

https://dspace.org/

# Transethnic genomic informed risk assessment for Alzheimers disease An International Hundred K Cohorts Consortium study

**P M Sleiman; H Q Qu; J J Connolly; F Mentch; Stephen Tollman ...**

https://hdl.handle.net/10539/39316
WOS

Downloaded from a DSpace repository. https://dspace.org/

Check for updates

Received: 12 January 2023 | Revised: 26 April 2023 | Accepted: 5 May 2023

DOI: 10.1002/alz.13378

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

# 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[11,12] | Naaheed Mukadam[13] | Isabelle F. Foote[13,14] | Charles R. Marshall[13,14] | Adam Butterworth[15,16] | Bram P. Prins[15,16] | Joseph T. Glessner[1,2,3] | Hakon Hakonarson[1,2,3,17,18] | on behalf of the Davos Alzheimer Collaborative and IHCC consortium**

[1]The Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

[2]Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[3]Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 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, Japan

[9]Medical Genome Center, Research Institute, National Center for Geriatrics and Gerontology (NCGG), Obu, Japan

[10]Korea Biobank Project, Korea National Institute of Health, Osong, South Korea

[11]Department of Psychiatry, Ajou University School of Medicine, Suwon, South Korea

[12]Department of Biostatistics, Epidemiology and Informatics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[13]Preventive Neurology Unit, Wolfson Institute of Population Health, Queen Mary University of London, London, UK

[14]Genes & Health, Blizard Institute, Queen Mary University of London, London, UK

[15]MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

[16]NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

[17]Division of Pulmonary Medicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

[18]Faculty of Medicine, University of Iceland, Reykjavik, Iceland

**Correspondence**
Hakon Hakonarson, Center for Applied Genomics, 3615 Civic Center Blvd, Abramson Building, Philadelphia, PA 19104, USA.
E-mail: hakonarson@chop.edu

## Abstract

**BACKGROUND:** As a collaboration model between the International HundredK+ Cohorts Consortium (IHCC) and the Davos Alzheimer's Collaborative (DAC), our aim

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2023 The Authors. *Alzheimer's & Dementia* published by Wiley Periodicals LLC on behalf of Alzheimer's Association.

was to develop a trans-ethnic genomic informed risk assessment (GIRA) algorithm for Alzheimer's disease (AD).

**METHODS:** The GIRA model was created to include polygenic risk score calculated from the AD genome-wide association study loci, the apolipoprotein E haplotypes, and non-genetic covariates including age, sex, and the first three principal components of population substructure.

**RESULTS:** We 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.

**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.

**KEYWORDS**
Alzheimer's disease, data sharing, female infertility, genomic informed risk assessment, minority population, polygenic risk score, thyroid, trans-ethnic

## 1 | BACKGROUND

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 according to the Centers for Disease Control and Prevention (CDC; 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 (*APOE*) gene 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] While the *APOE* action remains unclear in AD, the observed associations with the two pathological hallmarks of the disease, including extracellular amyloid beta (A$\beta$) plaques and intraneuronal tau pathology, as well as contributions to neural network deficits and neuroinflammation, suggest an important role for *APOE* in AD.[7–9]

Human *APOE* has three common genetic isoforms, *APOE* $\varepsilon$2 (112Cys-158Cys), *APOE* $\varepsilon$3 (112Cys-158Arg), and *APOE* $\varepsilon$4 (112Arg-158Arg), with different amino acid residues at positions 112 and 158 in the N-terminal domain.[9,10] The *APOE* genetic isoforms are determined by two single nucleotide polymorphisms (SNPs): (1) rs429358 (hg38, chr19:44908684:T:C) causes amino acid substitution of Cys to Arg at residue 112 and (2) rs7412 (hg38, chr19:44908822:C:T) causes amino acid substitution of Arg to Cys at residue 158.[11] Among the three *APOE* isoforms encoded by three haplotypes of rs429358 and rs7412 SNPs, the *APOE* $\varepsilon$3 haplotype has the most common frequency in all populations studied to date, including African populations.[12,13] Compared to the *APOE* $\varepsilon$3 haplotype, *APOE* $\varepsilon$2 is protective against AD, and *APOE* $\varepsilon$4 is predisposing to AD with an odds ratio (OR) value about two to four for heterozygous carriers.[12,14,15] 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, for example, 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.[16] and Bellenguez et al.[17] 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,[16,17] 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.[18–20] In particular, a PRS with selected loci has been demonstrated for precise prediction of type 1 diabetes (T1D).[21] 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.[21] Nowadays, the major challenge in GIRA models is potential health disparities due to the lack of genomic information in minority or understudied populations.[22] In many cases, effective GIRA models relying on data from large-sample GWAS are only available for the European population as the majority of GWASs have been done in people with European ancestry.[23] With the aim to

15525279, 2023, 12, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.13378 by South African Medical Research, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**RESEARCH IN CONTEXT**

1. **Systematic Review**: By literature review, the authors identified the major genetic risk factors of Alzheimer's disease (AD) to develop a genomic informed risk assessment (GIRA) algorithm for AD, including both the apolipoprotein E isoforms and the polygenic risk score calculated by the multiple AD loci.
2. **Interpretation**: We validated the performance of the GIRA model in different populations; the proteomic study in the participant sites gained insights into the molecular mechanisms of AD.
3. **Future Directions**: The International HundredK+ Cohorts Consortium (IHCC) has brought together large-scale cohorts of diverse populations from around the world, and presents a rich resource of data for collaborative research with trans-ethnic focus.

eliminate genomics-related health disparities, we leveraged an international effort supported by the International HundredK+ Cohorts Consortium (IHCC),[24] which has brought together large-scale cohorts of diverse populations from around the world, who have genome-wide genotype data.

## 2 | METHODS

This study was exempted by the institutional review board of the Children's Hospital of Philadelphia, USA. Human participant personal information was not shared with the research group and participants were de-identified. 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.

### 2.1 | Genotyping

The genotyping data was imputed with the TOPMed (version R2 on GRC38) Reference Panel at the Trans-Omics for Precision Medicine 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 Service (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.

### 2.2 | *APOE* haplotyping

The *APOE* haplotypes were inferred from the two SNPs rs429358 and rs7412, that is, *APOE* ε2 = rs429358/T-rs7412/T; *APOE* ε3 = rs429358/T-rs7412/C; *APOE* ε4 = rs429358/C-rs7412/C.

### 2.3 | GWAS loci

The GWAS loci were identified by the Stage I studies from Jun et al.[16] and Bellenguez et al.[17] 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).[16] The study by Bellenguez et al. included 2447 diagnosed cases, 46,828 proxy cases of dementia and 338,440 controls, all with European ancestry.[17] Multi-allelic variants, indels, and rare SNPs with minor allele frequency <3% were excluded from further analysis. The remaining variants from the combined summary stats were linkage disequilibrium pruned using an $R^2$ threshold of 0.3 resulting in a final list of 74 variants (Table S1 in supporting information).

### 2.4 | GIRA model

The GIRA model includes three components, that is, *APOE* haplotype; PRS by genomic markers; and covariates including age, sex, and the first three 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 ORs.

### 2.5 | 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 to 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 ORs rather than hazard ratios for the phenotype outcomes (Table 1). The cohort validations were performed at four participant sites, including Korea pheno1/Korea pheno2, East London, Japan, and UK Biobank (UKB) white. Without gold standard for AD diagnosis,[25] we acknowledge the phenotypic heterogeneity in this multinational study, and we also highlight the potential clinical importance of a transethnic PRS to assist with the AD diagnosis. Korea pheno1 and Korea pheno2 were from the Biobank Innovations for Chronic Cerebrovascular Disease with Alzheimer's Disease Study (BICWALZS) as a Korea Biobank Project.[26] Korea pheno1 defined

15525279, 2023, 12, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.13378 by South African Medical Research, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 1** Sites participating in the validation.

| Site | Genetic ancestry | Phenotypic outcome | *N* (case/control) |
|---|---|---|---|
| **Validation of PRS models** | | | |
| National Center for Geriatrics and Gerontology (NCGG) Biobank | Japanese (East Asian) | AD, MCI | Case:1000 Normal Cognitive:1000 |
| East London Genes and Health cohort | British-Pakistani/British-Bangladeshi (South Asian) | All-cause dementia (from secondary/primary care records); MCI/cognitive decline cases excluded | 104 cases; 614 controls |
| Korean Biobank Project | Korean (East Asian) | Phenotype 1: cortical amyloid positivity (by flutemetamol PET imaging) (control: cortical amyloid negativity) | 191:337 (total 528) |
| | | Phenotype 2: CDR global score ≥1 (control: CDR global ≤0.5) | 157:539 (total 696) |
| UKBiobank | European (White British) | ICD-9/10 diagnoses of AD/dementia derived from hospital in-patient records or patients with date of AD/dementia report | 2923 cases; 226,342 controls |
| **Proteomics study** | | | |
| The Brazilian Longitudinal Study of Adult Health (ELSA-Brasil) | Brazilian (admixed) | | |
| INTERVAL (UK Blood Donors) | European (White British) | Stroop Test (attention and reaction times), Trail Making Test (executive function), Pairs Test (episodic memory), Reasoning Tests (intelligence), > 3K proteins on the SomaLogic proteomics platform | ≈9k cognitive measure; 1140 proteomics |
| **PRS scores in African populations** | | | |
| Africa Wits-INDEPTH partnership for genomic studies (AWI-Gen) | African (different ethnolinguistic and geographic groups) | Not assessed for AD | Population cross-sectional cohort: n = 10,700 |

Abbreviations: AD, Alzheimer's disease; CDR, Clinical Dementia Rating; ICD, International Classification of Diseases; MCI, mild cognitive impairment; NA, not applicable; PET, positron emission tomography; PRS, polygenic risk score.

cases by cortical amyloid positivity in flutemetamol positron emission tomography (PET) imaging.[26] Brain amyloid PET positivity is sensitive and specific for the diagnosis of amyloid pathology,[27] which was detected in 72% of AD cases in the BICWALZS study.[26] Korea pheno2 defined cases by Clinical Dementia Rating (CDR) global score ≥1.[28] The Japan cohort included cases with AD or mild cognitive impairment (MCI) from Japan's National Center for Geriatrics and Gerontology (NCGG).[29] The AD diagnosis was based on the clinical criteria by the National Institute of Neurological and Communicative Diseases and Stroke--Alzheimer's Disease and Related Disorders Association Work Group,[30] and the MCI diagnosis was based on expert recommendations by Petersen et al.[31] The East London cohort was based on a community-based, British Bangladeshi and British Pakistani population. The cases included all-cause dementia from secondary/primary care records with MCI/cognitive decline cases excluded.[32] The UKB White cohort defined cases by International Classification of Diseases, 9th/10th revision diagnoses of AD/dementia (https://www.ukbiobank.ac.uk/).[33]

Each validation site was required to return:

1. The OR 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 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 versus everybody else. That is, the participants in the top 2.5% of the PRS versus the bottom 97.5%.
4. The sensitivity/specificity as well as negative predictive values (NPV) and positive predictive values (PPV) at the proposed cutoff (split by ancestry if appropriate for each cohort).

The NPV/PPV used prevalence-adjusted metrics, that is, PPV = (Sn * Pr) / ([Sn * Pr] + [(1 − Sp) * (1 − Pr)]) and NPV = (Sp * [1 − Pr]) / ([Sp

15525279, 2023, 12, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.13378 by South African Medical Research, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 2** Validation of PRS models (95% CIs).

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

[a]The full GIRA model: the model with PRS, *APOE*, and non-genetic covariates.
Abbreviations: *APOE*, apolipoprotein E; AUC, area under the curve; CI, confidence interval; GIRA, genomic informed risk assessment; PRS, polygenic risk score; SD, standard deviation; UKB, UK Biobank.

* (1 – Pr)] + [(1 – Sn) * Pr]) where Sn = sensitivity, Sp = specificity, and Pr = population-based prevalence reflective of the study population.

## 2.6 | 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, 9333 DNA samples were analyzed for genetic ancestry using a software tool for maximum likelihood estimation of individual ancestries from multi-locus SNP genotype datasets.[34] The INTERVAL BioResource recruited 45,263 whole blood donors (22,466 men and 22,797 women) between June 11, 2012, and June 15, 2014.[35] Donors were aged ≥18 years from 25 National Health Service Blood and Transplant (NHSBT) blood donation centers distributed across England. In addition to the GIRA validation at these two sites, the PRS scores were tested for association with 3282 plasma protein targets from the SomaLogic proteomics platform (SomaLogic Inc.).

## 2.7 | 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.[36] African populations from three geographic regions (south—South Africa; east—Kenya; and west—Ghana/Burkina Faso) were assessed.

# 3 | RESULTS

The genetic markers for the PRS model were polymorphic and informative in 4145 ancestrally diverse dementia patients from the eMERGE consortium (Table S2 in supporting information). Four cohorts from three sites, Korea Biobank, NCGG, and East London Genes & Health, reported results of the PRS in dementia cases versus controls (Table 2).

**TABLE 3** Correlation of *APOE* haplotypes with apoE protein levels.

| Proteomics ID | $R^2$ | *P* | Beta | SE |
|---|---|---|---|---|
| **APOE.2937.10.2 (isoform ε3)** | 6.55E-03 | .0057 | 16.9 | 6.11 |
| **APOE.5312.49.3 (isoform ε2)** | 0.00475 | .0194 | 15.03 | 6.42 |

Abbreviations: *APOE*, apolipoprotein E; SE, standard error.

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 (https://www.wits.ac.za/research/sbimb/research/awi-gen/)[37,38] to determine whether 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 (Figure S1 in supporting information), 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 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 $P < .05$ (Table S3 in supporting information). Using the WebGestalt (WEB-based Gene Set Analysis Toolkit) web tool, over-representation analysis (ORA) of the correlated genes by the DisGeNET approach, 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

15525279, 2023, 12, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.13378 by South African Medical Research, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

| Gene symbol | PRS_$R^2$ | *P* | Beta | SE | Gene name |
|---|---|---|---|---|---|
| ***CEBPB*** | 0.004 | .032 | 15.645 | 7.274 | CCAAT enhancer binding protein beta |
| ***FSHB*** | 0.003 | .018 | −13.862 | 5.871 | Follicle stimulating hormone subunit beta |
| ***LEP*** | 0.004 | .035 | −15.106 | 7.147 | Leptin |
| ***LHB*** | 0.002 | .036 | −12.101 | 5.765 | Luteinizing hormone beta polypeptide |
| ***LIF*** | 0.003 | .049 | −13.995 | 7.109 | LIF, interleukin six family cytokine |

Abbreviations: AD, Alzheimer's disease; PRS, polygenic risk score; SE, standard error.

GLAD4U approach highlighted the gene sets PA445859:Thyroiditis, Autoimmune (FDR = 0.017306, including *CD3G*, *CGA*, *CXCL9*, *FAS*, *LILRB1*), PA444172:Fetal Growth Retardation (FDR = 0.044508, including *IGF1*, *IGFBP3*, *LEP*, *SELP*, *TMEM70*), and PA443588:Cachexia (FDR = 0.044508, including *IGF1*, *LEP*, *LIF*, *SELP*). 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.

## 4 | 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 throughout the 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.

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. In addition to its clinical potential, this study uncovered associations between the AD PRS scores and other disease-related gene sets in the ELSA-Brasil and INTERVAL studies, 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.[39] 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 five 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.[40] Genetic variants of *FSHB* have been shown to cause hypogonadotropic hypogonadism in women and men.[41,42] 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.[43] More importantly, *FSHB* may represent a new opportunity to develop hormone therapy for AD.[44,45] In addition, previous studies have demonstrated correlations between abnormal levels of thyroid-stimulating hormone and the risk of AD.[46] While further studies are warranted, our study identified five genes that may mediate this correlation.

Given the increased recognition of limitations of PRS,[47] we examined GIRA and PRS in under-represented populations, and show the GIRA model performed better than the PRS model. Moreover, we were able to leverage the proteomics biomarker data to support the molecular mechanisms of the underlying pathogenesis of AD through the PRS associations with proteomic biomarkers, using statistical measures. In conclusion, 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.

### ACKNOWLEDGMENTS

We thank all the participants who contributed to and enabled this study. The study was 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's 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

15525279, 2023, 12, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.13378 by South African Medical Research, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 center 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 center thanks blood donor center 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.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest. Author disclosures are available in the supporting information.

## DATA AVAILABILITY STATEMENT

Supporting data from this study can be obtained by emailing the corresponding author Dr. Hakon Hakonarson.

## CONSENT FOR PUBLICATION

All authors have provided consent for publication of the manuscript.

## ORCID

*Hui-Qi Qu* https://orcid.org/0000-0001-9317-4488
*Hakon Hakonarson* https://orcid.org/0000-0003-2814-7461

## REFERENCES

1. Desai AK, Grossberg GT. Diagnosis and treatment of Alzheimer's disease. *Neurology*. 2005;64:S34-S9.
2. Dartigues J-F. Alzheimer's disease: early diagnosis. *Rev Prat*. 2011; 61:926-930.
3. Riedel BC, Thompson PM, Brinton RD. Age, APOE and sex: triad of risk of Alzheimer's disease. *J Steroid Biochem Mol Biol*. 2016;160:134-147.
4. Cacabelos R, Fernández-Novoa L, Lombardi V, Kubota Y, Takeda M. Molecular genetics of Alzheimer's disease and aging. *Methods Finds Exp Clin Pharmacol*. 2005;27:1-573.
5. Roses AD, Saunders AM. APOE is a major susceptibility gene for Alzheimer's disease. *Curr Opin Biotechnol*. 1994;5:663-667.
6. Blum CB. Type III hyperlipoproteinemia: still worth considering? *Prog Cardiovasc Dis*. 2016;59:119-124.
7. Koutsodendris N, Nelson MR, Rao A, Huang Y. Apolipoprotein E and Alzheimer's disease: findings, hypotheses, and potential mechanisms. *Annu Rev Pathol*. 2022;17:73-99.
8. Huang Y, Mahley RW. Apolipoprotein E: structure and function in lipid metabolism, neurobiology, and Alzheimer's diseases. *Neurobiol Dis*. 2014;72:3-12.
9. Kanekiyo T, Xu H, Bu G. ApoE and A$\beta$ in Alzheimer's disease: accidental encounters or partners? *Neuron*. 2014;81:740-754.
10. Phillips MC. Apolipoprotein E isoforms and lipoprotein metabolism. *IUBMB Life*. 2014;66:616-623.
11. Belbin O, Dunn JL, Ling Y, et al. Regulatory region single nucleotide polymorphisms of the apolipoprotein E gene and the rate of cognitive decline in Alzheimer's disease. *Hum Mol Genet*. 2007;16:2199-2208.
12. Martens YA, Zhao N, Liu C-C, et al. ApoE cascade hypothesis in the pathogenesis of Alzheimer's disease and related dementias. *Neuron*. 2022;110:1304-1317.
13. Stevenson-Hoare J, Heslegrave A, Leonenko G, et al. Plasma biomarkers and genetics in the diagnosis and prediction of Alzheimer's disease. *Brain*. 2023;146:690-699.
14. Emrani S, Arain HA, DeMarshall C, Nuriel T. APOE4 is associated with cognitive and pathological heterogeneity in patients with Alzheimer's disease: a systematic review. *Alzheimers Res Ther*. 2020;12:141.
15. Fasolino M, Schwartz GW, Patil AR, et al. Single-cell multi-omics analysis of human pancreatic islets reveals novel cellular states in type 1 diabetes. *Nat Metab*. 2022;4:284-299.
16. Jun GR, Chung J, Mez J, et al. Transethnic genome-wide scan identifies novel Alzheimer's disease loci. *Alzheimers Dement*. 2017;13:727-738.
17. Bellenguez C, Küçükali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet*. 2022;54:412-436.
18. Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. *Hum Mol Genet*. 2019;28:R133-R42.

15525279, 2023, 12, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.13378 by South African Medical Research, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50:1219-1224.
20. Patel AP, Khera AV. Advances and applications of polygenic scores for coronary artery disease. *Annu Rev Med*. 2023;74:141-154.
21. Qu HQ, Qu J, Glessner J, et al. Improved genetic risk scoring algorithm (GRS2') for type 1 diabetes prediction. *Pediatr Diabetes*. 2022;23:320-323.
22. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51:584-591.
23. Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. *Cell*. 2019;177:26-31.
24. Manolio TA, Goodhand P, Ginsburg G. The International Hundred Thousand Plus Cohort Consortium: integrating large-scale cohorts to address global scientific challenges. *Lancet Digi health*. 2020;2:e567-e8.
25. Scheltens P, Rockwood K. How golden is the gold standard of neuropathology in dementia? *Alzheimers Dement*. 2011;7:486-489.
26. Roh HW, Kim NR, Lee DG, et al. Baseline clinical and biomarker characteristics of biobank innovations for chronic cerebrovascular disease with Alzheimer's disease study: BICWALZS. *Psychiatry Investig*. 2022;19:100-109.
27. Chételat G, Arbizu J, Barthel H, et al. Amyloid-PET and 18F-FDG-PET in the diagnostic investigation of Alzheimer's disease and other dementias. *Lancet Neurol*. 2020;19:951-962.
28. Khan TK. Chapter 2 - clinical diagnosis of Alzheimer's disease. In: Khan TK, ed. *Biomarkers in Alzheimer's disease*. Academic Press; 2016: 27-48.
29. Ogama N, Yoshida M, Nakai T, Niida S, Toba K, Sakurai T. Frontal white matter hyperintensity predicts lower urinary tract dysfunction in older adults with amnestic mild cognitive impairment and A lzheimer's disease. *Geriatr Gerontol Int*. 2016;16:167-174.
30. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984;34:939.
31. Petersen RC, Doody R, Kurz A, et al. Current concepts in mild cognitive impairment. *Arch Neurol*. 2001;58:1985-1992.
32. Finer S, Martin HC, Khan A, et al. Cohort Profile: east London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. *Int J Epidemiol*. 2020;49:20-21.
33. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12:e1001779.
34. Chor D, Pereira A, Pacheco AG, et al. Context-dependence of race self-classification: results from a highly mixed and unequal middle-income country. *PLoS One*. 2019;14:e0216653.
35. Di Angelantonio E, Thompson SG, Kaptoge S, et al. Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors. *Lancet (London, England)*. 2017;390:2360-2371.
36. Ramsay M, Sankoh O. African partnerships through the H3Africa Consortium bring a genomic dimension to longitudinal population studies on the continent. *Int J Epidemiol*. 2016;45:305-308.
37. Ramsay M, Crowther N, Tambo E, et al. H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. *Glob Health Epidemiol Genom*. 2016;1:e20.
38. Ali SA, Soo C, Agongo G, et al. Genomic and environmental risk factors for cardiometabolic diseases in Africa: methods used for Phase 1 of the AWI-Gen population cross-sectional study. *Glob Health Action*. 2018;11:1507133.
39. Bae JB, Lipnicki DM, Han JW, et al. Does parity matter in women's risk of dementia? A COSMIC collaboration cohort study. *BMC Med*. 2020;18:1-11.
40. Orlowski M, Sarao MS. Physiology, follicle stimulating hormone. *Follicle Stimulating Horm*. 2018;6.
41. Matthews CH, Borgato S, Beck-Peccoz P, et al. Primary amenorrhoea and infertility due to a mutation in the beta-subunit of follicle-stimulating hormone. *Nat Genet*. 1993;5:83-86.
42. Lindstedt G, Nyström E, Matthews C, Ernest I, Janson PO, Chatterjee K. Follitropin (FSH) deficiency in an infertile male due to FSHbeta gene mutation. A syndrome of normal puberty and virilization but underdeveloped testicles with azoospermia, low FSH but high lutropin and normal serum testosterone concentrations. *Clin Chem Lab Med*. 1998;36:663-665.
43. Tan R, Pu S, A pilot study on the effects of testosterone in hypogonadal aging male patients with Alzheimer's disease. Aging Male. 2003;6:13-17.
44. Shao H, Breitner JC, Whitmer RA, et al. Hormone therapy and Alzheimer disease dementia: new findings from the Cache County Study. *Neurology*. 2012;79:1846-1852.
45. Resnick SM, Henderson VW. Hormone therapy and risk of Alzheimer disease: a critical time. *JAMA*. 2002;288:2170-2172.
46. Tan ZS, Beiser A, Vasan RS, et al. Thyroid function and the risk of Alzheimer disease: the Framingham Study. *Arch Intern Med*. 2008;168:1514-1520.
47. Wray NR, Lin T, Austin J, et al. From basic science to clinical application of polygenic risk scores: a primer. *JAMA Psych*. 2021;78:101-109.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Sleiman PM, Qu H, Connolly JJ, et al. Trans-ethnic genomic informed risk assessment for Alzheimer's disease: An International Hundred K+ Cohorts Consortium study. *Alzheimer's Dement*. 2023;19:5765–5772. https://doi.org/10.1002/alz.13378

15525279, 2023, 12, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.13378 by South African Medical Research, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License