Genetic risk for Alzheimer’s disease is distinct from genetic risk for amyloid deposition

Running head: genetic risk for amyloid deposition

Ganna Leonenko1+, Maryam Shoai2+, Eftychia Bellou1, Rebecca Sims1, Julie Williams1,3, John Hardy2,4, Valentina Escott-Price1,3, & and the Alzheimer’s Disease Neuroimaging Initiative ^

1 MRC Centre for Neuropsychiatric Genetics and Genomics and UK Dementia Research Institute, Cardiff University, UK
2 Reta Lilla Research Laboratories, Department of Neurodegeneration and UK Dementia Research Institute, UCL Institute of Neurology, London, UK
3 UK Dementia Research Institute at Cardiff University, Cardiff, UK
4 UK Dementia Research Institute at UCL, London, UK

^ Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). As such, 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. ADNI investigators include (complete listing available at http://adni.loni.ucla.edu/wpcontent/uploads/how_to_apply/ADNI_Authorship.pdf)

& Corresponding author
e-mail: EscottPriceV@cardiff.ac.uk
Address: Dementia Research Institute, Hadyn Ellis Building, Cardiff University, Maindy Road, Cardiff, CF24 4HQ
Phone: +(44)2920688429
+ joint first author
Abstract

Objectives: Alzheimer's disease (AD) is the most common form of dementia and is responsible for a huge and growing health care burden in the developed and developing world. The Polygenic Risk Score (PRS) approach has shown 75%-84% prediction accuracy of identifying individuals with AD risk.

Methods: In this study we tested the prediction accuracy of AD, MCI and amyloid deposition risks with PRS, including and excluding APOE genotypes in a large publicly available data set with extensive phenotypic data: the Alzheimer's Disease Neuroimaging Initiative cohort. Among MCI individuals with amyloid positive status we examined PRS prediction accuracy in those who converted to AD. In addition, we divided polygenic risk score by biological pathways and tested them independently for distinguishing between AD, MCI and amyloid deposition.

Results: We found that AD and MCI are predicted by both APOE genotype and PRS (AUC=0.82% and 68%, respectively). Amyloid deposition is predicted by APOE only (AUC=79%). Further progression to AD of individuals with MCI and amyloid positive status is predicted by PRS over and above APOE (AUC=67%). In pathway-specific PRSs analyses the protein-lipid complex has the strongest association with AD and amyloid deposition even when genes in APOE region were removed (p=0.0055 and p=0.0079, respectively).

Interpretation: The results showed different pattern of APOE contribution in PRS risk predictions of AD/MCI and amyloid deposition. Our study suggests that APOE mostly contributes to amyloid accumulation and the PRS affects risk of further conversion to AD.
Introduction

Alzheimer’s disease is the most common form of dementia in the elderly people and is a major health problem world-wide\(^1\). The clinical diagnosis is typically characterized by progressive loss of memory and cognitive function. In the last decade numerous relevant susceptibility loci, genes and pathways have been identified\(^2\)–\(^6\) that have improved the understanding of this complex disease. However the risk for developing AD involves multiple genetic and environmental components, with the \textit{APOE} genotype\(^7\) having the strongest genetic effect\(^2\).

Amyloid-beta (A\(\beta\)) plays a key role in the pathogenesis of AD but little is known about the process of its formation in a brain. Identification of earliest pathological signature of Alzheimer’s disease requires longitudinal measurements of A\(\beta\) deposition in the brain by positron-emission tomography (PET) imaging or by measurements of A\(\beta\) reduction in cerebrospinal fluid (CSF). Although A\(\beta\) is necessary for the pathologic diagnosis of AD it is not sufficient in itself to cause cognitive dysfunction and clinical AD. It has been shown that amyloid deposition has low specificity for predicting development of AD\(^8\)\(^9\).

The pre-clinical stage of AD starts with mild impairment in cognitive domains (MCI) and includes a syndrome featuring relatively isolated memory deficits\(^10\). In 2011, the National Institute on Aging and Alzheimer’s Association (NIA-AA) created separate sets of diagnostic guidelines for the symptomatic or ”clinical” stages of AD\(^11\)\(^12\), where AD represents the ”disease”, and ”dementia” represents the clinical syndrome. Thus a person may progress from MCI to dementia (due to AD); but both MCI and dementia cases may or may not be AD.

Studying individuals that develop MCI and then further progress to AD requires detailed longitudinal datasets. ADNI is a multicentre study designed to assess the utility of various biomarkers for detecting early changes associated with MCI and AD. It includes collection of neuroimaging data, clinical and cognitive assessments, along with information on demographics and individual genetic profiles.

The PRS approach aggregates the effects of multiple genetic markers identified through Genome-Wide Association Studies (GWAS)\(^2\) and has shown great potential in identifying an individual’s risk of developing AD\(^13\)\(^14\). A few studies...
have recently used AD PRS to predict mild cognitive functions and clinical MCI\textsuperscript{15}, however, only one has suggested that PRS could identify MCI in middle aged adults\textsuperscript{16} more effectively than the \textit{APOE} locus alone. The PRS approach has also been applied to biological pathways related to AD but was not more predictive than \textit{APOE} alone\textsuperscript{17}. The implementation of the Polygenic Hazard Score (PHS) (which is closely related to PRS\textsuperscript{18}) analysis in the ADNI data showed that PHS is associated to AD biomarkers (CSF and PET) in individuals without AD\textsuperscript{19}, and that higher PHS were associated with greater rates of cognitive and clinical decline, even after controlling for \textit{APOE} status\textsuperscript{20}: however, its predictive value was not quantified.

In this study we estimate the predictive accuracy of PRS differentiating a) AD cases versus controls, b) MCI cases versus controls, c) amyloid positive versus amyloid negative individuals. We also investigate whether d) the AD PRS can predict individuals with MCI who will progress to AD and those who will remain MCI, with positive amyloid deposition.

Recently GWAS studies and exome/genome sequencing have implicated, with varying degrees of confidence, lipid metabolism, the innate immune system and endosomal vesicle recycling in late-onset AD pathogenesis\textsuperscript{21,22}. Therefore we also examined the pathway-specific PRS association using these recently identified pathway\textsuperscript{6} related to AD risk.

\textbf{Methods}

\textbf{ADNI: Setting/Clinical description}

Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) is a publicly available database (adni.loni.usc.edu). The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. The data were collected for about 900 individuals between ages 55-90. The initial five-year study (ADNI1) followed participants for 2-3 years with repeated imaging scans and psychometric measurements every 6 or 12 months. All ADNI participants
provided written informed consent. The ADNI project was extended as ADNI-GO and ADNI2 studies with a proportion of new and original ADNI1 participants. Clinical diagnosis and genetic information were available for 770 individuals from ADNI1, ADNI-GO and ADNI2 studies. Longitudinal data contained information of clinical assessments from the first visit (baseline) to the latest available visit with mean follow up time approximately 5 years. Details of the ADNI design, participant recruitment, clinical testing, and additional methods have been previously reported elsewhere\textsuperscript{23,24}.

Table 1 shows the classification of diagnosis and number of individuals whose diagnosis remained stable during the study. It also presents the diagnostic categories and the numbers of individuals within those diagnostic categories at the latest assessment, which were used for the analyses.

To assess amyloid deposition, the latest MRI PET scans from 663 participants were used in the analysis (AV45 ligand threshold of 1.11). In this study we used the individuals’ diagnosis at the latest point of assessment. We then tested whether AD polygenic risk scores are associated with AD, MCI and amyloid status in three main analyses a) AD vs controls, b) MCI vs controls, and c) amyloid positive vs amyloid negative status (Table 2).

**ADNI: genotyping and QC**

770 samples from ADNI1/GO/2 set were whole-genome sequenced (WGS) and genotyped using the Ilumina Omni 2.5M BeadChip (42,732,452 variants). WGS calls were made using the Broad Institute best practices (BWA & GATK HaplotypeCaller).

Basic quality control checks were performed using standard procedure\textsuperscript{25}. SNPs were excluded where genotype missingness was greater than 0.02, Hardy-Wienberg equilibrium p-value was <1e-6 and SNP minor allele frequency was smaller than 0.01. This retained 7,808,548 SNPs for the analyses. Matching those SNPs with the latest publically available GWAS AD summary statistics\textsuperscript{2} reduced that number to 5,771,686.

**Generating PRS**

Generation of PRS requires two independent datasets: summary statistics of association with AD in a discovery sample, and a test sample, which is independent of the discovery sample and contains genotypes for each
individual\textsuperscript{26}. As the discovery sample we used summary statistics from the powerful study genome-wide association study (17,008 AD cases and 37,154 controls) of the International Genomics of Alzheimer’s Project (IGAP Stage 1)\textsuperscript{2}. PRS were generated using SNPs with AD association p-value ≤ 0.5 in the IGAP dataset, as it has been reported as having the best prediction accuracy\textsuperscript{13}. The SNPs were then Linkage Disequilibrium (LD) pruned (r\textsuperscript{2}=0.1 and 1000kb window), keeping the SNPs most associated with AD. The number of SNPs after the LD pruning was 162,957. We included APOE ε2 and ε4 allele genotypes directly to the PRS with effect sizes B=-1.04 and B=1.55 for ε2 and ε4, respectively, calculated in the ADNI data, whilst excluding the APOE region (chromosome 19:44,400 -19:46,500 kb)\textsuperscript{13}. Prior to all analyses, the PRS were adjusted for the 8 principal components and then standardised.

441 ADNI participants were part of original IGAP summary statistics\textsuperscript{2}. To overcome a potential bias in PRS analysis due to overlapping samples, we used a simulation approach we previously described\textsuperscript{14}. In brief, first we assessed the variation in the SNPs’ effect sizes using 1\textsuperscript{,}000 simulations when randomly excluding 266 cases and 173 controls (matching the numbers of overlapping samples). The variation in the IGAP effect sizes due to the overlap was estimated in terms of standard deviation (SD\textsubscript{IGAP}=0.053) from the mean (i.e. the original IGAP SNP beta-coefficient (Beta\textsubscript{IGAP})). Then, new IGAP genome-wide summary statistics were simulated 10,000 times with adjusted effect sizes (Beta\textsubscript{adjusted}) and p-values for each SNP. Beta\textsubscript{adjusted} was sampled from a Normal distribution with mean=Beta\textsubscript{IGAP} and SD=0.053*SE\textsubscript{IGAP}, p-values\textsubscript{adjusted} were redefined accordingly. At each simulation SNPs were re-selected and re-pruned based on LD r\textsuperscript{2}=0.1 and 1000kb window. The prediction accuracies (AUCs) reported in the Results section are presented as means from these 10,000 simulations.

**Genome-wide and pathway specific PRS predictions**

Initially we tested whether PRSs are associated with AD risk (AD cases vs controls) in the ADNI data set. Then we assessed whether the AD PRS can distinguish individuals with MCI from cognitively normal controls and amyloid positive from amyloid negative individuals. Finally, we were interested to assess whether PRS can predict AD risk over and above APOE in MCI individuals who have had positive amyloid deposition (to be precise, the MCI individuals who
converted to AD between the baseline and final time of assessment vs non-converters). All analyses were performed using logistic regression models with the following predictors: a) \( APOE \) (\( ε2 + ε4 \)), b) PRS without \( APOE \) and c) full PRS model (a. and b. together). Gender and age we used as covariates in all analyses. We tested whether the PRS significantly improves the model fit over and above \( APOE \) alone with anova() function in R. We report the accuracy of the models in terms of Area Under the receiver operator Curve (AUC). In addition, we calculated PRS prediction accuracies (AUCs) in the extremes of PRS distribution for individuals whose PRS score was greater or smaller than ±1.5 standard deviations from the PRS mean.

For the pathway-specific analyses we chose latest published nine pathways which have been reported as playing a role in AD pathogenesis, namely: (1) protein-lipid complex assembly, (2) regulation of beta-amyloid formation, (3) protein-lipid complex, (4) regulation of amyloid precursor protein catabolic process, (5) reverse cholesterol transport, (6) protein-lipid complex subunit organization, (7) plasma lipoprotein particle assembly, (8) tau protein binding and (9) activation of immune response. Finally, to quantify the proportion of variance which remains unexplained by the pathways together, we calculated and tested PRS for the whole genome excluding these nine pathways.

Pathway specific PRS were generated in ADNI dataset for each individual as described above with and without \( APOE \) region. The PRS scores in this case were adjusted not only for 8 principal components but also for age and gender and then standardised.

The results were considered as significant if the resulting p-value was smaller than or equal to 1.85x10\(^{-3}\)=0.05/(3 scenarios × 9 pathways), corresponding to the Bonferroni correction for multiple comparisons.

**Results**

The prediction accuracy of AD cases (N=174) vs controls (N=224) at the last assessment point was AUC\(_{APOE}\)=76\% and AUC\(_{PRS}\)=75\%, for \( APOE \) alone and for PRS without \( APOE \), respectively (Table 3, first row). The best prediction accuracy (AUC\(_{FULL}\)=82\%) was achieved with the full model, which includes both \( APOE \) and PRS. An anova test (last column of Table 3) confirmed that PRS significantly improves the prediction accuracy of the model over and above \( APOE \) (p=1.7 x10\(^{-3}\)}.
A similar pattern of results was observed when we compared MCI individuals at the last point of assessment (N=344, see Table 1 for details) with controls, however, the accuracy has reduced (AUC$_{APOE}$ = 62%, AUC$_{FULL}$ = 68%). Again, PRS significantly improves the prediction accuracy of MCI risk over and above $APOE$ (p=2.5 x10$^{-11}$). Figure 1 shows standardised density plots of polygenic risk scores in AD cases (red line), controls (blue line) and MCI (orange line), where the mean of the PRS for the latter is between the means of the PRS for AD cases and controls. Interestingly, the results for prediction of amyloid deposition by PRS follows a different pattern: $APOE$ alone significantly predicted amyloid deposition with AUC$_{APOE}$ = 76%, and PRS did not improve the prediction accuracy further.

When we tested the full PRS model for prediction of individuals at the extremes of polygenic score distribution (±1.5sd from the PRS mean), the prediction accuracy as expected, increased (AUC=94% for AD vs controls and AUC=91% for MCI vs controls).

We were interested to test whether the PRS can predict progression to AD in individuals with MCI. Out of 459 individuals with MCI at the baseline assessment, 441 had known amyloid deposition status (270 were amyloid positive and 171 were amyloid negative). The prediction accuracy of amyloid deposition in this subsample was AUC$_{APOE}$=79% by $APOE$ alone and PRS did not improve the prediction accuracy (p=0.48, Figure 2). Out of 270 amyloid positive individuals, 112 have progressed to AD and 150 individuals remained MCI as at the last point of assessment. In this case PRS did predict AD progression (AUC$_{APOE}$=63% and AUC$_{FULL}$=69%), significantly improving the prediction over and above $APOE$ (p=0.0002, Figure 2).

Finally, we calculated pathway specific PRSs and tested them for association with risk for AD, MCI and amyloid deposition. The results are presented in Table 4. The majority of pathways were significantly associated with AD risk however this association was mostly driven by $APOE$ region. Two pathways (protein-lipid complex, protein-lipid complex subunit) remained significant after removing genes in the $APOE$ region. When we excluded all pathways from the whole genome PRS, we observed that a substantial part of variance still remains unexplained (p=2.2x10$^{-14}$, last row of Table 4). Comparing amyloid positive vs
amyloid negative individuals, the same two protein-lipid-related pathways and additionally reverse cholesterol transport were significant after removing genes in the APOE region. The association results of the nine pathways’ PRS with MCI risk were nominally significant for all pathways and the association was mostly attributed to APOE. This clearly demonstrates that the pathways which contain APOE region are strong predictors of amyloid deposition. Protein-lipid complex has shown the strongest association with AD and amyloid deposition risk in all the analyses. The overlap of genes in the three pathways above is presented in Figure 3.

Finally, we tested these pathways’ PRS for association with amyloid deposition in individuals with MCI and with their further progression to AD when their amyloid deposition status was positive. We found that protein lipid complex, protein-lipid complex subunit organization and reverse cholesterol transport pathways are also associated with amyloid deposition even after exclusion the APOE region (Table 5, 4th column).

Discussion

The pathological process related to AD starts long before clinical onset and lasts approximately 15-20 year\(^{28}\). It is widely believed that identifying individuals that have high risk of AD earlier is essential for therapeutic strategies for AD prevention and intervention\(^{29}\). Due to the diagnostic heterogeneity of MCI and different length of follow up assessments, the conversion rate to AD or other types of dementia varies widely over different studies\(^{30,31}\). Identifying individuals with MCI and monitoring them through biomarker measurements should provide a better understanding of the process of progression from MCI to AD. Although there is no generally accepted diagnostic criteria that specifies MCI individuals who will convert to AD it is notable that increased amyloid plaques that starts many years before clinical symptoms appear, plays an important role in brain degenerative changes.

A reasonable prediction accuracy can be achieved with a PRS approach that uses genetic profile information and relates it to AD risk\(^{13,14}\). The PRS and its modifications have been assessed for association with AD and AD-related phenotypes in a number of studies, however the reported prediction accuracies
have not been entirely consistent. In this study we examined prediction accuracy that can be achieved with \textit{APOE} alone and with the full PRS model differentiating between AD, MCI, controls and amyloid status. We have shown that the best prediction accuracy can be achieved with the PRS which includes \textit{APOE} for both AD vs controls and MCI vs controls analyses (AUC=82\% and AUC=68\%, respectively). In both analyses the PRS improves the prediction accuracy by about 8-9\% as compared to \textit{APOE} alone, which replicates the analyses in independent datasets published elsewhere\textsuperscript{13,14,16}. Of course, GWA studies indicate that \textit{APOE} is the strongest risk factor and other common genetic variants have smaller effect sizes. However, \textit{APOE} region explains \textasciitilde5\% of SNP heritability whilst the whole genome explains \textasciitilde24\%\textsuperscript{32}. In addition, PRS prediction accuracy shows a substantial increase in AUC which makes the PRS potentially clinically useful for disease risk prediction. Furthermore, AD GWAS risk loci have greatly expanded our understanding of the disease mechanisms. As expected, the accuracy of MCI prediction is lower than AD, which can be explained by the inclusion of a subset of MCI individuals who will not develop AD. For individuals with extreme PRS the AUC reaches 90\% and above for both AD and MCI.

The prediction of amyloid deposition showed a different pattern. In the whole sample the prediction accuracy with \textit{APOE} alone was 76\% and the PRS did not improve the accuracy any further (AUC remained 76\%). Similar results were obtained when we tested the prediction accuracy of amyloid deposition in individuals with MCI. However, when we looked at individuals who have already had positive amyloid deposition and attempted to predict their progression to AD, the best accuracy was observed with the full PRS model which includes \textit{APOE} region however also requires the PRS component.

Note, that for all the used models the best prediction accuracy was achieved with p-value threshold of 0.5 of AD associated SNPs. The same threshold was previously reported in studies that were done on different genotyping arrays\textsuperscript{13,14}. For the best prediction accuracy in clinical practice PRS should be generated on set of SNPs in a way that captures genetic liability of the whole genome.
The potential implication of these findings is that the APOE gene affects amyloid deposition but that much of the rest of the risk of disease is involved in the rate at which amyloid deposition causes a neurodegenerative response. Clinical trials have previously shown that there is little correlation between AD progression and accumulation of amyloid plaques supporting a hypothesis that AD development may have two separated stages: amyloid dependent and amyloid independent. It is also known that the APOE gene influences the deposition of amyloid in the brain and that this is necessary but not sufficient for development of clinical AD. Moreover it has been shown that neuronal loss and tangle numbers increase as AD progresses unlike the number of amyloid plaques reaches its maximum with the onset of clinical symptoms.

While analysis of early onset AD firmly implicated APP metabolism and Aβ production in the aetiology of the disease, genome wide association studies and exome and genome sequencing have implicated with varying degrees of confidence a number of potentially biologically relevant pathways in late-onset AD pathogenesis. Of course, pathway construction is an imperfect art both because of the knowledge base used in the generation of the pathways and because proteins may have more than one function in more than one cell type. Nevertheless, it is valuable to divide polygenic risk by pathways both in terms of modeling the disease through iPSC technologies, one might like to assign high or low risk by pathway and, eventually, it is possible that one might wish to tailor therapies to pathway deficits. To dissect AD PRS by biologically relevant gene sets, we tested pathways enriched in AD identified by International Genomic Alzheimer’s Project (IGAP). All pathways, except “activation of immune response” were highly significantly associated with AD risk and amyloid deposition risk, however most of the signal was attributed to APOE region alone. Protein-lipid complex has shown the strongest association to AD and amyloid deposition risk in all the analyses.

In conclusion, our results imply that APOE contributes to disease risk in a manner that is mechanistically different from the other genetic contributors to disease risk. We speculate that APOE affects amyloid deposition and that the PRS affects conversion from amyloid positivity to AD. Therefore, in the context of the amyloid cascade hypothesis, APOE acts prior to amyloid deposition and the
remaining genetic risk factors identified through GWAS act between amyloid deposition and clinical onset of Alzheimer’s disease.

**Acknowledgements**

We thank the MRC Centre for Neuropsychiatric Genetics and Genomics and UK Dementia Research Institute (DRI), Dementia Platform UK (DPUK), InnovateUK. JH additionally thanks the Dolby Foundation and an anonymous donor for support.

We thank the International Genomics of Alzheimer’s Project (IGAP) for providing summary results data for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report.

Also we thank ADNI database. Collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute of Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug discovery foundation; Aralcon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpin, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmum; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer’s’ Immunotherapy Research and Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research
Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Author Contribution

GL, MS, EB, RS contributed to the acquisition and data analysis; GL, MS, EB, RS, JH, VE-P contributed to the drafting of the manuscript or part of it; JH, JW, VE-P contributed to the critical review of the manuscript; JH, JW, VE-P contributed to the conception and design of the study.

Potential Conflicts of Interest

Nothing to report.

References


34. Musiek ES, Holtzman DM. Three dimensions of the amyloid hypothesis:


Figure 1: Density plots of PRS for AD, MCI and cognitively normal participants.

Legend: Standardised individual PRS scores for three phenotypes (AD, MCI and controls).

Figure 2: Diagram of prediction of Amyloid deposition and further prediction of conversion of MCI individuals to Alzheimer’s Diseased in sample that were first clinically diagnosed with MCI using APOE and AD PRS.

Legend: PRS predictions were first made for individuals who had baseline diagnosis of MCI. APOE alone and full PRS model were used to predict amyloid deposition. The same models were used to predict which MCI individuals converted to AD versus those individuals who had an MCI diagnosis using the latest clinical diagnosis.

Figure 3. Overlap between three pathways. 1: protein-lipid complex (40 genes), 2: protein-lipid complex subunit organization (35 genes), 3: reverse cholesterol transport (17 genes).
### Table 1: Clinical classification of diagnosis in ADNI dataset

<table>
<thead>
<tr>
<th>Diagnosis description</th>
<th>Number of samples with diagnosis at the first time point</th>
<th>Number of samples with diagnosis at the last time point</th>
<th>N samples stable over time</th>
<th>Usage for the analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stable control to control</td>
<td>262</td>
<td>224</td>
<td>200</td>
<td>Controls</td>
</tr>
<tr>
<td>Stable MCI to MCI</td>
<td>459</td>
<td>289</td>
<td>267</td>
<td>MCI</td>
</tr>
<tr>
<td>Stable AD to AD</td>
<td>47</td>
<td>174</td>
<td>46</td>
<td>AD</td>
</tr>
<tr>
<td>Conversion control to MCI</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>Exclude</td>
</tr>
<tr>
<td>Conversion MCI to AD</td>
<td>1</td>
<td>50</td>
<td>0</td>
<td>MCI</td>
</tr>
<tr>
<td>Conversion MCI to control</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>Exclude</td>
</tr>
<tr>
<td>Conversion AD to MCI</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>MCI</td>
</tr>
</tbody>
</table>

Diagnosis description - classification of clinical diagnosis made for each participant and each time-point; 2nd column - number of participants of baseline diagnosis and 3rd column - numbers of participants at the last point of diagnosis. 4th column shows number of participants who did not change their diagnosis at the last assessment from baseline diagnosis. Last column shows clinical classification of individuals based on the last available diagnosis for our analyses.
Table 2. ADNI phenotypes and PET amyloid status.

<table>
<thead>
<tr>
<th></th>
<th>Number of samples</th>
<th>N (% of MCI)</th>
<th>N (% of AD)</th>
<th>N (% of controls)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amyloid positive</td>
<td>357</td>
<td>162 (47%)</td>
<td>120 (69%)</td>
<td>65 (29%)</td>
</tr>
<tr>
<td>Amyloid negative</td>
<td>304</td>
<td>149 (43%)</td>
<td>18 (10%)</td>
<td>128 (57%)</td>
</tr>
<tr>
<td>NA</td>
<td>89</td>
<td>34 (10%)</td>
<td>36 (21%)</td>
<td>31 (14%)</td>
</tr>
<tr>
<td>All samples</td>
<td>770</td>
<td>344</td>
<td>174</td>
<td>224</td>
</tr>
</tbody>
</table>

Shows number of individuals with positive/negative amyloid for clinically diagnosed samples (MCI, AD, and controls).
### Table 3. PRS and APOE predictions of AD/MCI/Controls/ Amyloid phenotypes in ADNI.

<table>
<thead>
<tr>
<th>Model</th>
<th>Statistical Characteristics</th>
<th>AD vs Controls [174/224]</th>
<th>MCI vs Controls [344/224]</th>
<th>Amyloid positive vs Amyloid negative [357/304]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>APOE</strong></td>
<td>Beta(1,3,4) [se]</td>
<td>0.99 [0.13]</td>
<td>0.3 [0.1]</td>
<td>1.08 [0.01]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.58 [0.22]</td>
<td>-0.5 [0.17]</td>
<td>0.2 [0.17]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.03 [0.01]</td>
<td>-0.02 [0.01]</td>
<td>0.04 [0.01]</td>
</tr>
<tr>
<td></td>
<td><strong>p</strong></td>
<td>1.06e-18</td>
<td>9.6e-5</td>
<td>&lt;2.2e-16</td>
</tr>
<tr>
<td></td>
<td><strong>AUC/ AUC(^c)</strong></td>
<td>0.72/</td>
<td>0.58/</td>
<td>0.72/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.76</td>
<td>0.62</td>
<td>0.76</td>
</tr>
<tr>
<td><strong>PRS (p&lt;0.5) without APOE</strong></td>
<td>Beta(2,3,4) [se]</td>
<td>0.93 [0.12]</td>
<td>0.68 [0.1]</td>
<td>0.3 [0.08]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.7 [0.2]</td>
<td>-0.47 [0.18]</td>
<td>0.13 [0.16]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.016 [0.015]</td>
<td>-0.007 [0.01]</td>
<td>0.023 [0.01]</td>
</tr>
<tr>
<td></td>
<td><strong>p</strong></td>
<td>2.7e-18</td>
<td>6.2e-12</td>
<td>1.4e-3</td>
</tr>
<tr>
<td></td>
<td><strong>AUC/ AUC(^c)</strong></td>
<td>0.74/</td>
<td>0.66/</td>
<td>0.58/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.75</td>
<td>0.67</td>
<td>0.58</td>
</tr>
<tr>
<td><strong>FULL PRS model</strong></td>
<td>Beta(1-4) [se]</td>
<td>0.93 [0.13]</td>
<td>0.26 [0.1]</td>
<td>1.06 [0.1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.88 [0.13]</td>
<td>0.66 [0.1]</td>
<td>0.22 [0.09]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.63 [0.24]</td>
<td>-0.47 [0.18]</td>
<td>0.22 [0.17]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.04 [0.02]</td>
<td>-0.002 [0.01]</td>
<td>0.05 [0.01]</td>
</tr>
<tr>
<td></td>
<td><strong>p</strong></td>
<td>1.9e-30</td>
<td>1.1e-12</td>
<td>2.3e-29</td>
</tr>
<tr>
<td></td>
<td><strong>AUC/ AUC(^c)</strong></td>
<td>0.81/</td>
<td>0.67/</td>
<td>0.75/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.82</td>
<td>0.68</td>
<td>0.76</td>
</tr>
</tbody>
</table>

**Anova p (PRS above APOE)**

1.7e-13 1.8e-10 0.038

Beta\(^1\)=beta (e2+e4), Beta\(^2\)=beta(PRS), Beta\(^3\)=beta(sex), Beta\(^4\)=beta(age);
AUC\(^-\)area under the curve without taking gender and age into account;
AUC\(^c\) area under the curve where gender and age were used as predictors;

Legend: 1\(^{st}\) column –three scenarios where PRS predictions were made: APOE alone, PRS without APOE and full model (APOE plus PRS (p<0.5)); 2\(^{nd}\) column-statistical characteristics that were calculated for each model, these includes effect size (beta) with standard error (se), p-values and AUC (with and without gender and age) and p-value of significance PRS above APOE model; columns 3-5 represent three analyses with number of samples where different models were tested.
Table 4: Prediction of AD and amyloid deposition risk with pathway-specific PRSs

<table>
<thead>
<tr>
<th>Pathways</th>
<th>Number of Genes</th>
<th>AD (174) vs Controls (224)</th>
<th>Amyloid positive (357) vs negative (304)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beta</td>
<td>p-value</td>
<td>p-value (no APOE region)</td>
</tr>
<tr>
<td>protein-lipid complex assembly</td>
<td>20</td>
<td>0.87</td>
<td>3e-13</td>
</tr>
<tr>
<td>regulation of beta-amyloid formation</td>
<td>10</td>
<td>0.79</td>
<td>1.1e-11</td>
</tr>
<tr>
<td>protein-lipid complex</td>
<td>40</td>
<td>0.91</td>
<td>8.14e-14</td>
</tr>
<tr>
<td>regulation of amyloid precursor protein catabolic process</td>
<td>12</td>
<td>0.79</td>
<td>1.1e-11</td>
</tr>
<tr>
<td>tau protein binding</td>
<td>11</td>
<td>0.77</td>
<td>3.1e-11</td>
</tr>
<tr>
<td>reverse cholesterol transport</td>
<td>17</td>
<td>0.84</td>
<td>2.4e-12</td>
</tr>
<tr>
<td>protein-lipid complex subunit organization</td>
<td>35</td>
<td>0.92</td>
<td>8.e-14</td>
</tr>
<tr>
<td>plasma lipoprotein particle assembly</td>
<td>18</td>
<td>0.89</td>
<td>2e-13</td>
</tr>
<tr>
<td>activation of immune response</td>
<td>432</td>
<td>0.18</td>
<td>0.06</td>
</tr>
<tr>
<td>Whole genome without all pathways</td>
<td>-</td>
<td>0.93</td>
<td>2.2e-14</td>
</tr>
</tbody>
</table>

Legend: 1st column-name of pathways that were analysed, 2nd column-number of genes in each pathway, PRS pathways-specific effect sizes with p-values and p-values (no APOE region) of the models are presented in columns 3-8 for AD vs controls and amyloid deposition status.
Table 5: Prediction of amyloid deposition in individuals with MCI and of progression to AD in individuals with MCI and positive amyloid deposition with pathway-specific PRSs

<table>
<thead>
<tr>
<th>Pathways</th>
<th>Amyloid positive (270) vs amyloid negative (171)</th>
<th>MCI and amyloid positive (AD (112) vs MCI (150))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beta</td>
<td>P</td>
</tr>
<tr>
<td>protein-lipid complex assembly</td>
<td>1.11</td>
<td>1.92e-17</td>
</tr>
<tr>
<td>regulation of beta-amyloid formation</td>
<td>0.95</td>
<td>7.6e-14</td>
</tr>
<tr>
<td>protein-lipid complex</td>
<td>1.12</td>
<td>1.1e-17</td>
</tr>
<tr>
<td>regulation of amyloid precursor protein catabolic process</td>
<td>0.95</td>
<td>8.4e-14</td>
</tr>
<tr>
<td>tau protein binding</td>
<td>0.99</td>
<td>2.2e-14</td>
</tr>
<tr>
<td>reverse cholesterol transport</td>
<td>1.05</td>
<td>1.9e-15</td>
</tr>
<tr>
<td>protein-lipid complex subunit organization</td>
<td>1.1</td>
<td>1.2e-17</td>
</tr>
<tr>
<td>plasma lipoprotein particle assembly</td>
<td>1.09</td>
<td>3.4e-17</td>
</tr>
<tr>
<td>activation of immune response</td>
<td>0.18</td>
<td>0.068</td>
</tr>
<tr>
<td>Whole genome PRS without pathways</td>
<td>0.36</td>
<td>2.1e-3</td>
</tr>
</tbody>
</table>

Legend: 1st column-name of pathways that were analysed, PRS pathways-specific effect sizes with p-values and p-values (no APOE region) of the models are presented in columns 2-7.
Baseline clinical MCI

441 samples

Amyloid positive (PET)

270 samples

AD/MCI predictions
AUC

APOE 0.63
PRS full 0.69
p* = 0.0002

Amyloid prediction
AUC

APOE 0.79
PRS full 0.79
p* = 0.48

Amyloid negative (PET)

171 samples

AD

112 samples

MCI

150 samples

Clinical diagnosis at the last time-point