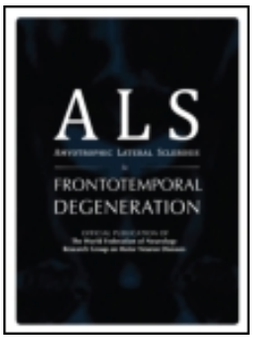

Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration

ISSN: 2167-8421 (Print) 2167-9223 (Online) Journal homepage: informahealthcare.com/journals/iafd20

Genetic and epigenetic studies of amyotrophic lateral sclerosis

Ammar Al-Chalabi, Shin Kwak, Mark Mehler, Guy Rouleau, Teepu Siddique, Michael Strong & Peter Nigel Leigh

To cite this article: Ammar Al-Chalabi, Shin Kwak, Mark Mehler, Guy Rouleau, Teepu Siddique, Michael Strong & Peter Nigel Leigh (2013) Genetic and epigenetic studies of amyotrophic lateral sclerosis, Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 14:sup1, 44-52, DOI: 10.3109/21678421.2013.778571

To link to this article: https://doi.org/10.3109/21678421.2013.778571

Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 2013; 14(Suppl. 1): 44–52

informa
healthcare

ORIGINAL ARTICLE

Genetic and epigenetic studies of amyotrophic lateral sclerosis

AMMAR AL-CHALABI[1], SHIN KWAK[2], MARK MEHLER[3], GUY ROULEAU[4], TEEPU SIDDIQUE[5], MICHAEL STRONG[6] & PETER NIGEL LEIGH[7]

[1]*King's College London, Institute of Psychiatry, Department of Clinical Neuroscience, London, UK,* [2]*Department of Neurology, Center for Disease Biology and Integrative Medicine, The University of Tokyo, Tokyo, Japan,* [3]*Departments of Neurology, Neuroscience and Psychiatry, and Behavioral Sciences, Albert Einstein College of Medicine, Bronx, New York, USA,* [4]*Montreal Neurological Institute and Hospital, Department of Neurology and Neurosurgery, McGill University, Canada,* [5]*Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA,* [6]*Department of Clinical Neurological Sciences, Western University, London, Ontario, Canada, and* [7]*Department of Neurology, Brighton and Sussex Medical School, University of Sussex, East Sussex, UK*

Abstract

The identification of genetic and epigenetic factors that are associated with an increased risk of developing amyotrophic lateral sclerosis (ALS), or that modify the age of onset or rate of progression, requires a multimodal research strategy, facilitated through international collaboration. The discovery of several ALS genes strongly linked to RNA biology, the proteasome pathway, and axonal transport suggest they have an important role in pathogenesis, but the immense complexity of these processes is also apparent. The increasing rate of genetic discoveries brings the hope of designing more targeted and efficacious therapies.

Key words: *ALS, genetic, epigenetic, familial ALS, modifier genes*

Introduction

Amyotrophic lateral sclerosis (ALS) is a devastating neurological disease in which neurodegeneration, predominantly of motor neurons, leads to progressive paralysis of voluntary muscles until death follows, usually from respiratory failure, typically within about five years of symptom onset. Research into the causes of ALS is essential for the rational design of therapies. A simple view is that there can only be two causes of ALS: genetic variation and environmental exposures. A third possibility is also important to consider and is known as epigenetic variation: environmental exposures, random processes or genetic variation result in changes to DNA, principally, but not exclusively, by methylation. The study of epigenetic factors is in its infancy but making rapid progress. Environmental factors are important and the subject of epidemiological studies, but have the difficulty that an environmental factor may have acted well before the onset of symptoms and could therefore be difficult to capture accurately. Genetic studies are feasible and, with the rapid advances in genetic technology, becoming easier.

It is known that ALS has a genetic component because a family history of ALS is reported by about 5% of patients (1); twin studies show that the heritability is about 60% (2), and disease-causing mutations have been identified in up to 60% of those with a family history, and between 5% and 15% of those without to date (3).

One problem is that genes described as ALS genes may produce atypical phenotypes including very young age of onset, very slow progression, or purely upper or lower motor neuron signs, while other genes regarded as typical ALS genes may give rise to ALS with frontotemporal dementia (ALS-FTD). Another issue is the distinction made by many between familial ALS (FALS) and apparently sporadic ALS (SALS), because this is useful when gene-hunting. However, FALS is not well defined by the research community (4); the identification of another affected individual depends on the recall and knowledge of the patient, and there is a major statistical bias in whether a large effect gene will affect only one person in a pedigree, dependent on the family size and how penetrant the disease

Correspondence: A. Al-Chalabi, King's College London, Institute of Psychiatry P 041, London SE5 8AF, UK. Fax: 44 20 7848 5190. E-mail: ammar.al-chalabi@kcl.ac.uk

ISSN 2167-8421 print/ISSN 2167-9223 online
DOI: 10.3109/21678421.2013.778571

gene is (5). Thus, the distinction is artificial and it is therefore expected that genes mutated in FALS should also be found to cause a proportion of what is called SALS.

ALS genes

How to find ALS genes

The frequency of a genetic variation is a measure of how large an effect it has on a disease. This is because when a mutation first arises, if it has a very deleterious effect it will be strongly selected against and is unlikely to rise to a high frequency. If it has a neutral or small effect it can become more frequent by chance. A direct consequence of this is that strongly deleterious genes, which may be sufficient to cause ALS alone, are more likely to be found in a few families only, while small effect genes, which only increase the risk of ALS a little, need to act in concert and may therefore lead to familial clustering without necessarily resulting in a strong family history.

There are two broad strategies for identifying disease genes – linkage studies and association studies. In linkage analysis, a statistical test is performed in a family to determine if a genetic variant occurs in affected individuals more often than would be expected by chance. If it does, this is called linkage (6). In association analysis, the frequency of a genetic variant is compared between a group of unrelated affected individuals and a group of unrelated controls. If the frequency differs between the groups then the variant shows association with disease (7). Thus, linkage studies are useful in pedigrees and therefore in finding rare, highly deleterious genes, while association studies are useful in large groups of unrelated individuals.

Linkage studies use an indirect approach, employing known genetic variants as markers for the true disease-causing gene variant, and produce a robust estimate of the likely location of a disease gene (8). The resolution of linkage is low, so the estimate of the location may be very many megabases long. On the other hand, association studies are prone to false positive findings but produce an estimate of the true disease gene location that is precise and often just a few kilobases in length.

Association studies historically used direct sequencing of genes thought to be good candidates for disease causation, and the general burden of variants in cases compared with controls. With advances in genetic technology, an indirect method that interrogates the entire genome became possible in which hundreds of thousands of genetic markers could be tested in thousands of individuals. This is known as a genome-wide association study (GWAS). It is now becoming possible to directly assay all genes by sequencing, so-called whole genome or whole exome sequencing.

Genes causing ALS

A database of ALS genes (ALSoD) (9) with their associated phenotypes and a measure of their credibility is available at the World Federation of Neurology and European Network for the Cure of ALS (ENCALS) project website http://alsod.iop.kcl.ac.uk, generously supported by ALSA, ALS Therapy Alliance, MNDA and ALS Canada.

Familial ALS genes, focusing on UBQLN2

The genes most widely accepted as causing typical ALS when mutated are *SOD1* (10), *TARDBP* (coding for TDP-43) (11–13), and *FUS* (14,15). Genes that may have an effect confined to specific populations are *ANG* (16) and *OPTN* (17). Genes for ALS and ALS-FTD include *UBQLN2* (18) and a gene identified within a few days of this meeting, *C9orf72* (19,20). The *C9orf72* gene causes ALS through a massive expansion of a hexanucleotide repeat motif between non-coding exons 1a and 1b $(GGGCC)_n$. In unaffected individuals, the sequence is usually repeated just once or twice, although repeat sizes up to the 20s and 30s have been recorded. In ALS or ALS-FTD the expansion is hundreds of repeats in length. The mechanism by which this results in disease is unknown, and it is not obvious if the gene itself is relevant or the expansion would be pathological if inserted into other genes. The *C9orf72* gene expansion probably comes from a single founder, but it is also possible that the haplotype on which it occurs is inherently unstable and results in spontaneous expansion. Because it is a relatively common cause of ALS, and is found in apparently sporadic ALS as well, mutation in *C9orf72* is a landmark discovery of great clinical importance.

The *UBQLN2* gene is of particular interest because it is involved in the ubiquitin pathway, and aberrant ubiquitination is a common feature of all pathological protein inclusions in ALS. A family with ALS showing X-linked inheritance was studied by linkage, and a locus identified in the pericentromeric region (18). A mutation was found in the *UBQLN2* gene that segregated with disease. Testing other families, including families with ALS-FTD, identified mutations in four other pedigrees; these mutations all occurred in the same part of the gene.

The existing major ALS genes are associated with distinct pathology, unrelated to TDP-43 or FUS (*SOD1*), related to FUS (*FUS)* or related to TDP-43 (*TARDBP,* other genes and ALS without an identified genetic basis). Interestingly, antibodies to UBQLN2 show that it colocalizes in ALS in inclusions containing TDP-43, FUS and OPTN. The pathology of individuals with ALS caused by *UBQLN2* mutation shows a novel aggregate in radial neurites with a membrane bound inclusion in the hippocampus. In patients with dementia or ALS not caused by *UBQLN2* mutation, UBQLN2 aggregates are still found, but TDP-43 aggregates are not always

present, suggesting that the UBQLN2 pathway represents a more unifying process than that of TDP-43.

Thus, mutations in *UBQLN2* cause both ALS and dementia, in adults and juvenile patients, with a novel pathology extending to sporadic ALS and ALS dementia and other forms of non-ubiquilin familial diseases, even when TDP-43 pathology is not present.

Genes for apparently sporadic ALS

Several non-overlapping GWAS have been conducted for SALS, the first reporting in 2007 (21–32), including genome-wide association studies of copy number variation (33–36). The early studies failed to find replicable associations, even with quite large sample sizes. Furthermore, as sample sizes increase, the effect of any genetic variant identified tends to be smaller, which some would argue makes it of less relevance. Thus, it is perhaps not unreasonable to conclude that GWAS has failed in ALS. However, a large scale GWAS identified two susceptibility loci for SALS; one on chromosome 19p13.3 that maps to a haplotype block within the boundaries of the *UNC13A* (unc-13 homolog A) gene and one on chromosome 9p21.2 (28). The second locus was of particular interest since it coincided with a known linkage region for ALS-FTD and was therefore a more robust finding. It was then replicated in two further GWAS studies, one in UK samples and one in Finnish samples (29,30). Thus, the weakness of association studies, that they are prone to false positive results, was not at play here, and there was only the benefit of association, that the disease gene location is more precisely known. The disease gene was identified within a few months of the publications and a few days of this conference, as *C9orf72* (19,20), now known to be a moderate effect gene, involved in up to 60% of FALS and up to 15% of SALS (37,38), varying considerably by population.

Three other genes are of note here: *UNC13A*, *ELP3*, and *ATXN2* (28,32,39). The first two were found by GWAS with corroborating evidence, and implicate disease pathways already thought to be involved in ALS. The ALS-associated SNP within *UNC13A* is, by meta-analysis, the most strongly associated ALS genetic variant known. It is of interest because the ubiquitin pathway is implicated in ALS by pathology, mutation in the *SQSTM1* (p62) gene, and the findings of the UBQLN2 mutations mentioned above. The genetic variants associated with ALS in the *ELP3* gene are of interest because they are microsatellite alleles rather than SNPs. ELP3 is a component of RNA polymerase II and RNA processing is emerging as a common theme in ALS pathogenesis; mutagenesis experiments in flies have shown that ELP3 is essential for correct neuronal development.

Recently, following a genetic screen in yeast, ataxin 2 protein (coded for by *ATXN2*) was found to modulate TDP-43 toxicity. Because of the central role played by TDP-43 in several neurodegenerative diseases including ALS, it then became the subject of a genetic study. ATXN2 and TDP-43 interact in an RNA-dependent complex, and both become mislocalized in spinal cord neurons of patients with ALS.

The ATXN2 (12q24.1) polyQ tract shows a normal size range that extends between 14 and 31 repeats, 22 and 23 repeats being the most frequent (40). Expansions over 34 repeats are associated with the development of spinocerebellar ataxia type 2 (SCA2), an autosomal dominant disorder characterized by progressive cerebellar gait and limb ataxia, slow saccadic eye movements, supranuclear ophthalmoplegia, and hyporeflexia (41), and there have been a few case reports of motor neuron degeneration associated with SCA2. Alleles with 32 and 33 repeats are also known to cause SCA2 but these are associated with exceptionally late onset disease (42). The genetic study found that intermediate repeat lengths predispose to ALS, a finding now replicated in several other studies.

Thus, there are now two repeat sequences definitively associated with ALS: the trinucleotide repeat in ATXN2, and the hexanucleotide repeat in C9orf72, suggesting that repeat variation may be a common theme with more to be found in the future.

Genes that modify the phenotype of ALS

Although genes for susceptibility to ALS will reveal the pathway leading to neurodegeneration, other factors such as those that modify the age of onset, the site of first symptoms, the burden of upper and lower motor neuron disease, and the rate of disease progression are also very important. Modifying the rate of disease progression is the goal of all therapeutic trials in ALS, and finding gene variants that have the same effect would have profound implications for the design of rational therapies. Similarly, factors that raise the age of onset would be useful in delaying disease in gene carriers. It is likely that there is considerable overlap between the effects of different modifying factors, as well as between modifier genes and susceptibility genes. For example, genes that act to reduce the age of onset will have a similar effect to susceptibility genes, and some susceptibility genes are known to be associated with specific patterns of upper or lower motor neuron disease, or rates of disease progression.

Modifiers of survival

Some mutations confer a survival effect in a predictable manner. This is best characterized for SOD1 mutations. The A4V, G41S, G93A and R115G variants of SOD1 have consistently poor survival while D90A homozygosity, G41D and H46R have a much slower course than average (3). Homozygosity or compound heterozygosity for ALS-causing mutations is also associated with a more rapid disease course.

The *UNC13A* gene, in which a SNP confers susceptibility to ALS, also appears to modulate survival, with the risk allele leading to shorter survival (43).

A GWAS that used a prevalent clinic based sample identified variation in the *KIFAP3* gene associated with survival, confirmed in multiple populations within the study. The effect was similar to that of riluzole (27). A subsequent GWAS that used an incident population based sample did not replicate the original finding, but this may be a result of the younger, generally longer lived population that is seen in prevalence rather than incidence cohorts (44). Primary lateral sclerosis (PLS) has also been associated with the same variant in *KIFAP3* and it is conceivable that since PLS has a generally longer survival than ALS, different inclusion criteria between the studies account for the different results (45).

Modifiers of age of onset

A study of heritability for age of ALS onset that examined the age at which members of pedigrees with two different SOD1 mutations (A4V or D90A) were affected, found heritability to be 0.29 (0–0.42), suggesting that genetic factors may play a role (46). In other words, the implication is that there are genes that increase the risk of developing ALS (susceptibility genes) and other genes that only modify phenotype without apparently increasing risk.

Homozygosity for SOD1 mutations or compound heterozygosity for ALS-causing mutations is associated with younger onset age. SOD1 mutations specifically associated with a young onset age include I104F, G37R, L38V and G114A. *APOE* genotype is reported to modify age of onset in ALS (47).

No modifiers of limb or bulbar onset have yet been found. Similarly, in ALS-FTD, no modifier that determines ALS rather than FTD has yet been found, although international efforts are underway to explore this.

Despite the increasing success in identifying genes that underlie familial ALS (FALS), which results from penetrant monogenic mutations, the etiology of most cases of ALS (sporadic ALS or SALS) remains unknown, and more complex genetic models may need to be considered. Such models could, for example, take into account mutations in several genes, possibly with variable penetrance, with additive effects in the presence or absence of environmental factors. Therefore, variations in a modifier gene could affect disease parameters such as age of onset, duration and site of onset. Several modifier genes have been proposed for ALS, including *SMN1*, the paraoxonase cluster (*PON1-PON3*), *VEGF*, *ANG* and others (48–51).

RNA processing, epigenetics and ALS

Non-coding RNA, RNA networks and epigenetics

Epigenetics is a new genomic science that involves DNA methylation, histone, nucleosome and chromatin remodeling, non-coding RNAs, RNA editing and DNA recoding. These processes work together and regulate expression of the genome. Epigenetics is a mechanism through which gene-environment interactions are mediated.

RNA is the 'sister' molecule of DNA. Although about 2% of the genome is transcribed into mRNA that encodes proteins, the rest is not. Much of it is transcribed into RNA that is thought to have sophisticated regulatory functions; up to 98% of the genome is non-coding RNAs, which are frequently transcribed, functional, and often biologically context-specific. Our knowledge of this area is still limited, but the discovery of many different functional subclasses of small and long RNAs and the ongoing studies of their function mean this is a very fruitful direction of research.

There is a rapidly evolving body of evidence that ALS is associated with complex and nuanced alterations in RNA regulatory networks (52,53). These include processes associated with transcriptional regulation, post-transcriptional processing, deployment and function of different subclasses of non-coding RNAs, local control of translation and additional epigenetic processes. Post-transcriptional processing include pre-RNA splicing; RNA editing; RNA stabilization, degradation, quality control, nuclear export, intracellular transport, and intercellular trafficking (54). Epigenetic processes include DNA methylation, histone, nucleosome and higher-order chromatin remodeling (55). The epigenome differs from the genome in that every cell has its own unique epigenome and this changes continuously as the environment interacts with it. Three common methods to examine the epigenome are: direct RNA sequencing to examine different RNA subclasses, species and splice variants; chip sequencing in which antibodies are used to immunoprecipitate a complex of interest and the DNA it contains is sequenced to examine histone, nucleosome and higher order chromatin remodeling factors and associated DNA methylation cofactors; and the third method is to examine DNA methylation using bisulfite sequencing or methylation-sensitive endonucleases.

The non-coding RNAs, which include microRNAs and so-called long non-coding RNAs, exert their dynamic, versatile and environmentally-responsive local as well as long-distance genomic effects, in large part, through genomic site-specific targeting of more generic DNA methylation, histone and chromatin modifying proteins and associated epigenetic regulatory complexes and epigenetic remodeling and reprogramming modules (56,57). Associated non-coding RNA regulatory networks have preferential roles in nervous system development, homeostasis, plasticity, connectivity and aging (58).

RNA is increasingly being recognized as a universal biosensor of environmental, interoceptive and homeostatic signals, as a transducer of biological processes utilizing a fraction of the bioenergetic

costs of equivalent protein species and as a unique signaling molecule. It is able to function as an integral component of both digital, sequence-specific DNA/RNA codes and analogue, conformational protein-based codes as well as a reservoir and interpreter of embedded genetic information. This is because RNA has a digital code in the sequence of nucleotides, so it can interact with DNA in a digital manner, and also has an analogue code in having a tertiary structure and therefore can interact with proteins in an analogue manner.

The ability of RNA to participate in seminal biological functions at a molecular, genetic and cellular systems level, endows it with a uniquely broad range of complex and nuanced functional properties, including complex epigenetic memory states.

Alterations in RNA processing are implicated in ALS/MND pathogenesis through many pathways, including defects in axonal transport and cytoskeletal stabilization, DNA replication and repair, mitochondrial dysfunction, connectivity, protein degradation and stress responses (59–61). The field of RNA biology, in general, and RNA regulatory networks, in particular, represents an important, challenging and rapidly evolving scientific discipline for helping to decipher the molecular pathogenesis of ALS (62).

A number of issues remain unresolved. What are the roles of long non-coding RNAs in ALS pathogenesis? The main roles appear to be regulating adjacent protein coding genes as well as site-specific targeting of multiple genes throughout the genome that form functional gene networks. What are the roles of newly identified non-coding RNAs transcribed from the mitochondrial genome? What are the roles of alterations in developmental processes in the selective vulnerability of the motor unit? What are the implications of abnormal RNA biology in the identification of novel molecular targets and the design of innovative therapeutic strategies?

RNA editing

A feature of some RNA transcripts is that they undergo RNA editing. This is a process in which the nucleotide sequence is changed enzymatically resulting in a different sequence from that encoded by the original DNA. RNA editing occurs most efficiently in the central nervous system and the commonest change in mammals is deamination from adenosine to inosine. When editing occurs in the coding regions of mRNA, this may change the amino acid sequence from that encoded by the gene.

Glutamate receptors come in a variety of types. The AMPA receptor is a non-NMDA ionotropic receptor mediating fast synaptic neurotransmission. In vertebrate AMPA receptor proteins a section of the pore lining domain has two alternative structures with either a glutamine (Q) or arginine (R) amino acid. This variation, known as a Q/R site, has a major influence on the receptor properties. The R containing variant of the mammalian AMPA receptor subunit is not coded by the DNA sequence, but is created by post-transcriptional RNA editing activities.

Motor neurons express abnormal Q/R site-unedited GluA2 or GluR2 mRNA in sporadic ALS patients in a disease-specific and motor neuron-selective manner (63–65). GluA2 plays a regulatory role in Ca^{2+} permeability of the AMPA receptor after conversion of adenosine for the Q/R site to inosine (A-to-I conversion) by the RNA editing enzyme called adenosine deaminase acting on RNA 2 (ADAR2).

In AR2 mice in which the *ADAR2* gene is conditionally knocked out in motor neurons, the motor neurons undergo slow death by failure to edit the GluA2 Q/R site (66). Notably, the ADAR2-lacking motor neurons are rescued when they are genetically engineered to express only edited GluA2 in AR2 mice. These lines of evidence suggest that complete editing of GluA2 Q/R site is necessary for neuronal survival and that expression of unedited GluA2 resulting from the reduction of ADAR2 activity may be a cause of death of motor neurons in sporadic ALS. Together with the finding that ADAR2 expression is always reduced in ALS motor neurons that exhibit TDP-43 pathology (67), a hallmark of sporadic ALS, one hypothesis is that there is a progressive down-regulation of ADAR2 activity in ALS motor neurons, in which the pathological process commences once unedited GluA2 is expressed. Normalization of GluA2 RNA editing in motor neurons may therefore be a therapeutic strategy for sporadic ALS (68).

Next steps

Every method for gene discovery in ALS has limitations. The problem comes down to the relationship between frequency and effect size. A large effect gene is rare and relevant to a few people, while a small effect gene is common and therefore of importance for ALS in general, but of such small impact that the therapeutic potential is more limited. Moderate effect genes with a penetrance in the range of 20–40% would be ideal targets in being relevant to a significant proportion of people and having a relevant effect size, and such variants could be detected with either linkage or association studies.

For GWAS, the problem is that several thousand well phenotyped cases are needed to find genetic factors of modest effect. Any genetic heterogeneity in SALS might limit the identification of causative alleles, since different disease-causing alleles may exist within the same gene in different people. If this is the case, the strength of association signals for that gene will be diluted. Although rare highly penetrant mutations are expected to cause a proportion

of SALS, the identification of such variations by GWAS will inherently be more difficult since the variant needs to be frequent enough for detection.

For linkage, the problem is that large well phenotyped pedigrees are needed to find genetic causes of limited relevance since in most cases the variant will be of high penetrance and therefore rare. Since ALS is generally a disease of older age groups, atypical ALS affecting younger people or progressing more slowly than normal is more commonly studied.

For whole exome or whole genome sequencing, the problem is that huge datasets are generated, consisting of petabytes of data, and very many genetic variants are identified in each person without necessarily a simple way to tell which ones are relevant to ALS.

Nevertheless, each of these approaches has advantages too. GWAS with larger sample sizes is likely to uncover many more genes and therefore increase our understanding of the network of problems causing ALS. Linkage techniques are improving and generate robust results. Whole genome and whole exome sequencing will become easier as computing and statistical methods catch up with genetic technology. Each genetic discovery provides another view of the pathway to neurodegeneration that we hope will at some point come into sharp focus, allowing the design of targeted treatments and possibly personalized medicine.

Controversies and discussion points

There remains a strong argument in favour of simple candidate gene studies, focused on a single gene. The model is one in which biology is used to identify the gene of interest and genetic variants are then sought. This is the technique that led to the identification of *ATXN2*, and combined with linkage, led to the identification of *TARDBP*.

Similarly, combining high throughput, whole exome sequencing with association studies examining the burden of genetic variants in each gene in cases and controls allows a hybrid study in which a non-biologically based approach can still identify genes in which multiple alleles in a single gene contribute to SALS. For the reasons described above, such a scenario is likely to be less common than rare variants in families and common variants in SALS, but it is still likely to occur and will not be easily detected by other methods.

A key problem faced by researchers is that innovative ideas may be seen as high-risk projects in terms of funding, even by agencies that express a wish to consider innovative ideas. Most funders require a precedent, or evidence to directly support a hypothesis. For example, there are about 500 kinases in the genome and, even though studies show that all are important, only a few have ever been studied, possibly because it is not easy to obtain funding for research into kinases that have no existing data. There may be other methods that would find other types of genetic variation, but currently these are hidden.

A major issue is in establishing the pathogenic mechanism of ALS in pathways revealed by genetic studies. For example, is TDP-43 protein toxic when accumulated, or is there a loss of function caused by mislocalization away from the nucleus? These ideas are critical for the design of rational therapies, since it may or may not be necessary to remove the inclusions seen pathologically in ALS. One view is that these are not mutually exclusive since it is quite possible that a protein develops a novel toxic function for example by accumulating, and in being removed by the aggregate also results in a loss of function. The aggregates themselves may be toxic, may be simple gravestone-type passive markers, may be protective and part of cell defence, or may be aggregating some other protein.

Another key issue is whether future investment should be targeted towards genetic or environmental studies of ALS. The view of the geneticists in the room was that the genetic question represents a finite space and therefore a tractable question, whereas the environment is not only an infinite research space, but also impossible to address in the same detail and with the same reliability as a genome. Nevertheless, it is meaningless to understand genes without the context of the environment in which they act. Another issue is that it is possible to examine the genome without a prior hypothesis using linkage, GWAS or whole genome sequencing, but it is difficult to examine all environmental factors in the same way, so any environmental study is potentially prone to bias in the selection of factors to examine. If clusters of cases could be mapped geographically with a common factor identified, or some other such environmental clue found without a prior hypothesis, then such bias would disappear, but this is not straightforward. On the other hand, genetic research has been funded for about 20 years and can only explain about 15% of ALS, and so perhaps the time has come for environmental studies to be given priority. A solution is to study gene-environment interactions (although these require huge sample sizes), or genetic epidemiology, or to devote some resources to developing new disease models, e.g. examining the initial regional CNS nidus and subsequent spread of ALS, or cell models that might highlight the multiple environmental impacts that could trigger degeneration.

The identification of genetic factors runs into an issue to do with frequency and effect size, as discussed above. Rare genes have small impact on the disease numerically because they only affect a few people, but have a large effect biologically. On the other hand, small effect common genes may impact many people but only contribute a little individually. A gene of moderate effect might have the 'best' of both worlds. The main reason to identify genes is

therefore not to design a specific gene therapy but to understand the biology of ALS, and so to design a rational therapy.

Acknowledgements

We thank the Motor Neurone Disease Association of Great Britain and Northern Ireland, the ALS Association, the Les Turner ALS Foundation, the ALS Society of Canada, and the Canadian Institutes of Health Research (CIHR) for support.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

The research leading to these results has received funding from the European Community 's Health Seventh Framework Programme. AAC receives salary support from the National Institute for Health Research (NIHR) Dementia Biomedical Research Unit at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author and not necessarily those of the NHS, the NIHR or the Department of Health. GAR holds the Canada Research Chair and a Jeanne-et-J-Louis-Levesque Chair for the Genetics of Brain Diseases.

References

1. Byrne S, Walsh C, Lynch C, Bede P, Elamin M, Kenna K, et al. Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis. Journal of Neurology, Neurosurgery, and Psychiatry. 2011;82:623–7.
2. Al-Chalabi A, Fang F, Hanby MF, Leigh PN, Shaw CE, Ye W, et al. An estimate of amyotrophic lateral sclerosis heritability using twin data. Journal of Neurology, Neurosurgery, and Psychiatry. 2010;81:1324–6.
3. Andersen PM, Al-Chalabi A. Clinical genetics of amyotrophic lateral sclerosis: what do we really know? Nature reviews. Neurology. 2011;7:603–15.
4. Byrne S, Bede P, Elamin M, Kenna K, Lynch C, McLaughlin R, et al. Proposed criteria for familial amyotrophic lateral sclerosis. Amyotrophic lateral sclerosis: official publication of the World Federation of Neurology Research Group on Motor Neuron Diseases. 2011;12: 157–9.
5. Al-Chalabi A, Lewis CM. Modelling the effects of penetrance and family size on rates of sporadic and familial disease. Human Heredity. 2011;71:281–8. Epub 2011/08/19.
6. Morton NE. Sequential tests for the detection of linkage. Am J Hum Genet. 1955;7:277–318.
7. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996;273:1516–7.
8. Ott J. Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. Am J Hum Genet. 1974;26:588–97.
9. Abel O, Powell JF, Andersen PM, Al-Chalabi A. ALSoD: a user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. Human Mutation. 2012. Epub 2012/07/04.
10. Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature. 1993;362:59–62.
11. Kabashi E, Valdmanis PN, Dion P, Spiegelman D, McConkey BJ, Velde CV, et al. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat Genet. 2008;40:572–4.
12. Van Deerlin VM, Leverenz JB, Bekris LM, Bird TD, Yuan W, Elman LB, et al. TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. The Lancet Neurology. 2008;7:409–16.
13. Sreedharan J, Blair IP, Tripathi VB, Hu X, Vance C, Rogelj B, et al. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science. 2008;319:1668–72.
14. Kwiatkowski TJ Jr, Bosco DA, LeClerc AL, Tamrazian E, Vanderburg CR, Russ C, et al. Mutations in the FUS/TLS Gene on Chromosome 16 Cause Familial Amyotrophic Lateral Sclerosis. Science. 2009;323:1205–8.
15. Vance C, Rogelj B, Hortobagyi T, de Vos KJ, Nishimura AL, Sreedharan J, et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science. 2009;323:1208–11.
16. Greenway MJ, Andersen PM, Russ C, Ennis S, Cashman S, Donaghy C, et al. ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. Nat Genet. 2006;38:411–3.
17. Maruyama H, Morino H, Ito H, Izumi Y, Kato H, Watanabe Y, et al. Mutations of optineurin in amyotrophic lateral sclerosis. Nature. 2010;465:223–6.
18. Deng HX, Chen W, Hong ST, Boycott KM, Gorrie GH, Siddique N, et al. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. Nature. 2011;477:211–5.
19. Renton AE, Majounie E, Waite A, Simon-Sanchez J, Rollinson S, Gibbs JR, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron. 2011;72:257–68.
20. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC hexanucleotide repeat in non-coding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. 2011;72:245–56.
21. Schymick JC, Scholz SW, Fung HC, Britton A, Arepalli S, Gibbs JR, et al. Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data. Lancet Neurology. 2007;6:322–8.
22. Dunckley T, Huentelman MJ, Craig DW, Pearson JV, Szelinger S, Joshipura K, et al. Whole genome analysis of sporadic amyotrophic lateral sclerosis. The New England Journal of Medicine. 2007;357:775–88.
23. Cronin S, Berger S, Ding J, Schymick JC, Washecka N, Hernandez DG, et al. A genome-wide association study of sporadic ALS in a homogenous Irish population. Hum Mol Genet. 2008;17:768–74.
24. van Es MA, van Vught PW, Blauw HM, Franke L, Saris CG, van den Bosch L, et al. Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis. Nat Genet. 2008;40:29–31.
25. Chio A, Schymick JC, Restagno G, Scholz SW, Lombardo F, Lai SL, et al. A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. Hum Mol Genet. 2009;18:1524–32.
26. van Es MA, van Vught PW, Blauw HM, Franke L, Saris CG, Andersen PM, et al. ITPR2 as a susceptibility gene in sporadic amyotrophic lateral sclerosis: a genome-wide association study. Lancet Neurology. 2007;6:869–77.
27. Landers JE, Melki J, Meininger V, Glass JD, van den Berg LH, van Es MA, et al. Reduced expression of the kinesin-associated protein 3 (KIFAP3) gene increases survival in sporadic amyotrophic lateral sclerosis. Proc Natl Acad Sci U S A. 2009;106:9004–9.
28. van Es MA, Veldink JH, Saris CG, Blauw HM, van Vught PW, Birve A, et al. Genome-wide association study identifies

19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. Nat Genet. 2009;41:1083–7.
29. Shatunov A, Mok K, Newhouse S, Weale ME, Smith B, Vance C, et al. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. Lancet Neurology. 2010; 9:986–94.
30. Laaksovirta H, Peuralinna T, Schymick JC, Scholz SW, Lai SL, Myllykangas L, et al. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. Lancet Neurology. 2010;9:978–85.
31. Kwee LC, Liu Y, Haynes C, Gibson JR, Stone A, Schichman SA, et al. A high-density genome-wide association screen of sporadic ALS in US veterans. PloS One. 2012;7:32768.
32. Simpson CL, Lemmens R, Miskiewicz K, Broom WJ, Hansen VK, van Vught PW, et al. Variants of the elongator protein 3 (ELP3) gene are associated with motor neuron degeneration. Hum Mol Genet. 2009;18:472–81.
33. Blauw HM, Al-Chalabi A, Andersen PM, van Vught PW, Diekstra FP, van Es MA, et al. A large genome scan for rare CNVs in amyotrophic lateral sclerosis. Hum Mol Genet. 2010;19:4091–9.
34. Wain LV, Pedroso I, Landers JE, Breen G, Shaw CE, Leigh PN, et al. The role of copy number variation in susceptibility to amyotrophic lateral sclerosis: genome-wide association study and comparison with published loci. PloS One. 2009;4:8175.
35. Cronin S, Blauw HM, Veldink JH, van Es MA, Ophoff RA, Bradley DG, et al. Analysis of genome-wide copy number variation in Irish and Dutch ALS populations. Hum Mol Genet. 2008;17:3392–8.
36. Blauw HM, Veldink JH, van Es MA, van Vught PW, Saris CG, van der Zwaag B, et al. Copy number variation in sporadic amyotrophic lateral sclerosis: a genome-wide screen. Lancet Neurology. 2008;7:319–26.
37. Smith BN, Newhouse S, Shatunov A, Vance C, Topp S, Johnson L, et al. The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. European Journal of Human Genetics: EJHG. 2012.
38. Majounie E, Renton AE, Mok K, Dopper EG, Waite A, Rollinson S, et al. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. Lancet Neurology. 2012;11:323–30.
39. Elden AC, Kim H-J, Hart MP, Chen-Plotkin AS, Johnson BS, Fang X, et al. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature. 2010;466:1069–75.
40. Pulst SM, Nechiporuk A, Nechiporuk T, Gispert S, Chen XN, Lopes-Cendes I, et al. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nat Genet. 1996;14:269–76.
41. Imbert G, Saudou F, Yvert G, Devys D, Trottier Y, Garnier JM, et al. Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. Nat Genet. 1996;14:285–91.
42. Cancel G, Durr A, Didierjean O, Imbert G, Burk K, Lezin A, et al. Molecular and clinical correlations in spinocerebellar ataxia 2: a study of 32 families. Hum Mol Genet. 1997;6:709–15.
43. Diekstra FP, van Vught PW, van Rheenen W, Koppers M, Pasterkamp RJ, van Es MA, et al. UNC13A is a modifier of survival in amyotrophic lateral sclerosis. Neurobiol Aging. 2012;33:630, e3–8.
44. Traynor BJ, Nalls M, Lai SL, Gibbs RJ, Schymick JC, Arepalli S, et al. Kinesin-associated protein 3 (KIFAP3) has no effect on survival in a population based cohort of ALS patients. Proc Natl Acad Sci U S A. 2010;107: 12335–8.
45. Orsetti V, Pegoraro E, Cima V, d'Ascenzo C, Palmieri A, Querin G, et al. Genetic variation in KIFAP3 is associated with an upper motor neuron-predominant phenotype in amyotrophic lateral sclerosis. Neurodegenerative Diseases. 2011;8:491–5.
46. Fogh I, Rijsdijk F, Andersen PM, Sham PC, Knight J, Neale B, et al. Age at onset in SOD1-mediated amyotrophic lateral sclerosis shows familiality. Neurogenetics. 2007; 8:235–6.
47. Zetterberg H, Jacobsson J, Rosengren L, Blennow K, Andersen PM. Association of APOE with age at onset of sporadic amyotrophic lateral sclerosis. J Neurol Sci. 2008; 273:67–9.
48. Lambrechts D, Storkebaum E, Morimoto M, del-Favero J, Desmet F, Marklund SL, et al. VEGF is a modifier of amyotrophic lateral sclerosis in mice and humans and protects motor neurons against ischemic death. Nat Genet. 2003;34:383–94.
49. Millecamps S, Salachas F, Cazeneuve C, Gordon P, Bricka B, Camuzat A, et al. SOD1, ANG, VAPB, TARDBP, and FUS mutations in familial amyotrophic lateral sclerosis: genotype-phenotype correlations. Journal of Medical Genetics. 2010;47:554–60.
50. Wills AM, Cronin S, Slowik A, Kasperaviciute D, van Es MA, Morahan JM, et al. A large-scale international meta-analysis of paraoxonase gene polymorphisms in sporadic ALS. Neurology. 2009;73:16–24.
51. Veldink JH, Kalmijn S, van der Hout AH, Lemmink HH, Groeneveld GJ, Lummen C, et al. SMN genotypes producing less SMN protein increase susceptibility to and severity of sporadic ALS. Neurology. 2005;65:820–5.
52. Kolb SJ, Sutton S, Schoenberg DR. RNA processing defects associated with diseases of the motor neuron. Muscle Nerve. 2010;41:5–17.
53. Strong MJ. The evidence for altered RNA metabolism in amyotrophic lateral sclerosis (ALS). J Neurol Sci. 2010; 288:1–12.
54. Lemmens R, Moore MJ, Al-Chalabi A, Brown RH Jr, Robberecht W. RNA metabolism and the pathogenesis of motor neuron diseases. Trends in Neurosciences. 2010; 33:249–58.
55. Mehler MF. Epigenetic principles and mechanisms underlying nervous system functions in health and disease. Progress in Neurobiology. 2008;86:305–41.
56. Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, et al. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature. 2008;454: 126–30.
57. Qureshi IA, Mattick JS, Mehler MF. Long non-coding RNAs in nervous system function and disease. Brain Research. 2010;1338:20–35.
58. Qureshi IA, Mehler MF. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. Nat Rev Neurosci. 2012;13:528–41.
59. Williams AH, Valdez G, Moresi V, Qi X, McAnally J, Elliott JL, et al. MicroRNA-206 delays ALS progression and promotes regeneration of neuromuscular synapses in mice. Science. 2009;326:1549–54.
60. Chestnut BA, Chang Q, Price A, Lesuisse C, Wong M, Martin LJ. Epigenetic regulation of motor neuron cell death through DNA methylation. The Journal of Neuroscience: the official journal of the Society for Neuroscience. 2011; 31:16619–36.
61. Laffita-Mesa JM, Bauer PO, Kouri V, Pena Serrano L, Roskams J, Almaguer Gotay D, et al. Epigenetics DNA methylation in the core ataxin-2 gene promoter: novel physiological and pathological implications. Hum Genet. 2012;131:625–38.
62. Da Cruz S, Cleveland DW. Understanding the role of TDP-43 and FUS/TLS in ALS and beyond. Current Opinion in Neurobiology. 2011;21:904–19.

63. Kawahara Y, Ito K, Sun H, Aizawa H, Kanazawa I, Kwak S. Glutamate receptors: RNA editing and death of motor neurons. Nature. 2004;427:801.
64. Kwak S, Kawahara Y. Deficient RNA editing of GluR2 and neuronal death in amyotropic lateral sclerosis. J Mol Med. 2005;83:110–20.
65. Takuma H, Kwak S, Yoshizawa T, Kanazawa I. Reduction of GluR2 RNA editing, a molecular change that increases calcium influx through AMPA receptors, selective in the spinal ventral gray of patients with amyotrophic lateral sclerosis. Annals of Neurology. 1999;46: 806–15.
66. Hideyama T, Yamashita T, Suzuki T, Tsuji S, Higuchi M, Seeburg PH, et al. Induced loss of ADAR2 engenders slow death of motor neurons from Q/R site-unedited GluR2. The Journal of Neuroscience : the official journal of the Society for Neuroscience. 2010;30:11917–25.
67. Aizawa H, Sawada J, Hideyama T, Yamashita T, Katayama T, Hasebe N, et al. TDP-43 pathology in sporadic ALS occurs in motor neurons lacking the RNA editing enzyme ADAR2. Acta Neuropathol. 2010;120:75–84.
68. Kwak S, Nishimoto Y, Yamashita T. Newly identified ADAR-mediated A-to-I editing positions as a tool for ALS research. RNA Biol. 2008;5:193–7.