

# Modelling population genetic screening in rare neurodegenerative diseases

Thomas P Spargo MSc<sup>1,\*</sup>, Alfredo Iacoangeli PhD<sup>1,2,3,\*,#</sup>, Mina Ryten PhD<sup>4,5,6</sup>, Francesca Forzano MD<sup>7</sup>, Neil Pearce PhD<sup>8</sup>, and Ammar Al-Chalabi PhD<sup>1,9,#</sup>.

<sup>1</sup>Maurice Wohl Clinical Neuroscience Institute, King's College London, Department of Basic and Clinical Neuroscience, London, UK

<sup>2</sup>Department of Biostatistics and Health Informatics, King's College London, London, UK

<sup>3</sup>NIHR Maudsley Biomedical Research Centre (BRC) at South London and Maudsley NHS Foundation Trust and King's College London, London, UK

<sup>4</sup>Genetics and Genomic Medicine, Great Ormond Street Institute of Child Health, University College London, London, UK

<sup>5</sup>NIHR Great Ormond Street Hospital Biomedical Research Centre, University College London, London, UK

<sup>6</sup>Department of Clinical Genetics, Great Ormond Street Hospital, London, UK

<sup>7</sup>Department of Clinical Genetics, Guy's and St Thomas NHS Foundation Trust, London, UK

<sup>8</sup>Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK.

<sup>9</sup>King's College Hospital, Bessemer Road, London, SE5 9RS, UK.

\*co-first author

#correspondence should be addressed to [alfredo.iacoangeli@kcl.ac.uk](mailto:alfredo.iacoangeli@kcl.ac.uk) and [ammar.al-chalabi@kcl.ac.uk](mailto:ammar.al-chalabi@kcl.ac.uk)

## Abstract

**Importance:** Genomic sequencing enables rapid identification of a breadth of genetic variants. For clinical purposes, sequencing for small genetic variations is considered a solved problem, while challenges remain for structural variants given the lower sensitivity and specificity. Interest has recently risen among governing bodies in developing protocols for population-wide genetic screening. However, usefulness is constrained when the probability of being affected by a rare disease remains low despite a positive genetic test. This is a common scenario in neurodegenerative disorders. The problem is recognised among statisticians and statistical geneticists but less well understood by clinicians and researchers who will act on these results, and by the general public who might access screening services directly without the appropriate support for interpretation.

**Observations:** We explore the probability of subsequent disease following genetic screening of several of variants, both single nucleotide variants (SNVs) and larger repeat expansions, for two neurological conditions, Huntington's disease (HD) and amyotrophic lateral sclerosis (ALS), comparing with screening for phenylketonuria which is well established. The risk following a positive screening test was 0.5% for *C9orf72* in ALS and 0.4% for *HTT* in HD, when testing repeat expansions for which the test had sub-optimal performance (sensitivity=99% and specificity =90%), and 12.7% for phenylketonuria and 10.9% for ALS *SOD1*, when testing pathogenic SNVs (sensitivity=99.96% and

specificity=99.95%). Subsequent screening confirmation via PCR for *C9orf72* led to a 2% risk of developing ALS as a result of the reduced penetrance (44%).

**Conclusions and Relevance:** We show that risk following a positive screening test result can be strikingly low for rare neurological diseases even for fully penetrant variants such as *HTT* if the test has sub-optimal performance. Accordingly, to maximise the utility of screening, it is vital to prioritise protocols of very high sensitivity and specificity, careful selection of markers for screening, giving regard to clinical interpretability, actionability, high penetrance, and secondary testing to confirm positive findings.

Keywords: genetic screening, genomics, next-generation sequencing, health informatics, mathematical model, Bayesian probability tools

## Introduction

Single nucleotide variants (SNVs) and small insertions and deletions (indels) are now genotyped in next-generation sequencing data with over 99% accuracy (1-3), while genotyping larger, structural, variants is often less reliable (4). For clinical purposes, the validity of sequencing for small genetic variations is considered a solved problem, while challenges remain for structural variants.

Widened availability of sequencing data has favoured our understanding of the role of genetic factors in various phenotypes (5-8). Genetic testing has therefore become valuable for healthcare and gained popularity among consumers who can access it directly, without advice from trained clinicians to guide interpretation of results (9-11).

Interest has recently risen among governing bodies in developing protocols for population-wide genetic screening; such initiatives are being rolled out in the UK and considered in the US (12-18). Genetic screening involves testing a population for genetic variants indicative of risk for specific diseases to identify people with either higher predisposition of developing that disease or the potential to pass it on to their offspring. This approach utilises modern sequencing techniques to evaluate multiple genes associated with selected traits. In contrast 'targeted' tests are those performed because of some suggestion that a person may harbour disease variants (e.g., symptoms or family history of disease). Although screening is relevant to liabilities ranging between monogenic and polygenic (cf. (16, 19)), we focus here on screening for pathogenic variants with monogenic associations with rare diseases, particularly as applied to neurodegenerative disorders.

Although no widespread implementation of genetic screening protocols currently exist internationally, comparable metabolic screening, testing neonates for metabolite markers of metabolic diseases, is routine in many countries (17, 20, 21). Positive metabolic tests are typically validated with secondary testing, including targeted genetic tests (9, 22, 23).

The utility of genetic screening can be assessed by the extent to which action can be taken following a positive test: its actionability (24, 25). One key tenet of actionability is the probability of having or later developing a disease following a positive test. Yet post-test disease probability can be strikingly low where disease risk prior to testing is low, as would be for population screening for rare diseases (26).

Bayesian inference, which is routine within clinical decision making (27), can be used to understand post-test ('posterior') disease risk. Under this logic, disease probability following a test can be inferred given existing knowledge of the probability of other relevant events. Key considerations to understanding post-test disease risk, beyond pre-test (also known as 'prior') disease risk, include the genetic marker penetrance, its frequency among people displaying disease symptoms, and the sensitivity and specificity of the test (analytic validity). This reasoning is therefore highly relevant to screening for rare neurodegenerative diseases, for which genetic causes are typically rare variants of variable penetrance (28).

This article overviews important considerations for genetic screening of rare disorders and it presents several case studies focused on neurodegenerative diseases. Considering conditional probability in medical decision making is not novel but these concepts must be

emphasised in the genomic medicine era, since many results that could be obtained within large-scale indiscriminate testing of genetic variation across a population will not be actionable and may be misinterpreted. We modelled genetic screening for Huntington's disease (29), HD, and amyotrophic lateral sclerosis (30), ALS, using Bayesian logic to examine the probability of disease following a positive test result for a dichotomous ALS and HD genetic markers. We additionally modelled screening for phenylketonuria (PKU) (31), to compare genetic and metabolic screening.

## Methods:

### Definition of probability in the Bayesian framework

We use Bayesian logic to calculate the probability of having or subsequently manifesting disease  $D$  following a test result indicating presence or absence of marker  $M$  (the genetic variant).  $M$  is associated with increased liability of  $D$ , and positive test result  $T$  indicates presence of  $M$ , while negative test result  $T'$  indicates its absence, denoted  $M'$ . Disease risk following a positive result is denoted  $P(D|T)$ , using  $P(D|T')$  for a negative result.

Supplementary Materials 1 summarises the underlying logic. We assume that all model parameters represent binary events. This was a necessary simplification of reality as, for example, disease severity is not considered.

We estimate  $P(D|T)$  and  $P(D|T')$  using the following input parameters:

- $P(D)$ , probability of a person having or later manifesting disease  $D$  prior to testing
- $P(M|D)$ , frequency of marker  $M$  among those affected by  $D$
- $P(D|M)$ , penetrance, probability of having or later manifesting  $D$  for people harbouring  $M$
- $P(T|M)$ , sensitivity (true positive rate) of the testing procedure for detecting  $M$
- $P(T'|M')$ , specificity (true negative rate) of the testing procedure for identifying the absence of  $M$

Bayes theorem is applied to derive the total probability of harbouring disease marker  $M$ ,

*Equation 1*

$$P(M) = \frac{P(D) \times P(M|D)}{P(D|M)}$$

and of disease  $D$  manifesting given the absence of  $M$ ,

*Equation 2*

$$P(D|M') = \frac{P(D) \times (1 - P(M|D))}{(1 - P(M))}.$$

We next calculate the total probability of positive test result  $T$ ,  $P(T)$ , according to the sensitivity and specificity of the test and the probabilities of  $M$  being present versus absent:

*Equation 3*

$$P(T) = P(T|M) \times P(M) + (1 - P(T'|M')) \times (1 - P(M)).$$

Bayes theorem is then used to derive the probabilities of  $M$  being present after positive,

Equation 4

$$P(M|T) = \frac{P(M) \times P(T|M)}{P(T)},$$

and negative,

Equation 5

$$P(M|T') = \frac{P(M) \times (1 - P(T|M))}{(1 - P(T))},$$

results.

The probabilities of manifesting disease  $D$  (which has conditional independence from  $T$  when considering  $M$ ) after receiving positive test result  $T$ ,

Equation 6

$$P(D|T) = P(D|M) \times P(M|T) + P(D|M') \times (1 - P(M|T)),$$

and negative test result  $T'$ ,

Equation 7

$$P(D|T') = P(D|M) \times P(M|T') + P(D|M') \times (1 - P(M|T')),$$

can then be determined.

## Case studies

The Project MinE (32), ALS variant server (33), ClinVar (34), and gnomAD databases were searched alongside the previous databases and next-generation sequencing tool benchmarking reports to obtain data for the included case studies. Most input parameters were defined using data from published literature and online databases. Sensitivity ( $P(T|M)$ ) and specificity ( $P(T'|M')$ ) were defined by performance benchmarks for variant calling with state-of-the-art genomic sequencing techniques specialised for genotyping particular variant types (see Supplementary Materials 2; Table S1). Although analytical accuracy will vary across the genome and other sources of error exist, these heuristics are sufficient for our purposes.

Table 1 presents the parameter estimates and post-test disease risks calculated across various scenarios. A comprehensive description of parameter ascertainment, including penetrance estimation, is given in Supplementary Materials 3; Table S2 summarises the assumptions and corresponding reality.

### Case 1 – Huntington's disease

HD is a late-onset Mendelian disease with autosomal dominant inheritance caused by a trinucleotide, CAG, short tandem repeat expansion (STRE) in the *HTT* gene (OMIM: 613004). We let  $M$  be a CAG expansion of >40 repeat units, which would have complete penetrance in a normal lifespan (29).

We modelled two scenarios for this example, (1) as in genetic screening, defining pre-test disease probability by baseline risk of HD in a general population and (2) as a targeted test, considering pre-test disease probability for an individual whose parent harbours the fully penetrant *HTT* STRE and who has a 0.5 probability of inheriting *M* (we have not modelled genetic anticipation (35)).

#### *Case 2 – amyotrophic lateral sclerosis*

ALS is a late-onset disease with locus and allelic heterogeneity; genetic associations with risk and phenotype modification range between monogenic and polygenic. Variants in at least 40 genes are associated with ALS (30, 36-38). *SOD1* (OMIM: 147450) and *C9orf72* (OMIM: 614260) are the most frequently implicated genes, where variants of each account for fewer than 10% of cases. Autosomal dominant inheritance is typical for most people with a known genetic disease cause.

We modelled several definitions of markers for ALS risk, drawing from three of the commonest ALS genes: *SOD1*, *C9orf72*, and *FUS* (OMIM: 137070). *SOD1*- and *FUS*-linked ALS is typically attributed to SNVs and many pathogenic variants with varying strength of supporting evidence have been reported in these genes (39). The pathogenic form of *C9orf72* is a hexanucleotide, GGGGCC, STRE associated principally with the onset of either one or both of ALS and frontotemporal dementia (40). Known variants in these genes have typically incomplete penetrance; examples include ~90-100% penetrance for *SOD1* p.A5V and ~45% for the *C9orf72* STRE (41, 42).

The definitions of *M* modelled in this case study were:

- *SOD1* (all) – *M* includes any rare variant reported in people with ALS of European ancestry contained within the meta-analysis sample set from which the variant frequencies were derived (see Supplementary Materials 3.2) (37).
- *SOD1* (A5V) – *M* represents the pathogenic *SOD1* p.A5V variant, one of the most common *SOD1* variants among North American ALS populations, characterised by high penetrance (41, 43).
- *FUS* (all) – *M* includes any rare variant reported in people with ALS of European ancestry contained within the meta-analysis sample set from which the variant frequencies were derived (see Supplementary Materials 3.2) (37).
- *FUS* (ClinVar) – *M* includes any of 21 *FUS* variants reported as pathogenic or likely pathogenic for ALS within ClinVar and present within databases of familial and sporadic ALS (see Table S3) (32-34).
- *C9orf72* – *M* represents a pathogenic *C9orf72* STRE of  $30 \leq$  repeat GGGGCC units within the first intron of the *C9orf72* gene.

In the *SOD1* (all) and *FUS* (all) scenarios, *M* encompasses variants classed as pathogenic, likely pathogenic, and variants of uncertain significance (44). It is appropriate to model these scenarios since many variants of uncertain significance in ALS-implicated genes have a high probability of being deleterious and should not necessarily be ignored (38).

For the *C9orf72* marker, we modelled two scenarios: (1) genetic screening with sensitivity and specificity defined by performance of existing tools for genotyping STREs in sequencing data; (2) using repeat-primed polymerase chain reaction with amplicon-length analysis (45) as a secondary test to validate a positive result from screening via sequencing in scenario 1.

240

241 *Case 3 – phenylketonuria*

242 PKU is an autosomal recessive disease with infantile onset, caused by variants in the *PAH*  
243 gene (OMIM: 612349) (31); most variants pathogenic for PKU are SNVs. *M* represents being  
244 homozygous or compound heterozygous for any of the three most common *PAH* variants  
245 recorded in European populations of people with PKU: p.Arg408Trp, c.1066-11G>A,  
246 p.Arg261Gln (31).

247

248 We modelled two testing scenarios for PKU: (1) genetic screening, with pre-test disease  
249 probability defined per the baseline population risk of PKU, and (2) secondary testing to  
250 confirm positive results obtained using tandem mass spectrometry (46) as in established  
251 metabolic screening protocols (see Supplementary Materials 3.3).

Table 1. Input parameters and disease risk estimates following testing for all case study scenarios.

Parameter ascertainment is comprehensively described within Supplementary Materials 3. HD = Huntington's disease; ALS = amyotrophic lateral sclerosis; PKU = phenylketonuria. SNV = single nucleotide variant; STRE = short tandem repeat expansion. <sup>§</sup>Estimates are based on populations of predominantly European ancestry – 95% confidence intervals shown for newly derived estimates in the ALS case study; \*includes FUS variants classified as pathogenic or likely pathogenic for ALS in the ClinVar database and recorded within ALS population databases (see Table S3, Table S4); <sup>¶</sup>defined by variant calling performance benchmarks of tools for genotyping in sequencing data by variant type (see Table S1) and, where marked <sup>†</sup>, by aggregate laboratory accuracy for genotyping C9orf72 STRE with repeat-primed polymerase chain reaction and amplicon-length analysis (45). <sup>⊲</sup>Risk following positive results in primary screening and confirmatory tests relative to a negative screening result (probability of PKU given a negative metabolic screening result is approximated as  $1 \times 10^{-6}$ ). Probabilities are defined in the methods.

Case study	Gene containing marker (case study scenario)	Variant type	Pre-test disease probability <sup>§</sup>	Marker frequency in people affected <sup>§</sup>	Penetrance <sup>§</sup>	Test sensitivity <sup>¶</sup>	Test specificity <sup>¶</sup>	Disease risk after positive test	Disease risk after negative test	Relative disease risk after positive rather than negative test
-	-	-	P(D)	P(M D)	P(D M)	P(T M)	P(T' M')	P(D T)	P(D T')	-
<b>1: HD</b>	HTT (screening)	STRE	0.000410	1.000	1.000	0.990	0.900	0.00404	0.00000456	887
	HTT (targeted)	STRE	0.500	1.000	1.000	0.990	0.900	0.908	0.011	82.7
<b>2: ALS</b>	SOD1 (all)	SNV	0.00333	0.0188 (0.0138, 0.0238)	0.701 (0.491, 0.926)	0.9996	0.9995	0.109	0.00327	33.3
	SOD1 (A5V)	SNV	0.00333	0.000529 (4.43x10 <sup>-5</sup> , 0.00101)	0.91	0.9996	0.9995	0.00683	0.00333	2.05
	FUS (all)	SNV	0.00333	0.00425 (0.0023, 0.0062)	0.579 (0.291, 0.884)	0.9996	0.9995	0.0302	0.00332	9.09
	FUS (ClinVar*)	SNV	0.00333	0.00251 (0.00125, 0.00377)	0.536 (0.211, 0.877)	0.9996	0.9995	0.0194	0.00333	5.84
	C9orf72	STRE	0.00333	0.0635 (0.0538, 0.0732)	0.439 (0.358, 0.520)	0.990	0.900	0.00519	0.00313	1.66
	C9orf72 (positive sequencing screening confirmation)	STRE	0.0052	0.0635 (0.0538, 0.0732)	0.439 (0.358, 0.520)	0.95 <sup>†</sup>	0.98 <sup>†</sup>	0.0198	0.00489	4.06 (6.35 <sup>⊲</sup> )
<b>3: PKU</b>	PAH (screening)	SNV	0.000100	0.743	0.892	0.9996	0.9995	0.127	0.0000257	4,961
	PAH (positive metabolic screening confirmation)	SNV	0.167	0.743	0.892	0.9996	0.9995	0.889	0.0497	17.9 (889,000 <sup>⊲</sup> )



## Results and Discussion

### Post-test disease probability

#### *Screening versus diagnostic testing*

Across case studies, we showed low probability of disease following positive results within contextually-blind genetic screening scenarios; risk ranged between 12.7% and 0.4% (see Table 1). Disease risk was always negligible following a negative test result, indicating absence of the tested variant, as would be seen for any rare trait.

The HD case study illustrates the distinction between contextually-blind screening and diagnostic testing for rare diseases: following a positive test result, lifetime HD risk was high (90.8%) using targeted testing but low (0.4%) in screening. This difference reflects the effect of the sub-optimal performance (sensitivity=99% and specificity =90%) of the test for repeat expansions generating a large number of false positives in population screening despite the complete penetrance of *HTT*. Indeed, unlike screening, targeted testing is performed based on some indication of a person's elevated disease risk (e.g., existing disease symptoms or family history). Accordingly, pre-test disease probability is greater. Inherently low pre-test disease probability will be a pervasive issue in screening for rare diseases.

#### *Relative risk and secondary testing*

The utility of a test for identifying at-risk individuals can be examined based on relative disease risk following positive rather than negative test results: utility is limited when risk is only minimally greater for people testing positive rather than negative. This is observed in the ALS *C9orf72* case study, where risk was only 1.7 times greater (Table 1) following a positive screening from sequencing alone, despite this variant being the most common genetic cause of ALS (37).

We additionally observed 6.35 times greater ALS risk for a person testing positive on both screening for *C9orf72* and a secondary test than for a person testing negative on the initial screening. This increased relative risk reflects that a person testing positive on two independent measures of disease risk has greater absolute probability of disease than after the initial screening result alone.

We emphasise that secondary testing is important to increase certainty in positive tests. The PKU case study demonstrates its potentially sizable impact. A positive screening result using the established metabolic approach alone indicated 16.7% PKU risk, versus 12.7% within genetic screening. The metabolic marker, which is universal across people with PKU and indicates existing disease manifestation, eclipses need for genetic screening for PKU, marked by variants with incomplete penetrance that are not present for all people with PKU. However, the genetic test remains useful for validating the positive metabolic screening result (17): probability of PKU following a confirmatory genetic test conducted on the basis of a positive metabolic screening result was 88.9%.

The overall benefit of secondary testing will however differ by scenario. For ALS, the risk remained moderate (~2%) despite two positive test results for *C9orf72*.

#### *Constraints upon post-test disease probability*

Figure 1, Figure 2, and Figure 3 illustrate how the post-test disease probability reduces as the probability of any test, disease, or marker characteristic decreases. Sensitivity and specificity critically constrain certainty about post-test disease risk, and this role is amplified as the other parameter probabilities decrease. Figure 1 particularly demonstrates the increased role of specificity in rarer diseases, where disease risk following a positive test result will be moderated only by penetrance in a protocol of perfect specificity.

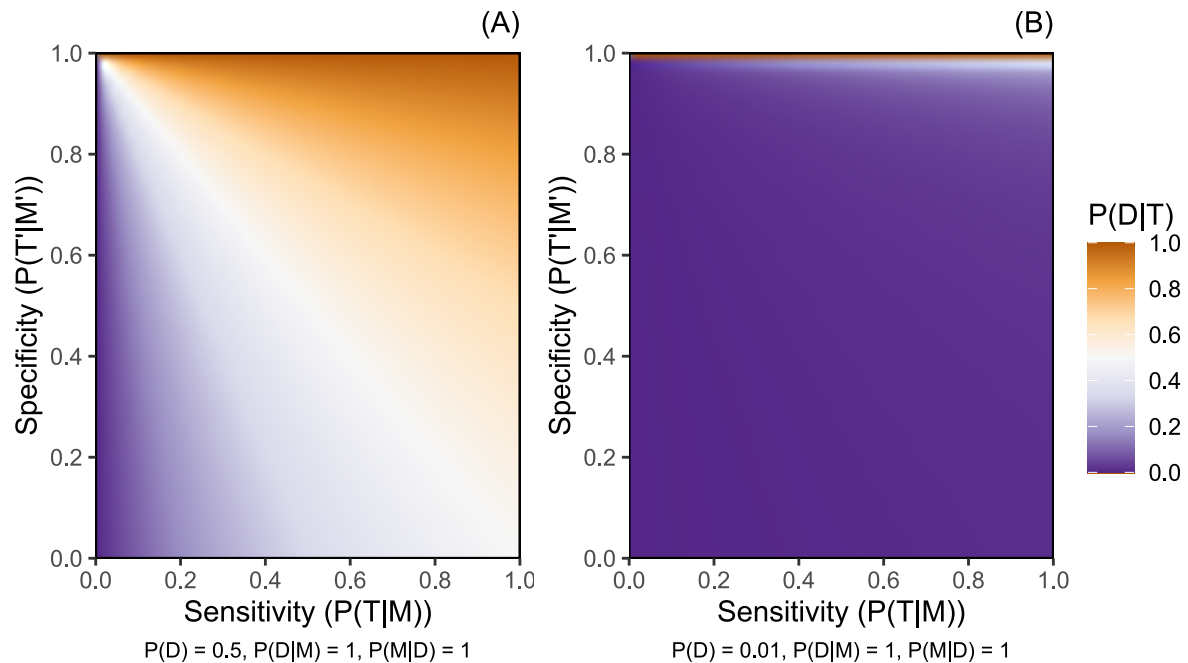


Figure 1. Probability of a disease given a positive genetic test result for a marker of increased disease risk ( $P(D|T)$ ) according to the sensitivity ( $P(T|M)$ ) and the specificity ( $P(T'|M')$ ) of the testing protocol

Panel A presents this for a disease with pre-test probability ( $P(D)$ ) of 0.5, while panel B presents a disease with  $P(D)$  of 0.01. Penetrance is complete ( $P(D|M) = 1$ ) and variant  $M$  is harboured by all people with disease  $D$  ( $P(M|D) = 1$ ) in both panels. Probabilities are defined in the methods.

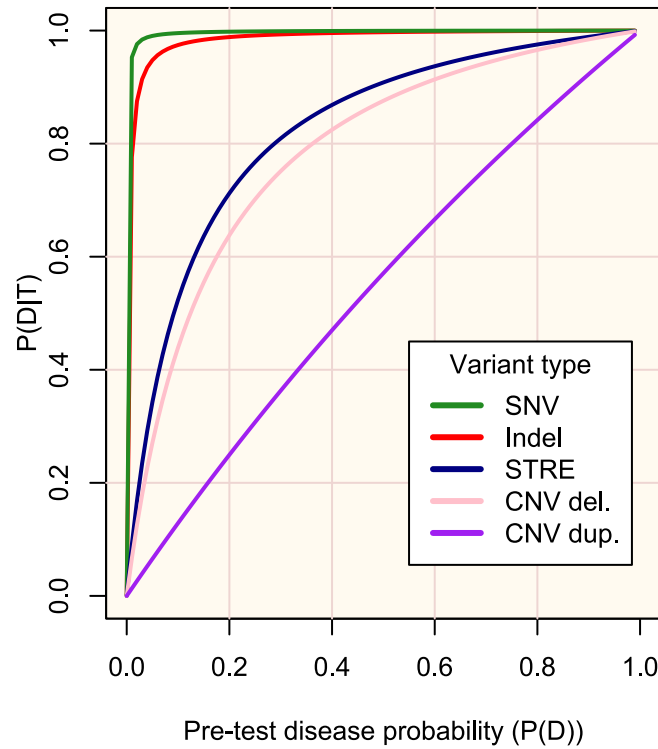
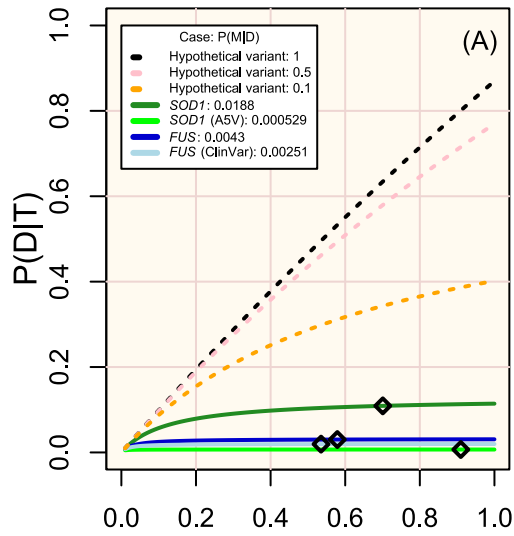
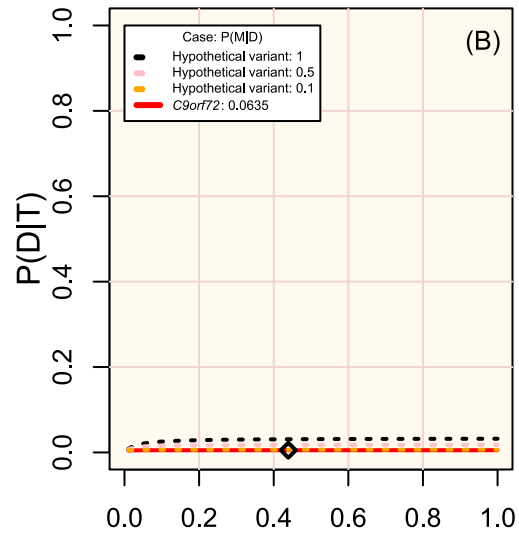


Figure 1. Probability of disease  $D$  following a positive genetic test result for marker  $M$  ( $P(D|T)$ ) according to pre-test disease probability ( $P(D)$ )

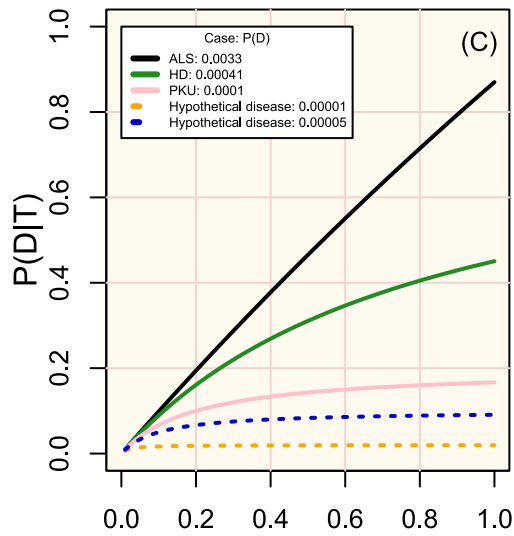
$M$  occurs in all people with  $D$  ( $P(M|D) = 1$ ) and has complete penetrance ( $P(D|M) = 1$ ). Plot lines are defined according to sensitivity ( $P(T|M)$ ) and specificity ( $P(T'|M')$ ) of existing protocols for genotyping variant types (see Supplementary Materials 2): single nucleotide variant (SNV), test sensitivity  $P(T|M) = 0.9996$ , test specificity  $P(T'|M') = 0.9995$ ; small insertion or deletion (Indel), test sensitivity  $P(T|M) = 0.9962$ , test specificity  $P(T'|M') = 0.9971$ ; short tandem repeat expansion (STRE), test sensitivity  $P(T|M) = 0.99$ , test specificity  $P(T'|M') = 0.90$ ; copy number variant (CNV) – del. (deletion), test sensitivity  $P(T|M) = 0.289$ , test specificity  $P(T'|M') = 0.959$ ; CNV – dup. (duplication), test sensitivity  $P(T|M) = 0.1020$ , test specificity  $P(T'|M') = 0.9233$ . Probabilities are defined in the methods.



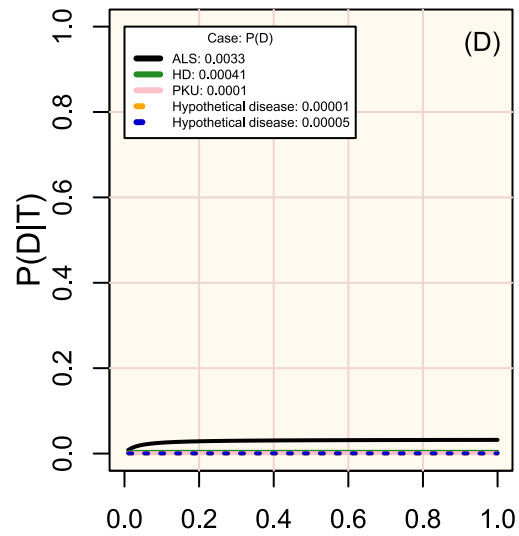
Penetrance ( $P(D|M)$ )  
 $P(D) = 0.0033$ ,  $P(T|M) = 0.9996$ ,  $P(T'|M') = 0.9995$



Penetrance ( $P(D|M)$ )  
 $P(D) = 0.0033$ ,  $P(T|M) = 0.99$ ,  $P(T'|M') = 0.90$



Penetrance ( $P(D|M)$ )  
 $P(M|D) = 1$ ,  $P(T|M) = 0.9996$ ,  $P(T'|M') = 0.9995$



Penetrance ( $P(D|M)$ )  
 $P(M|D) = 1$ ,  $P(T|M) = 0.99$ ,  $P(T'|M') = 0.90$

Figure 3. Change in disease risk following a positive test result for a marker of increased disease risk ( $P(D|T)$ ) according to penetrance ( $P(D|M)$ )

Panels A and B display modelled and hypothetical markers of ALS which differ in frequency across people affected by ALS ( $P(M|D)$ ), where pre-test disease probability ( $P(D)$ ) is 0.0033 and diamonds mark the penetrance estimated for non-hypothetical variants (see Table 1). Panels C and D display diseases in which  $P(M|D) = 1$  and with  $P(D)$  set in line with the modelled case studies or a hypothetical rare disease. Analytic validity parameters are defined according to the performance of sequencing tools for genotyping single nucleotide variants in A and C, and of short tandem repeat expansions in B and D (see Table S1; Figure 2). Probabilities are defined in the methods.

Figure 2 illustrates the impact of the genetic test's performance (sensitivity and specificity) in existing protocols for genotyping different variant types (SNVs, indels, STREs and CNVs) on the probability of disease, given a positive test result. As is well recognized, high sensitivity and specificity are crucial for maximizing the utility of testing, especially for rare diseases. The respective trade-offs between prioritising each of these must be regarded: high sensitivity is required to detect markers, while high specificity increases confidence in positive results. Established screening protocols prioritise high sensitivity to maximise detection of at-risk individuals, with confirmatory secondary testing being vital to minimise false-positive results (9, 22, 23).

Since the characteristics of diseases and associated variants are all pre-determined within a population, disease markers (variants) screened should be chosen carefully. Figure 3 illustrates how the prevalence of a marker among affected individuals, its penetrance, and the performance of the test (sensitivity and specificity) influence the probability of developing the disease given a positive test result. The most useful variants are those that are more prevalent among affected individuals, have high penetrance, and can be genotyped with high sensitivity and specificity (see examples of real and hypothetical diseases and markers in Figure 3, panels A-D).

## **Practical implementation of genetic screening**

### *Marker selection*

Before a marker is used in screening, its relevance across people must be evaluated, recognising that this may vary by ancestry. For instance, particular variants may be less common or only present in certain populations, and penetrance can also vary between them (36, 43). Screening protocols must therefore account for these differences to prevent systemic inequalities, especially for minorities which are often under-studied and therefore have limited genetic information available.

Regard must be given to the clinical interpretability of selected markers. We illustrated several approaches to defining markers in the ALS case study. Within the *SOD1* (all) scenario, disease risk is marked by an aggregation of putatively pathogenic *SOD1* variants. Without curation, the relationship to disease varies across them. E.g., *SOD1* p.I114T has substantially lower penetrance than p.A5V, and many potentially relevant variants have uncertain significance (34, 38, 39, 42). Curation could involve defining a positive result as presence of variants with sufficient evidence of pathogenicity (44), as in the *FUS* (ClinVar) scenario, or as harbouring specific variants, as in the *SOD1* (A5V) scenario.

*De novo* variants and variants of uncertain significance present a substantial challenge for screening since they will frequently be identified, yet must be set aside until variant interpretation is possible despite potentially being deleterious (12). PKU demonstrates the scale of this issue for rare diseases with multiple implicated variants, as 55% of deleterious *PAH* genotypes are observed uniquely (31).

### *Utility over time and actionability*

As genetic screening is possible from birth, while non-genetic methods may not be, age of viability for screening methods should be evaluated. For late-onset diseases, early genetic screening may enable preventative treatments to at-risk individuals or close monitoring for

prodromal disease markers. For instance, rapid eye movement sleep behaviour disorder is a prodromal feature for Parkinson's disease (47) and monitoring of at-risk individuals may enable early intervention. The influence of time upon treatment viability and effectiveness must also be considered. For example, genetic therapy has potential utility for preventing or delaying onset of degenerative disorders (48).

The ultimate benefit of early identification of disease risk through genetic screening is contingent upon the actionability of the result. A framework of actionability (24, 25), shown to align with laypersons' views on treatment acceptability (49), states that actionability is determined by: disease likelihood and severity, intervention effectiveness in disease minimisation or prevention, and the consequence of the intervention to a person and risk if not performed. Each of these elements are critical when selecting traits and markers for genetic screening and for the clinical interpretation of results.

### **Limitations**

The Bayesian logic in the case studies simplifies genotype-phenotype relationships and cannot address all considerations relevant to clinical genetic testing. Phenotype variability is not considered. Other factors include: polygenicity and oligogenicity, pleiotropy, the role of genetic and environmental modifying factors, and that of additive genetic effects in recessive conditions and heterozygous carriers of pathogenic variants. Such influences can fundamentally impact both the probability that a disease will manifest and its severity. For instance, although spinal muscular atrophy is caused by partial or complete biallelic deletion of the *SMN1* gene, additional copies of *SMN2* reduce disease expressivity by mitigating loss of *SMN1* function (50). Any results from genetic screening must be interpreted within the wider context of that disease and its modifiers.

### **Conclusion**

We have shown that risk following a positive screening test result can be strikingly low for rare neurological diseases. Accordingly, to maximise the utility of screening, we suggest prioritising protocols of very high sensitivity and specificity, careful selection of markers for screening, giving regard to clinical interpretability, and secondary testing to confirm positive findings.

A key advantage of a genetic screening approach for late-onset diseases is that these markers can be examined across the lifespan. Hence, positive test results could be useful for targeting people for prevention, and for monitoring of prodromal features.

Although considering disease risk within a Bayesian context is not novel, it is important to stress the considerations raised here at a time when governments evaluate implementation of genomic sequencing for population screening and as access to genetic testing outside healthcare settings increases. While genetic screening has many potential benefits, the limitations of such an approach should be properly understood. Policy makers must consider the impact of a positive test result on large numbers of people that will never develop a given disease, a particularly salient issue for late-onset diseases. Although not the present focus, the substantial ethical and social considerations raised in conjunction to screening must also be regarded (13).

## **Data Availability**

Data sharing not applicable

## **Acknowledgements**

We wish to thank Professor Naomi Wray of the Queensland Brain Institute for her contribution to project conceptualisation and methodology. We also thank Dr Leslie Biesecker of the National Human Genome Research Institute of the National Institutes of Health for contributing to our discussions around this work.

## **Author contributions**

**TPS:** Conceptualisation, Methodology, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review and Editing, Visualisation. **AI:** Conceptualisation, Methodology, Writing – Original Draft, Writing – Review and Editing, Supervision. **MR:** Conceptualisation, Writing – Review and Editing. **FF:** Conceptualisation, Writing – Review and Editing. **NP:** Conceptualisation, Writing – Review and Editing. **AAC:** Conceptualisation, Methodology, Writing – Review and Editing, Supervision.

## **Funding**

AAC is an NIHR Senior Investigator. This is an EU Joint Programme-Neurodegenerative Disease Research (JPND) project. The project is supported through the following funding organizations under the aegis of JPND (United Kingdom, Medical Research Council MR/L501529/1 and MR/R024804/1; Economic and Social Research Council ES/L008238/1) and through the Motor Neurone Disease Association. This study represents independent research part funded by the NIHR Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The work leading up to this publication was funded by the European Community's Horizon 2020 Programme (H2020-PHC-2014-two-stage; grant 633413). This work was part funded by my Name's5 Doddie Foundation, and Alan Davidson Foundation. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, King's College London, or the Department of Health and Social Care. AI is funded by South London and Maudsley NHS Foundation Trust, MND Scotland, Motor Neurone Disease Association, National Institute for Health and Care Research, Spastic Paraplegia Foundation, Rosetrees Trust, Darby Rimmer MND Foundation, the Medical Research Council (UKRI) and Alzheimer's Research UK.

## **Role of the funding source**

The funding sources had no role in the writing of this manuscript or the decision to submit it for publication.

## **Declaration of interests**

AI, MR, FF, and NP declare no competing interests. TPS declares employment with AstraZeneca. AAC declares royalties from The Brain, A Beginner's Guide: Oneworld Publications, and The Genetics of Complex Human Diseases, A Laboratory Manual. AAC reports consultancies or advisory boards for Amylyx, Apellis, Biogen, Brainstorm, Cytokinetics, GenieUs, GSK, Lilly, Mitsubishi Tanabe Pharma, Novartis, OrionPharma,

Quralis, Sano, Sanofi, and Wave Pharmaceuticals. AAC is the deputy editor for Brain. AAC has planned patents for 'Use of CSF-Neurofilament determinations and CSF-Neurofilament thresholds of prognostic and stratification value with regards to response to therapy in neuromuscular and neurodegenerative diseases'. AAC is Scientific co-lead for United to End MND Partnership Coalition and Guarantor of Brain. AAC has stock options for Quralis

#### References:

1. Iacoangeli A, Al Khleifat A, Sproviero W, Shatunov A, Jones AR, Morgan SL, et al. DNAscan: personal computer compatible NGS analysis, annotation and visualisation. BMC Bioinform. 2019;20(1):213.
2. Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat Biotechnol. 2019;37(10):1155-62.
3. Illumina. Accuracy Improvements in Germline Small Variant Calling with the DRAGEN Platform 2019 07/09/2019. Available from: <https://science-docs.illumina.com/documents/Informatics/dragen-v3-accuracy-appnote-html-970-2019-006/Content/Source/Informatics/Dragen/dragen-v3-accuracy-appnote-970-2019-006/dragen-v3-accuracy-appnote-970-2019-006.html>.
4. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. Genome Biol. 2019;20(1):117.
5. Rhoades R, Jackson F, Teng S. Discovery of rare variants implicated in schizophrenia using next-generation sequencing. J transl genet genom. 2019;3:1.
6. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011;12(7):499-510.
7. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018;50(8):1112-21.
8. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.
9. Hörster F, Kölker S, Loeber JG, Cornel MC, Hoffmann GF, Burgard P. Newborn Screening Programmes in Europe, Arguments and Efforts Regarding Harmonisation: Focus on Organic Acidurias. JIMD Rep. 2017;32:105-15.
10. Perrone F, Cacace R, Van Mossevelde S, Van den Bossche T, De Deyn PP, Cras P, et al. Genetic screening in early-onset dementia patients with unclear phenotype: relevance for clinical diagnosis. Neurobiol Aging. 2018;69:292.e7-.e14.
11. Majumder MA, Guerrini CJ, McGuire AL. Direct-to-Consumer Genetic Testing: Value and Risk. Annu Rev Med. 2021;72(1):151-66.
12. Murray MF, Evans JP, Khoury MJ. DNA-Based Population Screening: Potential Suitability and Important Knowledge Gaps. JAMA. 2019.
13. Dickinson JA, Pimlott N, Grad R, Singh H, Szafran O, Wilson BJ, et al. Screening: when things go wrong. Can Fam Physician. 2018;64(7):502-8.
14. Jansen ME, Lister KJ, van Kranen HJ, Cornel MC. Policy Making in Newborn Screening Needs a Structured and Transparent Approach. Front Public Health. 2017;5(53).
15. Genome UK: The future of healthcare. <https://www.gov.uk/government/publications/genome-uk-the-future-of-healthcare> 2020.



16. Moorthie S, Hall A, Janus J, Brigden T, Villiers CBd, Blackburn L, et al. Polygenic scores and clinical utility. University of Cambridge; 2021.
17. Adhikari AN, Gallagher RC, Wang Y, Currier RJ, Amatuni G, Bassaganyas L, et al. The role of exome sequencing in newborn screening for inborn errors of metabolism. *Nat Med*. 2020;26(9):1392-7.
18. Centers for Disease Control and Prevention. Use of Genomics in Newborn Screening Programs: The Promise and Challenges 2021 [Available from: [https://www.cdc.gov/genomics/events/newborn\\_screening\\_2021.htm](https://www.cdc.gov/genomics/events/newborn_screening_2021.htm)].
19. Pain O, Gillett AC, Austin JC, Folkersen L, Lewis CM. A tool for translating polygenic scores onto the absolute scale using summary statistics. *Eur J Hum Genet*. 2022;30:339-48.
20. Loeber JG, Burgard P, Cornel MC, Rigter T, Weinreich SS, Rupp K, et al. Newborn screening programmes in Europe; arguments and efforts regarding harmonization. Part 1 – From blood spot to screening result. *J Inherit Metab Dis*. 2012;35(4):603-11.
21. Tarini BA, Christakis DA, Welch HG. State Newborn Screening in the Tandem Mass Spectrometry Era: More Tests, More False-Positive Results. *Pediatrics*. 2006;118(2):448.
22. Southern KW, Munck A, Pollitt R, Travert G, Zanolla L, Dankert-Roelse J, et al. A survey of newborn screening for cystic fibrosis in Europe. *J Cyst Fibros*. 2007;6(1):57-65.
23. Rinaldo P, Zafari S, Tortorelli S, Matern D. Making the case for objective performance metrics in newborn screening by tandem mass spectrometry. *Ment Retard Dev Disabil Res Rev*. 2006;12(4):255-61.
24. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, et al. ClinGen – The Clinical Genome Resource. *N Engl J Med*. 2015;372(23):2235-42.
25. Hunter JE, Irving SA, Biesecker LG, Buchanan A, Jensen B, Lee K, et al. A standardized, evidence-based protocol to assess clinical actionability of genetic disorders associated with genomic variation. *Genet Med*. 2016;18(12):1258-68.
26. Biesecker LG. Genomic screening and genomic diagnostic testing-two very different kettles of fish. *Genome Med*. 2019;11(1):75.
27. Hunink MGM, Weinstein MC, Wittenberg E, Drummond MF, Pliskin JS, Wong JB, et al. *Decision Making in Health and Medicine: Integrating Evidence and Values*. 2nd ed. Cambridge: Cambridge University Press; 2014.
28. Bertram L, Tanzi RE. The genetic epidemiology of neurodegenerative disease. *J Clin Invest*. 2005;115(6):1449-57.
29. Langbehn DR, Brinkman RR, Falush D, Paulsen JS, Hayden MR. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. *Clin Genet*. 2004;65(4):267-77.
30. Brown RH, Al-Chalabi A. Amyotrophic Lateral Sclerosis. *N Engl J Med*. 2017;377(2):162-72.
31. Hillert A, Anikster Y, Belanger-Quintana A, Burlina A, Burton BK, Carducci C, et al. The Genetic Landscape and Epidemiology of Phenylketonuria. *Am J Hum Genet*. 2020;107(2):234-50.
32. van der Spek RAA, van Rheenen W, Pulit SL, Kenna KP, van den Berg LH, Veldink JH. The project MinE databrowser: bringing large-scale whole-genome sequencing in ALS to researchers and the public. *Amyotroph Lateral Scler Frontotemporal Degener*. 2019;20:432-40.
33. ALS Variant Server [Internet]. [cited 02/2021]. Available from: <http://als.umassmed.edu/>.

34. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 2018;46(D1):D1062-D7.
35. Paulson H. Repeat expansion diseases. *Handb Clin Neurol.* 2018;147:105-23.
36. Iacoangeli A, Al Khleifat A, Sproviero W, Shatunov A, Jones AR, Opie-Martin S, et al. ALSgeneScanner: a pipeline for the analysis and interpretation of DNA sequencing data of ALS patients. *Amyotroph Lateral Scler Frontotemporal Degener.* 2019;20:207-15.
37. Zou Z-Y, Zhou Z-R, Che C-H, Liu C-Y, He R-L, Huang H-P. Genetic epidemiology of amyotrophic lateral sclerosis: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry.* 2017;88:540-9.
38. Mehta PR, Iacoangeli A, Opie-Martin S, van Vugt J, Al Khleifat A, Bredin A, et al. The impact of age on genetic testing decisions in amyotrophic lateral sclerosis. *Brain.* 2022;145(12):4440-7.
39. Abel O, Powell JF, Andersen PM, Al-Chalabi A. ALSod: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. *Hum Mutat.* 2012;33(9):1345-51.
40. 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(2):257-68.
41. Cudkovic ME, McKenna-Yasek D, Sapp PE, Chin W, Geller B, Hayden DL, et al. Epidemiology of mutations in superoxide dismutase in amyotrophic lateral sclerosis. *Annals of Neurology.* 1997;41(2):210-21.
42. Spargo TP, Opie-Martin S, Bowles H, Lewis CM, Iacoangeli A, Al-Chalabi A. Calculating variant penetrance from family history of disease and average family size in population-scale data. *Genome Med.* 2022;14:141.
43. Saeed M, Yang Y, Deng HX, Hung WY, Siddique N, Dellefave L, et al. Age and founder effect of SOD1 A4V mutation causing ALS. *Neurology.* 2009;72(19):1634-9.
44. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-23.
45. Akimoto C, Volk AE, van Blitterswijk M, Van den Broeck M, Leblond CS, Lumbroso S, et al. A blinded international study on the reliability of genetic testing for GGGGCC-repeat expansions in C9orf72 reveals marked differences in results among 14 laboratories. *J Med Genet.* 2014;51(6):419.
46. Schulze A, Lindner M, Kohlmüller D, Olgemöller K, Mayatepek E, Hoffmann GF. Expanded newborn screening for inborn errors of metabolism by electrospray ionization-tandem mass spectrometry: results, outcome, and implications. *Pediatrics.* 2003;111(6):1399-406.
47. Ferini-Strambi L, Marelli S, Galbiati A, Rinaldi F, Giora E. REM Sleep Behavior Disorder (RBD) as a marker of neurodegenerative disorders. *Arch Ital Biol.* 2014;152(2-3):129-46.
48. Amado DA, Davidson BL. Gene therapy for ALS: A review. *Mol Ther.* 2021;29(12):3345-58.
49. Paquin RS, Mittendorf KF, Lewis MA, Hunter JE, Lee K, Berg JS, et al. Expert and lay perspectives on burden, risk, tolerability, and acceptability of clinical interventions for genetic disorders. *Genet Med.* 2019;21(11):2561-8.

611 50. Wadman RI, Jansen MD, Stam M, Wijngaarde CA, Curial CAD, Medic J, et al.  
612 Intragenic and structural variation of the SMN locus and clinical variability of spinal muscular  
613 atrophy. Brain Commun. 2020;2(2):fcaa075.  
614