

1 Modelling population genetic screening in rare neurodegenerative diseases

2
3
4 Thomas P Spargo MSc^{1,*}, Alfredo Iacoangeli PhD^{1,2,3,*,#}, Mina Ryten PhD^{4,5,6}, Francesca
5 Forzano MD⁷, Neil Pearce PhD⁸, and Ammar Al-Chalabi PhD^{1,9,#}.

6
7 ¹Maurice Wohl Clinical Neuroscience Institute, King's College London, Department of Basic
8 and Clinical Neuroscience, London, UK

9 ²Department of Biostatistics and Health Informatics, King's College London, London, UK

10 ³NIHR Maudsley Biomedical Research Centre (BRC) at South London and Maudsley NHS
11 Foundation Trust and King's College London, London, UK

12 ⁴Genetics and Genomic Medicine, Great Ormond Street Institute of Child Health, University
13 College London, London, UK

14 ⁵NIHR Great Ormond Street Hospital Biomedical Research Centre, University College
15 London, London, UK

16 ⁶Department of Clinical Genetics, Great Ormond Street Hospital, London, UK

17 ⁷Department of Clinical Genetics, Guy's and St Thomas NHS Foundation Trust, London, UK

18 ⁸Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London
19 WC1E 7HT, UK.

20 ⁹King's College Hospital, Bessemer Road, London, SE5 9RS, UK.

21
22
23
24 *co-first author

25 #correspondence should be addressed to alfredo.iacoangeli@kcl.ac.uk and ammar.al-chalabi@kcl.ac.uk

28 Abstract

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

39
40 **Observations:** We explore the probability of subsequent disease following genetic
41 screening of several of variants, both single nucleotide variants (SNVs) and larger repeat
42 expansions, for two neurological conditions, Huntington's disease (HD) and amyotrophic
43 lateral sclerosis (ALS), comparing with screening for phenylketonuria which is well
44 established. The risk following a positive screening test was 0.5% for *C9orf72* in ALS and
45 0.4% for *HTT* in HD, when testing repeat expansions for which the test had sub-optimal
46 performance (sensitivity=99% and specificity=90%), and 12.7% for phenylketonuria and
47 10.9% for ALS *SOD1*, when testing pathogenic SNVs (sensitivity=99.96% and

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

50

51 **Conclusions and Relevance:** We show that risk following a positive screening test result
52 can be strikingly low for rare neurological diseases. Accordingly, to maximise the utility of
53 screening, it is vital to prioritise protocols of very high sensitivity and specificity, careful
54 selection of markers for screening, giving regard to clinical interpretability, actionability,
55 high penetrance, and secondary testing to confirm positive findings.

56

57

58 Keywords: genetic screening, genomic, next generation sequencing, health informatics,
59 mathematical model, Bayesian probability tools

60 Introduction

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

66
67 Widened availability of sequencing data has favoured our understanding of the role of
68 genetic factors in various phenotypes (5-8). Genetic testing has therefore become valuable
69 for healthcare and gained popularity among consumers who can it directly, without advice
70 from trained clinicians to guide interpretation of results (9-11).

71
72 Interest has recently risen among governing bodies in developing protocols for population-
73 wide genetic screening; such initiatives are being rolled out in the UK and considered in the
74 US (12-18). Genetic screening involves testing a population for genetic variants indicative of
75 risk for specific diseases to identify people with either higher predisposition of developing
76 that disease or the potential to pass it on to their offspring. This approach utilises modern
77 sequencing techniques to evaluate multiple genes associated with selected traits. In
78 contrast 'targeted' tests are those performed because of some suggestion that a person
79 may harbour disease variants (e.g., symptoms or family history of disease). Although
80 screening is relevant to liabilities ranging between monogenic and polygenic (cf. (16, 19)),
81 we focus here on screening for pathogenic variants with monogenic associations with rare
82 diseases, particularly as applied to neurodegenerative disorders.

83
84 Although no widespread implementation of genetic screening protocols currently exist
85 internationally, comparable metabolic screening, testing neonates for metabolite markers
86 of metabolic diseases, is routine in many countries (17, 20, 21). Positive metabolic tests are
87 typically validated with secondary testing, including targeted genetic tests (9, 22, 23).

88
89 The utility of genetic screening can be assessed by the extent to which action can be taken
90 following a positive test: its actionability (24, 25). One key tenet of actionability is the
91 probability of having or later developing a disease following a positive test. Yet post-test
92 disease probability can be strikingly low where disease risk prior to testing is low, as would
93 be for population screening for rare diseases (26).

94
95 Bayesian inference, which is routine within clinical decision making (27), can be used to
96 understand post-test ('posterior') disease risk. Under this logic, disease probability following
97 a test can be inferred given existing knowledge of the probability of other relevant events.
98 Key considerations to understanding post-test disease risk, beyond pre-test (also known as
99 'prior') disease risk, include the genetic marker penetrance, its frequency among people
100 displaying disease symptoms, and the sensitivity and specificity of the test (analytic validity).
101 This reasoning is therefore highly relevant to screening for rare neurodegenerative diseases,
102 for which genetic causes are typically rare variants of variable penetrance (28).

103
104 This article overviews important considerations for genetic screening of rare disorders and it
105 presents several case studies focused on neurodegenerative diseases. Considering
106 conditional probability in medical decision making is not novel but these concepts must be

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

114

115 **Methods:**

116 **Bayesian framework**

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

122

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

126

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

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

135

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

137

Equation 1

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

138

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

140

Equation 2

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

141

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

144

Equation 3

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

145

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

147

Equation 4

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

148

149 and negative,

150

Equation 5

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

151 results.

152

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

155

Equation 6

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

156

157 and negative test result T' ,

158

Equation 7

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

159

160 can then be determined.

161

162 Case studies

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

172

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

177

178

179 Case 1 – Huntington's disease

180 HD is a late-onset Mendelian disease with autosomal dominant inheritance caused by a
181 trinucleotide, CAG, short tandem repeat expansion (STRE) in the *HTT* gene (OMIM:
182 613004). We let M be a CAG expansion of >40 repeat units, which would have complete
183 penetrance in a normal lifespan (29).

184

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

190

191 *Case 2 – amyotrophic lateral sclerosis*

192 ALS is a late-onset disease with locus and allelic heterogeneity; genetic associations with risk
193 and phenotype modification range between monogenic and polygenic. Variants in at least
194 40 genes are associated with ALS (30, 36-38). *SOD1* (OMIM: 147450) and *C9orf72* (OMIM:
195 614260) are the most frequently implicated genes, where variants of each account for fewer
196 than 10% of cases. Autosomal dominant inheritance is typical for most people with a known
197 genetic disease cause.

198

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

207

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

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

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

227

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

232

233 *Case 3 – phenylketonuria*

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

239

240 We modelled two testing scenarios for PKU: (1) genetic screening, with pre-test disease
241 probability defined per the baseline population risk of PKU, and (2) secondary testing to
242 confirm positive results obtained using tandem mass spectrometry (46) as in established
243 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. [§]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); [¶] defined by variant calling performance benchmarks of tools for genotyping in sequencing data by variant type (see Table S1) and, where marked [†], by aggregate laboratory accuracy for genotyping C9orf72 STRE with repeat-primed polymerase chain reaction and amplicon-length analysis (45). [¶]Risk following positive results in primary screening and confirmatory

Case study	Gene containing marker (case study scenario)	Variant type	Pre-test disease probability [§]	Marker frequency in people affected [§]	Penetrance [§]	Test sensitivity [¶]	Test specificity [¶]	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')	-
1: HD	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
2: ALS	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 ⁻⁵ , 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 [†]	0.98 [†]	0.0198	0.00489	4.06 (6.35 [¶])
3: PKU	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 [¶])

tests relative to a negative screening result (probability of PKU given a negative metabolic screening result is approximated as 1x10⁻⁶).

244 **Results and Discussion**

245

246 **Post-test disease probability**

247 *Screening versus diagnostic testing*

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

252

253 The HD case study illustrates the distinction between contextually-blind screening and
254 diagnostic testing for rare diseases: following a positive test result, lifetime HD risk was high
255 (90.8%) using targeted testing but low (0.4%) in screening. This difference reflects that,
256 unlike screening, targeted testing is performed based on some indication of a person's
257 elevated disease risk (e.g., existing disease symptoms or family history). Accordingly, pre-
258 test disease probability is greater. Inherently low pre-test disease probability will be a
259 pervasive issue in screening for rare diseases.

260

261 *Relative risk and secondary testing*

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

268

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

274

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

284

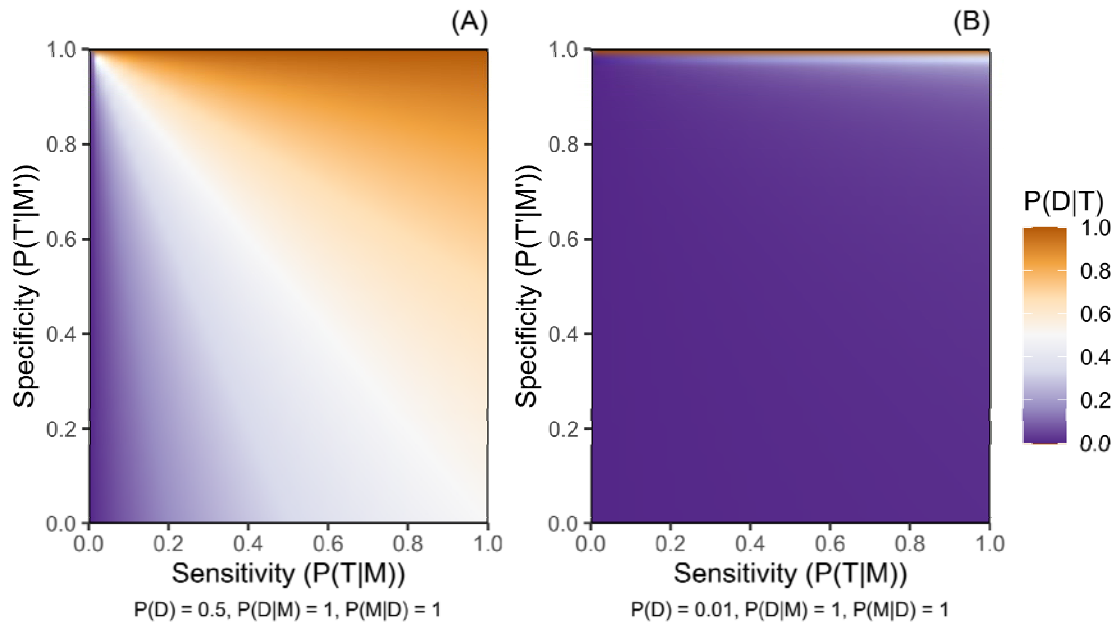
285 The overall benefit of secondary testing will however differ by scenario. For ALS, the risk
286 remained moderate (~2%) despite two positive test results for *C9orf72*.

287

288 *Constraints upon post-test disease probability*

289 Figure 1, Figure 2, and Figure 3 illustrate how the post-test disease probability reduces as
290 the probability of any test, disease, or marker characteristic decreases. Sensitivity and

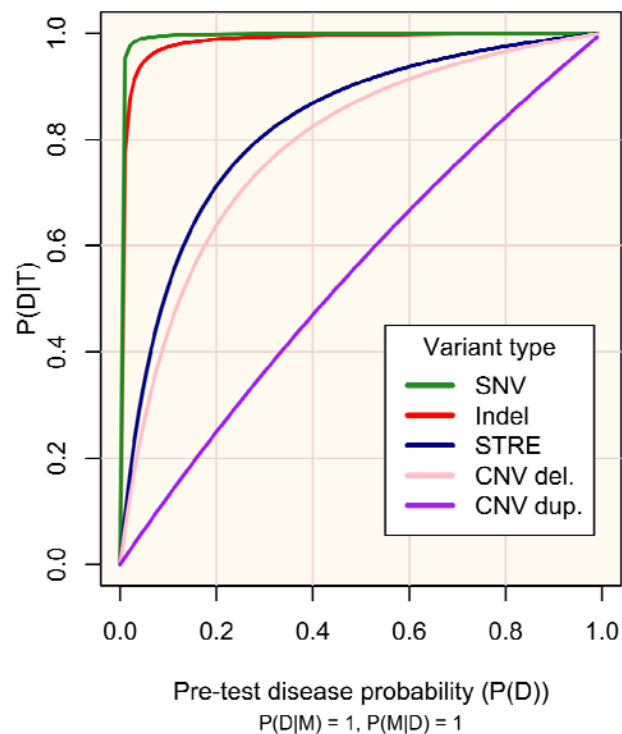
291 specificity critically constrain certainty about post-test disease risk, and this role is amplified
292 as the other parameter probabilities decrease. Figure 1 particularly demonstrates the
293 increased role of specificity in rarer diseases, where disease risk following a positive test
294 result will be moderated only by penetrance in a protocol of perfect specificity.
295



296
297
298
299
300
301

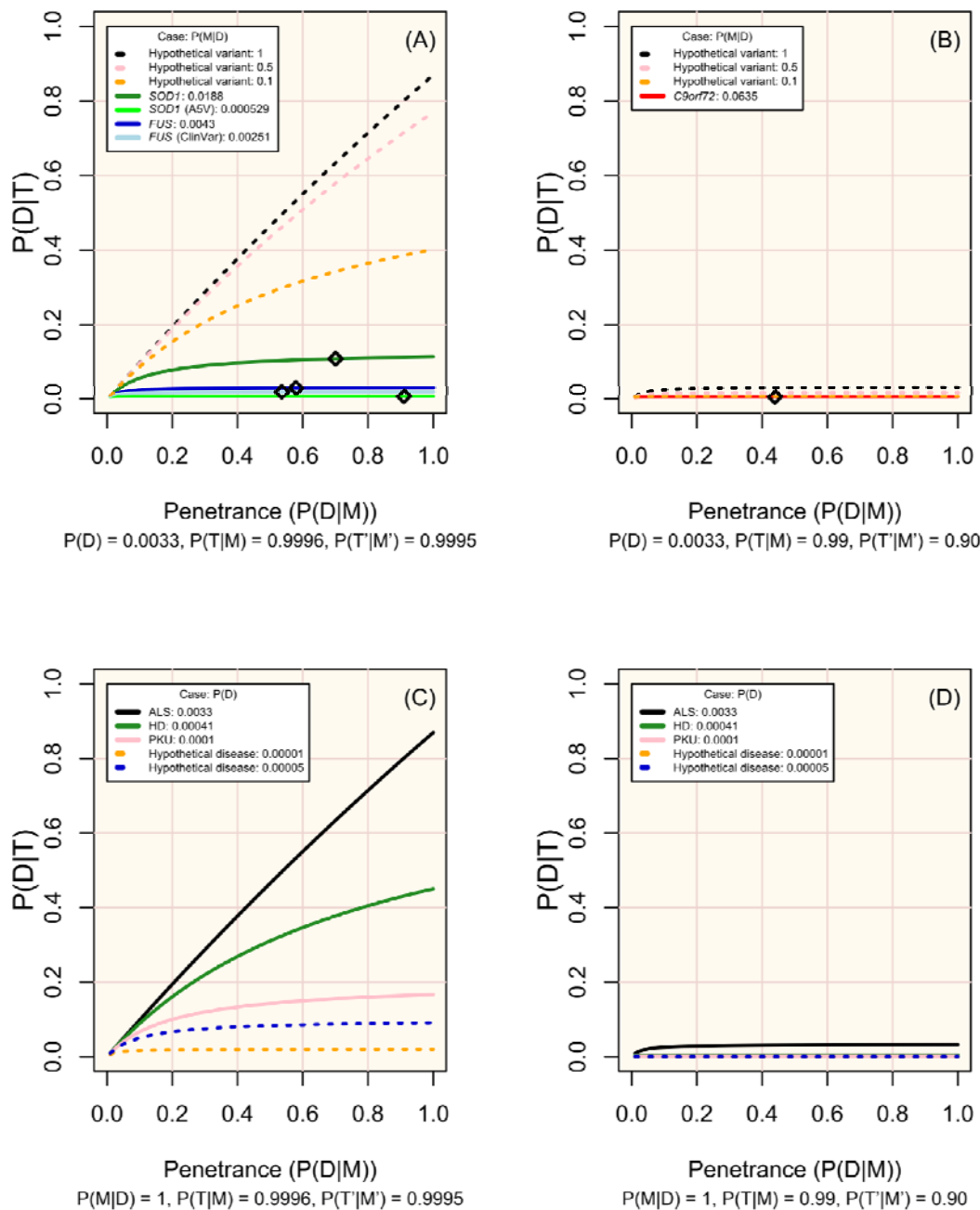
Figure 1. Probability of a disease given a positive genetic test result for a marker of increased disease risk ($P(D|T)$) according to the sensitivity ($P(T|M)$) and the specificity ($P(T'|M')$) of the testing protocol

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.



302
303 *Figure 2. Probability of disease D following a positive genetic test result for marker M ($P(D|T)$) according to pre-test disease*
304 *probability ($P(D)$)*

305 *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*
306 *sensitivity ($P(T|M)$) and specificity ($P(T'|M')$) of existing protocols for genotyping variant types (see Supplementary*
307 *Materials 2): single nucleotide variant (SNV), $P(T|M) = 0.9996, P(T'|M') = 0.9995$; small insertion or deletion (Indel), $P(T|M)$*
308 *$= 0.9962, P(T'|M') = 0.9971$; short tandem repeat expansion (STRE), $P(T|M) = 0.99, P(T'|M') = 0.90$; copy number variant*
309 *(CNV) – del. (deletion), $P(T|M) = 0.289, P(T'|M') = 0.959$; CNV – dup. (duplication), $P(T|M) = 0.1020, P(T'|M') = 0.9233$.*



310
311
312

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

313
314
315
316
317
318

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).

319

320 As is well-recognised, high sensitivity and specificity are essential to maximise utility of
321 testing. The respective trade-offs between prioritising each of these must be regarded: high
322 sensitivity is required to detect markers, while high specificity increases confidence in
323 positive results. Established screening protocols prioritise high sensitivity to maximise
324 detection of at-risk individuals, with confirmatory secondary testing being vital to minimise
325 false-positive results (9, 22, 23).

326
327 Since the characteristics of diseases and associated variants are all pre-determined within a
328 population, disease markers (variants) screened should be chosen carefully. The most useful
329 variants will be those more prevalent among people affected, of high penetrance, and
330 which can be genotyped with high sensitivity and specificity (see Table 1, ALS case study;
331 Figure 3, panels A and B).

332

333 **Practical implementation of genetic screening**

334 *Marker selection*

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

341

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

350

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

356

357 *Utility over time and actionability*

358 As genetic screening is possible from birth, while non-genetic methods may not be, age of
359 viability for screening methods should be evaluated. For late-onset diseases, early genetic
360 screening may enable preventative treatments to at-risk individuals or close monitoring for
361 prodromal disease markers. For instance, rapid eye movement sleep behaviour disorder is a
362 prodromal feature for Parkinson's disease (47) and monitoring of at-risk individuals may
363 enable early intervention. The influence of time upon treatment viability and effectiveness
364 must also be considered. For example, genetic therapy has potential utility for preventing or
365 delaying onset of degenerative disorders (48).

366

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

374

375 **Limitations**

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

386

387 **Conclusion**

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

393

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

397

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

407

408

409 **Data Availability**

410 Data sharing not applicable

411

412

413 **Acknowledgements**

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

418

419 **Author contributions**

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

426

427 **Funding**

428 AAC is an NIHR Senior Investigator. This is an EU Joint Programme-Neurodegenerative
429 Disease Research (JPND) project. The project is supported through the following funding
430 organizations under the aegis of JPND (United Kingdom, Medical Research Council
431 MR/L501529/1 and MR/R024804/1; Economic and Social Research Council ES/L008238/1)
432 and through the Motor Neurone Disease Association. This study represents independent
433 research part funded by the NIHR Maudsley Biomedical Research Centre at South London
434 and Maudsley NHS Foundation Trust and King's College London. The work leading up to this
435 publication was funded by the European Community's Horizon 2020 Programme (H2020-
436 PHC-2014-two-stage; grant 633413). This work was part funded by my Name's 5 Doddie
437 Foundation, and Alan Davidson Foundation. The views expressed are those of the author(s)
438 and not necessarily those of the NHS, the NIHR, King's College London, or the Department
439 of Health and Social Care. AI is funded by South London and Maudsley NHS Foundation
440 Trust, MND Scotland, Motor Neurone Disease Association, National Institute for Health and
441 Care Research, Spastic Paraplegia Foundation, Rosetrees Trust, Darby Rimmer MND
442 Foundation, the Medical Research Council (UKRI) and Alzheimer's Research UK.

443

444 **Role of the funding source**

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

447

448 **Declaration of interests**

449 AI, MR, FF, and NP declare no competing interests. TPS declares employment with
450 AstraZeneca. AAC declares royalties from The Brain, A Beginner's Guide: Oneworld
451 Publications, and The Genetics of Complex Human Diseases, A Laboratory Manual. AAC
452 reports consultancies or advisory boards for Amylyx, Apellis, Biogen, Brainstorm,
453 Cytokinetics, GenieUs, GSK, Lilly, Mitsubishi Tanabe Pharma, Novartis, OrionPharma,
454 Qoralis, Sano, Sanofi, and Wave Pharmaceuticals. AAC is the deputy editor for Brain. AAC
455 has planned patents for 'Use of CSF-Neurofilament determinations and CSF-Neurofilament
456 thresholds of prognostic and stratification value with regards to response to therapy in
457 neuromuscular and neurodegenerative diseases'. AAC is Scientific co-lead for United to End
458 MND Partnership Coalition and Guarantor of Brain. AAC has stock options for Qoralis.

459

460

461

462

463 **References:**

- 464 1. Iacoangeli A, Al Khleifat A, Sproviero W, Shatunov A, Jones AR, Morgan SL, et al.
465 DNAscan: personal computer compatible NGS analysis, annotation and visualisation. BMC
466 Bioinform. 2019;20(1):213.
- 467 2. Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, et al. Accurate
468 circular consensus long-read sequencing improves variant detection and assembly of a
469 human genome. Nat Biotechnol. 2019;37(10):1155-62.
- 470 3. Illumina. Accuracy Improvements in Germline Small Variant Calling with the DRAGEN
471 Platform 2019 07/09/2019. Available from: [https://science-](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)
472 [docs.illumina.com/documents/Informatics/dragen-v3-accuracy-appnote-html-970-2019-](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)
473 [006/Content/Source/Informatics/Dragen/dragen-v3-accuracy-appnote-970-2019-](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)
474 [006/dragen-v3-accuracy-appnote-970-2019-006.html](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).
- 475 4. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y. Comprehensive
476 evaluation of structural variation detection algorithms for whole genome sequencing.
477 Genome Biol. 2019;20(1):117.
- 478 5. Rhoades R, Jackson F, Teng S. Discovery of rare variants implicated in schizophrenia
479 using next-generation sequencing. J transl genet genom. 2019;3:1.
- 480 6. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-
481 wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev
482 Genet. 2011;12(7):499-510.
- 483 7. Lee JJ, Wedow R, Okbay A, Kong E, Maghziyan O, Zacher M, et al. Gene discovery and
484 polygenic prediction from a genome-wide association study of educational attainment in 1.1
485 million individuals. Nat Genet. 2018;50(8):1112-21.
- 486 8. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank
487 resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.
- 488 9. Hörster F, Kölker S, Loeber JG, Cornel MC, Hoffmann GF, Burgard P. Newborn
489 Screening Programmes in Europe, Arguments and Efforts Regarding Harmonisation: Focus
490 on Organic Acidurias. JIMD Rep. 2017;32:105-15.
- 491 10. Perrone F, Cacace R, Van Mossevelde S, Van den Bossche T, De Deyn PP, Cras P, et al.
492 Genetic screening in early-onset dementia patients with unclear phenotype: relevance for
493 clinical diagnosis. Neurobiol Aging. 2018;69:292.e7-.e14.
- 494 11. Majumder MA, Guerrini CJ, McGuire AL. Direct-to-Consumer Genetic Testing: Value
495 and Risk. Annu Rev Med. 2021;72(1):151-66.
- 496 12. Murray MF, Evans JP, Khoury MJ. DNA-Based Population Screening: Potential
497 Suitability and Important Knowledge Gaps. JAMA. 2019.
- 498 13. Dickinson JA, Pimlott N, Grad R, Singh H, Szafran O, Wilson BJ, et al. Screening: when
499 things go wrong. Can Fam Physician. 2018;64(7):502-8.
- 500 14. Jansen ME, Lister KJ, van Kranen HJ, Cornel MC. Policy Making in Newborn Screening
501 Needs a Structured and Transparent Approach. Front Public Health. 2017;5(53).
- 502 15. Genome UK: The future of healthcare.
503 <https://www.gov.uk/government/publications/genome-uk-the-future-of-healthcare> 2020.
- 504 16. Moorthie S, Hall A, Janus J, Brigden T, Villiers CBd, Blackburn L, et al. Polygenic scores
505 and clinical utility. University of Cambridge; 2021.
- 506 17. Adhikari AN, Gallagher RC, Wang Y, Currier RJ, Amatuni G, Bassaganyas L, et al. The
507 role of exome sequencing in newborn screening for inborn errors of metabolism. Nat Med.
508 2020;26(9):1392-7.

- 509 18. Centers for Disease Control and Prevention. Use of Genomics in Newborn Screening
510 Programs: The Promise and Challenges 2021 [Available from:
511 https://www.cdc.gov/genomics/events/newborn_screening_2021.htm.
512 19. Pain O, Gillett AC, Austin JC, Folkersen L, Lewis CM. A tool for translating polygenic
513 scores onto the absolute scale using summary statistics. *Eur J Hum Genet.* 2022;30:339-48.
514 20. Loeber JG, Burgard P, Cornel MC, Rigter T, Weinreich SS, Rupp K, et al. Newborn
515 screening programmes in Europe; arguments and efforts regarding harmonization. Part 1 –
516 From blood spot to screening result. *J Inher Metab Dis.* 2012;35(4):603-11.
517 21. Tarini BA, Christakis DA, Welch HG. State Newborn Screening in the Tandem Mass
518 Spectrometry Era: More Tests, More False-Positive Results. *Pediatrics.* 2006;118(2):448.
519 22. Southern KW, Munck A, Pollitt R, Travert G, Zanolla L, Dankert-Roelse J, et al. A
520 survey of newborn screening for cystic fibrosis in Europe. *J Cyst Fibros.* 2007;6(1):57-65.
521 23. Rinaldo P, Zafari S, Tortorelli S, Matern D. Making the case for objective performance
522 metrics in newborn screening by tandem mass spectrometry. *Ment Retard Dev Disabil Res*
523 *Rev.* 2006;12(4):255-61.
524 24. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, et al. ClinGen
525 — The Clinical Genome Resource. *N Engl J Med.* 2015;372(23):2235-42.
526 25. Hunter JE, Irving SA, Biesecker LG, Buchanan A, Jensen B, Lee K, et al. A standardized,
527 evidence-based protocol to assess clinical actionability of genetic disorders associated with
528 genomic variation. *Genet Med.* 2016;18(12):1258-68.
529 26. Biesecker LG. Genomic screening and genomic diagnostic testing—two very different
530 kettles of fish. *Genome Med.* 2019;11(1):75.
531 27. Hunink MGM, Weinstein MC, Wittenberg E, Drummond MF, Pliskin JS, Wong JB, et
532 al. *Decision Making in Health and Medicine: Integrating Evidence and Values.* 2nd ed.
533 Cambridge: Cambridge University Press; 2014.
534 28. Bertram L, Tanzi RE. The genetic epidemiology of neurodegenerative disease. *J Clin*
535 *Invest.* 2005;115(6):1449-57.
536 29. Langbehn DR, Brinkman RR, Falush D, Paulsen JS, Hayden MR. A new model for
537 prediction of the age of onset and penetrance for Huntington's disease based on CAG
538 length. *Clin Genet.* 2004;65(4):267-77.
539 30. Brown RH, Al-Chalabi A. Amyotrophic Lateral Sclerosis. *N Engl J Med.*
540 2017;377(2):162-72.
541 31. Hillert A, Anikster Y, Belanger-Quintana A, Burlina A, Burton BK, Carducci C, et al. The
542 Genetic Landscape and Epidemiology of Phenylketonuria. *Am J Hum Genet.*
543 2020;107(2):234-50.
544 32. van der Spek RAA, van Rheenen W, Pulit SL, Kenna KP, van den Berg LH, Veldink JH.
545 The project MinE databrowser: bringing large-scale whole-genome sequencing in ALS to
546 researchers and the public. *Amyotroph Lateral Scler Frontotemporal Degener.* 2019;20:432-
547 40.
548 33. ALS Variant Server [Internet]. [cited 02/2021]. Available from:
549 <http://als.umassmed.edu/>.
550 34. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar:
551 improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.*
552 2018;46(D1):D1062-D7.
553 35. Paulson H. Repeat expansion diseases. *Handb Clin Neurol.* 2018;147:105-23.

- 554 36. Iacoangeli A, Al Khleifat A, Sproviero W, Shatunov A, Jones AR, Opie-Martin S, et al.
555 ALSgeneScanner: a pipeline for the analysis and interpretation of DNA sequencing data of
556 ALS patients. *Amyotroph Lateral Scler Frontotemporal Degener.* 2019;20:207-15.
- 557 37. Zou Z-Y, Zhou Z-R, Che C-H, Liu C-Y, He R-L, Huang H-P. Genetic epidemiology of
558 amyotrophic lateral sclerosis: a systematic review and meta-analysis. *J Neurol Neurosurg*
559 *Psychiatry.* 2017;88:540-9.
- 560 38. Mehta PR, Iacoangeli A, Opie-Martin S, van Vugt J, Al Khleifat A, Bredin A, et al. The
561 impact of age on genetic testing decisions in amyotrophic lateral sclerosis. *Brain.*
562 2022;145(12):4440-7.
- 563 39. Abel O, Powell JF, Andersen PM, Al-Chalabi A. ALSod: A user-friendly online
564 bioinformatics tool for amyotrophic lateral sclerosis genetics. *Hum Mutat.* 2012;33(9):1345-
565 51.
- 566 40. Renton AE, Majounie E, Waite A, Simon-Sanchez J, Rollinson S, Gibbs JR, et al. A
567 hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-
568 FTD. *Neuron.* 2011;72(2):257-68.
- 569 41. Cudkowicz ME, McKenna-Yasek D, Sapp PE, Chin W, Geller B, Hayden DL, et al.
570 Epidemiology of mutations in superoxide dismutase in amyotrophic lateral sclerosis. *Annals*
571 *of Neurology.* 1997;41(2):210-21.
- 572 42. Spargo TP, Opie-Martin S, Bowles H, Lewis CM, Iacoangeli A, Al-Chalabi A. Calculating
573 variant penetrance from family history of disease and average family size in population-
574 scale data. *Genome Med.* 2022;14:141.
- 575 43. Saeed M, Yang Y, Deng HX, Hung WY, Siddique N, Dellefave L, et al. Age and founder
576 effect of SOD1 A4V mutation causing ALS. *Neurology.* 2009;72(19):1634-9.
- 577 44. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and
578 guidelines for the interpretation of sequence variants: a joint consensus recommendation of
579 the American College of Medical Genetics and Genomics and the Association for Molecular
580 Pathology. *Genet Med.* 2015;17(5):405-23.
- 581 45. Akimoto C, Volk AE, van Blitterswijk M, Van den Broeck M, Leblond CS, Lumbroso S,
582 et al. A blinded international study on the reliability of genetic testing for GGGGCC-repeat
583 expansions in C9orf72 reveals marked differences in results among 14 laboratories. *J Med*
584 *Genet.* 2014;51(6):419.
- 585 46. Schulze A, Lindner M, Kohlmüller D, Olgemöller K, Mayatepek E, Hoffmann GF.
586 Expanded newborn screening for inborn errors of metabolism by electrospray ionization-
587 tandem mass spectrometry: results, outcome, and implications. *Pediatrics.*
588 2003;111(6):1399-406.
- 589 47. Ferini-Strambi L, Marelli S, Galbiati A, Rinaldi F, Giora E. REM Sleep Behavior Disorder
590 (RBD) as a marker of neurodegenerative disorders. *Arch Ital Biol.* 2014;152(2-3):129-46.
- 591 48. Amado DA, Davidson BL. Gene therapy for ALS: A review. *Mol Ther.*
592 2021;29(12):3345-58.
- 593 49. Paquin RS, Mittendorf KF, Lewis MA, Hunter JE, Lee K, Berg JS, et al. Expert and lay
594 perspectives on burden, risk, tolerability, and acceptability of clinical interventions for
595 genetic disorders. *Genet Med.* 2019;21(11):2561-8.
- 596 50. Wadman RI, Jansen MD, Stam M, Wijngaarde CA, Curial CAD, Medic J, et al.
597 Intragenic and structural variation of the SMN locus and clinical variability of spinal muscular
598 atrophy. *Brain Commun.* 2020;2(2):fcaa075.
- 599