NEUROLOGY/2017/866988

The multistep hypothesis of ALS revisited: the role of genetic mutations

Adriano Chiò, MD; Letizia Mazzini, MD; Sandra D'Alfonso, PhD; Lucia Corrado, PhD; Antonio Canosa, MD, PhD; Cristina Moglia, MD, PhD; Umberto Manera, MD; Enrica Bersano, MD; Maura Brunetti, BSc; Marco Barberis, BSc, PhD; Jan H. Veldink, MD, PhD; Leonard H. van den Berg, MD; PhD; Neil Pearce, DSc; William Sproviero, PhD; Russell McLaughlin, PhD; Alice Vajda, PhD; Orla Hardiman, MD, PhD; James Rooney, MSc; Gabriele Mora, MD; Andrea Calvo, MD; PhD; Ammar Al-Chalabi, PhD FRCP

From the 'Rita Levi Montalcini' Department of Neuroscience, University of Torino, Turin, Italy (Prof. Adriano Chiò, Antonio Canosa, Cristina Moglia, Umberto Manera, Maura Brunetti, Marco Barberis, Andrea Calvo); Institute of Cognitive Sciences and Technologies, National Research Council (CNR), Rome, Italy (Prof. Adriano Chiò); ALS Center, Department of Neurology, Azienda Ospedaliera Universitaria Maggiore della Carità, Novara, Italy (Letizia Mazzini, Enrica Bersano); Department of Health Sciences, Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), 'Amedeo Avogadro' University of Eastern Piedmont, Novara, Italy (Sandra D'Alfonso, Lucia Corrado); Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom (Neil Pierce); Centre for Public Health Research, Massey University Wellington Campus, Wellington, New Zealand (Neil Piece); Department of Neurology and Neurosurgery, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands (Leonard van den Berg, Prof. Jan Veldink); Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Ireland (Orla Hardiman; James Rooney; Russell McLaughlin; Alice Vajda); Istituti Clinici Scientifici Maugeri, IRCCS Milano, Milan, Italy (Gabriele Mora); King's College London, Institute of Psychiatry, Psychology and Neuroscience, Maurice Wohl Clinical Neuroscience Institute, London, United Kingdom (William Sproviero; Ammar Al-Chalabi).

Corresponding author: Adriano Chiò, MD, ALS Center, "Rita Levi Montalcini" Department of Neuroscience, University of Torino, via Cherasco 15, 10126 Torino, Italy (achio@usa.net).

Abstract word count: 246

Text work count: 2782

Character count for title: 72

Number of references: 37

Running title: Genetic mutations in the six-steps process in ALS

Author contributions. ACh, LM, JHV, LHvdB, NP, OH, JR, GM, ACal, AA-C contributed to the literature search, figures, study design, data collection, data analysis, data interpretation, and writing the manuscript. SDA, ACan, CM, UM, EB, LC, MBr, MBa, WS, RM, AV contributed to the data collection and data analysis. All authors critically revised the manuscript.

Conflict of interest disclosures. Dr Chio reports grants from Italian Ministry of Health (Ricerca Finalizzata), grants from EU Joint Programme–Neurodegenerative Disease Research (JPND) through the Ministry of Education, University and Research; grants from the Italy-Israel Scientific Collaboration thought the Italian Foreign Ministry during the study; personal fees from Biogen Idec, personal fees from Cytokinetics, personal fees from Italfarmaco, personal fees from Mitsubishi Tanabe, and personal fees from Neuraltus outside the submitted work. Dr Mazzini reports no disclosures. Dr D'Alfonso reports grant from Agenzia Italiana per la Ricerca sulla SLA during the study. Dr Corrado reports no disclosures. Dr Canosa reports to disclosures. Dr Moglia reports a grant from the Italian Ministry of Health (Ricerca Finalizzata), and a grant from Compagnia di San Paolo. Dr Manera reports no disclosures. Dr Bersano reports no disclosures. Dr

Brunetti reports no disclosures. Dr Barberis reports no disclosures. Dr Veldink reports no disclosures. Dr van den Berg reports grants from The Netherlands ALS Foundation, grants from Netherlands Organisation for Health Research and Development (Vici scheme), grants from VSB fonds, grants from H Kersten and M Kersten (Kersten Foundation), grants from Prinses Beatrix Fonds (PB 0703), grants from Adessium Foundation, grants from Netherlands Organisation for Health Research through the JPND, during the conduct of the study; personal fees from travel grants and consultancy fees from Baxter, personal fees from Scientific Advisory Board Biogen Idec, outside the submitted work. Dr Pearce reports no disclosures. Dr Sproviero reports no disclosures. Dr McLaughlin reports no disclosures. Dr Vadja reports no disclosures. Dr Hardiman reports no disclosures. Dr Rooney reports no disclosures. Dr Mora reports grants from Italian Ministry of Health (Ricerca Finalizzata), outside the submitted work. Dr Calvo reports no disclosures. Dr Al-Chalabi reports grants from the EU Joint Programme–Neurodegenerative Disease Research (JPND) through the Medical Research Council and through the Economic and Social Research Council during the study; consultancy for Biogen Idec, Cytokinetics Inc, Treeway Inc, Mitsubishi-Tanabe Pharma, and OrionPharma outside the submitted work.

The Article Processing Charge was funded by Research Councils UK.

Abstract

Objective. ALS incidence rates are consistent with the hypothesis that ALS is a multistep process. We tested the hypothesis that carrying a large effect mutation might account for one or more steps through the effect of the mutation, thus leaving fewer remaining steps before ALS begins.

Methods. We generated incidence data from an ALS population register in Italy (2007-2015) for which genetic analysis for *C9orf72, SOD1, TARDBP* and *FUS* genes was performed in 82% of incident cases. As confirmation, we used data from ALS cases diagnosed in the Republic of Ireland (2006-2014). We regressed the log of age-specific incidence against the log of age with least squares regression for the subpopulation carrying disease-associated variation in each separate gene.

Results. Of the 1077 genetically-tested cases, 74 (6.9%) carried *C9orf72* mutations, 20 (1.9%) *SOD1* mutations, 15 (1.4%) *TARDBP* mutations and 3 (0.3%) *FUS* mutations. In the whole population there was a linear relationship between log incidence and log age (r^2=0.98) with a slope estimate of 4.65 (4.37-4.95), consistent with a 6-step process. The analysis for *C9orf72* mutated patients confirmed a linear relationship (r^2 = 0.94) with a slope estimate of 2.22 (1.74-2.29), suggesting a 3-step process. This estimate was confirmed by data from the Irish ALS register. The slope estimate was consistent with a 2-step process for *SOD1* and with a 4-step process for *TARDBP*.

Conclusions. The identification of a reduced number of steps in ALS patients with genetic mutations compared to those without mutations supports the idea of ALS as a multistep process and is an important advance for dissecting the pathogenic process in ALS.

Introduction

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disorder characterized by a progressive loss of cortical, bulbar and spinal motor neurons, often associated with an involvement of prefrontal cortex. There are indications that the degenerative process in ALS is the consequence of a combination of genetic and environmental factors. More than 20 genes have been detected as causes of ALS.[1] Several environmental factors have been proposed, but none of them, with the possible exception of cigarette smoking and military service, are consistently associated with ALS.[2-5] About 20% of ALS heritability is attributable to common genetic variation compared with an overall heritability of 60% in studies based on concordance of monozygotic twin pairs.[6-8]

In a previous study, we utilized the Armitage-Doll model derived from cancer research to assess whether ALS incidence is consistent with a multistep process, and if so, to estimate the number of steps, *n*, required for ALS to develop.[9, 10] The model can be briefly conceptualized as follows: if we assume that ALS is caused in a single step molecular process, then the incidence in a particular year will be proportional to the risk of undergoing the step, which in turn depends on exposure to the relevant disease-causing factor. The probability a second molecular step has occurred by that year is dependent on the risk of exposure to the relevant factor per year and the number of years of exposure, or age, and this is true for any subsequent step. Thus incidence is proportional to the product of the risks of undergoing the first step and the subsequent steps. This concept implies a logarithmic increase in incidence with age, obeying a power law in which one less than the number of steps, n-1, relates to the rate of increase. As a result, taking logs of the age of onset and incidence rates has the form of a straight line equation with slope n-1 if a multistep model applies. Our previous study found a linear relationship, with a slope estimate of 5, indicating that the process leading to ALS needs on average 6 steps.[9] Considering the large heterogeneity of ALS in terms of clinical

presentation, progression and outcome, it is likely that the number of steps varies in specific subgroups of patients. For example, those carrying a large effect mutation might have one or more steps accounted for by the effect of the mutation, and thus have fewer remaining steps before ALS is established. We therefore tested this hypothesis using the Armitage-Doll model in genetically defined patient subgroups from a population-based cohort.

Methods

All people with ALS diagnosed in Piemonte and Valle d'Aosta, Italy, in the period 2007-2015 were eligible to be enrolled in the study. Cases were identified through the Piemonte and Valle d'Aosta register for ALS (PARALS). PARALS is a prospective epidemiological register based on the collaboration of the neurological departments of the two Italian regions. ALS cases are ascertained through several concurrent sources (hospital admission, etc.). ALS diagnosis is based on El Escorial revised criteria. Cases with definite, probable and probable laboratory supported El Escorial diagnosis during the course of the disease are included in the register. A detailed description of register methodology is reported elsewhere.[11] The cohort included in this study is different from that included in the previous one, which was based on patients incident in the 1995-2004 period.

As a confirmation cohort, we used the data from ALS cases diagnosed in the Republic of Ireland in the period 2006-2014. ALS cases were identified though the Irish ALS register.[12] Although similar cohorts exist for the other registers studied in our original report, the genetic data are either not complete enough or do not overlap enough with the population data to allow similar analysis.

Genetic analysis. All cases were tested for mutations in *SOD1* (all exons), *TARDBP* (exon 6), *FUS* (exons 14 and 15), and *C9orf72* using standard methodology described elsewhere.[13] *C9orf72* repeat length was determined using repeat primed PCR. Normal was defined as 28 or fewer repeats.

Statistical analysis. The Armitage-Doll methodology was used,[10] under the same assumptions as our previous paper.[9] In brief, a plot of the log of ALS incidence against log age will be linear if a multistep model applies, and will have slope $n-1$, i.e. one less than the number of steps needed for disease onset. According to the pattern identified in cancer, the model predicts that the slope will be approximately linear, but will decrease (and therefore will be less than linear)

at older age groups due to a substantial proportion of the population having undergone one or more of the earlier steps.

Following this model, we calculated the incidence rates per 100,000 person-years in 5-year age-groups for people aged 35-74 years. We excluded the youngest age-groups (those younger than 35) because of the small number of patients, and the older age-groups (those over 74) because of the risk of under-ascertainment or cohort-effect; this reflects also the finding in some cancers where the log incidence and log age association is non-linear in the older age groups.[10]

We then performed a preliminary analysis of the log incidence against log age on all cases (i.e., both mutated and non-mutated), in order to verify if our population followed a multistep model and to replicate our previous findings. Second, we assessed separately familial and non-familial ALS patients. Third, we assessed the incidence of ALS for cases involving each single gene. To correctly calculate incidence, the population used for the denominator should correspond to the population used for the numerator. For example, for ALS incidence in those carrying a *C9orf72* mutation, the correct denominator to use would be the count of all people in the population carrying a *C9orf72* mutation. This information was only available for the cases but not for the general population. However, since the relevant mutations do not in general markedly increase mortality apart from their effects on ALS, we assumed that the proportions of the population carrying a specific mutation would not differ substantially by age-group (e.g. the proportions with the *C9orf72* mutation would be similar in the 40-44 and 60-64 years age-groups). Under this assumption, it is then reasonable to use the total population as the denominators in the analyses for specific genes (e.g. cases involving the *C9orf72* mutation), since this would involve multiplying the relevant age-specific population denominator by an unknown but fixed constant (e.g. if 5% of the population carry a particular mutation, then the total population denominator would be 20 times that of the unknown population subgroup carrying this

mutation). Thus, all of the age-specific incidence rates would be overestimated by an unknown but fixed multiplying factor; this in turn would affect the age-specific incidence rates, but would have no effect on the slope of the graph of log incidence against log age.

Standard Protocol Approvals, Registrations, and Patient Consents

The Piedmont regional government has recognized the Piemonte ALS Registry as a 'Registry of High Sanitary Interest' (Regional Law, April 11, 2012, n. 4). Accordingly, PARALS has the right to access to all the existing databases owned by the regional administration and to obtain clinical information about ALS patients from public and private hospitals, and general practitioners. The study was approved by the Ethical Committee of the Città della Salute e della Scienza of Turin. The register database is anonymized and treated according to Italian Data Protection Code. Patients sign a written informed consent. The Irish ALS Register complies with Irish Data protection legislation (1988 and 2003), and has been approved by the Beaumont Hospital Ethics Committee (02/28 and 05/49).

Data availability

Anonymized data will be shared by request from any qualified investigator.

Results

Of the 1309 cases incident during the 2007-2015 period, 1077 (82.2%) underwent genetic analysis of all four genes, 93.5% (1030) of those followed by the two ALS multidisciplinary centers and 21.7% (47) of those followed by general neurology departments. Patients who did not undergo genetic analysis were older and more frequently had a bulbar onset than those who were tested (Table 1). *C9orf72* mutations were detected in 74 cases (6.9%), *SOD1* in 20 (1.9%), *TARDBP* in 15 (1.4%) and *FUS* in 3 (0.3%). One patient carried both a *C9orf72* expansion and the p.Asn390Ser heterozygous missense mutation of the *TARDBP* gene. A list of *SOD1, TARDBP* and *FUS* mutations is reported in Table 2.

In the 1077 patients with genetic test data, there was a linear relationship between log incidence and log age (r^2=0.98) with a slope estimate of 4.65 (95% CI 4.37-4.95), consistent with a 6-step process (Figure 1a), replicating our previous findings. A similar result (r^2=0.99) was obtained when including all 1309 incident cases (Figure 1b). There was no effect of sex (data not shown).

When considering the 109 patients with definite or probable familial ALS (10.1% of the total),[14] there was a linear relationship between log incidence and log age, with a slope estimate of 2.95 (2.43-3.57), consistent with a 4-step process (Figure 2).

The analysis for *C9orf72* mutated patients confirmed a linear relationship (r^2 = 0.94) with a slope estimate of 2.22 (1.74-2.79) suggesting a 3-step process (Figure 3a). Similarly, a linear relationship was found for *SOD1* mutated patients (r^2 = 0.53; n-1 = 0.76, 95% CI 0.46-1.17) consistent with a 2-step process (Figure 3b) and for *TARDBP* (r^2 = 0.93; n-1 = 3.24, 95% CI 2.21-4.13) consistent with a 4-step process (Figure 3c). Due to the very small number of cases carrying *FUS* mutations we did not estimate the slope. When considering the 45 patients with familial ALS but negative for the 4 tested genes, the linear relationship was confirmed (r^2 =

0.95; *n*-1 = 3.71, 95% CI 2.67-4.53) consistent with a 5-step process (Figure not shown). These data are summarized in Table 3.

We next analyzed patients from the Irish ALS register. The register includes 597 genetically-tested patients (56.3% of incident patients in the 2006-2014 period), of whom 67 carried a *C9orf72* expansion. In the 597 patients with genetic test data, there was a linear relationship between log incidence and log age (r^2=0.93) with a slope estimate of 5.09 (4.69-5.52), consistent with a 6-step process. In the *C9orf72* expanded cases results were similar to those of the Piemonte register (r^2 = 0.66; slope estimate 2.47, 95% CI 1.91-3.13, consistent with a 3 step process). Finally, the 530 Irish patients without a *C9orf72* expansion had a slope estimate of 5.35 (4.92-5.82) (r^2=0.95). No patients with *SOD1* mutations and only two with *TARDBP* missense mutations were identified in the Irish ALS register, making it impossible to assess the effects of these genes.

For comparison, we assessed the slope for type 1 and type 2 diabetes mellitus, using the data of the Piemonte register for diabetes for the age-groups 30 to 49 years (Table 4).[15] In keeping with our findings on ALS, the slope estimate for diabetes type 1, a highly genetically determined disease, was 0.96 (0.62-1-13) (r^2=1.0) consistent with a 2 step process, while that of diabetes type 2, a multifactorial disease with a polygenic architecture, was 5.27 (4.50-6.18) (r^2=0.98), consistent with a 6 step process.

Discussion

We have found that in patients carrying a genetic mutation, the slope of the graph of log incidence and log age is lower than that of cases who do not carry these mutations. This in turn implies that the number of steps necessary to start the neurodegenerative process in genetically mediated ALS is reduced compared to non-mutated cases. The number of steps varies according to the mutated gene, and is lower for *SOD1* (2), intermediate for *C9orf72* (3) and higher for *TARDBP* (4). The number of steps identified in non-mutated patients is 6, consistent with our previous paper.[9] In particular, the slope in the patients from the Piemonte register reported in that paper, which was based on incident cases in the 1995-2004 period, is almost identical to that found in the present paper, which was based on the incident cases in the 2007-2015 period. Furthermore, the slope for *C9orf72,* as well the overall slope of genetically-tested patients, was confirmed in the Irish ALS population. These findings suggest that a genetic lesion alone might account for up to four molecular steps, leaving only two further, likely environmental, steps for those with *SOD1* mutation for example. This argues for the concentration of efforts in dissecting environmental risk factors in individuals with identified mutations rather than those with apparently sporadic ALS, since such environmental factors will be fewer in number per person, and likely of larger effect size as a result.

It is generally recognized that ALS is a multifactorial disease, characterized by interplay between genetic and environmental factors. Although several ALS-related genes are known, it is increasingly clear that genetic mutations alone cannot fully explain the pathological process in ALS, but that genes can rather be considered triggers of the degenerative process. A similar role can be attributed to environmental toxins. However, we have very little information about the possible exogenous factors involved in ALS. Cigarette smoking may be a risk factor in ALS;[5, 16] other suggested factors are physical activity, participating in professional sports, and physical trauma.[4] Protective factors have also been hypothesized, such as diabetes mellitus[17]

and an unfavorable lipid profile.[18, 19] All these factors could act on the genome through epigenomic interactions. For example, smoking induces DNA hypermethylation in specific CpG sites, which persists for years after cessation of smoking,[20] or may induce somatic nucleic acid changes.[5, 21] It is likely that the remaining steps in different genetic subgroups may originate from one or more of these risk factors.

Besides our current results, there are more observations that fit the multistep hypothesis in ALS. First, there are indications that ALS can be an oligogenic disease. In fact, there are several reports of patients carrying two or more mutations of different ALS-related genes.[22] In a study of 391 ALS patients which assessed variants in 17 genes, 3.8% had variants in more than one gene.[23] In that series, the burden of rare variants in known ALS genes significantly reduced the age of onset of symptoms.[23, 24] In the present series one patient had both a *C9orf72* expansion and a heterozygous mutation of the *TARDBP* gene, even though we assessed only four genes. Second, besides 'causative' genes, several other genes have been reported to modify ALS phenotype, such as *UNC13A*, *ATXN2* and *CAMTA1*;[25-27] suggesting that variants in these genes modify the sequential process, either accelerating or slowing it.

Non-genetic elements such as environmental factors[28] and aging also likely trigger molecular steps. However, consistent with other reports, the slope is the same between sexes in all analyses suggesting there is no effect of sex on the assumed cascade.[29]

There appears to be some relationship between the number of remaining steps identified for each mutated gene and the penetrance of mutation in the gene. Such a relationship is consistent with a multi-step model, since a greater number of remaining steps will correspond to a lower probability of exposure to all the steps and therefore reduce the probability of disease given a specific genotype. For example, *C9orf72* expansion mutation penetrance has been estimated to be 60% at the age of 60 and 91% at the age of 80 years[30, 31] and corresponds to 3 remaining steps. At least three mechanisms might regulate *C9orf72* penetrance: first, the size of the

GGGGCC expansion; second, DNA methylation and transcriptional downregulation of the promoter,[32] and third the presence of additional mutations.[33]

The penetrance of *TARDBP* mutation is much lower than that of *C9orf72* (60% at 80 years for the p.Ala382Thr mutation)[34] and it leaves 4 remaining steps, more than for the other two genes. The lowest number of remaining steps, 2, has been estimated for *SOD1* mutation. However, *SOD1* penetrance varies across the different mutations. For example, in a study on pedigrees dating back to the 18th century, carriers of p.Glu101Gly, p.Ile114Thr and p.Val149Gly *SOD1* mutations were reported to have a penetrance of more than 95% at the age of 78.[35] Similarly, the penetrance of the p.Ala5Val mutation, the commonest in the USA, has been estimated to be 91% at the age of 80.[36] Other mutations have a much reduced penetrance; an example is the p.Asp91Ala mutation, which is transmitted with a recessive inheritance in people of Scandinavian ancestry and with a dominant inheritance, albeit with a low penetrance, in the other populations.[37, 38] Most of the *SOD1* mutations we identified are regarded as having very high penetrance, and would therefore be expected to account for more steps than low penetrance mutations.

This study has some weaknesses. First, it was not possible to genotype all incident ALS patients. Non-tested patients were older and more frequently had bulbar onset than those who were tested. However, we could obtain the DNA of >80% of incident patients, a high proportion in an epidemiological setting. Second, only the four more commonly mutated ALS genes were assessed. However, non-tested genes account for only a fraction of ALS patients in European-derived populations. Third, the estimation of the slope was performed on the relatively small number of genetic cases, in particular for *SOD1* and *TARDBP*, and the slope estimates may therefore be imprecise. Finally, population denominators were not available for specific mutations; however, as noted above, this would have affected our age-specific incidence estimates, but not the slope of the graph of log incidence against log age. It is

therefore important that our findings be replicated in other populations with larger cohorts of patients to confirm our results and determine the extent to which they can be generalized.

The identification of a reduced number of steps in ALS patients with genetic mutations compared to those without mutations strongly supports the idea of ALS as a multistep process and represents a first clue for uncovering the pathogenic process of ALS. Similar patterns have previously been observed in studies of specific cancers in which the relevant mutations and other environmentally-induced steps have been able to be identified, and postulated as being also relevant to neurodegeneration.[21] Our findings support the idea of parallels between the processes leading to carcinogenesis and those leading to ALS. The fact that only 2, 3 or 4 steps are required before disease onset in genetically-mediated ALS is consistent with the concept that up to 4 of the six steps required for disease onset are already accounted for by inherited mutation. This idea is also consistent with the observation that penetrance corresponds to the number of steps accounted for. An alternative explanation is that the underlying etiology must differ in at least one step between genetic and other forms of ALS. The analysis of the influence of non-genetic risk factors should therefore also be performed, to clarify their contribution to the multistep process of ALS. The relatively limited number of steps leading to ALS, as compared, for example, to the complexity of the mechanisms at the base of other multifactorial diseases such as schizophrenia,[39] provides hope for the development of an effective therapy for this devastating disease.

References

1. Al-Chalabi A, Hardiman O. The epidemiology of ALS: a conspiracy of genes, environment and time. Nat Rev Neurol 2013;9:617-628.

2. Al-Chalabi A, Pearce N. Commentary: Mapping the Human Exposome: Without It, How Can We Find Environmental Risk Factors for ALS? Epidemiology 2015;26:821-823.

3. Pearce N, Kromhout H. Occupational causes of amyotrophic lateral sclerosis: where to from here? Occup Environ Med 2017;74:83-84.

4. Ingre C, Roos PM, Piehl F, Kamel F, Fang F. Risk factors for amyotrophic lateral sclerosis. Clin Epidemiol 2015;7:181-193.

5. Armon C. Smoking may be considered an established risk factor for sporadic ALS. Neurology 2009;73:1693-1698.

6. Al-Chalabi A, Fang F, Hanby MF, et al. An estimate of amyotrophic lateral sclerosis heritability using twin data. J Neurol Neurosurg Psychiatry 2010;81:1324-1326.

7. Keller MF, Ferrucci L, Singleton AB, et al. Genome-wide analysis of the heritability of amyotrophic lateral sclerosis. JAMA Neurol 2014;71:1123-1134.

8. McLaughlin RL, Vajda A, Hardiman O. Heritability of Amyotrophic Lateral Sclerosis: Insights From Disparate Numbers. JAMA Neurol 2015;72:857-858.

9. Al-Chalabi A, Calvo A, Chio A, et al. Analysis of amyotrophic lateral sclerosis as a multistep process: a population-based modelling study. Lancet Neurol 2014;13:1108-1113.

10. Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer 1954;8:1-12.

11. Chio A, Mora G, Calvo A, et al. Epidemiology of ALS in Italy: a 10-year prospective population-based study. Neurology 2009;72:725-731.

12. O'Toole O, Traynor BJ, Brennan P, et al. Epidemiology and clinical features of amyotrophic lateral sclerosis in Ireland between 1995 and 2004. J Neurol Neurosurg Psychiatry 2008;79:30-32.

13. Chio A, Calvo A, Mazzini L, et al. Extensive genetics of ALS: a population-based study in Italy. Neurology 2012;79:1983-1989.

14. Byrne S, Bede P, Elamin M, et al. Proposed criteria for familial amyotrophic lateral sclerosis. Amyotroph Lateral Scler 2011;12:157-159.

15. Bruno G, Runzo C, Cavallo-Perin P, et al. Incidence of type 1 and type 2 diabetes in adults aged 30-49 years: the population-based registry in the province of Turin, Italy. Diabetes Care 2005;28:2613-2619.

16. Alonso A, Logroscino G, Hernan MA. Smoking and the risk of amyotrophic lateral sclerosis: a systematic review and meta-analysis. J Neurol Neurosurg Psychiatry 2010;81:1249-1252.

17. Kioumourtzoglou MA, Rotem RS, Seals RM, Gredal O, Hansen J, Weisskopf MG. Diabetes Mellitus, Obesity, and Diagnosis of Amyotrophic Lateral Sclerosis: A Population-Based Study. JAMA Neurol 2015;72:905-911.

18. Dupuis L, Corcia P, Fergani A, et al. Dyslipidemia is a protective factor in amyotrophic lateral sclerosis. Neurology 2008;70:1004-1009.

19. Sutedja NA, van der Schouw YT, Fischer K, et al. Beneficial vascular risk profile is associated with amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry 2011;82:638-642.

20. Ambatipudi S, Cuenin C, Hernandez-Vargas H, et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. Epigenomics 2016;8:599-618.

21. Frank SA. Evolution in health and medicine Sackler colloquium: Somatic evolutionary genomics: mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. Proc Natl Acad Sci U S A 2010;107 Suppl 1:1725-1730.

22. Lattante S, Ciura S, Rouleau GA, Kabashi E. Defining the genetic connection linking amyotrophic lateral sclerosis (ALS) with frontotemporal dementia (FTD). Trends Genet 2015;31:263-273.

23. Cady J, Allred P, Bali T, et al. Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. Ann Neurol 2015;77:100-113.

24. van Blitterswijk M, van Es MA, Hennekam EA, et al. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. Hum Mol Genet 2012;21:3776-3784.

25. Chio A, Calvo A, Moglia C, et al. ATXN2 polyQ intermediate repeats are a modifier of ALS survival. Neurology 2015;84:251-258.

26. Diekstra FP, van Vught PW, van Rheenen W, et al. UNC13A is a modifier of survival in amyotrophic lateral sclerosis. Neurobiol Aging 2012;33:630 e633-638.

27. Fogh I, Lin K, Tiloca C, et al. Association of a Locus in the CAMTA1 Gene With Survival in Patients With Sporadic Amyotrophic Lateral Sclerosis. JAMA Neurol 2016;73:812-820.

28. Cronin S, Greenway MJ, Prehn JH, Hardiman O. Paraoxonase promoter and intronic variants modify risk of sporadic amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry 2007;78:984-986.

29. Siu J, Perkins E, Cashman NR. Effects of sex and family history on the amyotrophic lateral sclerosis (ALS) multistep model. Amyotroph Lateral Scler 2016;17:94.

30. Murphy NA, Arthur KC, Tienari PJ, Houlden H, Chio A, Traynor BJ. Age-related penetrance of the C9orf72 repeat expansion. Sci Rep 2017;7:2116.

31. Van Langenhove T, van der Zee J, Gijselinck I, et al. Distinct clinical characteristics of C9orf72 expansion carriers compared with GRN, MAPT, and nonmutation carriers in a Flanders-Belgian FTLD cohort. JAMA Neurol 2013;70:365-373.

32. Gijselinck I, Van Mossevelde S, van der Zee J, et al. The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. Mol Psychiatry 2016;21:1112-1124.

33. Kramer NJ, Carlomagno Y, Zhang YJ, et al. Spt4 selectively regulates the expression of C9orf72 sense and antisense mutant transcripts. Science 2016;353:708-712.

34. Chio A, Borghero G, Pugliatti M, et al. Large proportion of amyotrophic lateral sclerosis cases in Sardinia due to a single founder mutation of the TARDBP gene. Arch Neurol 2011;68:594-598.

35. Aggarwal A, Nicholson G. Age dependent penetrance of three different superoxide dismutase 1 (sod 1) mutations. Int J Neurosci 2005;115:1119-1130.

36. Cudkowicz ME, McKenna-Yasek D, Sapp PE, et al. Epidemiology of mutations in superoxide dismutase in amyotrophic lateral sclerosis. Ann Neurol 1997;41:210-221.

37. Andersen PM, Nilsson P, Ala-Hurula V, et al. Amyotrophic lateral sclerosis associated with homozygosity for an Asp90Ala mutation in CuZn-superoxide dismutase. Nat Genet 1995;10:61-66.

38. Robberecht W, Aguirre T, Van den Bosch L, Tilkin P, Cassiman JJ, Matthijs G. D90A heterozygosity in the SOD1 gene is associated with familial and apparently sporadic amyotrophic lateral sclerosis. Neurology 1996;47:1336-1339.

39. Elert E. Aetiology: Searching for schizophrenia's roots. Nature 2014;508:S2-3.

Funding/Support. This is an EU Joint Programme–Neurodegenerative Disease Research (JPND) project. The project is supported through the following funding organizations under the aegis of JPND: UK, Medical Research Council (MR/L501529/1; MR/R024804/1) and Economic and Social Research Council (ES/L008238/1); Ireland, Health Research Board; Netherlands, ZonMw; Italy, Ministry of Health and Ministry of Education, University and Research. The work leading up to this publication was funded by the European Community's Health Seventh Framework Programme (FP7/2007–2013; grant agreement number 259867). OH and JR are funded by grants from the Irish Health Research Board, and by the charity Research Motor Neurone. AAC receives salary support from the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre.

Role of the Funder/Sponsor. The funding sources had no role in the design and conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Table 1. Comparison of ALS patients who underwent genetic assessment vs. patients who were not assessed

	Tested patients (n=1077)	Non-tested patients (n=232)	p
Gender (female, %)	508 (47.3%)	101 (43.2%)	0.28
Site of onset (bulbar, %)	348 (32.4%)	106 (45.3%)	<0.001
Mean age at onset (years, SD)	65.8 (11.0)	70.9 (10.0)	<0.001
Family history of ALS (n, %)	109 (10.1%)*	2 (0.9%)	<0.001
Multidisciplinary ALS center (n and % of genetically tested)	1030 (93.8%)	70 (6.5%)	<0.001

* 39 patients (3.6% of those tested) carried a mutation but did not report a family history for ALS. They are not included in the count of those with a family history.

Table 2. List of missense genetic mutations of *SOD1, TARDBP* and *FUS* genes

Gene	**Mutation**	**Family history for ALS or FTD**
C9orf72 (n=74)	74 cases with GGGGCC expansion*	Yes = 47; No = 27
SOD1 (n=20)	Leu145Phe (6 cases)	Yes = 5; No = 1
	GLy94Asp (4 cases)	Yes = 4
	Asp91Ala heterozygous (2 cases)	No = 2
	Asp110Tyr (2 cases)	No = 2
	Asn20Ser (1 case)	No
	Leu39Val (1 case)	Yes
	Val48Phe (1 case)	Yes
	Asn66Ser (1 case)	No
	Gly73Ser (1 case)	No
	Asp91Asn (1 case)	No
TARDBP (n=15)	Ala382Thr (8 cases)	Yes = 4; No = 4
	Asn276Ser (2 cases)	No = 2
	Asn390Ser (2 cases)*	Yes = 1 ; No = 1
	Ser393Leu (2 cases)	Yes = 1; No = 2
	Gly368Ser (1 case)	No
FUS (n=3)	Arg495X (1 case)	De novo mutation
	Arg514Ser (1 case)	Yes
	Gln519Glu (1 case)	De novo mutation

* 1 patient carried both a *C9orf72* GGGGCC expansion and the p.N390S missense mutation of the *TARDBP* gene

Table 3. Comparison of log incidence vs. log age for different ALS subgroups. Only the 1077 genetically tested cases are included

	Number included	r^2	*n*-1 (95% CI)	*n* steps
All tested cases	1077	0.98	4.652 (4.368-4.950)	6
All non-familial non-mutated cases	921	0.98	5.001 (4.681-5.341)	6
All familial cases	109	0.92	2.945 (2.430-3.566)	4
All familial cases negative for tested genes	45	0.95	3.710 (2.674-4.532)	5
SOD1 mutated	20	0.53	0.758 (0.463-1.167	2
*C9orf72** mutated	74	0.94	2.215 (1.738-2.791	3
*TARDBP** mutated	14	0.93	3.241 (2.209-4.131)	4

*1 case with both *C9orf72* and *TARDBP* mutation is counted for both genes

Figure legends

Figure 1. **Slope estimation for all ALS patients.**

A. Log incidence vs. log age for all incident ALS patients who have been genetically-tested (n=1077) (y=4.65x -7.60; r^2=0.98).

B. Log incidence vs. log age for all incident ALS patients in the register (n=1309) (y = 4.83x– 7.85; r^2= 0.99).

Figure 2. **Slope estimation for all those with familial ALS.** Log incidence vs. log age for all incident familial ALS patients (n=111) (y=2.95x -5.45; r^2=0.92).

Figure 3. Slope estimation for all ALS patients carrying a mutation in one of four tested genes. Log incidence vs log age for *C9orf72* ALS (74 cases) (y=2.22x – 4.33, r^2=0.94) (red line) for *SOD1* ALS (20 cases) (y=0.758x – 2.43, r^2=0.53) (green line) and for *TARDBP* ALS (15 cases) (y=3.24x – 6.76, r^2=0.93) (blue line). The fit to a straight line is good, consistent with a multistep model.

Figure 4. **Slope estimation for patients with type 1 and type 2 diabetes mellitus.** Data from Piemonte diabetes register.[14] Log incidence vs. log age for type 1 diabetes mellitus (y=0.96x – 0.66; r^2=0.98) (red line) and type 2 diabetes mellitus (y=5.28x – 6.78; r^2=1.0) (blue line).