Neurobiology of Aging 77 (2019) 178–182

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging

# Genetic analysis suggests high misassignment rates in clinical Alzheimer's cases and controls

Valentina Escott-Price [a,b], Emily Baker [b], Maryam Shoai [c], Ganna Leonenko [a], Amanda J. Myers [d], Matt Huentelman [e], John Hardy [c,*]

[a] Institute of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK
[b] Dementia Research Institute, Cardiff University, Cardiff, UK
[c] Department of Neurodegenerative Disorders, Institute of Neurology, Queen Square, London, UK
[d] Department of Psychiatry and Behavioral Sciences, Programs in Neuroscience and Human Genetics and Genomics and Center on Aging, Miller School of Medicine, University of Miami, Miami, FL, USA
[e] Neurogenomics Division, The Translational Genomics Research Institute (TGen), Phoenix, AZ, USA

ARTICLE INFO

*Article history:*
Received 30 July 2018
Received in revised form 10 December 2018
Accepted 10 December 2018
Available online 22 January 2019

*Keywords:*
Alzheimer
Genetics
Diagnosis
Biomarkers

ABSTRACT

Genetic case-control association studies are often based on clinically ascertained cases and population or convenience controls. It is known that some of the controls will contain cases, as they are usually not screened for the disease of interest. However, even clinically assessed cases and controls can be misassigned. For Alzheimer's disease (AD), it is important to know the accuracy of the clinical assignment. The predictive accuracy of AD risk by polygenic risk score analysis has been reported in both clinical and pathologically confirmed cohorts. The genetic risk prediction can provide additional insights to inform classification of subjects to case and control sets at a preclinical stage. In this study, we take a mathematical approach and aim to assess the importance of a genetic component for the assignment of subjects to AD-positive and -negative groups, and provide an estimate of misassignment rates (MARs) in AD case/control cohorts accounting for genetic prediction modeling results. The derived formulae provide a tool to estimate MARs in any sample. This approach can also provide an estimate of the maximal and minimal MARs and therefore could be useful for statistical power estimation at the study design stage. We illustrate this approach in 2 independent clinical cohorts and estimate misdiagnosis rate up to 36% in controls unscreened for the *APOE* genotype, and up to 29% when E3 homozygous subjects are used as controls in clinical studies.



## 1. Introduction

Genetic case-control association studies are often based on clinical assessment of cases and population or convenience controls. It is clearly the case that some of the controls can potentially contain patients in the early stage of disease, as they are not typically screened for the disease. It is assumed that the number of controls, who are actually cases, is relatively small and can be estimated by the prevalence of the disease in the population (e.g., ~3% lifetime prevalence of AD).

Polygenic risk score (PRS) analysis enhances the predictability of the diagnosis of AD (Escott-Price et al., 2015). The largest contributors to AD risk analysis, the E4 allele (risk) and the E2 allele (protective), gave area under the curve (AUC) of 0.68 (E4 alone) and 0.69 (E4+E2) as compared with overall PRS AUC = 0.75 in clinical cohorts [*ibid*]. In a recent PRS analysis, we showed that the AUC in a pathologically confirmed case/control series was 0.84 (Escott-Price et al., 2017). In addition, in a case/control sample of pathologically confirmed individuals who carry neither the E4 nor E2 allele (i.e., E3 homozygotes), the PRS gave AUC ~0.83 (95% CI: 0.80–0.86) (Escott-Price et al., 2019). When this was tested in clinical series, the AUC was reduced from 0.75 in the whole data set to 0.65 in E3 homozygotes [*ibid*]. This reduction in PRS in the clinical but not pathological series is indicative of a substantial misassignment rate (MAR) in the former.

A study at the National Institute on Aging Alzheimer Disease Centers (Beach et al., 2012) had reported measures of agreement between stratified levels for the clinical and neuropathologic diagnosis of AD in a sample of 919 subjects, who were classified based on their clinical categorization as "probable AD," "possible

* Corresponding author at: Department of Neurodegenerative Disorders, Institute of Neurology, Queen Square, London WC1N3BG, UK. Tel.: +44 2078298722; fax: +44 2031087490.
*E-mail address:* j.hardy@ucl.ac.uk (J. Hardy).

0197-4580/$ – see front matter © 2018 Published by Elsevier Inc.
https://doi.org/10.1016/j.neurobiolaging.2018.12.002

AD," or "not AD." The "not AD" group included non-AD dementias, and subjects with no dementia were excluded. The highest sensitivity (87.3%) reported in the study by Beach et al. (2012) was when the clinical diagnosis was defined as clinically probable or possible AD, and neuropathologic AD definition was defined as "frequent neuritic plaque density score" and Braak neurofibrillary tangle stage V or VI. In practice, most of the cases in clinical case/control samples are collected with "probable AD" diagnosis. For this combination of clinical and neuropathologic criteria, analysis of mismatched clinical and neuropathologic diagnoses provides sensitivity of 76.6% (Beach et al., 2012). This means that when the clinical diagnosis was defined as probable AD and the neuropathologic diagnosis as frequent neuritic plaques with Braak stage V-VI, 23.4% of people did not have frequent neuritic plaque density, despite their positive clinical diagnoses. Furthermore, more than a third of *APOE* E4 noncarriers with clinical diagnosis of mild to moderate Alzheimer's dementia had minimal AD plaque accumulation in cerebral cortex (Monsell et al., 2015).

In this study, we aim to estimate MAR in controls based on genetic prediction accuracy in clinical and neuropathology-confirmed samples of AD cases and controls. This is necessitated by the frequently asked question "what proportion of controls are actually early cases", when dealing with GWAS results? In this analysis, we seek to answer that question. We derive mathematical formulae to compare case/control classification by clinical diagnosis and true pathology status accounting for a hidden layer of genetic classification between diseased subjects and controls. These formulae were used to illustrate the potential MARs in clinical data samples using the reported values of prediction (by PRS) accuracy in AD pathology-confirmed samples of cases and controls (Escott-Price et al., 2017).

## 2. Methods

### 2.1. Derivation of misassignment rate estimates in a clinical sample

MAR was calculated using derived analytical formulae based on sensitivity and specificity. We first constructed three 2 × 2 contingency tables (also known as confusion matrices in the prediction modeling field), describing (1) clinical AD diagnosis (case/control) versus PRS prediction (yes/no) in a clinical sample, (2) pathologically confirmed AD status (yes/no) versus PRS prediction (yes/no), and (3) pathologically confirmed AD status versus clinical diagnosis. The latter table was expressed in terms of prediction accuracy measures (sensitivity, specificity, positive predictive value [PPV] and negative predictive value [NPV]), estimated from clinical and pathologically confirmed samples (see Appendix 1). The numerical results that we provide in this article are entirely derived from estimates made in previous publications. The estimates can be entirely populated using the clinical case/control numbers of the study.

### 2.2. Samples used for illustration of misassignment rates estimation

We applied the derived formulae and estimated MARs in 2 independent clinical cohorts. The first is the Genetic and Environmental Risk in Alzheimer's Disease (GERAD) consortium data (Harold et al., 2009). This is the first account where the prediction utility of AD PRS was reported. The best prediction accuracy using PRS was achieved when SNPs were pruned for linkage disequilibrium with parameters $r^2 = 0.1$ and a window of 1000 kb, and the most strongly associated AD SNPs with $p$-values $\leq 0.5$ were included in the individual PRS. The other independent data set was the Alzheimer's Disease Neuroimaging Initiative (ADNI) data set. This is a publically available database (http://www.loni.ucla.edu/ADNI/) which was started in 2004 and contains genetic, imaging, and biomarker data for about 900 individuals between the age of 55 and 90 years. Clinical diagnosis and genetic information were available for 770 individuals, who were either already diagnosed with AD (N = 47) or had mild cognitive impairment (MCI) (N = 459) or healthy controls (N = 262) at the baseline. We generated PRS for the ADNI participants in the same way as for GERAD data, using IGAP stage 1 (Lambert et al., 2013) summary statistics to inform AD PRS.

For prediction accuracy estimates in a pathologically confirmed sample, we used sensitivity/specificity/PPV/NPV estimates reported in the study by Escott-Price et al. (2017) for a pathologically confirmed sample of 1011 cases and 583 controls. This series was obtained from 21 National Alzheimer's Coordinating Center (NACC) brain banks and from the Miami Brain Bank as previously described [Corneveaux et al., 2010; Myers et al., 2007; Petyuk et al., 2018; Webster et al., 2009]. Our criteria for inclusion were as follows: self-defined ethnicity of European descent (in an attempt to control for the known allele frequency differences between ethnic groups), neuropathologically confirmed AD or no neuropathology present, and age of death greater than or equal to 65 years. Neuropathological diagnosis was defined by board-certified neuropathologists according to the standard NACC protocols [Beekly et al., 2004]. Samples derived from subjects with a clinical history of stroke, cerebrovascular disease, Lewy bodies, or comorbidity with any other known neurological disease were excluded. AD or control neuropathology was confirmed by plaque and tangle assessment with 45% of the entire series undergoing Braak staging (Braak and Braak, 1995). Samples were deidentified before receipt, and the study met human studies institutional review board and HIPPA regulations. This work is declared not human-subjects research and is IRB exempt under regulation 45 CFR 46.

To estimate the MAR in controls, the analytical formulae require us to fix the parameter of AD misdiagnosis rate in cases. Since most cases in clinical case/control samples are collected with clinically "probable AD" or "probable or possible AD" diagnosis, and in the pathology confirmed study (Escott-Price et al., 2017), the neuropathologic criterion for cases was Braak stage V or VI, we used sensitivity of 76.6% and 87.3% for AD misdiagnosis rates as reported in the study by Beach et al. (2012). In addition, according to Escott-Price et al. (2019), among *APOE* E4 noncarriers with the clinical diagnosis of mild to moderate AD, 37% had minimal neuritic plaques, and we used this value as an approximation of the misdiagnosis rate in the E3 homozygous cases.

## 3. Results

### 3.1. Estimation of misdiagnosis rates in a clinical sample

Assume that in a sample of $N$ clinically screened subjects ($N_{cas}^{(c)}$ cases and $N_{con}^{(c)}$ controls), $N_{cas}^{(p)}$ and $N_{con}^{(p)}$ are the numbers of true cases and controls that will be pathology confirmed (we use superscripts "(c)" and "(p)" to distinguish between the numbers of clinically and pathology-based classifications, respectively). The range for the number of subjects who were clinically and neuropathologically confirmed as having AD are between $\max\{0, N_{cas}^{(p)} - N_{con}^{(c)}\}$ and $\min\{N_{cas}^{(c)}, N_{cas}^{(p)}\}$. This means that in the worst-case scenario, all clinical cases are in fact unaffected (zero overlap), and in the best-case scenario, all clinical cases were given the correct diagnosis and will be confirmed neuropathologically. Similarly, the range for the number of controls who were also neuropathologically confirmed as "no AD" is between $\max\{0, N_{con}^{(c)} - N_{cas}^{(p)}\}$ and $\min\{N_{con}^{(c)}, N_{con}^{(p)}\}$. In reality, the number will lie somewhere in this range. To calculate these numbers in real data, we use values of prediction/classification accuracy reported in actual case/control studies.

For a clinical sample, the best PRS prediction accuracy (AUC) was reported as AUC = 0.75 with sensitivity and specificity $Se^{(c)} = Sp^{(c)} = 0.69$ (Escott-Price et al., 2015). The PRS prediction accuracy values in a pathologically confirmed sample of cases and controls were published in the study by Escott Price et al. (2017) as $Se^{(p)} = Sp^{(p)} = 0.79$, and $NPV^{(p)} = 0.69$. (The latter numbers, however, might be marginally overestimated, due to the 3% overlap of the discovery and test samples used in the study by Escott Price et al., 2017). Using these prediction accuracy values, we construct the confusion matrices (Tables A1 and A2 in Appendix 1) in the clinical sample (Escott-Price et al., 2015) of the total of $N = 4{,}603$ (3049 AD cases and 1554 controls) individuals, as shown in the following:

From these 2 tables we cannot simply imply that of 3049 clinical cases, 2892 cases will be pathologically confirmed, as some subjects, who are unaffected according to the clinical assessment, may actually have AD. Using sensitivity of 76.6% reported in the study by Beach et al. (2012), we estimate the number of true cases (which were clinically diagnosed as AD and expected also be pathologically confirmed) 3049*0.766 ≈ 2336 (denoted as $x$ in Appendix). Then the number of controls which expected to be pathologically confirmed is $N_{con}^{(c)} - N_{cas}^{(p)} + x = 1554 - 2892 + 2335 = 998$ (denoted as $y$ in the equation (1) in Appendix). Finally, in this sample, we obtain the MAR in controls MAR = 557/1554 = 0.36 (see equation (2) in Appendix 1).

For E3 homozygous subjects in the clinical cohort (Escott Price et al., 2015), the genetic based prediction AUC was lower (AUC = 0.65) with sensitivity and specificity $Se^{(c)} = Sp^{(c)} = 0.60$ (N cases = 1090 and N controls = 947). The values of the genetic prediction accuracy measures in pathologically confirmed sample (Escott-Price et al., 2017) were $Se^{(p)} = Sp^{(p)} = 0.745$, and $NPV^{(p)} = 0.768$. Clinical AD misdiagnosis rates in non-carriers of the apolipoprotein E4 allele are higher for subjects who are unscreened for E4 alleles. Using 37% as the approximation to AD misdiagnosis rate for E3 homozygous individuals (Monsell et al., 2015), gives the MAR in controls of about 29% clinical samples (Escott-Price et al., 2015). That is, about 29% of persons assigned as controls in the clinical series at the age of these series (late 70s) are in the early stages of disease.

In an attempt to replicate our result in an independent sample, we used the ADNI data. The ADNI cohort is older than GERAD; the mean age in the GERAD sample was 73.8 [SD = 8.6] and 71.4 [SD = 11.1], and the mean age in the ADNI sample at the last point of assessment was 78.4 [SD = 7.1] and 78.9 [SD = 7.6], in cases and controls, respectively. Similar to Tables 1–4 present the results for ADNI data. In this data set, we estimated the PRS for each individual as described in the study by Escott-Price et al. (2015) and calculated $Se^{(c)} = Sp^{(c)} = 0.678$, PPV = 0.621, NPV = 0.731, and AUC = 0.747. The values of the genetic prediction accuracy measures in pathologically confirmed sample (Escott-Price et al., 2017) were $Se^{(p)} = Sp^{(p)} = 0.79$ and $NPV^{(p)} = 0.686$. To estimate the MAR in controls with our analytical approach, we used sensitivity value 87.3%, which corresponds to the oldest group of people (83.2 years) with "clinically probable or possible" AD in the article of Beach et al. (2012). Our analytical approach gives the MAR in controls of 44.6%. The R code detailing these analyses is presented in Appendix 2.

In these data, we have also attempted to directly calculate MARs in controls. There were 262 controls available at the baseline assessment. On average within 4.7 years, 15 people have progressed to AD, 47 people have developed MCI, and 200 individuals did not change their diagnosis. This suggests the current MAR is in between 5.7% and 21.7%. The mean age of the progressors was 75.2 [4.0] years, and for those who did not progress the average age was 74.1 [SD = 5.7] years at the baseline of assessment. However, because AD is age dependent, it is expected that more controls will progress to AD when they reach age 85+ years. The incidence rate of AD increases almost exponentially with increasing age until 85 years of age. It is still debated whether the incidence will further increase at more advanced ages or will reach a plateau at a certain age (Qiu et al., 2009; Yesavage et al., 2002). Because there were only 5 individuals of age 85+ years in the ADNI data at the baseline, we were unable to estimate incidence rates directly. Here we used incident rates estimates (~55 persons per 1000-years at age 85+) reported by Qiu et al. (2009). Thus, we can expect an additional 55% of the sample to develop AD after 10 years, which is slightly above of the analytical estimate (44.6%).

**Table 2**
Pathologically confirmed status (derived estimates)

| Genetic test | Yes | No |
|---|---|---|
| Yes | $A = 2285$ | $B = 359$ |
| No | $C = 607$ | $D = 1352$ |
| Total | $N_{cas}^{(p)} = 2892$ | $N_{con}^{(p)} = 1711$ |

## 4. Discussion

It has been reported that AD misclassification rates range between 14% and 37% depending on the exact clinical and neuropathologic criteria used and whether the individuals were screened for *APOE* E4 alleles (Beach et al., 2012; Monsell et al., 2015). In addition, recent clinical trials show that 20% of all patients (and more than 33% of those who were noncarriers of the apolipoprotein E4 allele) with mild to moderate Alzheimer's dementia did not show an elevation in amyloid on positron emission tomography imaging (Doody et al., 2014; Salloway et al., 2014).

Conducting an actual autopsy-based study on unaffected individuals, aiming to identify AD cases among them, is difficult to justify unless it is a part of a large population screening study. Here we use the genetic prediction findings to mathematically estimate the misassignment in controls. Our earlier results show that the prediction accuracy of PRS in the pathologically confirmed sample of E3 homozygotes carriers is high and equivalent to the prediction accuracy in the samples of the whole data set (Escott Price et al., 2017 and under review), indicating that *APOE* is an independent risk factor for the disease. Therefore, we argue that it is not sufficient just to screen for *APOE* to classify subjects, for example, in AD clinical trials.

In this study, we derive analytical formulae to estimate MARs in clinical studies. These formulae are based on sensitivity, specificity, PPV, and NPV estimated from clinical and pathologically confirmed studies. However, the PPV and NPV estimates must be adjusted according to the case/control ratio of the clinical study for which MARs are being estimated (unless the ratios are equal), and here we

**Table 1**
Clinical diagnosis (GERAD)

| Genetic test | Yes | No |
|---|---|---|
| Yes | $a = 2096$ | $b = 485$ |
| No | $c = 953$ | $d = 1069$ |
| Total | $N_{cas}^{(c)} = 3049$ | $N_{con}^{(c)} = 1554$ |

**Table 3**
Clinical diagnosis (ADNI)

| Genetic test | Yes | No |
|---|---|---|
| Yes | $a = 118$ | $b = 72$ |
| No | $c = 56$ | $d = 152$ |
| Total | $N_{cas}^{(c)} = 174$ | $N_{con}^{(c)} = 224$ |

**Table 4**
Pathologically confirmed status (derived estimates)

| Genetic test | Yes | No |
|---|---|---|
| Yes | $A = 199$ | $B = 31$ |
| No | $C = 53$ | $D = 116$ |
| Total | $N_{cas}^{(p)} = 252$ | $N_{con}^{(p)} = 147$ |

show how to calculate sample prevalence–adjusted PPV and NPV. To demonstrate how these equations can be used in practice, we calculate misalignment rates in 2 independent clinical cohorts. Our headline figures are of course dependent on the quantities reported in previous studies. However, the approach is generalizable to other studies, and the MARs can be easily recalculated.

Our results show that the MARs in controls in clinical case-control studies is likely to be high (~30%). It would be expected to see an increased number of actual controls among E3 homozygous subjects as those individuals do not carry the strongest AD predictor. Indeed, the negative predictive value, or the percentage of correctly predicted controls, in the pathology-confirmed sample is higher than in clinical cohort (NPV = 0.77 and 0.57 in pathology confirmed and clinical samples, respectively). However, the misdiagnosis rate of cases in E3 homozygotes is high (37%), which implies reduced but still relatively high rates of misassignments, as compared with the sample not screened for *APOE* (29% vs. 36%, respectively).

In the ADNI data, there were 262 controls available, of them 15 progressed to AD, 47 developed MCI and 200 did not. This suggests a MAR between 8.0% and 23.7%; however, as AD is age-dependent, it is expected that more controls will progress to AD when they reach age 85+ years (prevalence of AD is 18% and 33% in 70–85 and 85+ years, respectively). Projecting the latter prevalence to this data, the MAR expected is about 40%, which is similar to our estimates.

These levels of MARs in both cases and controls reduce not only the power of statistical analyses in case/control series but also the PRS prediction accuracy in clinical samples. In biomarker studies of AD, they suggest that no biomarker will be able to give clean separations between those diagnosed with disease and those designated as controls because considerable proportions of both categories will be misclassified. As CSF and blood biomarkers of disease are assessed in clinical series, this inevitable misclassification, with ~30% of both cases and ~30% of controls being categorized in the wrong group.

## Disclosure

JH and VEP are cograntees of Cytox from Innovate UK (UK Department of Business). MS has previously been employed on this grant.

## Acknowledgements

JH and MS's genetic work is supported by an Anonymous Donor and the Dolby Foundation. JH and VEP are partly supported by the UK Dementia Research Institute. GL is funded by Dementia Platform UK. Additional funding was from the National Institutes of Health United States, as well as NIH EUREKA grant R01-AG-034504 to AJM and AG041232 (National Institute of Aging) to AJM and MH. Data collection was funded by Kronos. Data and biomaterials were collected from several National Institute on Aging (NIA) and National Alzheimer's Coordinating Center (NACC, grant #U01 AG016976) funded sites. Marcelle Morrison-Bogorad, PhD., Tony Phelps, PhD, and Walter Kukull PhD are thanked for helping to coordinate this collection. Additional non-NIA funded sites and full acknowledgments for the sample collection are given in the article of Corneveaux et al., (2010).

Authors' contributions: VEP, EB, and GL carried out the data analysis. VEP and JH designed the study and wrote the original draft. MS carried out quality control analyses of the genetic data. AM, MH, and JH were responsible for sample collection and data generation.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.neurobiolaging.2018.12.002.

## References

Beach, T.G., Monsell, S.E., Phillips, L.E., Kukull, W., 2012. Accuracy of the clinical diagnosis of Alzheimer disease at national Institute on aging Alzheimer disease centers, 2005-2010. J. Neuropathol. Exp. Neurol. 71, 266–273.

Beekly, D.L., Ramos, E.M., van Belle, G., Deitrich, W., Clark, A.D., Jacka, M.E., Kukull, W.A., 2004. The national Alzheimer's coordinating center (NACC) database: an Alzheimer disease database. Alzheimer Dis. Assoc. Disord. 18, 270–277.

Braak, H., Braak, E., 1995. Staging of Alzheimer's disease-related neurofibrillary changes. Neurobiol. Aging 16, 271–278 discussion 278-284.

Corneveaux, J.J., Myers, A.J., Allen, A.N., Pruzin, J.J., Ramirez, M., Engel, A., Nalls, M.A., Chen, K., Lee, W., Chewning, K., Villa, S.E., Meechoovet, H.B., Gerber, J.D., Frost, D., Benson, H.L., O'Reilly, S., Chibnik, L.B., Shulman, J.M., Singleton, A.B., Craig, D.W., Van Keuren-Jensen, K.R., Dunckley, T., Bennett, D.A., De Jager, P.L., Heward, C., Hardy, J., Reiman, E.M., Huentelman, M.J., 2010. Association of CR1, CLU and PICALM with Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. Hum. Mol. Genet. 19, 3295–3301.

Doody, R.S., 2014. Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease (vol 370, pg 311, 2014). N. Engl. J. Med. 371, 584–684.

Escott-Price, V., Sims, R., Bannister, C., Harold, D., Vronskaya, M., Majounie, E., Badarinarayan, N., Morgan, K., Passmore, P., Holmes, C., Powell, J., Brayne, C., Gill, M., Mead, S., Goate, A., Cruchaga, C., Lambert, J.C., van Duijn, C., Maier, W., Ramirez, A., Holmans, P., Jones, L., Hardy, J., Seshadri, S., Schellenberg, G.D., Amouyel, P., Williams, J., PERADES, G., Consortia, I., 2015. Common polygenic variation enhances risk prediction for Alzheimer's disease. Brain 138, 3673–3684.

Escott-Price, V., Myers, A.J., Huentelman, M., Hardy, J., 2017. Polygenic risk score analysis of pathologically confirmed Alzheimer disease. Ann. Neurol. 82, 311–314.

Escott-Price, V., Myers, A.J., Huentelman, M., Hardy, J., 2019. 2018 Polygenic risk score analysis of Alzheimer's disease in cases without APOE4 or APOE2 alleles. J. Prev. Alzheimers Dis. 6, 16–19.

Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M.L., Pahwa, J.S., Moskvina, V., Dowzell, K., Williams, A., Jones, N., Thomas, C., Stretton, A., Morgan, A.R., Lovestone, S., Powell, J., Proitsi, P., Lupton, M.K., Brayne, C., Rubinsztein, D.C., Gill, M., Lawlor, B., Lynch, A., Morgan, K., Brown, K.S., Passmore, P.A., Craig, D., McGuinness, B., Todd, S., Holmes, C., Mann, D., Smith, A.D., Love, S., Kehoe, P.G., Hardy, J., Mead, S., Fox, N., Rossor, M., Collinge, J., Maier, W., Jessen, F., Schurmann, B., van den Bussche, H., Heuser, I., Kornhuber, J., Wiltfang, J., Dichgans, M., Frolich, L., Hampel, H., Hull, M., Rujescu, D., Goate, A.M., Kauwe, J.S., Cruchaga, C., Nowotny, P., Morris, J.C., Mayo, K., Sleegers, K., Bettens, K., Engelborghs, S., De Deyn, P.P., Van Broeckhoven, C., Livingston, G., Bass, N.J., Gurling, H., McQuillin, A., Gwilliam, R., Deloukas, P., Al-Chalabi, A., Shaw, C.E., Tsolaki, M., Singleton, A.B., Guerreiro, R., Muhleisen, T.W., Nothen, M.M., Moebus, S., Jockel, K.H., Klopp, N., Wichmann, H.E., Carrasquillo, M.M., Pankratz, V.S., Younkin, S.G., Holmans, P.A., O'Donovan, M., Owen, M.J., Williams, J., 2009. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nature genetics 41, 1088–1093.

Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., Russo, G., Thorton-Wells, T.A., Jones, N., Smith, A.V., Chouraki, V., Thomas, C., Ikram, M.A., Zelenika, D., Vardarajan, B.N., Kamatani, Y., Lin, C.F., Gerrish, A., Schmidt, H., Kunkle, B., Dunstan, M.L., Ruiz, A., Bihoreau, M.T., Choi, S.H., Reitz, C., Pasquier, F., Cruchaga, C., Craig, D., Amin, N., Berr, C., Lopez, O.L., De Jager, P.L., Deramecourt, V., Johnston, J.A., Evans, D., Lovestone, S., Letenneur, L., Moron, F.J., Rubinsztein, D.C., Eiriksdottir, G., Sleegers, K., Goate, A.M., Fievet, N., Huentelman, M.W., Gill, M., Brown, K., Kamboh, M.I., Keller, L., Barberger-Gateau, P., McGuiness, B., Larson, E.B., Green, R., Myers, A.J., Dufouil, C., Todd, S., Wallon, D., Love, S., Rogaeva, E., Gallacher, J., St George-Hyslop, P., Clarimon, J., Lleo, A., Bayer, A., Tsuang, D.W., Yu, L., Tsolaki, M., Bossu, P., Spalletta, G., Proitsi, P., Collinge, J., Sorbi, S., Sanchez-Garcia, F., Fox, N.C., Hardy, J., Deniz Naranjo, M.C., Bosco, P., Clarke, R., Brayne, C., Galimberti, D., Mancuso, M., Matthews, F., European Alzheimer's Disease I, Genetic, Environmental Risk in Alzheimer's D, Alzheimer's Disease Genetic C, Cohorts for H, Aging Research in

Genomic E, Moebus, S., Mecocci, P., Del Zompo, M., Maier, W., Hampel, H., Pilotto, A., Bullido, M., Panza, F., Caffarra, P., Nacmias, B., Gilbert, J.R., Mayhaus, M., Lannefelt, L., Hakonarson, H., Pichler, S., Carrasquillo, M.M., Ingelsson, M., Beekly, D., Alvarez, V., Zou, F., Valladares, O., Younkin, S.G., Coto, E., Hamilton-Nelson, K.L., Gu, W., Razquin, C., Pastor, P., Mateo, I., Owen, M.J., Faber, K.M., Jonsson, P.V., Combarros, O., O'Donovan, M.C., Cantwell, L.B., Soininen, H., Blacker, D., Mead, S., Mosley Jr., T.H., Bennett, D.A., Harris, T.B., Fratiglioni, L., Holmes, C., de Bruijn, R.F., Passmore, P., Montine, T.J., Bettens, K., Rotter, J.I., Brice, A., Morgan, K., Foroud, T.M., Kukull, W.A., Hannequin, D., Powell, J.F., Nalls, M.A., Ritchie, K., Lunetta, K.L., Kauwe, J.S., Boerwinkle, E., Riemenschneider, M., Boada, M., Hiltunen, M., Martin, E.R., Schmidt, R., Rujescu, D., Wang, L.S., Dartigues, J.F., Mayeux, R., Tzourio, C., Hofman, A., Nothen, M.M., Graff, C., Psaty, B.M., Jones, L., Haines, J.L., Holmans, P.A., Lathrop, M., Pericak-Vance, M.A., Launer, L.J., Farrer, L.A., van Duijn, C.M., Van Broeckhoven, C., Moskvina, V., Seshadri, S., Williams, J., Schellenberg, G.D., Amouyel, P., 2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat. Genet. 45, 1452–1458.

Monsell, S.E., Kukull, W.A., Roher, A.E., Maarouf, C.L., Serrano, G., Beach, T.G., Caselli, R.J., Montine, T.J., Reiman, E.M., 2015. Characterizing apolipoprotein E epsilon4 carriers and noncarriers with the clinical diagnosis of mild to moderate Alzheimer dementia and minimal beta-amyloid peptide plaques. JAMA Neurol. 72, 1124–1131.

Myers, A.J., Gibbs, J.R., Webster, J.A., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., Leung, D., Bryden, L., Nath, P., Zismann, V.L., Joshipura, K., Huentelman, M.J., Hu-Lince, D., Coon, K.D., Craig, D.W., Pearson, J.V., Holmans, P., Heward, C.B., Reiman, E.M., Stephan, D., Hardy, J., 2007. A survey of genetic human cortical gene expression. Nat. Genet. 39, 1494–1499.

Petyuk, V.A., Chang, R., Ramirez-Restrepo, M., Beckmann, N.D., Henrion, M.Y.R., Piehowski, P.D., Zhu, K., Wang, S., Clarke, J., Huentelman, M.J., Xie, F., Andreev, V., Engel, A., Guettoche, T., Navarro, L., Jager, P., Schneider, J.A., Morris, C.M., McKeith, I.G., Perry, R.H., Lovestone, S., Woltjer, R.L., Beach, T.G., Sue, L.I., Serrano, G.E., Lieberman, A.P., Albin, R.L., Ferrer, I., Mash, D.C., Hulette, C.M., Ervin, J.F., Reiman, E.M., Hardy, J.A., Bennett, D.A., Schadt, E., Smith, R.D., Myers, A.J., 2018. The human brainome: network analysis identifies HSPA2 as a novel Alzheimer's disease target. Brain 141, 2721–2739.

Qiu, C., Kivipelto, M., von Strauss, E., 2009. Epidemiology of Alzheimer's disease: occurrence, determinants, and strategies toward intervention. Dialogues in clinical neuroscience 11, 111–128.

Salloway, S., Sperling, R., Fox, N.C., Blennow, K., Klunk, W., Raskind, M., Sabbagh, M., Honig, L.S., Porsteinsson, A.P., Ferris, S., Reichert, M., Ketter, N., Nejadnik, B., Guenzler, V., Miloslavsky, M., Wang, D., Lu, Y., Lull, J., Tudor, I.C., Liu, E.C., Grundman, M., Yuen, E., Black, R., Brashear, H.R., Bapineuzumab, C., 2014. Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. N. Engl. J. Med. 370, 322–333.

Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., McCorquodale, D.S., Cuello, C., Leung, D., Bryden, L., Nath, P., Zismann, V.L., Joshipura, K., Huentelman, M.J., Hu-Lince, D., Coon, K.D., Craig, D.W., Pearson, J.V., Heward, C.B., Reiman, E.M., Stephan, D., Hardy, J., Myers, A.J., Grp, NA-N., 2009. Genetic control of human brain transcript expression in Alzheimer disease. Am. J. Hum. Genet. 84, 445–458.

Yesavage, J.A., O'Hara, R., Kraemer, H., Noda, A., Taylor, J.L., Ferris, S., Gely-Nargeot, M.C., Rosen, A., Friedman, L., Sheikh, J., Derouesne, C., 2002. Modelling the prevalence and incidence of Alzheimer's disease and mild cognitive impairment. J. Psychiatr. Res. 36, 281–286.