

1 Does genetic risk help to predict amyloid burden in a non-demented
2 population? A Bayesian approach.

3 Nicola Voyle^a, Willemijn Jansen^b, Aoife Keohane^a, Hamel Patel^a, Amos Folarin^a, Stephen
4 Newhouse^a, Caroline Johnston^a, Kuang Lin^c, Pieter Jelle Visser^{d,e}, Angela Hodges^c,
5 Richard JB Dobson * ^{†a,f,g}, Steven J Kiddle * ^{†a,h}, for the Alzheimer's Disease
6 Neuroimaging Initiative ^{‡1}, and EDAR and DESCRIPA study groups¹

7 ^aMRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry,
8 Psychology and Neuroscience, King's College London, London, UK

9 ^bDepartment of Psychiatry and Neuropsychology, School for Mental Health and
10 Neuroscience, Alzheimer Center Limburg, Maastricht University, The Netherlands

11 ^cDepartment of Old Age Psychiatry, Institute of Psychiatry, Psychology and Neuroscience,
12 King's College London, London, UK

13 ^dDepartment of Neurology and Alzheimer Center, VU University Medical Center,
14 Amsterdam, The Netherlands

15 ^eHelen Wills Neuroscience Institute, University of California, Berkeley

16 ^fNIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for
17 Dementia at South London and Maudsley NHS Foundation, London, UK

18 ^gFarr Institute of Health Informatics Research, UCL Institute of Health Informatics,
19 University College London, London, UK

20 ^hMRC Biostatistics Unit, Cambridge Biomedical Campus, Cambridge Institute of Public
21 Health, Forvie Site, Robinson Way, Cambridge, UK.

* Joint senior author

[†]Corresponding authors: SGDP Centre, IoPPN, 16 De Crespigny Park, Denmark Hill, London, SE5 8AF; richard.j.dobson@kcl.ac.uk; steven.kiddle@kcl.ac.uk; Telephone 020 7848 0924; Fax 020 7848 0866

[‡]Data used in preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database

(adni.loni.usc.edu). The investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data, but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Abstract

INTRODUCTION: In this study we investigate the association between $A\beta$ levels in cerebrospinal fluid (CSF) and genetic risk in a non-demented population. This paper presents the first analysis to use a Bayesian methodology in this area.

METHODS: Data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the EDAR* and DESCRIPA** studies was used in a Bayesian logistic regression analysis. We modeled CSF $A\beta$ burden using age, diagnosis (healthy control or mild cognitive impairment), *APOE* and a polygenic risk score (PGRS) associated with Alzheimer's Disease (AD). We compared models built using informative priors on age, diagnosis and *APOE* with non-informative priors on all variables.

RESULTS: The use of informative priors did not improve model performance in the majority of cases. Models using only age, diagnosis and *APOE* genotype showed the best predictive ability.

DISCUSSION: A previous study indicated that a PGRS of AD case/control status was associated with CSF $A\beta$ burden in healthy controls. The current study suggests that this association does not lead to models that are more predictive of amyloid positivity than already known factors such as age and *APOE*.

* 'Beta amyloid oligomers in the early diagnosis of AD and as marker for treatment response'

** 'Development of screening guidelines and criteria for pre-dementia Alzheimers disease'

KEYWORDS: Amyloid; Alzheimer's Disease; Multi-modal; Polygenic Risk Score; Bayesian; Gene Expression Risk Score; Blood; Biomarker; Protein; Metabolite.

40 1 Introduction

41 It is hypothesized that late onset AD is caused by the presence of $A\beta$ plaques in the brain tissue and
42 hyperphosphorylated tau tangles in the neurons [1]. Hypothetical models and longitudinal studies indicate
43 that $A\beta$ pathology begins to develop up to 20 years prior to symptomatic changes [2, 3, 4]. This provides
44 a window of opportunity for disease-modifying treatments, provided accurate and sensitive diagnostic tools
45 are available. Existing tools for identifying the presence of pathology include measurements of amyloid
46 and tau in CSF or in brain by the use of positron emission tomography (PET) imaging. Although these
47 tests are reasonably accurate they are invasive and expensive. Peripheral biomarkers are sought as an
48 intermediate step to provide cost-effective enrichment of people with high risk profiles for clinical trials,
49 particularly secondary prevention trials of anti-amyloid therapeutics. In this study we aimed to address
50 this by investigating the degree to which a genetic risk score was predictive of CSF $A\beta$ in a non-demented
51 population.

52 AD has been identified as a complex disease meaning its presence or absence is determined by environ-
53 mental and genetic risk factors. Heritability is estimated to be 50-70% [5, 6, 7]. Over 20 risk loci have been
54 identified for AD with Lambert *et al.* providing the most comprehensive Genome Wide Association Study
55 (GWAS) to date [8, 9, 10, 11, 7]. However, it is estimated that currently identified SNPs can only explain
56 16-33% of phenotypic variation [9, 10, 7]. GWAS of $A\beta$ endpoints are also highlighting promising results
57 but sample sizes are considerably smaller so validation in larger numbers is needed [12]. An approach to
58 consolidating the combined effect of several smaller genetic contributions to disease is found in polygenic
59 risk scores (PGRS). A PGRS is calculated as a weighted sum and represents the cumulative effect of these
60 smaller genetic effects. They have been shown to be informative in understanding genetic contributions to
61 phenotypes beyond AD clinical diagnosis. For example, Sabuncu *et al.* found an AD case/control PGRS
62 associated with $A\beta$ burden in healthy controls ($p < 0.0001$). It has not been assessed whether this would
63 provide predictive utility for enriching populations for prevention trials [13, 14, 15].

64 It is well known that model estimates, and often predictive ability, become more reliable as the number
65 of individuals included in an analysis increases. However, in AD research the amalgamation of studies to
66 create large datasets is often unfeasible due to differences in study populations and data collection methods.
67 Initiatives such as the European Medical Information Framework (EMIF) are aiming to rectify this, but large
68 populations with both multi-modal biomarker data and amyloid pathology measures are not yet available
69 (www.emif.eu).

70 However, there is substantial information available on the associations between demographic variables
71 and prevalence of $A\beta$ burden, as discussed in meta-analyses by Jansen *et al.* and Ossenkoppele *et al.* [4, 16].
72 The former of these studies concentrated on persons without dementia ($N=7583$). The study concluded that
73 age, *APOE* genotype and presence of cognitive impairment were associated with $A\beta$ burden. No equivalent
74 studies are available for tau pathology and consequently this work focuses on amyloid alone.

75 The present study uses data from the ADNI, EDAR and DESCRIPA cohorts, to investigate genetic risk
76 as a blood biomarker of $A\beta$ burden using a Bayesian methodology [17]. This study includes older individuals
77 who do not have a clinical diagnosis of AD and are at a variable risk of developing the disease. This is
78 the first study to use a Bayesian framework in AD blood biomarker research with the aim of investigating
79 genetic risk as a marker in blood that could support strategies for identifying individuals at high risk of
80 developing disease for recruitment into clinical trials. The models created combine age, diagnosis (control or
81 MCI) and *APOE* genotype with a PGRS. A previous study has shown a case/control PGRS associated with
82 $A\beta$ in healthy controls. We aim to investigate whether this association translates to predictive ability. We
83 hypothesized that by informing estimates for demographic variables using the large meta-analysis presented
84 by Jansen *et al.* we would create more robust models. The Bayesian methodology used here has been made
85 accessible to future researchers through the development of a simple graphical user interface (GUI).

86 **2 Methods**

87 **2.1 Cohorts**

88 **2.1.1 EDAR**

89 EDAR is a prospective, longitudinal study with centers at multiple European sites. The study aims to
90 examine and evaluate biomarkers of early AD and treatment response [17]. For more information see
91 www.edarstudy.eu. Our access to samples and clinical and phenotypic information from the EDAR study
92 was enabled by EMIF.

93 **2.1.2 DESCRIPA**

94 DESCRIPA is a prospective, multi-center study based in Europe and coordinated by the European AD
95 Consortium. The study focused on collecting data from non-demented subjects with the aim of developing

96 screening guidelines and clinical criteria for AD in non-demented subjects. Further details of this study can
97 be found in Visser *et al.* [18].

98 **2.1.3 ADNI**

99 ADNI is a longitudinal cohort study aiming to validate the use of biomarkers in AD clinical trials and diag-
100 nosis. Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu).
101 The ADNI study was launched in 2003 as a public-private partnership, led by Principal Investigator Michael
102 W. Weiner, MD. The primary goal of ADNI has been to test whether biological markers and clinical and
103 neuropsychological assessment can be combined to measure the progression of mild cognitive impairment
104 (MCI) and AD. For information, see www.adni-info.org. ADNI was approved by the institutional review
105 boards of all participating institutions, and written informed consent was obtained for all participants.

106 The ADNI study comprises three stages. ADNI 1 is the initial study (target N = 800). ADNI GO
107 contains a subset of the controls and MCI participants from ADNI 1 and is supplemented by additional
108 individuals with MCI (target N = 700). ADNI 2 enhances ADNI 1 and ADNI GO further with the inclusion
109 of new participants in all diagnostic groups (target N = 1350).

110 This study uses data from ADNI 1 and the ADNI 2 and ADNI GO sub-groups, referred to as ADNI 2
111 from here onwards.

112 **2.2 Genetics**

113 Samples from EDAR and DESCRIPA were genotyped on the Illumina HumanOmniExpressExome-8v1.2
114 BeadChip and processed together (N = 336) [19]. This BeadChip contains 960,919 markers of which
115 273,000 represent functional exomic markers. The data was processed in Genome Studio (as described at
116 bit.ly/1VpRclH) before being run through the rare variant caller Zcall [20] (as described here: bit.ly/1YKHYhK).
117 ADNI 1 samples were run on the Human610-Quad BeadChip (N = 818) while ADNI 2 and ADNI GO samples
118 were run on the Illumina HumanOmniExpress BeadChip (N = 432). The HumanOmniExpress BeadChip
119 is similar to that used in the EDAR and DESCRIPA studies but does not include exomic markers, while
120 the Human610-Quad BeadChip is older. Details of the genotyping protocols followed in ADNI are given
121 elsewhere [21].

122 The cohorts were subject to quality control and imputation separately, as described in Coleman *et al.*

123 [22]. In short, the data was filtered to ensure a minor allele frequency of greater than 5% for all SNPs before
124 removal of rare variants and subjects with high levels of missing data. SNPs that differed significantly (p
125 < 0.00001) from the Hardy-Weinberg equilibrium were removed. The data was pruned for SNPs in linkage
126 disequilibrium and for genetically similar individuals. Finally, the data was imputed using reference files
127 from the 1000 Genomes Project. [23]

128 **2.3 Amyloid measurements**

129 Throughout this study amyloid measurements in cerebrospinal fluid (CSF) are used as the endpoint of
130 interest. All $A\beta$ measurements were dichotomized as detailed in the meta-analysis from Jansen *et al.*
131 The distribution of amyloid burden in all studies was bimodal (as expected) making this dichotomization
132 biologically relevant. Low CSF $A\beta$ is referred to as ‘abnormal’ $A\beta$ burden while high CSF $A\beta$ is referred to
133 as ‘normal’. The details for each study are as follows:

134 **2.3.1 EDAR**

135 Details of CSF collection and analysis can be found at www.edarstudy.eu. In brief, CSF measurements were
136 collected using the Alzbio3 Luminex assay in one batch at the end of the study. CSF amyloid measurements
137 were dichotomized at the previously published threshold of 389pg/ml.

138 **2.3.2 DESCRIPA**

139 Details of CSF measurements in DESCRIPA have been described elsewhere [19]. In brief, CSF measurements
140 were analyzed in one laboratory and collected using single-parameter ELISA kits (Innogenetics, Ghent,
141 Belgium). CSF amyloid samples were dichotomized using the previously published threshold of 550pg/ml.

142 **2.3.3 ADNI**

143 For ADNI, datasets used to extract CSF measures of amyloid were chosen to maximize sample size. The
144 dataset ‘UPENNBIOMK2’ was used for ADNI 1 and ‘UPENNBIOMK6’ for ADNI 2 and ADNI GO. Both
145 datasets contain CSF measurements collected using the xMAP Luminex platform and Innogenetics im-
146 munoassay kits. The CSF measures were dichotomized at the previously published threshold (192 pg/ml).

147 2.4 Statistical analysis

148 All statistical analysis was performed in R version 3.1.1 [24]. Models were built in ADNI 1 data and tested
149 in data from EDAR, DESCRIPA and ADNI 2. We built models including age, diagnosis (healthy control
150 or MCI) and *APOE* genotype as covariates ('basic model') and models including these variables with the
151 addition of a PGRS ('PGRS model'). The predictive ability of each model was quantified using accuracy,
152 sensitivity, specificity and area under the Receiver Operating Characteristic (ROC) curve [25, 26, 27].

153 2.4.1 PGRS

154 PGRS were created using the software package PRSice [28]. Effect sizes from stage 1 of the International
155 Genomics of Alzheimer's Project (IGAP) case/control GWAS were used as the weights to generate the risk
156 score ($N = 54,162$, number of SNPs = 7,055,881) [8]. We used 0.5 as the p-value threshold for inclusion in
157 the PGRS. This threshold showed the most significant association with case/control diagnosis in the large
158 IGAP PGRS study [15]. The genetic region coding for *APOE* was removed from all scores and included as
159 a covariate in modeling.

160 2.4.2 Data analysis

161 This study aimed to predict dichotomized amyloid burden using genetic risk in a Bayesian logistic regression.
162 The method was implemented using the R function 'MCMClogit' in the 'MCMCpack' package [29]. Models
163 were built using a Metropolis sampler with 100,000 MCMC iterations, with the first 3,000 discarded as
164 burn-in. This number of iterations ensured the ratio of standard deviation to Monte Carlo Standard Error
165 (MCSE) was less than 5% for all parameters.

166 The variables included in the 'basic model' were chosen based on the meta-analysis published by Jansen
167 *et al.* [4]. The study used generalized estimating equations (GEEs) to predict amyloid burden from age,
168 diagnosis (control or MCI) and *APOE* $\epsilon 4$ status (defined as the presence or absence of any number of $\epsilon 4$
169 alleles). Age is centered at 70 years. In this study we only consider healthy controls and people with a
170 diagnosis of MCI. The best model identified by Jansen *et al.* was:

$$171 \quad \textit{Final model} : A\beta \sim \textit{Age} + \textit{Diagnosis} + \textit{APOE} + \textit{Age} * \textit{Diagnosis} + \textit{Age} * \textit{APOE}$$

172 We have used the estimates from this meta-analysis to inform the regression estimates of variables in this

173 study where possible. It can be shown that the estimates from GEEs are normally distributed [30] and hence
174 we included these estimates as priors using a multivariate normal distribution (see Table 1). The PGRS had
175 a non-informative multivariate normal prior with mean 0 and variance 100. We also created models where
176 all variables had the non-informative Normal(0,100) priors for comparison.

Table 1: Informative prior distributions

| Variable | Estimate | SE |
|------------------------------|----------|--------|
| Intercept | 0.879 | 0.0967 |
| Age | 0.064 | 0.006 |
| Diagnosis = Control | -0.964 | 0.0793 |
| <i>APOE</i> status = 0 | -1.493 | 0.0772 |
| Age * <i>APOE</i> status = 0 | -0.021 | 0.0079 |
| Age * Diagnosis = Control | 0.019 | 0.0081 |

SE = Standard error.

MCI is used as the reference diagnosis.

APOE status 1 (at least one $\epsilon 4$ allele) is used as the reference level.

178 The tuning parameter of each model was adjusted to achieve an acceptance rate in the Metropolis sampler
179 of approximately 0.35. This tuning was performed over the 3,000 burn-in samples. This acceptance rate is
180 slightly higher than advised in the literature as lower acceptance rates were causing reduced mixing [31].

181 2.4.3 Graphical User Interface

182 The analysis methods used in this study have been packaged into a user-friendly application through Rshiny
183 [32]. The application is available to download from [https://github.com/KHP-Informatics/bayesian-logistic-](https://github.com/KHP-Informatics/bayesian-logistic-regression-r-shiny-app.git)
184 [regression-r-shiny-app.git](https://github.com/KHP-Informatics/bayesian-logistic-regression-r-shiny-app.git).

185 3 Results

186 3.1 PGRS

187 In the ADNI cohort the genetic data was imputed from 479,073 and 599,526 SNPs in ADNI 1 and ADNI
188 2 respectively, to 8,799,802 and 6,336,499 SNPs. In EDAR and DESCRIPA the data was increased from
189 619,609 SNPs to 5,409,779 by imputation. The PGRS was standardized by subtracting the mean and dividing
190 by the standard deviation, per cohort.

191 3.2 Cohort demographics

192 The demographics given below are from the individuals in ADNI 1, ADNI 2, EDAR and DESCRIPA with
193 CSF and GWAS data available.

194 In ADNI 1 (training data) there is 1 point difference in median MMSE between subjects with normal
195 and abnormal $A\beta$ (28 vs. 29). However, the larger sample size of ADNI 1 (compared to ADNI 2, EDAR
196 and DESCRIPA) means this difference is statistically significant. Diagnosis and *APOE* genotype also show
197 significant differences between groups, as we would expect. We see a significant difference (p-value < 0.05)
198 in the PGRS between groups. Similar associations are seen in the data from EDAR and DESCRIPA with
199 age also being nominally significantly associated with normal and abnormal $A\beta$. ADNI 2 data shows no
200 significant difference in any demographic variable. This is likely to be due to the small sample size (N=43)
201 of this population.

202 3.3 Data analysis

203 The test data indicated that, in most cases, the addition of informative priors on age, diagnosis and *APOE*
204 genotype did not improve the predictive ability of models. Furthermore, no model including a PGRS showed
205 higher accuracy than the basic models. The basic model with non-informative priors achieved the highest
206 accuracies at 54% and 49% with a high sensitivity of 81% and 63%, but low specificity (23% and 41%). See
207 Table 3 for full results.

208 4 Discussion

209 This study shows that the predictive ability of models including age, diagnosis and *APOE* genotype is not
210 improved by the addition of a PGRS despite previous studies showing an association [13]. The PGRS used
211 in this analysis was trained on a case/control endpoint. As GWA studies of amyloid endpoints become
212 available the predictive ability of a PGRS trained using this information is likely to improve [33].

213 This paper presents the first analysis to use a Bayesian methodology in AD blood biomarker research. We
214 aimed to inform estimates of well-researched risk variables (age, *APOE* status and control or MCI status)
215 by including prior information from a large meta-analysis [4]. In this study we see that this approach does
216 not improve predictive ability of models over those without informative prior information. However, we have
217 only used one type of Bayesian methodology. There is a risk that if ill-fitting priors are used in combination
218 with small sample sizes, model fit may be driven by the prior distributions. It is possible that the priors
219 used here are not optimal as our population demographics are slightly different from those seen by Jansen *et*
220 *al.* [4]. However, we believe informing model estimates with information from previous literature is likely to
221 reduce false positives in biomarker studies of a small size. Bayesian analysis is one way of doing that; other
222 methods should also be investigated.

223 The creation of the PGRS also created some limitations for this study. Firstly, the PGRS used here
224 only utilizes common genetic variation. This is because rare genetic variants, such as *TREM2*, may not
225 be significantly associated with disease in small populations. As larger studies become available inclusion
226 of such variants should be investigated. Further, the PGRS was created using a simple additive method.
227 This simplistic method is likely to be sub-optimal and as new methodologies for creating PGRS become
228 available they should be investigated in this setting. Additionally, the platforms used to measure genetics
229 differed between the discovery and test cohorts used in this study. It is well documented that there can be

230 inconsistencies between *omics* platforms which could have contributed to the reduced predictive ability seen
231 here [34, 35, 36, 37].

232 It is important to bear in mind in all further research that differences in normal/abnormal CSF $A\beta$
233 cut-offs can make replication and research in other cohorts difficult. Work is being done to investigate these
234 values with an aim of standardization [38]. This study also has the limitation that the IGAP data (used
235 to generate the PGRS) is not independent of the training and test data, as ADNI was included in IGAP.
236 However, we believe the benefit of larger sample size outweighs this. Finally, the cohorts used in this study
237 are still of a relatively small size. Although we tried to address this through the use of Bayesian methods,
238 studies of larger sample sizes will be vital for further investigations of blood based biomarkers of $A\beta$.

239 This study also presents opportunities for further work. Firstly, it is possible that markers identified in
240 case/control studies are associated with other AD related phenotypes such as tau burden. If large meta-
241 analyses of the risk factors affecting tau burden, and other endophenotypes, become available it would be
242 interesting to perform similar Bayesian analysis on these alternative endpoints. Secondly, in the setting of
243 preventative clinical trials the most common way to measure brain amyloid burden is through the use of a
244 PET scan. In this study we have used CSF. However, it has been shown that measurements from CSF and
245 PET are highly correlated meaning markers identified for one can reasonably be tested to see whether they
246 are transferable to the other. In this study the use of CSF allowed us to maximize the sample size with
247 measurements of amyloid and genetics. If promising blood markers of CSF measures are identified, future
248 studies should perform similar analysis using a measurement of $A\beta$ derived from a PET scan. Furthermore,
249 the effect of a measure of brain reserve on model accuracy could be investigated. It is well-known that some
250 people with high levels of brain amyloid burden at autopsy show no cognitive deficit during their lifetime. It
251 has been shown in recent studies that levels of ‘brain reserve’ may be driving the difference between people
252 who have high levels of pathology and no symptoms and those who show symptoms [39]. This is motivating
253 a theory that increased brain reserve may prevent development of symptoms even if pathology is present.
254 Although brain reserve itself may be hard to quantify it may be possible to measure and model associated
255 lifestyle, environmental and psychological factors such as social networks.

256 Although there are further candidates and alternative methods to be considered in the search for a blood
257 biomarker of amyloid burden it is imperative that appropriate populations are used. As in this work, the
258 search in asymptomatic individuals is likely to have the biggest impact for enrichment of clinical trials.
259 Furthermore, rigorous testing of biomarkers is essential. Without replication it is probable that model
260 performance is overestimated.

261 5 Conclusions

262 This paper presents the first analysis to use a Bayesian methodology in the search for a blood biomarker of
263 AD. We see that the Bayesian approach does not improve predictive ability in this setting, and that *omics*
264 measurements do not improve the predictive ability of models above that of demographics alone. We have
265 been unable to demonstrate any additional benefit over age diagnosis and *APOE* genotype by including
266 a case/control PGRS when predicting amyloid positivity in subjects without a clinical diagnosis of AD or
267 other dementia.

268 6 Acknowledgements

269 This work represents independent research part funded by the National Institute for Health Research (NIHR)
270 Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College
271 London. This study was supported by researchers at the National Institute for Health Research University
272 College London Hospitals Biomedical Research Centre. The research leading to these results has received
273 support from the Innovative Medicines Initiative Joint Undertaking under grant agreement number 115372,
274 resources of which are composed of financial contribution from the European Union's Seventh Framework
275 Programme (FP7/2007-2013) and EFPIA companies' in kind contribution. Steven Kiddle is supported
276 by an MRC Career Development Award in Biostatistics (MR/L011859/1). Nicola Voyle is funded by the
277 Alzheimer's Society. The views expressed are those of the author(s) and not necessarily those of the NHS,
278 the NIHR or the Department of Health.

279 The authors would like to thank Dr. Steven Hill and Dr. Robert Goudie from the MRC Biostatistics
280 Unit for their guidance in Bayesian methods.

281 This work was also supported by awards to establish the Farr Institute of Health Informatics Research at
282 UCL partners, London, from the Medical Research Council, Arthritis Research UK, British Heart Founda-
283 tion, Cancer Research UK, Chief Scientist Office, Economic and Social Research Council, Engineering and
284 Physical Sciences Research Council, National Institute for Health Research, National Institute for Social
285 Care and Health Research, and Wellcome Trust (grant MR/K006584/1).

286 Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initia-
287 tive (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense
288 award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National

289 Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following:
290 AbbVie, Alzheimers Association; Alzheimers Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.;
291 Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and
292 Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio;
293 GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research and Development, LLC.; John-
294 son and Johnson Pharmaceutical Research and Development LLC.; Lumosity; Lundbeck; Merck and Co.,
295 Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals
296 Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Ther-
297 apeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in
298 Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health
299 (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education,
300 and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California,
301 San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern
302 California.

303 References

- 304 [1] QuerfurthH, LaFerlaF (2011) Mechanisms of Disease: Alzheimer's Disease. *New England Journal of*
305 *Medicine* **362**, 329–344.
- 306 [2] VillemagneV L, PikeK E, ChételatG, EllisK A, MulliganR S, BourgeatP, AckermannU, JonesG,
307 SzoekC, SalvadoO, MartinsR, O'KeefeG, MathisC A, KlunkW E, AmesD, MastersC L, RoweC C
308 (2011) Longitudinal assessment of A β and cognition in aging and Alzheimer disease. *Annals of neuro-*
309 *logy* **69** (1), 181–192.
- 310 [3] JackC R, KnopmanD S, JagustW J, PetersenR C, WeinerM W, AisenP S, ShawL M, VemuriP, WisteH J,
311 WeigandS D, LesnickT G, PankratzV S, DonohueM C, TrojanowskiJ Q (2013) Tracking pathophysio-
312 logical processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *The*
313 *Lancet Neurology* **12** (2), 207–216.
- 314 [4] JansenW J, OssenkoppeleR, KnolD L, TijmsB M, ScheltensP, VerheyF R J, VisserP J, AaltenP, Aars-
315 landD, AlcoleaD, AlexanderM, AlmdahlI S, ArnoldS E, BaldeirasI, BarthelH, van BerckelB N M,
316 BibeauK, BlennowK, BrooksD J, van BuchemM A, CamusV, CavedoE, ChenK, ChételatG, CohenA D,
317 DrzezgaA, EngelborghsS, FaganA M, FladbyT, FleisherA S, van der FlierW M, FordL, FörsterS,

- 318 ForteaJ, FoskettN, FrederiksenK S, Freund-LeviY, FrisoniG B, FroelichL, GabryelewiczT, GillK D,
319 GkatzimaO, Gómez-TortosaE, GordonM F, GrimmerT, HampelH, HausnerL, HellwigS, HerukkaS-K,
320 HildebrandtH, IshiharaL, IvanoiuA, JagustW J, JohannsenP, KandimallaR, KapakiE, Klimkowicz-
321 MrowiecA, KlunkW E, KöhlerS, KoglinN, KornhuberJ, KramerbergerM G, Van LaereK, LandauS M,
322 LeeD Y, de LeonM, LisettiV, LleoA, MadsenK, MaierW, MarcussonJ, MattssonN, de MendonçaA, Meu-
323 lenbroekO, MeyerP T, MintunM A, MokV, MolinuevoJ L, MøllergårdH M, MorrisJ C, MroczkoB, Van
324 der MusseleS, NaD L, NewbergA, NordbergA, NordlundA, NovakG P, ParaskevasG P, ParnettiL, Per-
325 eraG, PetersO, PoppJ, PrabhakarS, RabinoviciG D, RamakersI H G B, RamiL, Resende de OliveiraC,
326 RinneJ O, RodrigueK M, Rodríguez-RodríguezE, RoeC M, RotU, RoweC C, RüttherE, SabriO, Sanchez-
327 JuanP, SantanaI, SarazinM, SchröderJ, SchütteC, SeoS W, SoeteweyF, SoinenH, SpuruL, StruyfsH,
328 TeunissenC E, TsolakiM, VandenbergheR, VerbeekM M, VillemagneV L, VosS J B, van Waalwijk van
329 DoornL J C, WaldemarG, WallinA, WallinÅ K, WiltfangJ, WolkD A, ZbochM, ZetterbergH (2015)
330 Prevalence of Cerebral Amyloid Pathology in Persons Without Dementia. *JAMA* **313** (19), 1915–1924.
- 331 [5] GatzM, PedersenN L, BergS, JohanssonB, JohanssonK, MortimerJ a, PosnerS F, ViitanenM, WinbladB,
332 AhlbomA (1997) Heritability for Alzheimer’s disease: the study of dementia in Swedish twins. *The*
333 *journals of gerontology Series A, Biological sciences and medical sciences* **52A** (2), M117–M125.
- 334 [6] PedersenN L, GatzM, BergS, JohanssonB (2004) How Heritable Is Alzheimer’s Disease Late in Life?
335 Findings from Swedish Twins. *Annals of Neurology* **55**, 180–185.
- 336 [7] RidgeP G, HoytK B, BoehmeK, MukherjeeS, CraneP K, HainesJ L, MayeuxR, FarrerL A, Pericak-
337 VanceM A, SchellenbergG D, KauweJ S K (2016) Assessment of the genetic variance of late-onset
338 Alzheimer’s disease. *Neurobiology of aging* **41**, 200.e13–200.e20.
- 339 [8] LambertJ-C, Ibrahim-VerbaasC A, HaroldD, NajA C, SimsR, BellenguezC, JunG, DeStefanoA L,
340 BisJ C, BeechamG W, Grenier-BoleyB, RussoG, Thornton-WellsT A, JonesN, SmithA V, ChourakiV,
341 ThomasC, IkramM A, ZelenikaD, VardarajanB N, KamataniY, LinC-F, GerrishA, SchmidtH, KunkleB,
342 DunstanM L, RuizA, BihoreauM-T, ChoiS-H, ReitzC, PasquierF, HollingworthP, RamirezA, HanonO,
343 FitzpatrickA L, BuxbaumJ D, ChampionD, CraneP K, BaldwinC, BeckerT, GudnasonV, CruchagaC,
344 CraigD, AminN, BerrC, LopezO L, De JagerP L, DeramecourtV, JohnstonJ A, EvansD, LovestoneS,
345 LetenneurL, MorónF J, RubinszteinD C, EiriksdottirG, SleegersK, GoateA M, FiévetN, HuentelmanM J,
346 GillM, BrownK, KambohM I, KellerL, Barberger-GateauP, McGuinnessB, LarsonE B, GreenR, My-
347 ersA J, DufouilC, ToddS, WallonD, LoveS, RogaevaE, GallacherJ, George-HyslopP S, ClarimonJ, LleoA,
348 BayerA, TsuangD W, YuL, TsolakiM, BossùP, SpallettaG, ProitsiP, CollingeJ, SorbiS, Sanchez-GarciaF,

- 349 FoxN C, HardyJ, NaranjoM C D, BoscoP, ClarkeR, BrayneC, GalimbertiD, MancusoM, MatthewsF,
350 MoebusS, MecocciP, Del ZompoM, MaierW, HampelH, PilottoA, BullidoM, PanzaF, CaffarraP, Nacmi-
351 asB, GilbertJ R, MayhausM, LannfeltL, HakonarsonH, PichlerS, CarrasquilloM M, IngelssonM, Beek-
352 lyD, AlvarezV, ZouF, ValladaresO, YounkinS G, CotoE, Hamilton-NelsonK L, GuW, RazquinC, Pas-
353 torP, MateoI, OwenM J, FaberK M, JonssonP V, CombarrosO, O'DonovanM C, CantwellL B, Soini-
354 nenH, BlackerD, MeadS, MosleyT H, BennettD A, HarrisT B, FratiglioniL, HolmesC, de BruijnR F A G,
355 PassmoreP, MontineT J, BettensK, RotterJ I, BriceA, MorganK, ForoudT M, KukullW A, HannequinD,
356 PowellJ F, NallsM A, RitchieK, LunettaK L, KauweJ S K, BoerwinkleE, RiemenschneiderM, BoadaM,
357 HiltunenM, MartinE R, SchmidtR, RujescuD, WangL-S, DartiguesJ-F, MayeuxR, TzourioC, HofmanA,
358 NöthenM M, GraffC, PsatyB M, JonesL, HainesJ L, HolmansP A, LathropM, Pericak-VanceM A,
359 LaunerL J, FarrerL A, van DuijnC M, Van BroeckhovenC, MoskvinA, SeshadriS, WilliamsJ, Schellen-
360 bergG D, AmouyelP (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for
361 Alzheimer's disease. *Nature Genetics* **45** (12), 1452–1458.
- 362 [9] LeeS H, HaroldD, NyholtD R, ConsortiumA, ConsortiumI E, disease ConsortiumG, for Alzheimer'sE R,
363 GoddardM E, ZondervanK T, WilliamsJ, MontgomeryG W, WrayN R, VisscherP M (2013) Estimation
364 and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple
365 sclerosis and endometriosis. *Human Molecular Genetics* **22** (4), 832–841.
- 366 [10] RidgeP G, MukherjeeS, CraneP K, KauweJ S K, ConsortiumA D G (2013) Alzheimer's Disease: Ana-
367 lyzing the Missing Heritability. *PLOS ONE* **8** (11), e79,771.
- 368 [11] RuizA, HeilmannS, BeckerT, HernándezI, WagnerH, ThelenM, MauleónA, BellenguezC (2014) Follow-
369 up of loci from the International Genomics of Alzheimer's Disease Project identifies TRIP4 as a novel
370 susceptibility gene. *Translational Psychiatry* **4** (January), e358.
- 371 [12] CruchagaC, EbbertM T W, KauweJ S K (2014) Genetic Discoveries in AD Using CSF Amyloid and
372 Tau. *Current Genetic Medicine Reports* **2** (1), 23–29.
- 373 [13] SabuncuM R, BucknerR L, SmollerJ W, LeeP H, FischlB, SperlingR A, for ADNI (2012) The Association
374 between a Polygenic Alzheimer Score and Cortical Thickness in Clinically Normal Subjects. *Cerebral*
375 *Cortex* **22** (11), 2653–2661.
- 376 [14] SleegersK, BettensK, De RoeckA, Van CauwenbergheC, CuyversE, VerheijenJ, StruyfsH, Van Don-
377 genJ, VermeulenS, EngelborghsS, VandenbulckeM, VandenberghR, De DeynP P, Van BroeckhovenC,
378 ConsortiumB (2015) A 22-single nucleotide polymorphism Alzheimer risk score correlates with family
379 history, onset age, and cerebrospinal fluid A β 42. *Alzheimer's & Dementia* **11** (12), 1452–1460.

- 380 [15] Escott-PriceV, SimsR, BannisterC, HaroldD, VronskayaM, MajounieE, BadarinarayanN, MorganK,
381 PassmoreP, HolmesC, PowellJ, BrayneC, GillM, MeadS, GoateA, CruchagaC, LambertJ C, Van DuijnC,
382 MaierW, RamirezA, HolmansP, JonesL, HardyJ, SeshadriS, SchellenbergG D, AmouyelP, WilliamsJ
383 (2015) Common polygenic variation enhances risk prediction for Alzheimer’s disease. *Brain* **138** (12),
384 3673–3684.
- 385 [16] OssenkuppeleR, JansenW J, RabinoviciG D, KnolD L, van der FlierW M, van BerckelB N M, ScheltensP,
386 VisserP J, VerfaillieS C J, ZwanM D, AdriaanseS M, LammertsmaA A, Barkhoff, JagustW J, MillerB L,
387 RosenH J, LandauS M, VillemagneV L, RoweC C, LeeD Y, NaD L, SeoS W, SarazinM, RoeC M, SabriO,
388 BarthelH, KoglinN, HodgesJ, LeytonC E, VandenbergheR, van LaereK, DrzezgaA, ForsterS, GrimmerT,
389 Sánchez-JuanP, CarrilJ M, MokV, CamusV, KlunkW E, CohenA D, MeyerP T, HellwigS, NewbergA,
390 FrederiksenK S, FleisherA S, MintunM A, WolkD A, NordbergA, RinneJ O, ChételatG, LleoA, BlesaR,
391 ForteaJ, MadsenK, RodrigueK M, BrooksD J (2015) Prevalence of Amyloid PET Positivity in Dementia
392 Syndromes. *JAMA* **313** (19), 1911–1939.
- 393 [17] BarnettJ H, BlackwellA, ScheltensP, WaldemarG, JohannsenP, TsolakiM, VandenbergheR, WahlundL-
394 O, VerheyF, VisserP J (2010) Cognitive function and cognitive change in dementia, mild cognitive
395 impairment, and healthy aging: The EDAR study. *Alzheimer’s & Dementia* **6** (4), S127.
- 396 [18] VisserP J, VerheyF R J, BoadaM, BullockR, De DeynP P, FrisoniG B, Frouml ichtL, HampelH, JollesJ,
397 JonesR, MinthonL, NobiliF, Olde RikkertM, OussetP J, RigaudA S, ScheltensP, SoininenH, SpuruL,
398 TouchonJ, TsolakiM, VellasB, WahlundL-O, WilcockG, WinbladB (2008) Development of Screening
399 Guidelines and Clinical Criteria for Predementia Alzheimer’s Disease. *Neuroepidemiology* **30** (4), 254–
400 265.
- 401 [19] VisserP J, VerheyF, KnolD L, ScheltensP, WahlundL-O, Freund-LeviY, TsolakiM, MinthonL,
402 WallinÅ K, HampelH, BürgerK, PirttilaT, SoininenH, RikkertM O, VerbeekM M, SpuruL, BlennowK
403 (2009) Prevalence and prognostic value of CSF markers of Alzheimer’s disease pathology in patients with
404 subjective cognitive impairment or mild cognitive impairment in the DESCRIPA study: a prospective
405 cohort study. *The Lancet Neurology* **8** (7), 619–627.
- 406 [20] GoldsteinJ I, CrenshawA, CareyJ, GrantG B, MaguireJ, FromerM, O’dushlaineC, MoranJ L, Cham-
407 bertK, StevensC, SklarP, HultmanC M, PurcellS, MccarrollS A, SullivanP F, DalyM J, NealeB M (2012)
408 Zcall: A rare variant caller for array-based genotyping. *Bioinformatics* **28** (19), 2543–2545.
- 409 [21] SaykinA J, ShenL, ForoudT M, PotkinS G, SwaminathanS, KimS, RisacherS L, NhoK, HuentelmanM J,
410 CraigD W, ThompsonP M, SteinJ L, MooreJ H, FarrerL A, GreenR C, BertramL, Jack JrC R, Wein-

- erM W (2010) Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. *Alzheimer's & Dementia* **6** (3), 265–273.
- [22] ColemanJ R I, EuesdenJ, PatelH, FolarinA A, NewhouseS, BreenG (2015) Quality control , imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray. *Briefings in Functional Genomics* 1–7.
- [23] AutonA, AbecasisG R, AltshulerD M, DurbinR M, BentleyD R, ChakravartiA, ClarkA G, DonnellyP, EichlerE E, FlicekP, GabrielS B, GibbsR A, GreenE D, HurlesM E, KnoppersB M, KorbelJ O, LanderE S, LeeC, LehrachH, MardisE R, MarthG T, McVeanG A, NickersonD A, SchmidtJ P, SherryS T, WangJ, WilsonR K, BoerwinkleE, DoddapaneniH, HanY, KorchinaV, KovarC, LeeS, MuznyD, ReidJ G, ZhuY, ChangY, FengQ, FangX, GuoX, JianM, JiangH, JinX, LanT, LiG, LiJ, LiY, LiuS, LiuX, LuY, MaX, TangM, WangB, WangG, WuH, WuR, XuX, YinY, ZhangD, ZhangW, ZhaoJ, ZhaoM, ZhengX, GuptaN, GharaniN, TojiL H, GerryN P, ReschA M, BarkerJ, ClarkeL, GillL, HuntS E, KellmanG, KuleshaE, LeinonenR, McLarenW M, RadhakrishnanR, RoaA, SmirnovD, SmithR E, StretterI, ThormannA, TonevaI, VaughanB, Zheng-BradleyX, GrocockR, HumphrayS, JamesT, KingsburyZ, SudbrakR, AlbrechtM W, AmstislavskiyV S, BorodinaT A, LienhardM, MertesF, SultanM, TimmermannB, YaspoM-L, FultonL, FultonR, AnanievV, BelaiaZ, BeloslyudtsevD, BoukN, ChenC, ChurchD, CohenR, CookC, GarnerJ, HefferonT, KimelmanM, LiuC, LopezJ, MericP, O'SullivanC, OstapchukY, PhanL, PonomarovS, SchneiderV, ShekhtmanE, SirotkinK, SlottaD, ZhangH, BalasubramaniamS, BurtonJ, DanecekP, KeaneT M, Kolb-KokocinskiA, McCarthyS, StalkerJ, QuailM, DaviesC J, GollubJ, WebsterT, WongB, ZhanY, CampbellC L, KongY, MarckettaA, YuF, AntunesL, BainbridgeM, SaboA, HuangZ, CoinL J M, FangL, LiQ, LiZ, LinH, LiuB, LuoR, ShaoH, XieY, YeC, YuC, ZhangF, ZhengH, ZhuH, AlkanC, DalE, KahveciF, GarrisonE P, KuralD, LeeW-P, Fung LeongW, StrombergM, WardA N, WuJ, ZhangM, DalyM J, DePristoM A, HandsakerR E, BanksE, BhatiaG, del AngelG, GenoveseG, LiH, KashinS, McCarrollS A, NemeshJ C, PoplinR E, YoonS C, LihmJ, MakarovV, GottipatiS, KeinanA, Rodriguez-FloresJ L, RauschT, FritzM H, StützA M, BealK, DattaA, HerreroJ, RitchieG R S, ZerbinoD, SabetiP C, ShlyakhterI, SchaffnerS F, VittiJ, CooperD N, BallE V, StensonP D, BarnesB, BauerM, Keira CheethamR, CoxA, EberleM, KahnS, MurrayL, PedenJ, ShawR, KennyE E, BatzerM A, KonkelM K, WalkerJ A, MacArthurD G, LekM, HerwigR, DingL, KoboldtD C, LarsonD, YeK, GravelS, SwaroopA, ChewE, LappalainenT, ErlichY, GymrekM, Frederick WillemsT, SimpsonJ T, ShriverM D, RosenfeldJ A, BustamanteC D, MontgomeryS B, De La VegaF M, ByrnesJ K, CarrollA W, DeGorterM K, LacrouteP, MaplesB K, MartinA R, Moreno-EstradaA, ShringarpureS S, ZakhariaF, HalperinE, BaranY, CerveiraE, HwangJ, MalhotraA, Plewczyn-

443 skiD, RadewK, RomanovitchM, ZhangC, HylandF C L, CraigD W, ChristoforidesA, HomerN, IzattT,
444 KurdogluA A, SinariS A, SquireK, XiaoC, SebatJ, AntakiD, GujralM, NoorA, YeK, BurchardE G, Her-
445 nandezR D, GignouxC R, HausslerD, KatzmanS J, James KentW, HowieB, Ruiz-LinaresA, DermitzakiS
446 E T, DevineS E, Min KangH, KiddJ M, BlackwellT, CaronS, ChenW, EmeryS, FritscheL, Fuchs-
447 bergerC, JunG, LiB, LyonsR, SchellerC, SidoreC, SongS, SliwerskaE, TaliunD, TanA, WelchR, Kate
448 WingM, ZhanX, AwadallaP, HodgkinsonA, LiY, ShiX, QuitadamoA, LunterG, MarchiniJ L, MyersS,
449 ChurchhouseC, DelaneauO, Gupta-HinchA, KretzschmarW, IqbalZ, MathiesonI, MenelaouA, RimmerA,
450 XifaraD K, OleksykT K, FuY, LiuX, XiongM, JordeL, WitherspoonD, XingJ, BrowningB L, Brown-
451 ingS R, HormozdiariF, SudmantP H, KhuranaE, Tyler-SmithC, AlbersC A, AyubQ, ChenY, ColonaV,
452 JostinsL, WalterK, XueY, GersteinM B, AbyzovA, BalasubramanianS, ChenJ, ClarkeD, FuY,
453 HarmanciA O, JinM, LeeD, LiuJ, Jasmine MuX, ZhangJ, ZhangY, HartlC, ShakirK, DegenhardtJ,
454 MeiersS, RaederB, Paolo CasaleF, StegleO, LameijerE-W, HallI, BafnaV, MichaelsonJ, GardnerE J,
455 MillsR E, DayamaG, ChenK, FanX, ChongZ, ChenT, ChaissonM J, HuddlestonJ, MaligM, NelsonB J,
456 ParrishN F, BlackburneB, LindsayS J, NingZ, ZhangY, LamH, SisuC, ChallisD, EvaniU S, LuJ, Na-
457 gaswamyU, YuJ, LiW, HabeggerL, YuH, CunninghamF, DunhamI, LageK, Berg JaspersenJ, HornH,
458 KimD, DesalleR, NarechaniaA, Wilson SayresM A, MendezF L, David PoznikG, UnderhillP A, CoinL,
459 MittelmanD, BanerjeeR, CerezoM, FitzgeraldT W, LouzadaS, MassaiaA, RitchieG R, YangF, KalraD,
460 HaleW, DanX, BarnesK C, BeiswangerC, CaiH, CaoH, HennB, JonesD, KayeJ S, KentA, KerasidouA,
461 MathiasR, OssorioP N, ParkerM, RotimiC N, RoyalC D, SandovalK, SuY, TianZ, TishkoffS, ViaM,
462 WangY, YangH, YangL, ZhuJ, BodmerW, BedoyaG, CaiZ, GaoY, ChuJ, PeltonenL, Garcia-MonteroA,
463 OrfaoA, DutilJ, Martinez-CruzadoJ C, MathiasR A, HennisA, WatsonH, McKenzieC, QadriF, LaRoc-
464 queR, DengX, AsogunD, FolarinO, HappiC, OmoniwaO, StrelauM, TariyalR, JallowM, Sisay JooffF,
465 CorrahT, RockettK, KwiatkowskiD, KoonerJ, Tnh Hiê'nT, DunstanS J, Thuy HangN, FonnierR, Gar-
466 ryR, KannehL, MosesL, SchieffelinJ, GrantD S, GalloC, PolettiG, SaleheenD, RasheedA, BrooksL D,
467 FelsenfeldA L, McEwenJ E, VaydylevichY, DuncansonA, DunnM, SchlossJ A (2015) A global reference
468 for human genetic variation. *Nature* **526 (7571)**, 68–74.

469 [24] R Core Team (2014). R: A Language and Environment for Statistical Computing.

470 [25] KuhnM, WingJ, WestonS, WilliamsA, KeeferC, EngelhardtA, CooperT, MayerZ, KenkelB, R Core
471 Team, BenestyM, LescarbeauR, ZiemA, ScruccaL (2015). caret: Classification and Regression Training.

472 [26] RobinX, TurckN, HainardA, TibertiN, LisacekF, SanchezJ-C, MüllerM (2011) pROC: an open-source
473 package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12 (1)**, 77.

- 474 [27] SingT, SanderO, BeerenwinkelN, LengauerT (2005) ROCR: visualizing classifier performance in R.
475 *Bioinformatics* **21** (20), 3940–3941.
- 476 [28] EuesdenJ, LewisC M, O’ReillyP F (2015) PRSice: Polygenic Risk Score software. *Bioinformatics* **31** (9),
477 1466–1468.
- 478 [29] MartinA, QuinnK (2006) Applied Bayesian inference in R using MCMCpack. *R News* **6** (March), 2–7.
- 479 [30] ZieglerA, KastnerC, BlettnerM (1998) The generalised estimating equations: an annotated bibliography.
480 *Biometrical Journal* **40**, 115–140.
- 481 [31] RobertsG O, GelmanA, GilksW R (1997) Weak convergence and optimal scaling of random walk
482 Metropolis algorithms. *Annals of Applied Probability* **7** (1), 110–120.
- 483 [32] ChangW, ChengJ, AllaireJ, XieY, McPhersonJ (2015). shiny: Web Application Framework for R.
- 484 [33] DudbridgeF (2013) Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genetics* **9** (3),
485 e1003348.
- 486 [34] DobbinK K, BeerD G, MeyersonM, YeatmanT J, GeraldW L, JacobsonJ W, ConleyB, BuetowK H,
487 HeiskanenM, SimonR M, MinnaJ D, GirardL, MisekD E, TaylorJ M, HanashS, NaokiK, HayesD N,
488 Ladd-AcostaC, EnkemannS A, VialeA, GiordanoT J (2005) Interlaboratory comparability study of
489 cancer gene expression analysis using oligonucleotide microarrays. *Clin Cancer Res* **11** (2 Pt 1),
490 565–572.
- 491 [35] LeekJ T, ScharpfR B, BravoH C, SimchaD, LangmeadB, JohnsonW E, GemanD, BaggerlyK, IrizarryR A
492 (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature*
493 *Reviews Genetics* **11** (10), 733–739.
- 494 [36] McShaneL M, CavenaghM M, LivelyT G, EberhardD A, BigbeeW L, WilliamsP M, MesirovJ P,
495 PolleyM-Y C, KimK Y, TricoliJ V, TaylorJ M G, ShumanD J, SimonR M, DoroshowJ H, ConleyB A
496 (2013) Criteria for the use of omics-based predictors in clinical trials. *Nature* **502** (7471), 317–320.
- 497 [37] O’BryantS E, ListaS, RissmanR A, EdwardsM, ZhangF, HallJ, ZetterbergH, LovestoneS, GuptaV, Graff-
498 RadfordN, MartinsR, JerominA, WaringS, OhE, KingM, BakerL, HampelH (2015) Comparing biological
499 markers of Alzheimer’s disease across blood fraction and platforms: Comparing apples to oranges.
500 *Alzheimer’s & Dementia: Diagnosis, Assessment & Disease Monitoring* **3**, 27–34.
- 501 [38] ZwanM D, RinneJ O, HasselbalchS G, NordbergA, LleoA, HerukkaS-K, SoininenH, LawI, BahlJ M C,
502 CarterS F, ForteaJ, BlesaR, TeunissenC E, BouwmanF H, van BerckelB N M, VisserP J (2015) Use

503 of amyloid-PET to determine cutpoints for CSF markers: A multicenter study. *American Academy of*
504 *Neurology* **86** (1), 50–58.

505 [39] BaucknehtM, PiccoA, NobiliF, MorbelliS (2015) Amyloid positron emission tomography and cognitive
506 reserve. *World Journal of Radiology* **7** (12), 475–483.

Table 2: Cohort demographics

| Demographic | ADNI 1 (N=272) | | | ADNI 2 (N=43) | | | EDAR and DESCRIPA * (N=127) | | |
|---------------------------------|-----------------------------------|-------------------------------------|---------|----------------------------------|------------------------------------|---------|----------------------------------|------------------------------------|---------|
| | Normal CSF A β (N = 105) | Abnormal CSF A β (N = 167) | P-value | Normal CSF A β (N = 27) | Abnormal CSF A β (N = 16) | P-value | Normal CSF A β (N = 60) | Abnormal CSF A β (N = 67) | P-value |
| Median age [IQR] | 74.1 [8.10] | 75.1 [8.05] | 0.837 | 67.8 [12.55] | 73.75 [11.075] | 0.087 | 66 [8.75] | 69.55 [11.83] | 0.047 |
| Gender (%): | | | 0.61 | | | 0.528 | | | 0.859 |
| Female | 36.2 | 39.5 | | 51.9 | 37.5 | | 46.7 | 49.3 | |
| Male | 63.8 | 60.5 | | 48.1 | 62.5 | | 53.3 | 50.7 | |
| Median years in education [IQR] | 16 [4] | 16 [4] | 0.955 | 16 [4] | 16[4] | 0.949 | 12 [7] | 10 [5] | 0.334 |
| Median MMSE [IQR] | 29 [3] | 28 [3] | <0.001 | 29 [2] | 27.5 [2.25] | 0.1 | 28.5 [3] | 27 [5] | 0.024 |
| Diagnosis (%) | | | <0.001 | | | 0.446 | | | 0.002 |
| MCI | 42.9 | 76.6 | | 74.1 | 87.5 | | 71.7 | 92.5 | |
| CTL | 57.1 | 23.4 | | 25.9 | 12.5 | | 28.3 | 7.5 | |
| <i>APOE</i> status (%) | | | <0.001 | | | 0.343 | | | 0.004 |
| 0 | 84.8 | 38.3 | | 66.7 | 50.0 | | 66.7 | 40.3 | |
| 1 | 15.2 | 61.7 | | 33.3 | 50.0 | | 33.3 | 59.7 | |
| Median PGRS [IQR] | -0.615 [2.196] | -0.560 [2.095] | 0.018 | -0.088 [1.362] | -0.221 [1.127] | 0.763 | -0.252 [1.387] | 0.226 [1.266] | 0.022 |

Kruskal Wallis chi-squared was used to test between normal and abnormal groups for continuous demographic variables.

Fishers exact was used to test between normal and abnormal groups for categorical demographic variables.

APOE status is 1 if an individuals genotype contains any $\epsilon 4$ alleles, and 0 otherwise.

IQR = Inter-quartile range; CSF = Cerebrospinal Fluid; MMSE = Mini Mental State Exam; MCI = Mild Cognitive Impairment; CTL = Control.

* One individual has missing MMSE and education information.

Table 3: Test data results

| Model | Informative priors? | Accuracy [95% CI] | Sensitivity | Specificity | AUC ROC |
|--------------------------|---------------------|----------------------|-------------|-------------|---------|
| EDAR and DESCRIPA | | | | | |
| Demographics | No | 0.535 [0.445; 0.624] | 0.806 | 0.233 | 0.412 |
| Demographics | Yes | 0.543 [0.453; 0.632] | 0.776 | 0.283 | 0.412 |
| PGRS | No | 0.457 [0.368; 0.547] | 0.657 | 0.233 | 0.391 |
| PGRS | Yes | 0.394 [0.308; 0.484] | 0.493 | 0.283 | 0.394 |
| ADNI 2 | | | | | |
| Demographics | No | 0.488 [0.333; 0.645] | 0.625 | 0.407 | 0.458 |
| Demographics | Yes | 0.488 [0.333; 0.645] | 0.625 | 0.407 | 0.461 |
| PGRS | No | 0.465 [0.312; 0.623] | 0.625 | 0.37 | 0.396 |
| PGRS | Yes | 0.442 [0.291; 0.601] | 0.5 | 0.407 | 0.4 |

CI = Confidence Interval; AUC ROC = Area under the Receiver Operating Characteristic Curve;

PGRS = Polygenic Risk Score