Trans-ethnic Genomic Informed Risk Assessment for Alzheimer’s disease: An International Hundred K+ Cohorts Consortium Study

Patrick M. Sleiman$^{1,2,3}$, Hui-Qi Qu$^1$, John J Connolly$^1$, Frank Mentch$^1$, Alexandre Pereira$^{4,5}$, Paulo A Lotufo$^{4,5}$, Stephen Tollman$^6$, Ananyo Choudhury$^7$, Michele Ramsay$^7$, Norihiro Kato$^8$, Kouichi Ozaki$^9$, Risa Mitsumori$^9$, Jae-Pil Jeon$^{10}$, Chang Hyung Hong$^{11}$, Sang Joon Son$^{11}$, Hyun Woong Roh$^{11}$, Dong-gi Lee$^{11}$, Naaheed Mukadam$^{12}$, Isabelle F Foote$^{12,13}$, Charles R Marshall$^{12,13}$, Adam Butterworth$^{14,15}$, Bram P Prins$^{14,15}$, Joseph T Glessner$^{1,2,3}$, Hakon Hakonarson$^{1,2,3,16,17}$†, on behalf of the Davos Alzheimer Collaborative and IHCC consortium

Affiliations:


2. Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, USA.

3. Division of Human Genetics, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA.

4. Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil.

5. Centro de Pesquisas Clínicas e Epidemiológicas, Hospital Universitário, Universidade de São Paulo, São Paulo, Brazil.

6. MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.

7. Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.

8. National Center for Global Health and Medicine, Tokyo, 1628655, Japan.

9. Medical Genome Center, Research Institute, National Center for Geriatrics and Gerontology (NCGG), Obu City, Aichi Prefecture, Japan.


11. Department of Psychiatry, Ajou University School of Medicine, Suwon, Korea.

12. Preventive Neurology Unit, Wolfson Institute of Population Health, Queen Mary University of London, UK.

13. Genes & Health, Blizard Institute, Queen Mary University of London, UK.

14. MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
15. NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

16. Division of Pulmonary Medicine, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA.

17. Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland.

† Corresponding author and reprint requests should be addressed to:
Dr. Hakon Hakonarson
Center for Applied Genomics
3615 Civic Center Blvd
Abramson Building
Philadelphia, PA 19104, United States of America
Telephone: 267-426-0088
Fax: 267-426-0363
Email: hakonarson@chop.edu

Word count: xxx

Number of tables and figures: x Tables and x figures.
Abstract

**Background** Alzheimer’s disease (AD) is a complex multifactorial progressive dementia affecting all human populations. As a collaboration model between the International Hundred K+ Cohorts Consortium (IHCC) and the Davos Alzheimer Collaborative (DAC), our aim was to develop a trans-ethnic genomic informed risk assessment (GIRA) algorithm for AD.

**Methods** The GIRA model was created to include a polygenic risk score (PRS) calculated from the AD GWAS loci, the APOE haplotypes, and non-genetic covariates including age, sex and first 3 principal components of population substructure. The model was first validated using an ancestrally diverse dataset from the eMERGE network, and subsequently validated in a South-Asian population in the UK and 3 East-Asian populations. The distributions of the PRS scores were also explored in populations from 3 African regions. In two validation sites, the PRS was tested for associated with the levels of plasma proteomics markers.

**Results** We created a trans-ethnic GIRA model for the risk prediction of AD and validated the performance of the GIRA model in different populations. The proteomic study in the participant sites identified proteins related to female infertility and autoimmune thyroiditis and associated with the risk scores of AD, highlighting molecular mechanisms underlying the previously observed correlations between these clinical phenotypes.

**Conclusions** As the initial effort by the IHCC to leverage existing large scale datasets in a collaborative setting with DAC, we developed a trans-ethnic GIRA for AD with the potential of identifying individuals at high risk of developing AD for future clinical applications. The PRS scores in this model also contribute new research discoveries for the molecular pathogenesis of AD as demonstrated by the proteomic data.

**Key words:** Alzheimer’s disease; data sharing; female infertility; genomic informed risk assessment; polygenic risk score; minority population; thyroid; trans-ethnic
Introduction

Alzheimer’s disease (AD) is characterized by irreversible neuronal degeneration with no cure to date\(^1\). AD affects all human populations with the prevalence of 8.4% to 13.8% in people aged 65 years and older according to the Centers for Disease Control and Prevention (CDC, [https://www.cdc.gov/media/releases/2018/p0920-alzheimers-burden-double-2060.html](https://www.cdc.gov/media/releases/2018/p0920-alzheimers-burden-double-2060.html)). Early diagnosis of AD may enable early intervention to minimize the damage to the central nervous system\(^2\). With older age and female sex as the known risk factors\(^3\), AD has been recognized as a complex multifactorial genetic disease\(^4\). The apolipoprotein E gene (APOE) variants have been established as a major susceptibility locus for AD\(^5\). APOE is a major apoprotein of the lipoproteins chylomicron and very low density lipoproteins (VLDL), with the function of lipid transportation\(^6\). Amyloid β (Aβ) forms the extracellular plaque in AD brains, and is the central mechanism of the AD pathogenesis\(^7\). The APOE is associated with Aβ aggregation and clearance by physical interaction and sharing common receptors\(^8\).

Human APOE has three common genetic isoforms APOE2 (112Cys-158Cys), APOE3 (112Cys-158Arg), and APOE4 (112Arg-158Arg), with different amino acid residues at positions 112 and 158 in the N-terminal domain\(^8\)\(^9\). The APOE genetic isoforms are determined by two single nucleotide polymorphisms: (1) rs429358 (hg38, chr19:44908684:T:C) causes amino acid substitution of Cys to Arg at residue 112; (2) rs7412 (hg38, chr19:44908822:C:T) causes amino acid substitution of Arg to Cys at residue 158\(^10\). Among the 3 APOE isoforms encoded by 3 haplotypes of rs429358 and rs7412 SNPs, the APOE3 haplotype has the most common frequency in all populations studied to date, including African populations\(^11\)\(^12\). Compared to the APOE3 haplotype, APOE2 is protective against AD, and APOE4 is predisposing to AD with an odds ratio (OR) value about 2 to 4 for heterozygous carriers\(^11\)\(^13\)\(^14\). Despite the large genetic effect, however, the APOE locus alone is insufficient to explain the genetic susceptibility to AD. A large number of genetic loci associated with AD, e.g. WASH complex subunit 3 (WASHC3), bridging integrator 1 (BIN1), complement C3b/C4b receptor 1 (Knops blood group) (CR1), and cathepsin B (CTSB), have recently been identified in the genome-wide association studies (GWAS) by Jun et al.\(^15\) and Bellenguez et al.\(^16\) with large sample sizes including participants with different ancestries. Additional genetic information from these AD loci may improve the risk prediction of AD and the identification of patients with high risk of AD.

In this study, we developed a genomic informed risk assessment (GIRA) algorithm for AD with both the APOE isoforms and the polygenic risk score (PRS) calculated by the multiple AD loci identified in previous GWAS\(^15\)\(^16\), as well as known non-genetic risk factors of AD including age and sex. The PRS aggregated the effects of GWAS loci other than the APOE locus into a single score, an approach that has been shown as an effective approach to identify patients with high risk of many complex diseases\(^17\)\(^19\). In particular, a PRS with selected loci has been demonstrated for precise prediction of type 1 diabetes (T1D)\(^20\). Compared to AD with APOE as a major genetic locus plus a number of minor loci, T1D has also the human leukocyte antigen (HLA) region as a major genetic locus plus a number of minor loci\(^20\). Enabled by risk prediction using GIRA, and with more attention given to the high risk patients, this will consequently enable early diagnosis and early intervention for the disease. Nowadays, the major challenge in GIRA models is potential health disparities due to the lack of genomic information in minority or understudied populations\(^21\). In many cases, effective GIRA models relying on data from large-sample genome-wide association studies (GWAS) are only available for the European population as the majority...
of GWASs have been done in people with European ancestry\textsuperscript{22}. With the aim to eliminate
genomics-related health disparities, we leveraged an international effort supported by the
International HundredK+ Cohorts Consortium (IHCC)\textsuperscript{23}, which has brought together large scale
cohorts of diverse populations from around the world, who have genome-wide genotype data.

\textbf{Methods}

\textbf{Genotyping.} The genotyping data was imputed with the TOPMed (Version R2 on GRC38)
Reference Panel at the TOPMed imputation server (https://imputation.biodatacatalyst.nhlbi.nih.gov), or the Haplotype Reference Consortium (HRC)
panel at the Michigan imputation server (https://imputationserver.sph.umich.edu), or the Sanger
Imputation Server (African reference panel), at each participant site. No quality issue was
reported for the imputation of any of the SNP markers in the GIRA model from any of
participating sites. In case of quality issues for the imputation, to find alternative SNP markers
based on the specific population and the genotyping arrays would be done in house upon request.

\textbf{APOE haplotyping.} The APOE haplotypes were inferred from the two SNPs rs429358 and
rs7412, i.e. APOE2=rs429358/T-rs7412/T; APOE3=rs429358/T-rs7412/C; APOE4=rs429358/C-
rs7412/C.

\textbf{GWAS loci.} The GWAS loci were identified by the Stage I studies from Jun et al.\textsuperscript{15} and
Bellenguez et al.\textsuperscript{16} The study by Jun et al. included cases/controls with European ancestry (EA,
n=13,100/13,220), African Americans (AA, n=1,472/3,511), Japanese (JPN, n=951/894), and
Israeli-Arab (IA, n=51/64).\textsuperscript{15} The study by Bellenguez et al. included 2,447 diagnosed cases,
46,828 proxy cases of dementia and 338,440 controls, all with European ancestry\textsuperscript{16}. Multi-allelic
variants, indels and rare SNPs with MAF < 3% were excluded from further analysis. The
remaining variants from the combined summary stats were LD pruned using an R\textsuperscript{2} threshold of
0.3 resulting in a final list of 74 variants (Supplementary Table 1).

\textbf{GIRA model.} The GIRA model includes 3 components, i.e. APOE haplotype, PRS by genomic
markers, and covariates including age, sex and the first 3 principal components of population
substructure for genetic ancestry correction. The PRS is calculated by the weighted sum of the
effect alleles, while weights are the reported odds ratios.

\textbf{Validation of PRS models.} The PRS markers were first validated for their allele frequencies in
4145 ancestrally diverse dementia patients from the electronic medical records and genomics
(eMERGE) consortium Phase I-III dataset. Subsequently, the scoring algorithm, SNPs and
weights, as well as detailed instructions were disseminated to participant sites (Table 1). Scores
were generated on a total of 25,786 participants across a variety of endpoints and returned to the
Children’s Hospital of Philadelphia (CHOP) for collation. Each site was requested to follow
standardized reporting metrics. As all groups did not have accurate age at onset data, we
requested odds ratios rather than hazard ratios for the phenotype outcomes (Table 1). Each
validation site was required to return:

1) The odds ratio per standard deviation of the PRS distribution with 95\% confidence interval (CI);

2) Estimate of model discrimination [Area under the receiver operating characteristic (ROC)]
Curve, AUC] with 95% CI of A) the non-APOE PRS alone; B) the PRS and APOE status; C) The non-genetic predictors alone; D) the full model.

3) Tail discrimination: We set the cutoff for the high risk group at the 97.5% of the PRS. The ORs and 95% CI (and the P-value for the OR) were calculated by comparing the high risk group vs. everybody else. i.e the participants in the top 2.5% of the PRS vs the bottom 97.5%.

4) The sensitivity / specificity as well as negative (NPV) and positive (PPV) predictive values at the proposed cutoff (split by ancestry if appropriate for each cohort).

The NPV/PPV used prevalence adjusted metrics, i.e. PPV = (Sn * Pr) / [ (Sn * Pr) + ((1 – Sp) * (1 – Pr)) ] and NPV = (Sp * (1 – Pr)) / [ (Sp * (1 – Pr)) + (( 1 – Sn) * Pr ) ] where Sn = sensitivity, Sp = specificity, and Pr = population based prevalence reflective of the study population.

**Proteomics study at two participant sites.** The Brazilian Longitudinal Study of Adult Health (ELSA-Brasil) enrolled 15,105 civil servants aged 35 to 74 years living in six cities, addressing the incidence of non-communicable diseases. From the 15,105 participants, 9,333 DNA samples were analyzed for genetic ancestry using a software tool for maximum likelihood estimation of individual ancestries from multi-locus SNP genotype datasets. The INTERVAL BioResource recruited 45,263 whole blood donors (22,466 men and 22,797 women) between June 11, 2012, and June 15, 2014. Donors were aged 18 years or older from 25 NHS Blood and Transplant (NSHBT) blood donation centers distributed across England, UK. In addition to the GIRA validation at these two sites, the PRS scores were tested for association with 3,282 plasma protein targets from the SomaLogic proteomics platform (SomaLogic Inc., Boulder, CO, USA).

**Exploration of the PRS model in African populations.** Without an African cohort of AD patients, we explored the distribution of the PRS scores in African populations in the Africa Wits-INDEPTH Partnership for genomic studies (AWI-Gen) leveraging its access via the IHCC resources. African populations from three geographic regions (south - South Africa; east – Kenya; and west - Ghana/Burkina Faso) were assessed.

**Results**

The genetic markers for the PRS model were polymorphic and informative in 4,145 ancestrally diverse dementia patients from the eMERGE consortium (Supplementary Table 2). Four cohorts from three sites, Korea Biobank, NCGG and East London Genes & Health, reported results of the PRS in dementia cases vs. controls (Table 2). Except the UKB white cohort, the best performance of prediction of dementia was by the full GIRA model with combination of the PRS component, the APOE haplotypes, and the non-genetic covariates. In the risk prediction, either genomic markers alone or non-genetic factors alone is insufficient, while the combined usage of genetic risk factors and non-genetic factors increased the performance of the prediction model. Most importantly, all four cohorts in Table 2 were non-European. A random effects restricted maximum likelihood (REML) meta-analysis of the four AUCs and their variance was 0.674 (0.643-0.706), indicating the potential of a prediction model in the non-European populations.

Further, to assess the potential application of the PRS in African populations, the PRS model was examined in the continental African AWI-Gen cohort including participants from across the continent, to determine if the SNP
markers forming the PRS are polymorphic and informative in African populations from different geographic regions. As shown, the score is normally distributed with similar variance across African populations (Supplementary Figure 1), indicating it is suitable for use in populations of African ancestry.

In addition to the validation of the prediction model, two sites, ELSA Brazil and INTERVAL UK reported on results of the score as well as results from plasma proteomic analysis. The proteomics analysis showed correlation with APOE haplotypes, including also with blood APOE levels across two separate peptides (Table 3). Correlation of 99 proteins with the full PRS model were also observed with $P<0.05$ (Supplementary Table 3). Using the WebGestalt (WEB-based Gene ScT AnaLysis Toolkit) web tool$^{30}$, over-representation analysis (ORA) of the correlated genes by the DisGeNET approach$^{31}$, highlighted the genes (CEBPB, FSHB, LEP, LHB, LIF) in the gene set C0021361:Female infertility with false discovery rate (FDR)=0.0066779 (Table 4). ORA by the GLAD4U approach$^{32}$ highlighted the gene sets PA445859:Thyroiditis, Autoimmune (FDR=0.017306, including CD3G, CGA, CXCL9, FAS, ILIRB1); PA444172:Fetal Growth Retardation (FDR=0.044508, including IGF1, IGFBP3, LEP, SELP, TMEM70); PA443588:Cachexia (FDR=0.044508, including IGF1, LEP, LIF, SELP). Taken together, these results demonstrate enrichment of several genes that influence AD risk, as well as risks for multiple other conditions, including but not limited to female infertility, thyroiditis and autoimmunity. These results warrant further evaluation for risk optimization.

**Discussion**

As one of the first IHCC multinational studies, we demonstrate the feasibility of developing a trans-ethnic AD GIRA model that is predictive of disease predisposition across diverse populations, globally. In this regard, the IHCC project served as an important resource to examine GIRA and PRS in under-represented populations in genomic studies. The GIRA model performed better than the PRS model in East Asian populations from Japan and Korea, and in South Asian populations of Pakistani/Bangladeshi origin recruited through UK. While lacking a well phenotyped AD cohort of African origin with genetic data, we demonstrate a normal distribution of the PRS scores in different regions in Africa, suggesting that the current PRS score system is potentially informative in African populations. Further assessment of the PRS model in African AD patients is warranted.

In addition to its clinical potential, this study uncovered associations between the AD PRS scores and other disease-related gene sets, leading to novel insights into the pathogenesis of AD. In this regard, we identified association of the AD PRS and genes related to female infertility. Clinically, it has been observed that parity is inversely associated with risk of AD$^{33}$. The genes related to female infertility identified in this study help explain the increased risk of AD in women, as well as the molecular mechanisms of the pathogenesis of AD in women and how parity decreases the risk of AD. Among the 5 genes related to female infertility identified (Table 4), the lowest $P$ value was seen between PRS and lower level of follicle stimulating hormone subunit beta (encoded by FSHB). The physiological function of follicle-stimulating hormone is to induce egg and sperm production$^{34}$. Genetic variants of FSHB have been shown to cause hypogonadotropic hypogonadism in women and men$^{35,36}$. With The FSHB gene as a potential mediator of the pathogenesis of AD, it may also help to explain the clinical correlation between hypogonadism and AD$^{37}$. More importantly, FSHB may represent a new opportunity to develop hormone
therapy for AD\textsuperscript{38 39}. In addition, previous studies have demonstrated correlations between abnormal levels of thyroid-stimulating hormone (TSH) and the risk of AD\textsuperscript{40}. While further studies are warranted, our study identified 5 genes that may mediate this correlation.

Given that this is the first effort by the IHCC to leverage large existing datasets that reside within the IHCC consortium for a trans-ethnic GIRA on AD, we envision an opportunity to scale this to other cohorts within the consortium and expand the number of traits that can be analyzed. As such, the IHCC presents a rich resource of data for collaborative research with trans-ethnic focus, where there is much unmet need at present as this is an area of research that has been largely neglected. We also need to emphasize that this is a proof-of-principle study for an international effort to develop a trans-ethnic GIRA model for precision medicine, whereas to improve the GIRA model as suggested by the currently limited AUC scores is still warranted through further international efforts. As our research efforts continue, we envision efficient data sharing across academia and industry, where we will focus on improving patient health care services for AD and other diseases as well as biomedical research discoveries, leveraging our established GIRA approach.

**Supplementary materials**

**Supplementary Table 1** SNPs and weights for the PRS model

**Supplementary Table 2** Frequencies of the genetic markers for the GIRA model in 4,145 ancestrally diverse dementia patients from the eMERGE consortium

**Supplementary Table 3** Correlation of 99 proteins with $P<0.05$ with the full GIRA model

**Supplementary Figure 1** Distribution of the PRS in the continental African AWI-GEN cohort from different geographic regions

**Acknowledgements** We thank all the participants who contributed to and enabled this study.

**Ethical Approval**

This study was exempted by the Institutional Review Board (IRB) of the Children’s Hospital of Philadelphia, USA. Human participant personal information was not shared with the research group and participants were de-identification. All human participants or their proxies provided written informed consent for their respective studies and all the studies were approved by their local and/or national ethics review boards.

**Funding**

The study was supported by:

All authors were supported by the International HundredK+ Cohorts Consortium (IHCC), which was created in collaboration with the Global Alliance for Genomics and Health (GA4GH) and the Global Genomics Medicine Collaborative (G2MC) with support from the National Institutes of Health and the Wellcome Trust.
The Davos Alzheimer Collaborative (DAC) provided funding to enable completion of the study.

The study was funded in part by an Institutional Development Fund from the Children’s Hospital of Philadelphia to the Center for Applied Genomics, and The Children’s Hospital of Philadelphia Endowed Chair in Genomic Research.

ELSA-Brasil is funded by the Brazilian Ministry of Health (Department of Science and Technology) and the Ministry of Science, Technology and Innovation (FINEP, Financiadora de Estudos e Projetos), and CNPq (the National Council for Scientific and Technological Development).

Genes & Health has recently been core-funded by Wellcome (WT102627, WT210561), the Medical Research Council (UK) (M009017), Higher Education Funding Council for England Catalyst, Barts Charity (845/1796), Health Data Research UK (for London substantive site), and research delivery support from the NHS National Institute for Health Research Clinical Research Network (North Thames). Genes & Health is/has recently been funded by Alnylam Pharmaceuticals, Genomics PLC; and a Life Sciences Industry Consortium of Bristol-Myers Squibb Company, GlaxoSmithKline Research and Development Limited, Maze Therapeutics Inc, Merck Sharp & Dohme LLC, Novo Nordisk A/S, Pfizer Inc, Takeda Development Centre Americas Inc.

Korea Biobank and Biobank Innovations for Chronic cerebrovascular disease With ALZheimer's disease Study (BICWALZS) are funded by the Korea Disease Control and Prevention Agency for the Korea Biobank Project (#6673-303).

Participants in the INTERVAL randomized controlled trial were recruited with the active collaboration of NHS Blood and Transplant (www.nhsbt.nhs.uk), which has supported fieldwork and other elements of the trial. DNA extraction and genotyping were co-funded by the National Institute for Health Research (NIHR), the NIHR BioResource (http://bioresource.nihr.ac.uk) and the NIHR Cambridge Biomedical Research Centre (BRC) (no. BRC-1215-20014). The academic coordinating centre for INTERVAL was supported by core funding from the NIHR Blood and Transplant Research Unit in Donor Health and Genomics (no. NIHR BTRU-2014-10024), UK Medical Research Council (MRC) (no. MR/L003120/1), British Heart Foundation (nos SP/09/002, RG/13/13/30194 and RG/18/13/33946) and the NIHR Cambridge BRC (no. BRC-1215-20014). A complete list of the investigators and contributors to the INTERVAL trial is provided in ref.17. The academic coordinating centre thanks blood donor centre staff and blood donors for participating in the INTERVAL trial. This work was supported by Health Data Research UK, which is funded by the UK MRC, Engineering and Physical Sciences Research Council (EPSRC), Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome. The views expressed in this manuscript are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

The AWI-Gen Study was funded by the National Human Genome Research Institute (NHGRI), Office of the Director (OD), Eunice Kennedy Shriver National Institute Of Child Health &
Human Development (NICHD), the National Institute of Environmental Health Sciences (NIEHS), the Office of AIDS research (OAR) and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), of the National Institutes of Health (NIH) under award number U54HG006938.

Declaration of Conflicting Interests

All authors declare no conflicts of interest with respect to the research, authorship, and/or publication of this article.

Consent for publication

All authors have provided consent for publication of the manuscript.

Availability of supporting data

Supporting data from this study can be obtained by emailing the corresponding author Dr. Hakon Hakonarson.
<table>
<thead>
<tr>
<th>Site</th>
<th>Genetic ancestry</th>
<th>Phenotypic outcome</th>
<th>N (case/control)</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Center for Geriatrics and Gerontology (NCGG) Biobank</td>
<td>Japanese (East Asian)</td>
<td>AD, MCI</td>
<td>Case:1000 Normal Cognitive:1000</td>
</tr>
<tr>
<td>East London Genes and Health cohort</td>
<td>British-Pakistani/British-Bangladeshi (South Asian)</td>
<td>All-cause dementia (from secondary/primary care records); MCI/cognitive decline cases excluded</td>
<td>104 cases; 614 controls</td>
</tr>
<tr>
<td>Korean Biobank Project</td>
<td>Korean (East Asian)</td>
<td>Phenotype 1: Cortical amyloid positivity (by Flutemetamol PET imaging) (Control: Cortical amyloid negativity)</td>
<td>191:337 (total 528)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Phenotype 2: CDR global Score 1 or over (Control: CDR global 0.5 or less)</td>
<td>157:539 (total 696)</td>
</tr>
<tr>
<td>Africa Wits-INDEPTH partnership for genomic studies (AWI-Gen)</td>
<td>African (Different ethnolinguistic and geographic groups)</td>
<td>Not assessed for AD</td>
<td>Population cross sectional cohort: n=10,700</td>
</tr>
<tr>
<td>The Brazilian Longitudinal Study of Adult Health (ELSA-Brasil)</td>
<td>Brazilian (Admixed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTERVAL (UK Blood Donors)</td>
<td>European (White British)</td>
<td>Stroop Test (attention and reaction times), Trail Making Test (executive function), Pairs Test (Episodic Memory), Reasoning Tests (intelligence), &gt;3K proteins on the SomaLogic proteomics platform</td>
<td>~9k Cognitive measure; 1140 proteomics</td>
</tr>
<tr>
<td>UKBiobank</td>
<td>European (White British)</td>
<td>ICD9/10 diagnoses of AD/dementia derived from hospital inpatient records or patients with date of Alzheimer’s disease/dementia report</td>
<td>2923 cases; 226342 controls</td>
</tr>
</tbody>
</table>

Abbreviations: AD, Alzheimer’s disease; CDR, Clinical Dementia Rating; ICD, International Classification of Diseases; MCI, mild cognitive impairment; NA, not applicable.
**Table 2. Validation of PRS models (95% CIs)**

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Odds ratio per SD</th>
<th>AUC with PRS only</th>
<th>AUC with genetic predictors (PRS and APOE haplotypes)</th>
<th>AUC with non-genetic covariates only</th>
<th>AUC with the full GIRA model*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korea pheno1</td>
<td>1.186 (0.992-1.418)</td>
<td>0.548 (0.497-0.600)</td>
<td>0.677 (0.628-0.726)</td>
<td>0.627 (0.576-0.677)</td>
<td>0.751 (0.705-0.796)</td>
</tr>
<tr>
<td>Korea pheno2</td>
<td>1.040 (0.871-1.243)</td>
<td>0.507 (0.456-0.559)</td>
<td>0.612 (0.560-0.664)</td>
<td>0.545 (0.493-0.597)</td>
<td>0.637 (0.586-0.689)</td>
</tr>
<tr>
<td>East London</td>
<td>1.110 (0.940-1.330)</td>
<td>0.530 (0.470-0.590)</td>
<td>0.540 (0.480-0.600)</td>
<td>0.680 (0.610-0.750)</td>
<td>0.690 (0.620-0.760)</td>
</tr>
<tr>
<td>Japan</td>
<td>1.120</td>
<td>0.545 (0.520-0.570)</td>
<td>0.607 (0.582-0.632)</td>
<td>0.616 (0.591-0.641)</td>
<td>0.625 (0.601-0.650)</td>
</tr>
<tr>
<td>UKB white</td>
<td>1.003 (0.990-1.007)</td>
<td>0.527 (0.516-0.537)</td>
<td>0.542 (0.532-0.553)</td>
<td>0.780 (0.773-0.787)</td>
<td>0.746 (0.738-0.754)</td>
</tr>
</tbody>
</table>

*The full GIRA model: the model with PRS, APOE and non-genetic covariates.
### Table 3. Correlation of *APOE* haplotypes with APOE protein levels

<table>
<thead>
<tr>
<th>Proteomics ID</th>
<th>R2</th>
<th>P</th>
<th>BETA</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>APOE.2937.10.2 (E3)</td>
<td>6.55E-03</td>
<td>0.0057</td>
<td>16.9</td>
<td>6.11</td>
</tr>
<tr>
<td>APOE.5312.49.3 (E2)</td>
<td>0.00475</td>
<td>0.0194</td>
<td>15.03</td>
<td>6.42</td>
</tr>
</tbody>
</table>

### Table 4. Genes in the gene-set C0021361: Female infertility associated with AD PRS

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>PRS_R2</th>
<th>P</th>
<th>BETA</th>
<th>SE</th>
<th>Gene Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEBPB</td>
<td>0.004</td>
<td>0.032</td>
<td>15.645</td>
<td>7.274</td>
<td>CCAAT enhancer binding protein beta</td>
</tr>
<tr>
<td>FSHB</td>
<td>0.003</td>
<td>0.018</td>
<td>-13.862</td>
<td>5.871</td>
<td>follicle stimulating hormone subunit beta</td>
</tr>
<tr>
<td>LEP</td>
<td>0.004</td>
<td>0.035</td>
<td>-15.106</td>
<td>7.147</td>
<td>leptin</td>
</tr>
<tr>
<td>LHB</td>
<td>0.002</td>
<td>0.036</td>
<td>-12.101</td>
<td>5.765</td>
<td>luteinizing hormone beta polypeptide</td>
</tr>
<tr>
<td>LIF</td>
<td>0.003</td>
<td>0.049</td>
<td>-13.995</td>
<td>7.109</td>
<td>LIF, interleukin 6 family cytokine</td>
</tr>
</tbody>
</table>
Reference:


34. Orlowski M, Sarao MS. Physiology, follicle stimulating hormone. *Follicle Stimulating Horm* 2018;6


