

## Keeping up with genetic discoveries in amyotrophic lateral sclerosis: The ALSod and ALSGene databases

Christina M. Lill, Olubunmi Abel, Lars Bertram & Ammar Al-Chalabi

**To cite this article:** Christina M. Lill, Olubunmi Abel, Lars Bertram & Ammar Al-Chalabi (2011) Keeping up with genetic discoveries in amyotrophic lateral sclerosis: The ALSod and ALSGene databases, Amyotrophic Lateral Sclerosis, 12:4, 238-249, DOI: [10.3109/17482968.2011.584629](https://doi.org/10.3109/17482968.2011.584629)

**To link to this article:** <https://doi.org/10.3109/17482968.2011.584629>



Published online: 27 Jun 2011.



Submit your article to this journal [↗](#)



Article views: 2260



View related articles [↗](#)



Citing articles: 8 View citing articles [↗](#)

ORIGINAL ARTICLE

## Keeping up with genetic discoveries in amyotrophic lateral sclerosis: The ALSoD and ALSGene databases

CHRISTINA M. LILL<sup>1,2</sup>, OLUBUNMI ABEL<sup>3</sup>, LARS BERTRAM<sup>1</sup> & AMMAR AL-CHALABI<sup>3</sup>

<sup>1</sup>Neuropsychiatric Genetics Group, Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany, <sup>2</sup>Department of Neurology, Johannes Gutenberg University, Mainz, Germany, and <sup>3</sup>MRC Centre for Neurodegeneration Research, King's College London Institute of Psychiatry, London, UK

### Abstract

Amyotrophic lateral sclerosis (ALS) is a genetically heterogeneous disorder that shows a characteristic dichotomy of familial forms typically displaying Mendelian inheritance patterns, and sporadic ALS showing no or less obvious familial aggregation. While the former is caused by rare, highly penetrant, and pathogenic mutations, risk for sporadic ALS is probably the result of the combined effects of common polymorphisms with minor to moderate effect sizes. Owing to recent advances in high-throughput genotyping and sequencing technologies, genetic research in both fields is evolving at a rapidly increasing pace making it more and more difficult to follow and evaluate the most significant progress in the field. To alleviate this problem, our groups have created dedicated and freely available online databases, ALSoD (<http://alsod.iop.kcl.ac.uk/>) and ALSGene (<http://www.alsgene.org>), which provide systematic and in-depth qualitative and quantitative overviews of genetic research in both familial and sporadic ALS. This review briefly introduces the background and main features of both databases and provides an overview of the currently most compelling genetic findings in ALS derived from analyses using these resources.

**Key words:** ALSoD, ALSGene, amyotrophic lateral sclerosis, ALS, database, genetic, *UNC13A*

### Introduction

For amyotrophic lateral sclerosis (ALS), as well as for many other neurodegenerative disorders such as Alzheimer's disease (1) and Parkinson's disease (2), familial aggregation was recognized as a salient feature decades before any of the underlying biological and biochemical properties were known (3). In many cases, the identification of specific, disease-segregating gene mutations has first directed the attention of molecular biologists to certain proteins and pathways. The breakthrough in understanding ALS pathogenesis started with the discovery of autosomal-dominant mutations in *SOD1* on chromosome 21q22 (encoding soluble copper/zinc superoxide dismutase 1) in 1993 (4). In the last decade, 12 additional genes causing various forms of ALS have been identified (Table I), and it can be anticipated that

the number of ALS-causing mutations will increase further with the application of recently developed massively parallel sequencing techniques.

Another similarity to other neurodegenerative diseases is the prominent dichotomy of these rare, familial forms of ALS following Mendelian inheritance (5–10% of all cases) alongside seemingly 'sporadic' forms of the disease showing no or less obvious familial aggregation, with no clear clinical or pathological distinction between the two. While a small percentage of the sporadic ALS cases have also been found to harbour mutations in the known Mendelian ALS genes, the aetiology for the majority remains largely unknown. It is likely that there is a contribution from common genetic variants (polymorphisms) that are also present in the healthy population (with frequencies > 1%). Each of these polymorphisms

Correspondence: The ALSGene database: L. Bertram, Max Planck Institute for Molecular Genetics, Ihnestr. 63, Berlin, Germany. E-mail: lbertram@molgen.mpg.de. The ALSoD database: A. Al-Chalabi, MRC Centre for Neurodegeneration Research, King's College London Institute of Psychiatry P 041, London SE5 8AF, UK. E-mail: ammar.al-chalabi@kcl.ac.uk

(Received 10 March 2011; accepted 5 April 2011)

ISSN 1748-2968 print/ISSN 1471-180X online © 2011 Informa Healthcare  
DOI: 10.3109/17482968.2011.584629

probably exerts only a small effect on ALS risk (odds ratios (ORs) in the order of  $\sim 1.2$ ), so that it is via their combined action that they impact ALS susceptibility. Because of the small individual genetic effect sizes, thousands of subjects have to be tested to uncover these risk genes with sufficient confidence (i.e. to achieve genome-wide significance which is commonly defined as  $p$ -values falling at or below  $\sim 5 \times 10^{-8}$ ). Several hundred candidate-gene association studies have been performed in the field of ALS genetics during the last two decades, yielding mostly inconclusive results. In recent years, genome-wide association studies (GWAS), which can assess several million polymorphisms across the genome in one experiment, have suggested several new ALS risk genes, most of which are still awaiting independent follow-up and replication.

The growing amount of information in the field of ALS genetics is becoming increasingly difficult to follow, evaluate and interpret. This is true for both Mendelian and non-Mendelian, genetically complex ALS. To address this problem, the ALSod and the ALSGene databases have been developed independently. Both databases aim to provide systematic and in-depth overviews of the respective ALS genetics fields. The aim of this review is to introduce both resources by outlining their complementary goals and methods, as well as by summarizing the most interesting recent findings of each project.

### Assessing Mendelian mutations and rare variants in ALS: the ALS Online Genetic Database (ALSod)

Although the first human disease-causing mutation was identified in haemoglobin in 1949, the pace of discovery was such that the relationship between gene variations and the diseases they caused could be easily remembered. As the numbers grew, it became important to store such information systematically, and so the first genetics database came about in the 1970s when McKusick created a catalogue of mutations and inherited diseases, known as "Mendelian Inheritance in Man". A version of this database was later developed for online access (5), and is now known as OMIM. By 1994 there were regularly updated online lists of gene mutations as well as gene-specific databases. Key to the success of such resources was a universal and logical mutation naming convention based on amino acid changes (6,7).

Mutation databases of human genes are now becoming more prominent and important in all areas of health care. Published and unpublished mutations are typically curated in locus-specific databases (LSDB) (8), and apart from providing vital information on protein function and structure, databases of mutations causing Mendelian disease play a fundamental role in research, diagnostics, and genetic health care (9,10). The existence and regular upkeep of these databases is often vital to a broad range of research fields.

Table I. Genes proposed to cause Mendelian forms of ALS.

Gene	Protein	Location	Inheritance	# Mutations	# Patients	Proposed molecular effects/pathogenic relevance
<i>ALS2</i>	amyotrophic lateral sclerosis 2	2q33.1	recessive	19	4	endosome/membrane trafficking
<i>ANG</i>	angiogenin	14q11.1	dominant	18	27	RNAase
<i>DAO</i>	D-amino-acid oxidase	12q24	dominant	1	1	D-serine pathway
<i>FIG4</i>	FIG4 homolog, SAC1 lipid phosphatase domain containing (S. cerevisiae)	6q21	dominant	10	9	Golgi complex maintenance, vesicle trafficking, cell signalling
<i>FUS</i>	fused in sarcoma	16p11.2	both	27	5	RNA processing
<i>NEFH</i>	neurofilament, heavy polypeptide 200kDa, heavy chain	22q12.1-q13.1	?	7	10	axonal transport
<i>OPTN</i>	optineurin	10p13	recessive	3	0	Golgi complex maintenance and membrane trafficking
<i>SETX</i>	senataxin	9q34.13	dominant	4	60	DNA and RNA processing
<i>SOD1</i>	superoxide dismutase 1	21q22.11	both	156	184	toxic aggregation of SOD1; oxidative damage; mitochondrial dysfunction; RNA destabilization; impairment of axonal transport; glutamate excitotoxicity
<i>TARDBP</i>	TAR DNA binding protein (TDP-43)	1p36.22	dominant	38	66	RNA processing
<i>VAPB</i>	VAMP (vesicle-associated membrane protein)-associated protein B and C	20q13.33	dominant	2	19	vesicle trafficking
<i>VCP</i>	valosin-containing protein	9p13	dominant	4	8	ubiquitination

For an up-to-date overview of these and other potential Mendelian ALS genes see the 'ALSod database' (<http://alsod.iop.kcl.ac.uk>). Note that mutations in other genes have also been proposed to cause familial forms of ALS, albeit with inconclusive evidence (see text for more details).

# Mutations = number of mutations reported as ALS-causing; # Patients = number of records on ALS patients submitted to ALSod that carry mutations in this gene.

Following the identification of ALS-causing mutations in *SOD1* there was a great focus of research on this gene. As the number of mutations increased and the range of phenotypes expanded, the need arose for a database that might reveal new genotype-phenotype relationships, keep track of the different disease-causing genes, and aid the recording of gene variations that might not be published. From this came the ALS Online Database, ALSoD, an open source data repository designed for sharing clinical and genetic data, developed by a consortium within the World Federation of Neurology (11,12). It provides both the scientific community and wider public with up-to-date information on ALS-related genes identified from family studies, candidate gene studies or genome-wide association (Figure 1). The functional importance of

such a database is underscored by the appearance of other similar websites, e.g. an ALS mutation database recently constructed as part of the Life Science Integrated Database Project conducted by the Japan Ministry of Education, Culture, Sports, Science, and Technology (13). This contains their original experimental results as well as published data extracted from scientific journals. This database is expected to play a complementary role to the ALSoD website, especially in collecting genetic variations commonly found in Asia. The ALS Online Database website was initially located at <http://www.alsod.org> and is now located at <http://alsod.iop.kcl.ac.uk>, but it remains a World Federation of Neurology resource, not tied to any institution. This change in web address was a simply the result of the need for huge data storage.

**ALSoD** ALS ONLINE GENETICS DATABASE

You are here ==> [Home](#)

Search

Home Analysis Summary GWAS News Data Contributors Feedback

Submit mutation and patient data now

Log In

User Name:

Password:

☐ Remember me

Forgot password?

If you wish to contribute data you must [REGISTER](#)

ALSoD is a joint project of World Federation of Neurology and European Network to Cure ALS

**ENCALS** European Network to Cure ALS

**WFN-ALS** World Federation of Neurology Research Group in Motor Neuron Diseases Neurologic Lateral Sclerosis

ALSoD is funded by

**ALS Association** Advancing Global Research

**ALS SOCIETY OF CANADA**

**mnda** motor neuron disease association

**MND Association in Iceland** [www.mnd.is](http://www.mnd.is)

**gsk** GlaxoSmithKline

**Chromosomes**

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y All

**Gene report of all ALS-Related genes in ALSoD (74)**

View Details	Gene	Gene name	Chromosome
Select	ALAD	d-Aminolevulinic Acid Dehydratase	9q33.1
Select	ALS2	amyotrophic lateral sclerosis 2 (juvenile) homolog (human), Alsin	2q33.2
Select	ANG	Angiogenin	14q11.1
Select	APEX1	Apurinic endonuclease	14q11.2
Select	APOE	Apolipoprotein E	19q13.2
Select	AR	Androgen receptor	Xq11.2
Select	ATXN2	ataxin 2	12q23-q24.1
Select	B4GALT6	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide	18q12.1
Select	C9orf72	chromosome 9 open reading frame 72	9p21.2
Select	CCS	Copper chaperone for superoxide dismutase	11q13
Select	CHMP2B	chromatin modifying protein 2B	3p12.1
Select	CNTF	Ciliary neurotrophic factor	11q12
Select	CNTN4	contactin 4	3p28.3
Select	CRYM	crystallin, mu	16p
Select	CSNK1G3	casein kinase 1, gamma 3	5q23
Select	CYP2D6	Cytochrome p450, subfamily IID, polypeptide 6	22q13.1
Select	DAO	D-amino-acid oxidase	12q24
Select	DCTN1	Dynactin	2p13
Select	DISC1	disrupted in schizophrenia 1	1q42.2
Select	DPP6	dipeptidyl-peptidase 6	7q38.2
Select	DYNC1H1	Dynein heavy chain	14q32
Select	EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1	2p16.1
Select	ELP3	elongation protein 3 homolog (S. cerevisiae)	8p21.1
Select	FGGY	FGGY carbohydrate kinase domain containing	1p32.1
Select	FIG4	FIG4 homolog, SAC1 lipid phosphatase domain containing (S. cerevisiae)	6q21
Select	FUS	fusion (involved in t(12;16) in malignant liposarcoma)	16p11.2
Select	GARS	Glycyl tRNA synthetase	7p15
Select	HEXA	Hexosaminidase A	15q23
Select	HFE	Haemochromatosis	6p21.3
Select	IFNK	interferon, kappa	9p21.2
Select	ITPR2	inositol 1,4,5-triphosphate receptor, type 2	12p11.23
Select	KIFAP3	kinesin-associated protein 3	1p38.13-q31.3
Select	LIF	Leukaemia inhibitory factor	22q12.2
Select	LIPC	lipase, hepatic	15q22.1
Select	LOX	Lysyl oxidase	5q23.2
Select	LUM	lumican	12q21.3
Select	MAOB	Monoamine oxidase B	Xp11.4
Select	MAPT	Microtubule-associated protein tau	17q21
Select	MOBK2B	MOB1, Mps One Binder kinase activator-like 2B (yeast)	9p21.2
Select	MT-ND2	Subunit 2 of mitochondrial NADH dehydrogenase (Complex I)	0

Figure 1. ALSoD homepage.



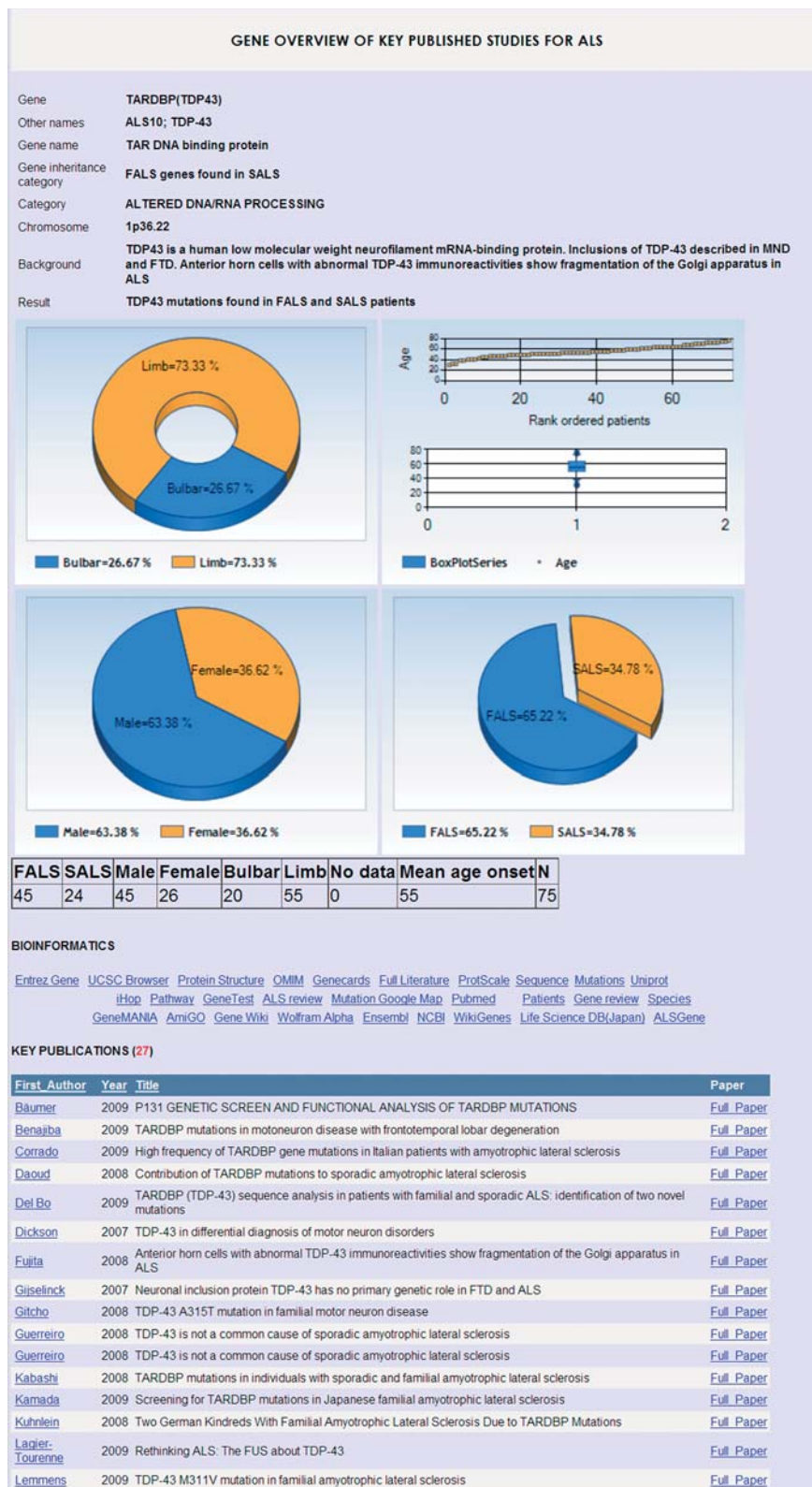


Figure 2. Genotype-phenotype summary for *TARDBP* (coding for TDP43). Gene summaries are displayed on selection. More detailed customised analyses are also possible allowing specific questions to be asked of the database.

### Aims

ALSod (freely available at <http://alsod.iop.kcl.ac.uk>) is designed to allow easy access to information about ALS-causing gene variants and their phenotypic effects. Because genotype-phenotype correlations are usually clearest for Mendelian disease genes, and

the database has been in existence for several years, much of the focus is on familial ALS, but all currently known ALS-causing genes are represented. Where relevant there are links out to an appropriate database resource. For example, the evidence for association at a given gene is provided by a link out



Figure 3. An example of a geographical overview of the mutation distribution of *SOD1*. Geographical analysis is available for most FALS genes. More sophisticated map overlays are planned.

to the ALSGene database where meta-analysis and other explorations are possible (see below), and predictions of *SOD1* mutation effects on structure and function are provided by a link to a *SOD1* protein structure database at UCL, whereas phenotype information on *TARDBP*, *FUS* or *OPTN* mutations is accessed locally. Mutation hot and cold spots, country-of-origin effects and literature and bioinformatics resources are also included within ALSod.

### Methods

ALSoD allows users to submit new gene, mutation and patient data. It also has various tables in the database schema, storing data on sequence mutation data on genes, and anonymized patient data including phenotypic features and country of origin. Published and unpublished data are continually curated by the database administrator and registered ALS researchers. Published data are obtained by regular PubMed searches, and screening of cross-references from relevant publications. Unpublished data are uploaded by users (as described above) or made available to the ALSoD investigators by collaborations, and are published on the ALSoD website after consistency checks performed by the database administrator. There are currently 12 contributors from various international institutions.

To allow integration of different databases through one gateway, ALSoD uses unique identifiers to systematically link to other bioinformatic tools and comprehensive databases. Selecting a link

automatically interrogates the relevant third party website with the appropriate genetic information. Thus, it is possible to identify the most up-to-date list of genes implicated in ALS, explore their interactions and pathways, and review the relevant literature. Customized and routine analyses of mutational relationships with phenotype are possible, either for genes with phenotypic data, such as *SOD1*, *TARDBP*, *FUS*, *SETX*, and others, or as a general overview of familial and sporadic phenotypic patterns (Figure 2).

The mapping of gene variants on Google Earth allows researchers or patients to view the geographical distribution of reported gene variations associated with ALS (Figure 3). The same computational method allows us to map the origin of users and show which countries predominantly access ALSoD.

For users with their own association data, an on-the-fly analysis is available to combine the data available in ALSoD with unpublished user data that can be confidentially uploaded. The user data is formatted accordingly before upload and the result is fed back in minutes without storing users' data on the database.

### Results

Currently, ALSoD displays information on 73 ALS genes (17 for familial ALS), 298 mutations and 419 patients (343 with age of onset data and described in analyses below). Typically there are more than 100 unique users visiting ALSoD each day, and in November 2010 the pages were accessed more than 18,000



To predict gene interactions accessed from [geneMANIA](#)

☐ 9p21 ☐ B4GALT6 ☐ DAO ☐ FIG4 ☐ KIFAP3 ☐ MOBKL2B ☐ PON2 ☐ SELL ☐ SNCG ☐ UNC13A  
☐ ALAD ☐ CCS ☐ DCTN1 ☒ FUS ☐ LIF ☐ MT-ND2 ☐ PON3 ☐ SEMA6A ☒ SOD1 ☐ VAPB  
☐ ALS2 ☐ CHMP2B ☐ DISC1 ☐ GARS ☐ LIPC ☐ NAIP ☐ PRPH ☐ SETX ☐ SOD2 ☐ VCP  
☐ ANG ☐ CNTF ☐ DPP8 ☐ GRN ☐ LOX ☐ NEFH ☐ PSEN1 ☐ SLC1A2 ☐ SPAST ☐ VDR  
☐ APEX1 ☐ CNTN4 ☐ DYNC1H1 ☐ HEXA ☐ LUM ☐ NT5C1A ☐ PVR ☐ SLC39A11 ☐ SPG7 ☐ VEGFA  
☐ APOE ☐ CRYM ☐ EFEMP1 ☐ HFE ☐ MAOB ☒ OPTN ☐ RBMS1 ☐ SMN1 ☐ SUSD1 ☐ ZFP84  
☐ AR ☐ CSNK1G3 ☐ ELP3 ☐ IFNK ☐ MAPT ☐ PON1 ☐ SCN7A ☐ SMN2 ☒ TARDBP ☐ ZNF748  
☒ ATXN2 ☐ CYP2D8 ☐ FGGY ☐ ITPR2

[Reset](#) [Execute](#) *Please wait to view result but works better and faster using [Chrome](#) or [Firefox](#)*

Selected gene(s):

ATXN2 FUS OPTN SOD1 TARDBP

**GENEMANIA** Find genes in  related to  [Show](#)

[Save](#) [Actions](#) [Networks legend](#) [Functions legend](#) [Networks](#) [Genes](#) [Functions](#) [Help](#)

Sort by: [name](#), [per cent weight](#)  
 Expand: [all](#), [only top level](#), [none](#)  
 Enable: [all](#), [none](#)

- ☒ Physical interactions 93.45 %
- ☒ Co-expression 1.77 %
- ☒ Pathway 1.68 %
- ☒ Predicted 1.57 %
- ☒ Co-localization 1.38 %
- ☐ Genetic interactions 0.16 %

[New features!](#) [Join our mailing list](#) [Contact us](#) [Donnelly Centre](#) [About GeneMANIA](#) [OpenHelix video tutorial](#) [Help](#)

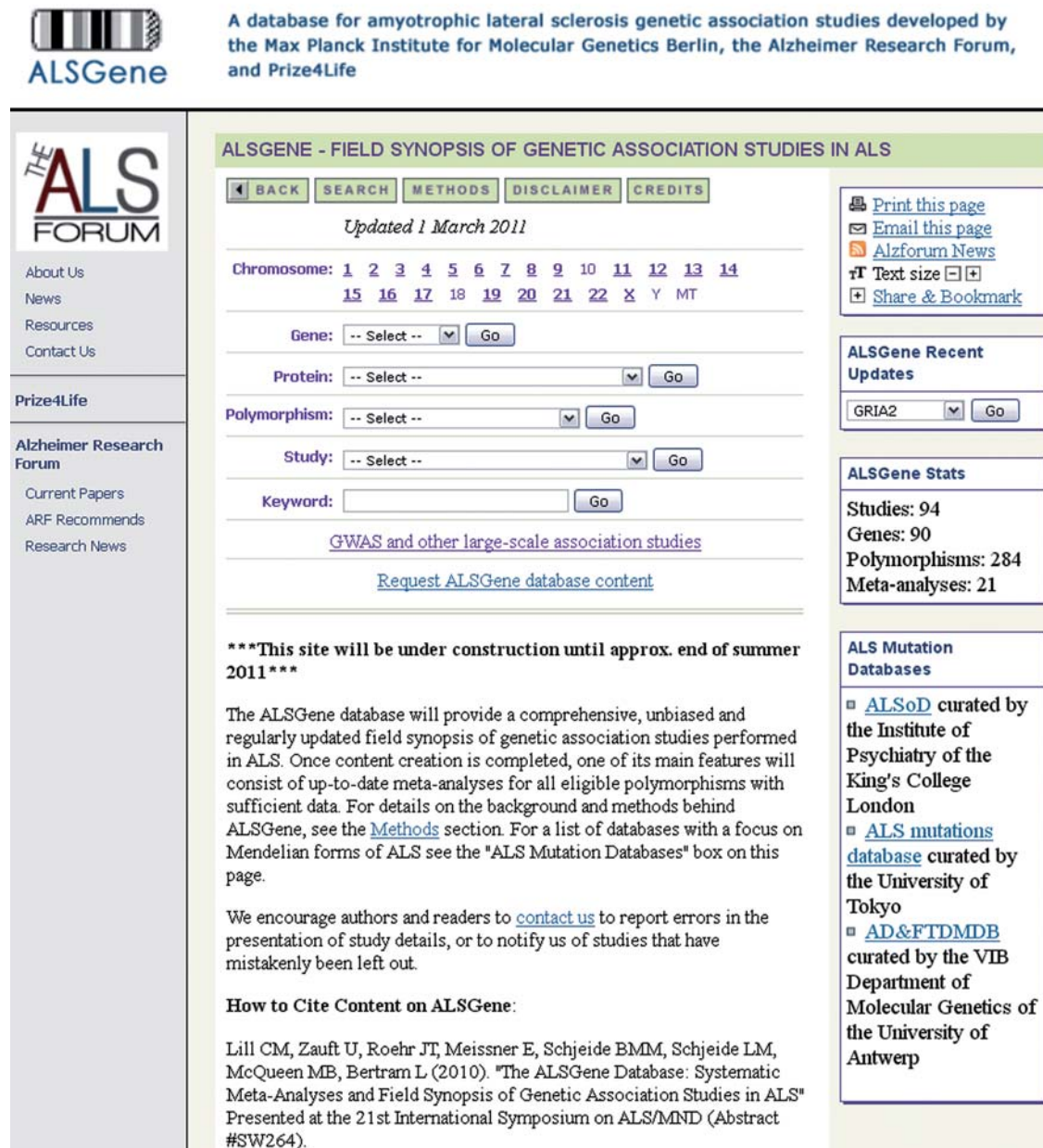
Figure 4. Gene interactions using a GeneMANIA interface at ALS<sub>o</sub>D. Potential gene pathways and networks can be explored via the GeneMANIA feature on ALS<sub>o</sub>D, e.g. by including established or proposed ALS genes in the functional predictions.

times. Data can be displayed or interrogated online or downloaded into a spreadsheet. For genes, the mutation type, sequence change, original and mutated amino acid, functional annotation and exon number are reported, with a link to the predicted mutant structure. For patients, the gene, mutation and mutation type are reported along with the patient's sex, country of residence, ethnicity, age at onset, clinical pattern, survival data and family history.

Analysis of the record of Mendelian mutations in which one would expect a 50:50 male: female ratio, shows that the actual recorded sex ratio is 172:119, which is almost exactly the 3:2 ratio commonly reported for ALS in general from clinic-based studies. While for 250 patients there is a family history of ALS, 84 patients have none despite

a Mendelian mutation, and nine record it as unknown. Those without a family history are not skewing the sex ratio, however, as for those with a family history the sex ratio is 120:88, which is still very close to 3:2.

The mean age of onset overall is 45.7 years. Excluding those with juvenile onset (for example with *ALS2* gene mutation), mean age of onset is 46.2 years. The mean age of onset for those with a family history is 42.8 years while for those without it is 53.3 years, which is very close to the mean age of onset recorded in clinic based studies, and suggests that a family history is associated with a younger age of onset. However, for those with *SOD1* mutation, mean age of onset is 47.2 years ( $n = 125$ ), which breaks down into 47.9 years for those with a



**ALSGENE - FIELD SYNOPSIS OF GENETIC ASSOCIATION STUDIES IN ALS**

Updated 1 March 2011

Chromosome: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT

Gene: -- Select -- Go

Protein: -- Select -- Go

Polymorphism: -- Select -- Go

Study: -- Select -- Go

Keyword: Go

[GWAS and other large-scale association studies](#)

[Request ALSGene database content](#)

\*\*\*This site will be under construction until approx. end of summer 2011\*\*\*

The ALSGene database will provide a comprehensive, unbiased and regularly updated field synopsis of genetic association studies performed in ALS. Once content creation is completed, one of its main features will consist of up-to-date meta-analyses for all eligible polymorphisms with sufficient data. For details on the background and methods behind ALSGene, see the [Methods](#) section. For a list of databases with a focus on Mendelian forms of ALS see the "ALS Mutation Databases" box on this page.

We encourage authors and readers to [contact us](#) to report errors in the presentation of study details, or to notify us of studies that have mistakenly been left out.

**How to Cite Content on ALSGene:**

Lill CM, Zaufu U, Roehr JT, Meissner E, Schjeide BMM, Schjeide LM, McQueen MB, Bertram L (2010). "The ALSGene Database: Systematic Meta-Analyses and Field Synopsis of Genetic Association Studies in ALS" Presented at the 21st International Symposium on ALS/MND (Abstract #SW264).

**ALSGene Recent Updates**

GRIA2 Go

**ALSGene Stats**

Studies: 94  
Genes: 90  
Polymorphisms: 284  
Meta-analyses: 21

**ALSGene Mutation Databases**

- ALSoD curated by the Institute of Psychiatry of the King's College London
- ALS mutations database curated by the University of Tokyo
- AD&FTDMDB curated by the VIB Department of Molecular Genetics of the University of Antwerp

Figure 5. Screenshot of the ALSGene homepage (www.alsgene.org).

family history ( $n = 105$ ) and, surprisingly, 41.7 for those without ( $n = 19$ ). For *TARDBP* mutations (codes for TDP43 protein), mean age of onset is 55.5 years ( $n = 75$ ) and for *FUS* mutations, 50.4 years ( $n = 5$ ). There are 31 different *TARDBP* mutations recorded, the commonest being A382T ( $n = 11$ ), G348C ( $n = 8$ ), and M337V ( $n = 7$ ). Five individuals with *SOD1* mutation have bulbar onset, whereas 116 have limb onset (30 upper limb, 83 lower limb, three not specified). Twenty individuals with *TARDBP* mutation have bulbar onset, compared with 41 limb onset (10 upper limb, four lower limb, rest not specified). This difference in recorded presentation patterns between those with *SOD1* and *TARDBP* mutations is highly significant ( $\chi^2$   $p$ -value  $4 \times 10^{-7}$ ).

A GeneMANIA web interface allows prediction of ALS gene interactions in ALSoD, which is a

powerful tool for the exploration of new pathways (14) (Figure 4).

#### Future implementations

ALSoD will widen its phenotypic remit to include frontotemporal dementia genes, develop more detailed geographical displays (for example, overlaying epidemiology study data on Google Earth), integrate genome-wide sequence data with appropriate display tools and, through collaborations, integrate transcriptomics data with existing gene and phenotype data. LOD scores and sequence data will be ported to a genome browser track for UCSC and Ensembl browsers allowing for easy analysis together with existing data as well as comparison with ALSGene results. Through collaboration with ALSGene, association evidence for phenotypic modifiers for age of onset, site of onset and survival will be displayed.



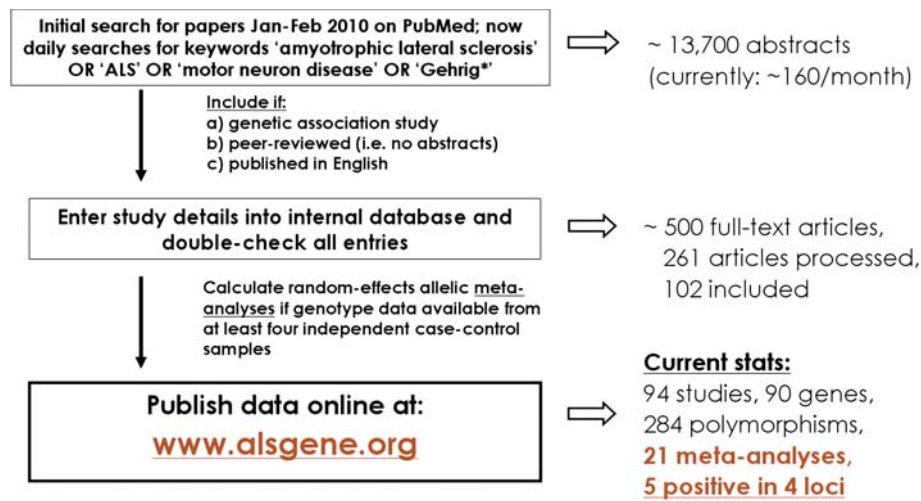


Figure 6. Summary of ALS*Gene* methodology. Flowchart of data selection, processing and analysis strategies applied for ALS*Gene*. Polymorphisms with non-overlapping data from at least four data sets are subjected to meta-analyses. Significant meta-analysis results are highlighted in a designated section on the ALS*Gene* homepage.

### Assessing genetically complex forms of ALS: the ALS*Gene* database

The methods to identify susceptibility variants for genetically complex traits differ from those aimed at finding new causative disease genes. Most commonly, susceptibility variants are identified by probing for significant differences in allele or genotype frequencies of a genetic variant across groups of affected ('cases') versus unaffected ('controls') individuals by performing genetic association analyses. As outlined in the introduction, several hundred such genetic association studies have been published in the field of sporadic ALS over the past two decades. As a result, more than 150 putative ALS risk genes have been reported. These findings have been met with inconclusive replication evidence, at best. More often than not, presumed risk effects were not observed in independent datasets, complicating an easy interpretation of these findings. More recently, genome-wide approaches have been applied to case-control datasets (15–24) which – at least initially – did not lead to an improvement in replicability of the top findings. Concerted efforts using increasingly larger datasets have somewhat alleviated this situation (e.g. References 20,24). With a steadily growing number of GWAS and smaller-scale association studies focusing on sporadic ALS, the field becomes increasingly more difficult to evaluate. To facilitate this situation, our group has developed a systematic meta-analytic approach to genetic association studies that has already successfully been applied to other neurodegenerative diseases, e.g. Alzheimer's disease (25) and Parkinson's disease (26). We have begun applying the same systematic methodology to the field of sporadic ALS in the form of the ALS*Gene* database.

#### Aims

ALS*Gene* (freely available via the URL: <http://www.alsgene.org> or Prize4Life's ALS portal <http://www.prize4life.org>)

ResearchALS.org; see Figure 5 for a screenshot of the homepage of ALS*Gene*) aims to serve as an exhaustive, unbiased, and regularly updated resource of genetic association studies in ALS. One of its key features will be up-to-date meta-analyses of all eligible genetic polymorphisms that have been investigated for association with ALS risk. This entails the inclusion of genome-wide association data, either in part or – if available – in full. All data and results are concisely summarized and displayed online. Upon completion of content curation, this will include an up-to-date list of 'Top Results', i.e. those showing the strongest evidence for association with risk for ALS when all available data are combined.

#### Methodology

For a visual summary of this paragraph see Figure 6. Eligible studies are identified via PubMed searches, or by screening the table of contents of neurological and genetic journals. Additional publications are also sought by systematically screening the references of relevant publications (e.g. primary association papers, but also reviews, etc.). To date, these searches have yielded ~500 articles on ALS genetics of which the full text is currently being screened for eligibility. A publication is included in ALS*Gene* if it represents a genetic association study (here defined as studies investigating DNA sequence variants with  $\geq 1\%$  frequency in the general population, a.k.a. 'polymorphisms'), and if it has been published in a peer-reviewed English language journal. Demographic details of included studies and genotype summary data of the investigated polymorphisms are extracted from each publication, and summarized on the ALS*Gene* website. Each study included and listed in ALS*Gene* is hyperlinked to the respective abstract in PubMed to facilitate identification and access to the original publication. In addition, an overview of all published ALS GWAS is provided in

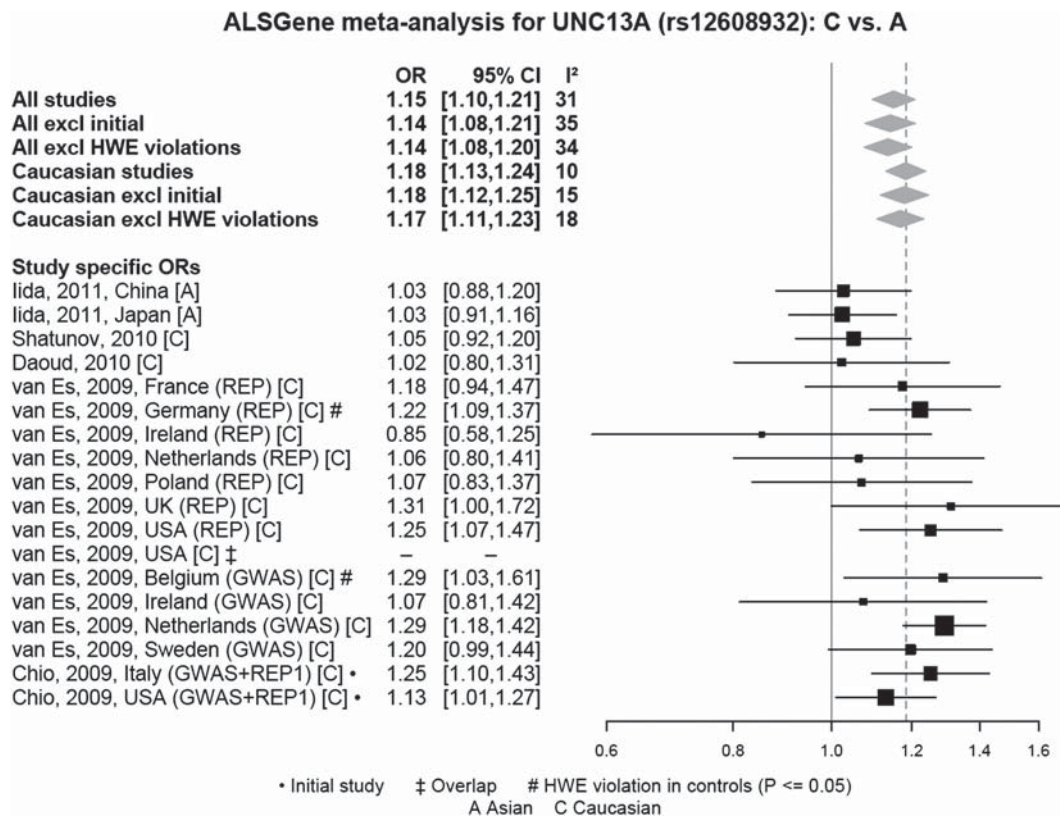


Figure 7. Random-effects meta-analysis of rs12608932 in *UNC13A* including all published studies (as of 1 March 2011). Example ALSGene forest plot providing a visual summary of meta-analysis results. Study-specific odds allelic ratios (ORs, bottom section) are displayed either as provided in the original publication or calculated de novo from allele or genotype summaries. Black squares represent point estimates of the OR (scaled to sample size), horizontal lines represent 95% confidence intervals. Summary ORs (top section) are then calculated from the study-specific ORs using random effects models for a number of different paradigms. Grey diamonds represent point estimates and 95% confidence intervals of the respective meta-analyses.  $I^2$  = a measure of between-study heterogeneity (see text for more details). Forest plots are only available for polymorphisms that have data available from four independent data sets.

a dedicated section on ALSGene (<http://www.alsgene.org/largescale.asp>). Alongside the main characteristics of each study, this GWAS overview section also provides a hyperlinked list of ‘featured genes’, i.e. those loci or pathways highlighted by the primary authors as the main outcome of their study after having completed all analyses. ALSGene meta-analyses are performed for all genetic polymorphisms with at least four independent case-control datasets using allelic contrasts. Whenever possible, meta-analyses are complemented by GWAS data for overlapping polymorphisms (either directly genotyped or determined by imputation). Results of these meta-analyses are visualized on the website as forest plots (representing an up-to-date summary of the current evidence) and cumulative meta-analyses (summarizing the development of effect size estimates over time). Additional analyses include an assessment of the ‘epidemiologic credibility’ of significant findings using the HuGENet ‘Venice’ criteria (27,28), as well as Bayesian analyses (29). Furthermore, meta-analyses are carried out after stratification for distinct ethnicities (e.g. Asian and Caucasian ethnicities), when at least three independent datasets for the respective ethnicity are available (for an example see Figure 7). More details on the methods behind data collection

and analyses for ALSGene can be found online. All significant meta-analysis results are summarized in a dedicated section of ALSGene (‘Top Results’; note that the Top Results feature will only become available after primary literature-based content creation has been completed, i.e. approx. summer/fall of 2011).

#### Preliminary results

Currently, the online version of ALSGene displays the details and genotype summary data of over 90 independent studies that have investigated > 250 polymorphisms in 90 genes. As this paper is going to press, 21 meta-analyses are available, of which five show nominally significant results in four genetic loci (*APEX1* (OR 0.78;  $p = 0.028$ ), *HFE* (OR 1.72;  $p = 0.0034$ ), *UNC13A* (OR 1.18 in Caucasians;  $p = 3.6 \times 10^{-13}$ ; Figure 7)), and a hitherto uncharacterized locus on chromosome 9p21.2 (OR 1.25;  $p = 5.3 \times 10^{-18}$ ). The chromosome 9p21.2 locus and *UNC13A*, both initially implicated by recent GWAS (20,23,24), currently show association with risk for ALS at genome-wide significance (i.e.  $p$ -values  $\leq 5 \times 10^{-8}$ ). The current *UNC13A* finding described here is a good example of why systematic data integration efforts are needed, even in the ‘GWAS era’

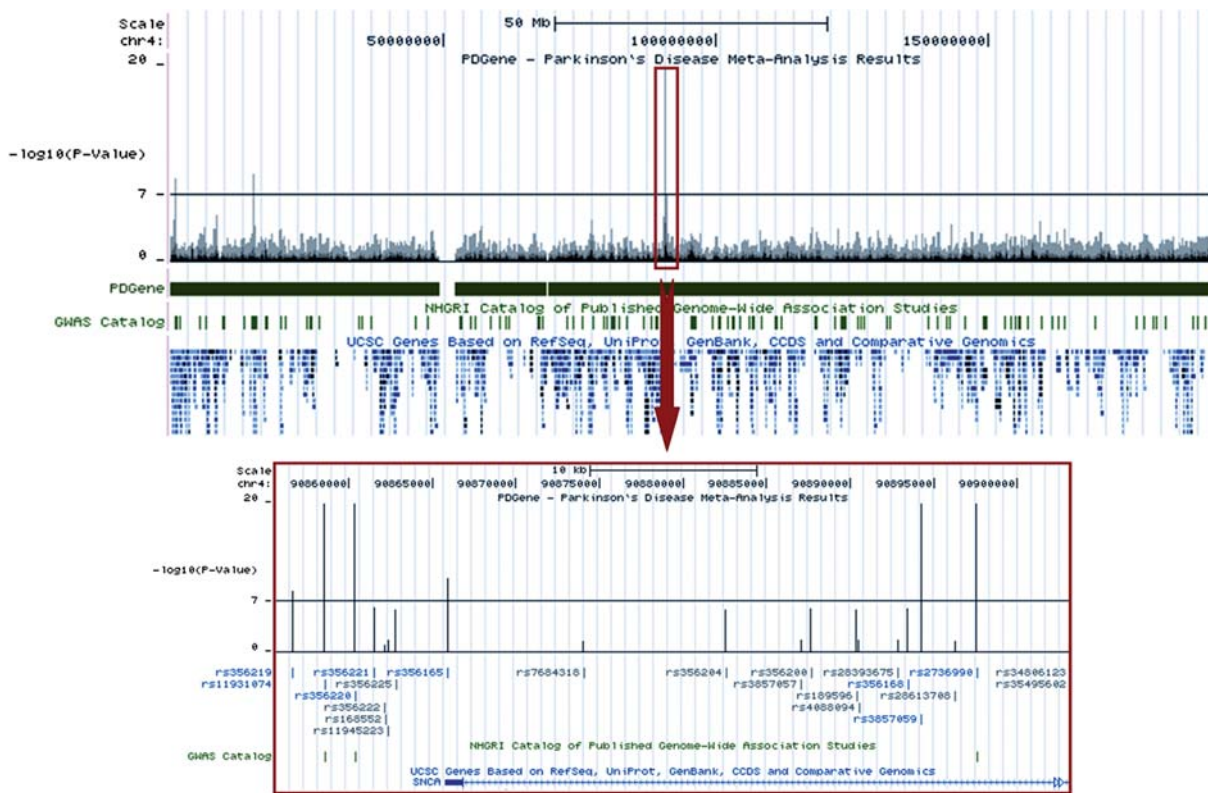


Figure 8. Example screenshot of a customized UCSC genome-browser track (PDGene database) displaying meta-analysis  $p$ -values. The UCSC genome-browser (<http://genome.ucsc.edu/cgi-bin/hgGateway>) presents an up-to-date, easy-to-navigate interface of the human genome. It integrates information from a broad variety of databases (e.g. gene-mapping, expression, evolutionary conservation) at base-pair resolution. Shown is an example screenshot for a customized browser track from the PDGene database maintained by our group (<http://www.pdgene.org>), which summarizes meta-analysis results ( $-\log(10)$   $p$ -values, based on data from both genome-wide and candidate-gene association studies) equivalently at base-pair resolution. SNPs (rs-numbers) displayed in grey are based on meta-analyses of available GWAS datasets, those displayed in blue are supplemented by candidate-gene/replication datasets. All rs-numbers cross-link back to the PDGene database to provide more details on the respective meta-analyses. Note that this feature will become available on ALS*Gene* once initial data entry and analysis is completed.

of complex genetics: as can be seen in Figure 7, the four most recent individual datasets investigating the rs12608932 polymorphism in this gene only show very modest support in favour of a genuine risk effect (24,30,31). In particular, the study by Iida et al. (31) is interesting as this is the first to investigate this marker in non-European populations. At least in the investigated Chinese and Japanese datasets this polymorphism may have no or a much smaller role than in Europeans, although additional studies (e.g. with other markers in *UNC13A*) are needed to assess this more definitively. A similar situation of different effect size estimates across distinct ethnicities is observed for several of the current Top Results genes in our sister-database on Parkinson's disease genetics (PDGene, [www.pdgene.org](http://www.pdgene.org)). Note that content for ALS*Gene* is still under construction and that these results will probably change over time.

#### Future implementations

After completion of the initial curation process, all results will be ported to a customized UCSC genome-browser track displaying meta-analysis  $p$ -values. This will also include the results of meta-analyses on full

GWAS datasets. This track will be hyperlinked to entries on ALS*Gene* proper, where details of the sample-specific effect size estimates can be found for the most significant findings. Figure 8 shows an example screenshot of this feature for the PDGene database. In addition to displaying a browsable summary of ALS*Gene* results on a genome-wide level, association signals can easily be put into the context of functional genetic studies, expression data, evolutionary conservation of genomic regions, etc., as annotated by the UCSC database (URL: <http://genome.ucsc.edu/>). This feature will also allow for an easy integration with content and results from the ALS*oD* database, as well as the comparison of genetic results emerging from our other database projects focusing on neurodegenerative disorders, such as Alzheimer's, Parkinson's, and frontotemporal dementia (FTD*Gene*; in preparation).

#### ALS*oD* and ALS*Gene*: joining forces from two different angles

As outlined above, the respective focuses of ALS*oD* and ALS*Gene* lie in different aspects of genetics research in ALS. The main focus for ALS*oD* is the summary and curation of genotype-phenotype correlations,



which is usually clearest for Mendelian ALS genes and mutational studies, while ALSGene is mainly centred on genetic association studies. Both efforts face different challenges owing to the advent of powerful high-throughput technologies now allowing a growing number of research laboratories to routinely perform whole-genome genotyping and sequencing. In ALS genetics, two of the most interesting unanswered questions to date are: 1) are the same genes causing Mendelian forms of ALS through private mutations also involved in modifying disease susceptibility of non-Mendelian forms of ALS via the action of common polymorphisms (as seems to be the case in Parkinson's disease and frontotemporal dementia); and 2) can any of the genetic associations with common polymorphisms be explained by DNA variants that are neither common nor private? While these questions can only be directly answered by performing additional laboratory experiments, the outcomes of these efforts will be tracked and fed back to the research community by ALSod and ALSGene. Combining the emerging evidence collected in both resources via an integrated interface (e.g. using the UCSC genome-browser architecture) is one efficient, yet simple, way of achieving this goal, in addition to hyperlinking related content and results across databases. In order to accomplish these objectives, the curatorial teams of both ALSod and ALSGene have joined forces to provide a seamless and easy-to-navigate user experience, which will hopefully allow researchers to gain new insights from apparently unrelated sets of data.

## Conclusions

With the advent of increasingly more powerful technologies, what we are learning about ALS is expanding every day. This is particularly true for research in human genetics, which has provided not only ground-breaking new insights into ALS pathogenesis, but also many insubstantial results. Online research databases such as ALSod and ALSGene serve a vital purpose in facilitating access to knowledge and generating new insights from existing disparate sets of data allowing both genetics experts and non-experts to more easily separate the wheat from the chaff in the growing mountain of raw genetic data.

## Acknowledgements

ALSod is a joint project of the World Federation of Neurology (WFN) and European Network for the Cure ALS (ENCALS). We are especially grateful for the long-standing and continued funding of this project from the ALS Association and the MND Association of Great Britain and Northern Ireland. We also thank ALS Canada, MNDA Iceland and the ALS Therapy Alliance. The research leading to these results has received funding from the European

Community's Health Seventh Framework Programme FP7/2007-2013) under grant agreement n° 259867. We also thank the NIHR Biomedical Research Centre for Mental Health. We thank Aleks Radunovic, Nigel Leigh, and Ian Gowrie, who originally conceived ALSod, and the ALSod team, especially Peter Andersen and John Powell.

ALSGene has been funded by Prize4Life as a critical functionality of their ALS Forum research portal ([www.ResearchALS.org](http://www.ResearchALS.org)), created to serve as a unified source of breaking ALS research news and valuable tools for the ALS research community. It has been developed in close collaboration with the Alzheimer Research Forum. We thank Gabriele Strobel, Colin Knep and Paula Noyes for the online adaptation and maintenance of ALSGene. We further thank the ALSGene database team, especially Ute Zauft, Esther Meissner, Johannes Roehr, Brit-Maren Schjeide, and Leif Schjeide.

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

## References

- Schottky J. Ueber praesenile Verblodungen. *Z Ges Neurol Psychiat.* 1932;140:333–97.
- Gowers WR. *A Manual of Diseases of the Nervous System.* Vol. I. Diseases of the nerves and spinal cord. 3rd edn. Philadelphia: P. Blakiston's Son & Co.; 1900.
- Aran FA. Recherches sur une maladie non encore decrite dy systeme musculaire (atrophie musculaire progressive). *Archs Gen Med.* 1850;3:5–35.
- 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.
- Hamosh A, Scott AF, Amberger J, Valle D, McKusick VA. Online Mendelian Inheritance in Man (OMIM). *Hum Mutat.* 2000;15:57–61.
- Cotton RG, McKusick V, Scriver CR. The HUGO Mutation Database Initiative. *Science.* 1998;279:10–1.
- Scriver CR, Nowacki PM, Lehtvaslaihio H. Guidelines and recommendations for content, structure, and deployment of mutation databases: II. Journey in progress. *Hum Mutat.* 2000;15:13–5.
- Claustres M, Horaitis O, Vanevski M, Cotton RG. Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases. *Genome Res.* 2002;12:680–8.
- George RA, Smith TD, Callaghan S, Hardman L, Pierides C, Horaitis O, et al. General mutation databases: analysis and review. *J Med Genet.* 2008;45:65–70.
- Soussi T, Ishioka C, Claustres M, Beroud C. Locus-specific mutation databases: pitfalls and good practice based on the p53 experience. *Nat Rev Cancer.* 2006;6:83–90.
- Radunovic A, Leigh PN. ALSODatabase: database of SOD1 (and other) gene mutations in ALS on the internet. European FALS Group and ALSOD Consortium. *Amyotroph Lateral Scler Other Motor Neuron Disord.* 1999;1:45–9.
- Wroe R, Wai-Ling Butler A, Andersen PM, Powell JF, Al-Chalabi A. ALSOD: the Amyotrophic Lateral Sclerosis Online Database. *Amyotroph Lateral Scler.* 2008;9:249–50.
- Yoshida M, Takahashi Y, Koike A, Fukuda Y, Goto J, Tsuji S. A mutation database for amyotrophic lateral sclerosis. *Hum Mutat.* 2010;31:1003–10.

14. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 2010;38: 214–20.
15. Dunckley T, Huentelman MJ, Craig DW, Pearson JV, Szelinger S, Joshipura K, et al. Whole-genome analysis of sporadic amyotrophic lateral sclerosis. *N Engl J Med.* 2007; 357:775–88.
16. 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 Neurol.* 2007;6:322–8.
17. 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 Neurol.* 2007;6:869–77.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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 Neurol.* 2010;9:978–85.
24. 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 Neurol.* 2010;9: 986–94.
25. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer's disease genetic association studies: the AlzGene database. *Nat Genet.* 2007;39:17–23.
26. Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, Schjeide BM, et al. The PDGene Database. *Alzheimer Research Forum.* Available at: <http://www.pdgene.org/>.
27. Ioannidis JP, Boffetta P, Little J, O'Brien TR, Uitterlinden AG, Vineis P, et al. Assessment of cumulative evidence on genetic associations: interim guidelines. *Int J Epidemiol.* 2008;37:120–32.
28. Khoury MJ, Bertram L, Boffetta P, Butterworth AS, Chanock SJ, Dolan SM, et al. Genome-wide association studies, field synopses, and the development of the knowledge base on genetic variation and human diseases. *Am J Epidemiol.* 2009;170:269–79.
29. Ioannidis JP. Effect of formal statistical significance on the credibility of observational associations. *Am J Epidemiol.* 2008;168:374–83; discussion 384–90.
30. Daoud H, Belzil V, Desjarlais A, Camu W, Dion PA, Rouleau GA. Analysis of the UNC13A gene as a risk factor for sporadic amyotrophic lateral sclerosis. *Arch Neurol.* 2010;67: 516–7.
31. Iida A, Takahashi A, Deng M, Zhang Y, Wang J, Atsuta N, et al. Replication analysis of SNPs on 9p21.2 and 19p13.3 with amyotrophic lateral sclerosis in East Asians. *Neurobiol Aging.* 2011;32:757. e13–4.