HHS Public Access

Author manuscript

Alzheimers Dement. Author manuscript; available in PMC 2023 April 01.

Published in final edited form as:

Alzheimers Dement. 2022 April ; 18(4): 858–862. doi:10.1002/alz.12583.

Replication study of AD-associated rare variants

Achal Neupane[1,2,3], **Brian Lenny**[1,2,3], **John P Budde**[1,2,3], **Fengxian Wang**[1,2,3], **Joanne Norton**[1,2,3], **John C Morris**[2,4], **NIA-LOAD family study group**, **NCRAD**, **ADSP project**, **Carlos Cruchaga**[1,2,3], **Maria Victoria Fernández**[1,2,3,*]

[1.]Department Psychiatry, Washington University School of Medicine (WUSM), 660 S. Euclid Ave. B8134, St. Louis, MO, 63110, USA.

[2.]Hope Center for Neurological Disorders, WUSM, 660 S. Euclid Ave. B8111, St. Louis, MO, 63110, USA.

[3.]Neurogenomics and Informatics Center, WUSM, 4444 Forest Park Ave, St. Louis, MO 63108.

[4.]Knight Alzheimer Disease Research Center, WUSM, 4488 Forest Park Ave, St. Louis, MO, 63108, USA

In the study of Alzheimer disease (AD) associated rare variants, Prokopenko et al. [1] presented a highly detailed analysis highlighting the utility of whole-genome sequencing (WGS) to identify AD-associated rare variants, including those outside the exonic regions. They used deep (>40x) WGS data from 2247 individuals (family-based discovery dataset from NIMH [n=1393] and NIA ADSP [n=854] cohorts) for rare variants (MAF ≤1%) analysis, and validated their findings in independent datasets (n=1650) of AD cases and controls from publicly available WGS (ADSP case-control population) datasets. Their single-variant association analysis identified novel intronic variants associated with four AD candidate genes, *SEL1L*, *FNBP1L*, *LINC00298*, and *C15orf41*. Through spatial clustering/region-based analysis, they identified nine new AD candidate gene regions (*PRKCH*, *C2CD3*, *KIF2A*, *APC*, *LHX9*, *NALCN*, *CTNNA2*, *SYTL3*, and *CLSTN2*). Nonetheless, they recommended that these results should be confirmed in additional datasets because of the possibility for the spurious association of some loci in their study. We also think it is in the best interest of the scientific community to confirm whether the association of these genes and regions only affects non-coding variants.

In direct extension to their work, we used independent datasets at Washington University (WashU) to confirm the associations described by Prokopenko et al. Specifically, we wanted to confirm whether their association is mostly driven by variants in non-coding regions or also conveys to exonic variants. We used our familial late-onset AD (fLOAD) dataset from the Familial Alzheimer Sequencing (FASe) project [2] (N=1803; 356 families; 1291 cases and 330 controls), as well as an independent unrelated dataset (N=1590; 667 cases and 651 controls) which includes non-overlapping ADNI [3] and WashU's MAP [4] samples recruited by Joanne Knight Alzheimer's Disease Research Center (Knight ADRC) at the WashU School of Medicine. The approval number for the Knight ADRC

*Corresponding author.

Genetics Core family studies is 201104178. We have processed all the data using the same bioinformatics pipeline as described previously [2] with the following changes: we used GRCh38 for sequence alignment and GATKv4.1.2 [5]. Our entire dataset was processed at the MGI center using docker images freely available at: https://github.com/NeuroGenomicsAndInformatics/dockerNGS.

We performed the same statistical analysis with the same statistical packages as described in Prokopenko et al [1]. Briefly, we performed an association analysis of rare variants (MAF ≤1%) on our familial dataset using FBAT [6], logistic regression on the unrelated dataset (MAF ≤1%), followed by a fixed-effects meta-analysis of the former two datasets. Our single-variant association analysis of familial datasets detected one novel intronic variant in *NALCN* (P <0.05). Our analysis on the unrelated dataset revealed one novel rare variant in *C2CD3* (Table 1). Both *NALCN* and *C2CD3* were detected by Prokopenko in the region-based analysis. We did not identify any significant variant in the genes that Prokopenko identified via single-variant association analysis, not even when we analyzed these regions via meta-analysis of the two cohorts. Both of the variants detected in our study were intronic or non-coding variants with a frequency of less than 0.5%. As of note, we are capable of capturing variants approximately 100 bp up/downstream of the coding regions, despite our dataset is restricted to exonic regions.

Since we did not detect major associations in the single-variant analysis, we conducted gene-based analysis for all the reported genes and regions by Prokopenko et al on our familial and unrelated datasets. Two sets of single-nucleotide variants (SNVs) were tested: first, using SNVs with a minor allele frequency MAF ≤1%, then using SNVs with CADD scores ≥20, as registered in ExAC [7]. We did not detect any genes associated with AD at P <0.05 in either of our datasets (Table 1). Consistent with Prokopenko et al, we also used the spatial clustering approach to systematically group our familial data into non-overlapping regions, albeit these cluster regions were remarkably different from the clusters obtained for WGS. Our multi-marker testing on both familial and unrelated datasets on these same regions using FBAT-RV [8] also did not detect the candidate genes reported in Prokopenko et al (P <0.05).

To summarize, we have found two novel nominally significant variants in two of the 13 genes reported by Prokopenko et al: *NALCN* and *C2CD3*. Our familial dataset is slightly underpowered compared to that of Prokopenko et al., but not the unrelated dataset. We understand that some variants in intronic regions might trigger abnormal splicing or enhance the expression of genes, but there is no direct way to understand their downstream biological consequences in absence of exonic variants. Therefore, we extended this study to look at variants in exonic regions that potentially correlate with intronic variants in these reported genes.

Acknowledgements:

We thank contributors who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible.

The Alzheimer's Disease Sequencing Project (ADSP) is comprised of two Alzheimer's Disease (AD) genetics consortia and three National Human Genome Research Institute (NHGRI) funded Large Scale Sequencing and

Analysis Centers (LSAC). The two AD genetics consortia are the Alzheimer's Disease Genetics Consortium (ADGC) funded by NIA (U01 AG032984), and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) funded by NIA (R01 AG033193), the National Heart, Lung, and Blood Institute (NHLBI), other National Institute of Health (NIH) institutes and other foreign governmental and non-governmental organizations. The Discovery Phase analysis of sequence data is supported through UF1AG047133 (to Drs. Schellenberg, Farrer, Pericak-Vance, Mayeux, and Haines); U01AG049505 to Dr. Seshadri; U01AG049506 to Dr. Boerwinkle; U01AG049507 to Dr. Wijsman; and U01AG049508 to Dr. Goate and the Discovery Extension Phase analysis is supported through U01AG052411 to Dr. Goate, U01AG052410 to Dr. Pericak-Vance and U01 AG052409 to Drs. Seshadri and Fornage.

Sequencing for the Follow Up Study (FUS) is supported through U01AG057659 (to Drs. PericakVance, Mayeux, and Vardarajan) and U01AG062943 (to Drs. Pericak-Vance and Mayeux). Data generation and harmonization in the Follow-up Phase is supported by U54AG052427 (to Drs. Schellenberg and Wang). The FUS Phase analysis of sequence data is supported through U01AG058589 (to Drs. Destefano, Boerwinkle, De Jager, Fornage, Seshadri, and Wijsman), U01AG058654 (to Drs. Haines, Bush, Farrer, Martin, and Pericak-Vance), U01AG058635 (to Dr. Goate), RF1AG058066 (to Drs. Haines, Pericak-Vance, and Scott), RF1AG057519 (to Drs. Farrer and Jun), R01AG048927 (to Dr. Farrer), and RF1AG054074 (to Drs. Pericak-Vance and Beecham).

The ADGC cohorts include: Adult Changes in Thought (ACT) (UO1 AG006781, UO1 HG004610, UO1 HG006375, U01 HG008657), the Alzheimer's Disease Centers (ADC) (P30 AG019610, P30 AG013846, P50 AG008702, P50 AG025688, P50 AG047266, P30 AG010133, P50 AG005146, P50 AG005134, P50 AG016574, P50 AG005138, P30 AG008051, P30 AG013854, P30 AG008017, P30 AG010161, P50 AG047366, P30 AG010129, P50 AG016573, P50 AG016570, P50 AG005131, P50 AG023501, P30 AG035982, P30 AG028383, P30 AG010124, P50 AG005133, P50 AG005142, P30 AG012300, P50 AG005136, P50 AG033514, P50 AG005681, and P50 AG047270), the Chicago Health and Aging Project (CHAP) (R01 AG11101, RC4 AG039085, K23 AG030944), Indianapolis Ibadan (R01 AG009956, P30 AG010133), the Memory and Aging Project (MAP) (R01 AG17917), Mayo Clinic (MAYO) (R01 AG032990, U01 AG046139, R01 NS080820, RF1 AG051504, P50 AG016574), Mayo Parkinson's Disease controls (NS039764, NS071674, 5RC2HG005605), University of Miami (R01 AG027944, R01 AG028786, R01 AG019085, IIRG09133827, A2011048), the Multi-Institutional Research in Alzheimer's Genetic Epidemiology Study (MIRAGE) (R01 AG09029, R01 AG025259), the National Cell Repository for Alzheimer's Disease (NCRAD) (U24 AG21886), the National Institute on Aging Late Onset Alzheimer's Disease Family Study (NIA- LOAD) (R01 AG041797), the Religious Orders Study (ROS) (P30 AG10161, R01 AG15819), the Texas Alzheimer's Research and Care Consortium (TARCC) (funded by the Darrell K Royal Texas Alzheimer's Initiative), Vanderbilt University/Case Western Reserve University (VAN/CWRU) (R01 AG019757, R01 AG021547, R01 AG027944, R01 AG028786, P01 NS026630, and Alzheimer's Association), the Washington Heights-Inwood Columbia Aging Project (WHICAP) (RF1 AG054023), the University of Washington Families (VA Research Merit Grant, NIA: P50AG005136, R01AG041797, NINDS: R01NS069719), the Columbia University HispanicEstudio Familiar de Influencia Genetica de Alzheimer (EFIGA) (RF1 AG015473), the University of Toronto (UT) (funded by Wellcome Trust, Medical Research Council, Canadian Institutes of Health Research), and Genetic Differences (GD) (R01 AG007584). The CHARGE cohorts are supported in part by National Heart, Lung, and Blood Institute (NHLBI) infrastructure grant HL105756 (Psaty), RC2HL102419 (Boerwinkle) and the neurology working group is supported by the National Institute on Aging (NIA) R01 grant AG033193.

The CHARGE cohorts participating in the ADSP include the following: Austrian Stroke Prevention Study (ASPS), ASPS-Family study, and the Prospective Dementia Registry-Austria (ASPS/PRODEM-Aus), the Atherosclerosis Risk in Communities (ARIC) Study, the Cardiovascular Health Study (CHS), the Erasmus Rucphen Family Study (ERF), the Framingham Heart Study (FHS), and the Rotterdam Study (RS). ASPS is funded by the Austrian Science Fond (FWF) grant number P20545-P05 and P13180 and the Medical University of Graz. The ASPS-Fam is funded by the Austrian Science Fund (FWF) project I904), the EU Joint Programme - Neurodegenerative Disease Research (JPND) in frame of the BRIDGET project (Austria, Ministry of Science) and the Medical University of Graz and the Steiermärkische Krankenanstalten Gesellschaft. PRODEM-Austria is supported by the Austrian Research Promotion agency (FFG) (Project No. 827462) and by the Austrian National Bank (Anniversary Fund, project 15435. ARIC research is carried out as a collaborative study supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). Neurocognitive data in ARIC is collected by U01 2U01HL096812, 2U01HL096814, 2U01HL096899, 2U01HL096902, 2U01HL096917 from the NIH (NHLBI, NINDS, NIA and NIDCD), and with previous brain MRI examinations funded by R01-HL70825 from the NHLBI. CHS research was supported by contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and grants U01HL080295 and U01HL130114 from the NHLBI with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629, R01AG15928, and R01AG20098 from the NIA. FHS research is supported by NHLBI contracts N01-HC-25195 and HHSN268201500001I. This study was also supported by additional grants from the NIA (R01s AG054076, AG049607 and AG033040 and NINDS (R01 NS017950). The ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported

by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002- 01254). High-throughput analysis of the ERF data was supported by a joint grant from the Netherlands Organization for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, the Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the municipality of Rotterdam. Genetic data sets are also supported by the Netherlands Organization of Scientific Research NWO Investments (175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), and the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project 050-060-810. All studies are grateful to their participants, faculty and staff. The content of these manuscripts is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the U.S. Department of Health and Human Services.

The FUS cohorts include: the Alzheimer's Disease Centers (ADC) (P30 AG019610, P30 AG013846, P50 AG008702, P50 AG025688, P50 AG047266, P30 AG010133, P50 AG005146, P50 AG005134, P50 AG016574, P50 AG005138, P30 AG008051, P30 AG013854, P30 AG008017, P30 AG010161, P50 AG047366, P30 AG010129, P50 AG016573, P50 AG016570, P50 AG005131, P50 AG023501, P30 AG035982, P30 AG028383, P30 AG010124, P50 AG005133, P50 AG005142, P30 AG012300, P50 AG005136, P50 AG033514, P50 AG005681, and P50 AG047270), Alzheimer's Disease Neuroimaging Initiative (ADNI) (U19AG024904), Amish Protective Variant Study (RF1AG058066), Cache County Study (R01AG11380, R01AG031272, R01AG21136, RF1AG054052), Case Western Reserve University Brain Bank (CWRUBB) (P50AG008012), Case Western Reserve University Rapid Decline (CWRURD) (RF1AG058267, NU38CK000480), CubanAmerican Alzheimer's Disease Initiative (CuAADI) (3U01AG052410), Estudio Familiar de Influencia Genetica en Alzheimer (EFIGA) (5R37AG015473, RF1AG015473, R56AG051876), Genetic and Environmental Risk Factors for Alzheimer Disease Among African Americans Study (GenerAAtions) (2R01AG09029, R01AG025259, 2R01AG048927), Gwangju Alzheimer and Related Dementias Study (GARD) (U01AG062602), Hussman Institute for Human Genomics Brain Bank (HIHGBB) (R01AG027944, Alzheimer's Association "Identification of Rare Variants in Alzheimer Disease"), Ibadan Study of Aging (IBADAN) (5R01AG009956), Mexican Health and Aging Study (MHAS) (R01AG018016), Multi-Institutional Research in Alzheimer's Genetic Epidemiology (MIRAGE) (2R01AG09029, R01AG025259, 2R01AG048927), Northern Manhattan Study (NOMAS) (R01NS29993), Peru Alzheimer's Disease Initiative (PeADI) (RF1AG054074), Puerto Rican 1066 (PR1066) (Wellcome Trust (GR066133/GR080002), European Research Council (340755)), Puerto Rican Alzheimer Disease Initiative (PRADI) (RF1AG054074), Reasons for Geographic and Racial Differences in Stroke (REGARDS) (U01NS041588), Research in African American Alzheimer Disease Initiative (REAAADI) (U01AG052410), Rush Alzheimer's Disease Center (ROSMAP) (P30AG10161, R01AG15819, R01AG17919), University of Miami Brain Endowment Bank (MBB), and University of Miami/Case Western/North Carolina A&T African American (UM/CASE/NCAT) (U01AG052410, R01AG028786).

The four LSACs are: the Human Genome Sequencing Center at the Baylor College of Medicine (U54 HG003273), the Broad Institute Genome Center (U54HG003067), The American Genome Center at the Uniformed Services University of the Health Sciences (U01AG057659), and the Washington University Genome Institute (U54HG003079).

Biological samples and associated phenotypic data used in primary data analyses were stored at Study Investigators institutions, and at the National Cell Repository for Alzheimer's Disease (NCRAD, U24AG021886) at Indiana University funded by NIA. Associated Phenotypic Data used in primary and secondary data analyses were provided by Study Investigators, the NIA funded Alzheimer's Disease Centers (ADCs), and the National Alzheimer's Coordinating Center (NACC, U01AG016976) and the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS, U24AG041689) at the University of Pennsylvania, funded by NIA This research was supported in part by the Intramural Research Program of the National Institutes of health, National Library of Medicine. Contributors to the Genetic Analysis Data included Study Investigators on projects that were individually funded by NIA, and other NIH institutes, and by private U.S. organizations, or foreign governmental or nongovernmental organizations.

Funding:

This work was possible thanks to the following governmental grants from the National institute of Health to CC (U01 AG058922, RF1 AG053303, RF1AG058501, R01AG044546, P01AG003991, RF1 AG058501), to JM (P30 AG066444) and to MVF (1K99AG061281-01).

Competing interests:

CC receives research support from: Biogen, EISAI, Alector and Parabon. The funders of the study had no role in the collection, analysis, or interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. CC is a member of the advisory board of Vivid genetics, Halia Therapeutics and ADx Healthcare.

References

[1]. Prokopenko D, Morgan SL, Mullin K, Hofmann O, Chapman B, Kirchner R, et al. Whole-genome sequencing reveals new Alzheimer's disease-associated rare variants in loci related to synaptic function and neuronal development. Alzheimers Dement. 2021.

[2]. Fernandez MV, Budde J, Del-Aguila JL, Ibanez L, Deming Y, Harari O, et al. Evaluation of Gene-Based Family-Based Methods to Detect Novel Genes Associated With Familial Late Onset Alzheimer Disease. Front Neurosci-Switz. 2018;12.

[3]. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology. 2010;74:201–9. [PubMed: 20042704]

[4]. Berg L, McKeel DW Jr., Miller JP, Storandt M, Rubin EH, Morris JC, et al. Clinicopathologic studies in cognitively healthy aging and Alzheimer's disease: relation of histologic markers to dementia severity, age, sex, and apolipoprotein E genotype. Arch Neurol. 1998;55:326–35. [PubMed: 9520006]

[5]. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303. [PubMed: 20644199]

[6]. Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. Genet Epidemiol. 2000;19 Suppl 1:S36–42. [PubMed: 11055368]

[7]. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–91. [PubMed: 27535533]

[8]. De G, Yip WK, Ionita-Laza I, Laird N. Rare variant analysis for family-based design. PLoS One. 2013;8:e48495. [PubMed: 23341868]

Table 1.

Nominally significant (P <0.05) rare variants detected from our analysis for the 13 genes described in Prokopenko et al. P-values for the 13 genes from our gene-based analysis on both familial and unrelated datasets are also shown.

Single-variant analysis						Gene-based analysis			
						Familial		Unrelated	
Prokopenko nearest genes	**Dataset**	**Variant**	**P-value**	**p.change**	**Effect**	**P-value (MAF ≤1 %)**	**P- value (CADD ≥20)**	**P-value (MAF ≤1 %)**	**P- value (CADD ≥20)**
FNBP1L	-	-	-	-	-	0.11 (23)	0.72 (18)	0.31 (11)	0.31 (8)
SEL1L	-	-	-	-	-	0.47 (40)	0.33 (31)	0.9 (18)	1 (14)
ID2 (LINC00298)	-	-	-	-	-	0.81 (10)	0.73 (8)	0.14 (2)	0.14 (2)
C15orf41	-	-	-	-	-	0.62 (25)	0.35 (15)	0.62 (9)	0.36 (7)
PRKCH	-	-	-	-	-	0.09 (41)	0.08 (30)	0.53 (17)	0.54 (14)
C2CD3	unrelated	11:74118187:C:T	0.047	-	intron_variant	0.99 (149)	0.54 (95)	0.48 (77)	0.16 (50)
KIF2A	-	-	-	-	-	0.25 (15)	0.29 (11)	0.25 (10)	0.26 (6)
APC	-	-	-	-	-	0.25 (184)	0.9 (121)	0.68 (89)	1 (57)
LHX9	-	-	-	-	-	0.24 (20)	0.24 (18)	0.55 (10)	0.54 (10)
NALCN	familial	13:101095770:G:A	0.024	-	intron_variant	0.33 (65)	0.66 (55)	0.7 (24)	0.81 (20)
CTNNA2	-	-	-	-	-	0.05 (39)	0.08 (33)	0.27 (8)	0.23 (7)
SYTL3	-	-	-	-	-	0.5 (77)	0.73 (44)	0.11 (39)	0.42 (25)
CLSTN2	-	-	-	-	-	0.83 (75)	0.7 (58)	0.44 (38)	0.26 (27)

• Numbers within the brackets are the total variants included in the gene-based analysis