox.ac.uk

ABSTRACT:

Background: Amyotrophic lateral sclerosis overlaps etiologically and genetically with frontotemporal dementia and occurs in both familial and apparently sporadic forms. The most commonly implicated genes are *C9orf72*, *SOD1*, *TARDBP* and *FUS*. Penetrance of disease-causing variants in these genes is known to be incomplete but has not been well-studied at population level.

Objective: We sought to determine the population-level penetrance of pathogenic and likely pathogenic variants in genes commonly causing amyotrophic lateral sclerosis.

Methods: Published epidemiological data for amyotrophic lateral sclerosis and frontotemporal dementia were used to calculate expected frequencies of disease-causing variants per gene at population level. Variant data from gnomAD and ClinVar databases were used to ascertain observed numbers of disease-causing variants and to estimate population-level penetrance per gene. Data for *C9orf72* were obtained from published literature.

Results: Maximum population penetrance for either amyotrophic lateral sclerosis or frontotemporal dementia was found to be 33% for *C9orf72* (95% CI [20.9, 53.2]), 54% for *SOD1* (95% CI [32.7, 88.6]), 38% for *TARDBP* (95% CI [21.1, 69.8]) and 19% for *FUS* (95% CI [13.0, 28.4]).

Conclusion: Population-level penetrance of amyotrophic lateral sclerosis disease genes is reduced. This finding has implications for the genetic testing and counselling of affected individuals and their unaffected relatives.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease affecting upper and lower motor neurons within the brain and spinal cord.[1] Despite significant advances in the understanding of its molecular pathogenesis, ALS remains an incurable condition with a peak of incidence between 60 and 75 years and a median survival of only 2-4 years after onset.[2,3] Between 5-10% of cases are reported to be familial, displaying autosomal dominant inheritance with incomplete penetrance.[4,5] Pathogenic variants in over 30 different genes have been implicated as causing familial ALS (fALS), with four genes in particular (*C9orf72* [MIM: 614260], *SOD1* [MIM: 147450], *TARDBP* [MIM: 605078], and *FUS* [MIM: 137070]) contributing to 55% of European ancestry ALS families.[6,7] However, a notably consistent finding has been the identification of pathogenic variants in the same genes contributing to apparently sporadic ALS (sALS), with around 7% of European cases linked to *C9orf72*, *SOD1*, *TARDBP* and *FUS*.[7]

ALS has a worldwide incidence of 1.75 per 100,000 and a lifetime risk of 1 in 347 for men and 1 in 436 for women.[8,9] It overlaps substantially in its disease spectrum with frontotemporal dementia (FTD), with up to 23% of ALS patients also having FTD and 12.5% of FTD patients having ALS.[10,11] A recent large cohort study across nine European countries has calculated the incidence of FTD to be 2.36 per 100,000, although previous estimates have ranged from 2.7-4.1 per 100,000.[12,13] Around 10% of FTD cases display clear autosomal dominant inheritance and the most commonly identified genes are *C9orf72*, *MAPT* and *GRN*, which again contribute to both familial (fFTD) and sporadic (sFTD) cases.[14,15]

The *C9orf72* $(GGGGCC)_n$ repeat expansion is known to be over-represented within European populations, where it has been identified in 0.15-0.2% of healthy controls.[16,17] Given the

relatively well-preserved proportions in contribution of *C9orf72*, *SOD1*, *TARDBP* and *FUS* to both fALS and sALS, it is reasonable to expect the presence of pathogenic variants in these and other ALS genes to be similarly over-represented within population cohorts and for disease penetrance to be reduced to a similar degree to *C9orf72*. In this study, we sought to estimate the expected frequencies of ALS-causing pathogenic variants at a population level based on their reported occurrences within familial and sporadic ALS cohorts and to compare this to the corresponding observed frequencies within a publicly available genomic database. Using these data, it has been possible to estimate the population-level penetrance of the most common ALS-causing genes.

METHODS

Figures for ALS and FTD incidence and proportions of familial and sporadic cases caused by variants in specific genes were obtained from published literature.[6,15] Observed allele frequencies and carrier frequencies (p_O) of pathogenic and likely pathogenic variants were calculated using the gnomAD v2.1.1 database (https://gnomad.broadinstitute.org/) with reference to variant classifications provided by the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/).[18,19] For each gene, the largest allele number from the set of pathogenic or likely pathogenic variants seen in gnomAD was used as the denominator for determining allele frequencies. The "non-neuro" dataset selection was applied to gnomAD variants so as to omit data from individuals ascertained as having a neurological condition in neurological case/control studies. Variants were only counted as disease-causing if they were listed as pathogenic or likely pathogenic in ClinVar (variants were disregarded if of uncertain significance or with conflicting interpretations or low-confidence flags).

If ALS genes were fully penetrant, the expected pathogenic variant carrier frequency in the population for a given gene (p_E) would equal the lifetime risk of ALS (R_A) multiplied by the sum of the proportion of cases that are fALS (F_A) multiplied by the reported frequency of variants causing fALS (p_{FA}) added to the proportion of cases that are sALS (1 - F_A) multiplied by the reported frequency of variants causing sALS (p_{SA}). In order to account for the additional overlap in contribution to FTD of many ALS genes, an identically formulated expression can be added to this equation but using figures for lifetime risk of FTD (R_D), proportion of fFTD (F_{FD}) and sFTD (1 - F_{FD}) cases and the associated frequencies of causative variants in fFTD and sFTD (p_{FD} and p_{SD}). Thus, the expected frequencies of disease-causing variants can be calculated as follows:

$$p_E = R_A\{F_A \cdot p_{FA} + p_{SA}(1 - F_A)\} + R_D\{F_{FD} \cdot p_{FD} + p_{SD}(1 - F_{FD})\}$$

Penetrance (K) can then be estimated as the reciprocal of the observed carrier frequencies divided by the expected carrier frequencies, which is more naturally expressed as the expected divided by the observed carrier frequencies:

$$K = 1 / (p_O / p_E) = p_E / p_O$$

In the calculations of penetrance undertaken, lifetime risks of 1 in 386.4 for ALS (combining both male and female incidence) and 1 in 305 for FTD have been used and fALS and fFTD are taken to represent 1 in 10 cases. The lifetime FTD risk has been estimated based on the upper limit incidence of 2.7-4.1 per 100,000, which if applied to an average population lifespan of 80 years would equate to 1 in 305-463 people.

RESULTS

The results of this study are summarised in Table 1. For *SOD1*, reported in 12% of fALS and 2% of sALS cases, disease-causing variants are expected to occur in 1 in 12,880 people. However, a population frequency of 1 in 6938 people in gnomAD were found to be heterozygous for such variants. Of note, the relatively common p.(Asp91Ala) variant responsible for a recessive form of ALS was excluded on account of its conflicting interpretations of pathogenicity on ClinVar (though no homozygotes were present). This result therefore suggests a maximum penetrance for *SOD1* of 54% at population level. For *TARDBP* and *FUS*, both reported to account for 4% of fALS, 1% of sALS and 1% of fFTD, disease-causing variants are expected to occur in 1 in 27,084 people. However, such variants are actually observed in at least 1 in 10,406 and 1 in 5202 people respectively, giving *TARDBP* an estimated population penetrance of 38% and *FUS* 19%.

Although gnomAD does not include data on the *C9orf72* repeat expansion, a similar calculation can be performed using previously reported figures for its prevalence.[16,17] Based on the contribution of *C9orf72* in up to 40% of fALS, 7% of sALS, 25% of fFTD and 6% of sFTD, the expected carrier frequency would be 1 in 1903 people. However, its reported prevalence in 5/2585 and 11/7579 healthy controls (totalling 17/10163 or 1 in 635 people) would suggest a maximum population penetrance of 33%.

Gene	fALS (sALS) %	fFTD (sFTD) %	Deleterious (Total) alleles	Expected DAF (Expected CF)	Observed DAF (Observed CF)	*p*-value (χ2)	Penetrance (% [95% CI])
C9orf72	40 (7)	25 (6)	16 (20328)	$2.63x10^{-4}$ (1 in 1903)	$7.87x10^{-4}$ (1 in 635)	$3.99x10^{-6}$	33.4 [20.9, 53.2]
SOD1	12 (2)	0 (0)	15 (208130)	$3.88x10^{-5}$ (1 in 12880)	$7.21x10^{-5}$ (1 in 6938)	$1.49x10^{-2}$	53.9 [32.7, 88.6]
TARDBP	4 (1)	1 (0)	10 (208126)	$1.85x10^{-5}$ (1 in 27084)	$4.80x10^{-5}$ (1 in 10406)	$1.68x10^{-3}$	38.4 [21.1, 69.8]
FUS	4 (1)	1 (0)	20 (208080)	$1.85x10^{-5}$ (1 in 27084)	$9.61x10^{-5}$ (1 in 5202)	$1.66x10^{-16}$	19.2 [13.0, 28.4]

Table 1. The expected and observed frequencies of disease-causing variants in ALS-related genes and the maximum estimated disease penetrance at population level. *p*-values have been generated from χ2 values (df = 1) and 95% confidence interval (CI) is given by penetrance$^{(1 \pm 1.96/\chi)}$. DAF, deleterious allele frequency; CF, carrier frequency.

DISCUSSION

Individuals predisposed to late-onset neurodegenerative diseases such as ALS are subject to age-related cumulative risks of developing these disorders. Genetically-determined forms of ALS will therefore always exhibit some degree of incomplete penetrance owing to confounding variables that may affect the lifespan of at-risk individuals. This can often be seen when assessing an affected individual's family history, for example where an intervening relative who is an obligate gene carrier has died at a relatively young age and thus would appear to have been unaffected, potentially leading to underestimation of penetrance. To complicate matters further, there is currently no reliable way to ascertain whether or not a deceased at-risk individual (e.g. a carrier of an apparently pathogenic *SOD1* variant) would have gone on to develop neurodegeneration. This makes measuring the true penetrance of genes predisposing to ALS problematic.

This study has sought to estimate ALS gene penetrance by reference to published and publicly available information. The calculated penetrance estimates depend heavily on the accuracy of the epidemiological figures employed. A number of additional inherent limitations also apply. Firstly, there are undoubtedly multiple and substantial non-genetic and environmental influences on ALS and we have not sought to address these factors in this purely genetic analysis. Secondly, this study assumes that the population in gnomAD is an accurate surrogate for the general population. Thirdly, no allowance has been made for different rates of ALS gene variants across different ethnic groups, for which good evidence exists.[7] Fourthly, the analysis assumes that ClinVar annotations are accurate and does not take into account the substantial numbers of variants present in gnomAD that are not present in ClinVar and which therefore have not been classified. Finally, by performing analysis at the

gene level, no consideration has been given to the possible variability in penetrance that might be attributable to specific gene variants.

In defence of the relatively simplistic approach used here, it should be noted that gnomAD v2.1.1 does indeed have a large preponderance of data from European ancestry individuals.[18] The ALS gene contribution figures used have also been largely ascertained in respect to European ancestry cohorts and so these are likely to be applicable.[6] We have sought to use generally accepted figures for percentage gene contributions to ALS and FTD but recognise that uncertainty exists, particularly for figures under 1% and so we have not included values below 1% and similarly have limited the analysis to only the four most commonly involved ALS genes. In addition, we have sought to maximise the expected number of variants (and hence to maximise the calculated gene penetrance) by using a fALS rate of 10% rather than 5% and by including lifetime risks of ALS (1 in 385) and FTD (1 in 305) that are towards the upper limits for what has been reported or calculated according to incidence. Furthermore, by only counting variants classified unambiguously on ClinVar, we believe that we have in fact underestimated the true frequencies of observed pathogenic and likely pathogenic variants, which if included would further increase the observed/expected variant ratios and therefore decrease the penetrance of each gene even further. By limiting our analysis to ALS and FTD, we have precluded the possibility of other forms of neurodegeneration being caused by these gene variants. However, widely accepted figures for such contributions are generally not available and we have therefore chosen to omit these. With regards to variant-level effects, such differences are indeed likely to exist and further research into the reasons for genotype-specific disease variability may prove fruitful in terms of understanding the pathogenic mechanisms underlying genetic forms of ALS and FTD.

However, the results of this study would tend to caution against the automatic attribution of penetrance figures to individual variants, since the implication of this work is that penetrance is not an intrinsic characteristic of a given gene or variant but rather is modifiable and dependent upon other genetic and non-genetic factors. In any case, the numbers of variants in the gnomAD cohort are too low to allow adequate analysis to this level of detail and this will only start to become possible with reference to genomic data cohorts that include millions of individuals.

In this study, we have identified that pathogenic and likely pathogenic variants in commonly identified genes linked to ALS are over-represented in control populations and therefore display reduced penetrance at the population level. This finding is of relevance to understanding the aetiology and pathogenesis of ALS and FTD and raises the question of what factors, genetic or otherwise, are influencing the penetrance in this way. It also has important clinical implications for those undertaking genetic testing of individuals affected by ALS and FTD and even more so for the predictive genetic testing of unaffected relatives. Now that clinical trials are taking place to target *SOD1* and *C9orf72* gene carriers, it is becoming critical to understand who will and who will not go on to develop disease and hence who is likely to benefit most and least from such treatments.[20] Importantly, it should be noted that the penetrance figures reported here apply only to the population as a whole, not to the individual. It would therefore not be appropriate to counsel affected individuals and their relatives based solely on these figures. Individual carriers of disease-causing variants in these genes are likely to have individualised risk profiles and this should be considered carefully when making clinical decisions about predictive testing. Until such time as the other relevant factors influencing ALS gene penetrance become known and their relative contributions

understood, an individual's risk will likely best be determined by reference to his or her own family history of disease.

ACKNOWLEDGMENTS

The authors would like to acknowledge all the contributors to the gnomAD and ClinVar databases, without which this study would not have been possible.

CONTRIBUTORS

Conceptualisation: A.G.L.D., D.B.; Data curation: A.G.L.D.; Formal analysis: A.G.L.D.; Funding acquisition: D.B.; Investigation: A.G.L.D.; Methodology: A.G.L.D.; Project administration: A.G.L.D.; Resources: A.G.L.D., D.B.; Software: A.G.L.D.; Supervision: D.B.; Validation: A.G.L.D.; Visualisation: A.G.L.D.; Writing-original draft: A.G.L.D.; Writing-review & editing: A.G.L.D., D.B.

FUNDING

This research has been funded by a National Institute for Health Research (NIHR) Research Professorship awarded to D.B. (RP-2016-07-011).

COMPETING INTERESTS

The authors declare no conflict of interest.

PATIENT CONSENT FOR PUBLICATION

Not required.

ETHICS APPROVAL

This research has been done solely using previously published information and free access to publicly available aggregate genomic data, which is not identifiable.

REFERENCES

1 Brown RH, Al-Chalabi A. Amyotrophic Lateral Sclerosis. *N Engl J Med* 2017;**377**:162–72. doi:10.1056/NEJMra1603471

2 Chiò A, Logroscino G, Traynor BJ, *et al.* Global epidemiology of amyotrophic lateral sclerosis: a systematic review of the published literature. *Neuroepidemiology* 2013;**41**:118–30. doi:10.1159/000351153

3 Chiò A, Logroscino G, Hardiman O, *et al.* Prognostic factors in ALS: A critical review. *Amyotroph Lateral Scler* 2009;**10**:310–23. doi:10.3109/17482960802566824

4 Byrne S, Walsh C, Lynch C, *et al.* Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry* 2011;**82**:623–7. doi:10.1136/jnnp.2010.224501

5 Fang F, Kamel F, Lichtenstein P, *et al.* Familial aggregation of amyotrophic lateral sclerosis. *Ann Neurol* 2009;**66**:94–9. doi:10.1002/ana.21580

6 Goutman SA, Hardiman O, Al-Chalabi A, *et al.* Emerging insights into the complex genetics and pathophysiology of amyotrophic lateral sclerosis. *Lancet Neurol* 2022;**21**:465–79. doi:10.1016/S1474-4422(21)00414-2

7 Zou Z, Zhou Z, Che C, *et al.* Genetic epidemiology of amyotrophic lateral sclerosis: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry* 2017;**88**:540–9. doi:10.1136/jnnp-2016-315018

8 Marin B, Boumédiene F, Logroscino G, *et al.* Variation in worldwide incidence of amyotrophic lateral sclerosis: a meta-analysis. *Int J Epidemiol* 2017;**46**:57–74. doi:10.1093/ije/dyw061

9 Ryan M, Heverin M, McLaughlin RL, *et al.* Lifetime Risk and Heritability of Amyotrophic Lateral Sclerosis. *JAMA Neurol* 2019;**76**:1367–74.

doi:10.1001/jamaneurol.2019.2044

10 Rippon GA, Scarmeas N, Gordon PH, *et al.* An observational study of cognitive impairment in amyotrophic lateral sclerosis. *Arch Neurol* 2006;**63**:345–52. doi:10.1001/archneur.63.3.345

11 Burrell JR, Kiernan MC, Vucic S, *et al.* Motor neuron dysfunction in frontotemporal dementia. *Brain* 2011;**134**:2582–94. doi:10.1093/brain/awr195

12 Logroscino G, Piccininni M, Graff C, *et al.* Incidence of Syndromes Associated With Frontotemporal Lobar Degeneration in 9 European Countries. *JAMA Neurol* 2023;**80**:279–86. doi:10.1001/jamaneurol.2022.5128

13 Onyike CU, Diehl-Schmid J. The epidemiology of frontotemporal dementia. *Int Rev Psychiatry* 2013;**25**:130–7. doi:10.3109/09540261.2013.776523

14 Rohrer JD, Beck J, Isaacs AM, *et al.* The heritability and genetics of frontotemporal lobar degeneration. *Neurology* 2009;**73**:1451–6. doi:10.1212/WNL.0b013e3181bf997a

15 Carr JS, Sirkis DW, Yokoyama JS. *Genetic contributions to sporadic frontotemporal dementia*. Elsevier Inc. 2020. doi:10.1016/B978-0-12-815868-5.00005-0

16 Majounie E, Renton AE, Mok K, *et al.* Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: A cross-sectional study. *Lancet Neurol* 2012;**11**:323–30. doi:10.1016/S1474-4422(12)70043-1

17 Beck J, Poulter M, Hensman D, *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**:345–53. doi:10.1016/j.ajhg.2013.01.011

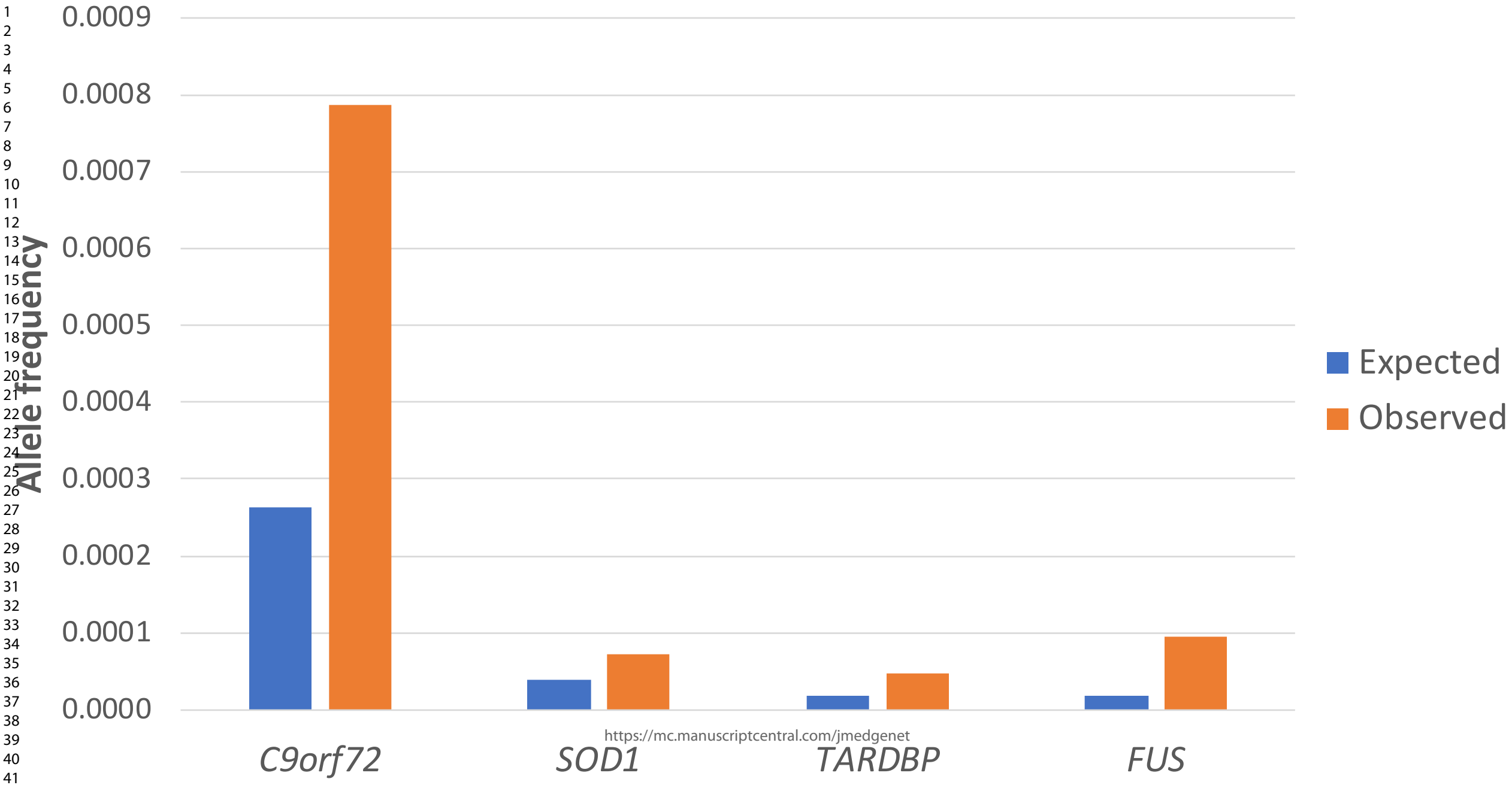
18 Karczewski KJ, Francioli LC, Tiao G, *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;**581**:434–43. doi:10.1038/s41586-020-2308-7

19 Landrum MJ, Lee JM, Benson M, *et al.* ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018;**46**:D1062–7. doi:10.1093/nar/gkx1153

20 Benatar M, Wuu J, Andersen PM, *et al.* Design of a Randomized, Placebo-Controlled, Phase 3 Trial of Tofersen Initiated in Clinically Presymptomatic SOD1 Variant Carriers: the ATLAS Study. *Neurotherapeutics* 2022;**19**:1248–58. doi:10.1007/s13311-022-01237-4

FIGURE LEGENDS

Figure 1. Expected and observed deleterious allele frequencies for genes commonly involved in ALS.


Observed vs expected ALS gene deleterious allele frequencies
Allele frequency
0.0009
0.0008
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
Expected
Observed
C9orf72
SOD1
TARDBP
FUS