Gary W. Beecham, PhD
J.C. Bis, PhD
E.R. Martin, PhD
S.-H. Choi, PhD
A. L. DeStefano, PhD
C.M. van Duijn, PhD
M. Fornage, PhD
S.B. Gabriel, PhD
D.C. Koboldt, PhD
D.E. Larson, PhD
A.C. Naj, PhD
B.M. Psaty, PhD
W. Salerno, PhD
W.S. Bush, PhD
T.M. Foroud, PhD
E. Wijsman, PhD
L.A. Farrer, PhD
A. Goate, PhD
J.L. Haines, PhD
Margaret A. Pericak-Vance, PhD
E. Boerwinkle, PhD
R. Mayeux, MD
S. Seshadri, MD
G. Schellenberg, PhD

Neurol Genet
2017;3:e194; doi: 10.1212/NXG.0000000000000194

Supplemental data at Neurology.org/ng

THE ALZHEIMER'S DISEASE SEQUENCING PROJECT: STUDY DESIGN AND SAMPLE SELECTION

OPEN

Late-onset Alzheimer disease (LOAD) is the leading cause of dementia worldwide, with substantial economic and public health implications.[1] LOAD is a neurodegenerative disease characterized by progressive dementia typically manifesting in the seventh to ninth decades. Neuropathological changes precede clinical symptoms by 10–20 years, resulting in clinically asymptomatic individuals carrying neuropathologic features of LOAD.[2] Much of the heritability of LOAD remains unexplained, despite LOAD having a high heritability (60%–80%) and despite the identification of the *APOE* locus, a major genetic determinant for LOAD.[3] Genetic analyses have identified more than 25 other variants associated with smaller individual effects on disease risk.[4]

To identify novel genetic variation influencing AD risk and protection, the Alzheimer's Disease Sequencing Project (ADSP) was implemented as a collaborative effort of the National Institutes on Aging, the National Human Genome Research Institute, and the Alzheimer disease research community. Individual contributors include the Alzheimer's Disease Genetics Consortium, the Neurology Phenotype Working Group of the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium, and the Large Scale Sequencing and Analysis Centers at Baylor University, the Broad Institute, and Washington University.

Study design and sample selection were conducted to address issues of phenotypic heterogeneity and maximize statistical power. The study design includes 2 primary phases: a whole-genome sequencing (WGS) family-based study and a whole-exome sequencing (WES) case-control study. The WGS study was designed to target rarer variation through allelic segregation and linkage analyses in multiplex AD families. The WES case-control study was designed to target low-frequency coding variation in genes that contribute to AD risk or protection.

ADSP family study design. Approximately 1,400 multiplex LOAD families were reviewed for inclusion. Families were required to have multiple members with LOAD, genomic DNA, and available *APOE* genotypes. Families meeting initial criteria were assigned a priority rank based on number and age at onset of affected individuals, number of generations affected, and presence of *APOE ε4* alleles. Priority was given to families heavily loaded for AD (≥4 affected members with DNA available) with minimal *APOE ε4* alleles. Cases met National Institute of Neurological Diseases–Alzheimer's NINCDS-ADRDA (National Institute of Neurological and Communicative Disorders and Stroke, and the Alzheimer's Disease and related Disorders Association; now, Alzheimer's Association) criteria for possible, probable, or definite AD. Controls were free of clinical AD on cognitive assessment. A detailed description of the family design is in Appendix 1 at Neurology.org/ng.

In total, we selected 582 individuals (498 affected and 84 unaffected) from 111 families for WGS to identify genomic regions associated with increased risk of LOAD. Selected individuals include 229 European ancestry and 353 Caribbean Hispanic (CH) individuals (table). The European ancestry families included 2 large Dutch families from the Erasmus Rucphen Family study.[5] Most of these families were recently analyzed for genetic linkage, an analysis that will be used in the analysis of the sequence data.[6,7] By design, no ε4/ε4 individuals were selected for sequencing, and we prioritized ε3/ε4 individuals with earlier disease onset. Twenty-seven percent of families had at least 1 case with autopsy confirmation.

ADSP case-control design. Over 30,000 samples were considered for inclusion in the case-control design. All cases met NINCDS-ADRDA criteria for possible, probable, or definite AD, had documented age at onset or age at death (for pathologically verified cases), and *APOE* genotyping. All controls were at least 60 years old and were free of dementia by direct, documented cognitive assessment. Three primary case-control selection strategies were evaluated, and ultimately, a design was chosen that targeted cases with minimal risk as predicted by known risk factors (age, sex, and *APOE*) and targeted controls with the least probability of conversion to AD by age 85 years. The details and rationale of the case-control selection process and the evaluation of alternate study designs are described in detail in Appendix 2.

In total, we selected 5,096 cases and 4,965 controls under the chosen design (table). We selected

Table Sample demographics for family and case-control studies

	Family study		Case-control study	
	Affected	Unaffected	Case	Control
N	498	84	5,778	5,136
Age at onset/examination (SD)	73.7 (9.4)	68.0 (11.0)	76.0 (9.2)	86.1 (5.2)
Male (%)	38.8	38.1	43.1	40.7
Hispanic/Latino (%)	60.6	59.5	3.7	3.4
Autopsy confirmation (%)	15.6	0.0	32.8	7.0
APOE genotype				
ε3ε3	268 (54%)	60 (71%)	2,915 (50%)	3,394 (66%)
ε3ε4	196 (39%)	9 (11%)	2,198 (38%)	679 (13%)
ε4ε4	0 (0%)	0 (0%)	161 (3%)	17 (<1%)
ε2ε2	1 (<1%)	1 (1%)	23 (<1%)	48 (1%)
ε2ε3	29 (6%)	11 (13%)	359 (6%)	925 (18%)
ε2ε4	4 (1%)	2 (2%)	122 (2%)	73 (1%)

682 additional unrelated cases from additional multiplex families that had a strong family history for LOAD. Because some of these 682 cases arose from CH multiplex families, we included 171 cognitively normal CH control samples in the WES.

The sequencing of the nearly 600 whole genomes and 11,000 whole exomes has been completed; the data sets are currently available to the research community through qualified access (dbGaP study phs000572.v7.p4). This data set will be used to identify genetic factors influencing AD risk and protection and will be a critical resource for the LOAD research community.

Standard protocol approvals, registrations, and patient consents. This study has the approval of the institutional review boards of participating institutions, and informed consent was obtained from all patients.

From the John P. Hussman Institute for Human Genomics (G.W.B., E.R.M., M.A.P.-V.) and Dr. John T. Macdonald Foundation Department of Human Genetics (G.W.B., E.R.M., M.A.P.-V.), Miller School of Medicine, University of Miami, FL; Cardiovascular Health Research Unit (J.C.B.), Department of Medicine, Cardiovascular Health Research Unit (B.M.P.), Departments of Medicine, Epidemiology, Health Services, Department of Biostatistics (E.W.), and Division of Medical Genetics (E.W.), Department of Medicine, University of Washington, Seattle; Department of Biostatistics (S.-H.C., A.D., L.A.F.), Boston University School of Public Health, MA; The Framingham Heart Study (A.D., S.S.), MA; Department of Neurology (A.D., L.A.F., S.S.), Boston University School of Medicine, MA; Department of Epidemiology (C.M.v.D), Erasmus MC, Rotterdam, Netherlands; Brown Foundation Institute of Molecular Medicine (M.F.) and Human Genetics Center (M.F.), University of Texas Health Science Center, Houston; The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology (S.B.G.), Cambridge; Harvard University (S.B.G.), Cambridge, MA; The McDonnell Genome Institute (D.C.K., D.E.L.) and Department of Genetics (D.E.L.), Washington University, St. Louis, MO; Department of Biostatistics and Epidemiology (A.C.N.) and Perelman School of Medicine (G.S.), University of Pennsylvania, Philadelphia; Group Health Research Institute (B.M.P.), Group Health Cooperative, Seattle, WA; Human Genome Sequencing Center (W.S., E.B.), Baylor College of Medicine, Houston, TX; Department of Epidemiology and Biostatistics (W.S.B., J.L.H.), Case Western Reserve University, Cleveland, OH; Department of Medical and Molecular Genetics (T.M.F.), Indiana University School of Medicine, Indianapolis; Department of Medicine (Biomedical Genetics) (L.A.F.), Department of Ophthalmology (L.A.F.), and Department of Epidemiology (L.A.F.), Boston University School of Medicine and Public Health, MA; Department of Neuroscience (A.G.), Icahn School of Medicine at Mount Sinai, New York, NY; Human Genetics Center (E.B.), UT Health School of Public Health, Houston, TX; Taub Institute for Research on Alzheimer's Disease and the Aging Brain (R.M.) and Gertrude H. Sergievsky Center (R.M.), Columbia University Medical Center, New York, NY; Department of Neurology (R.M.), Columbia University Medical Center and New York Presbyterian Hospital, NY; and Department of Epidemiology (R.M.), Mailman School of Public Health, Columbia University, New York, NY.

Author contributions: All authors contributed to the work presented in this article. Drafting: the primary manuscript was prepared by G.W.B., with significant contributions from S.S., E.B., G.S., M.A.P.-V., and J.C.B. All authors participated in the revision and editing of the manuscript. Concept and design: primary study concept and design was by G.W.B., with significant contributions from E.R.M., J.C.B., M.A.P.-V., J.L.H., R.M., S.S., E.B., G.S., L.A.F., A.G., C.M.v.D., A.C.N., and A.D. Analysis and interpretation: review of family data was performed by M.A.P.-V., R.M., E.B., S.S., C.M.v.D., and T.M.F. Primary statistical analyses were performed by G.W.B., with additions from J.C.B., A.C.N., E.R.M., S.-H.C., A.D., and S.S. All authors participated in the interpretation and discussion of results. Acquisition of data: sample data were contributed by C.M.v.D., A.D., T.M.F., L.A.F., A.G., J.L.H., M.A.P.-V., E.B., R.M., S.S., and G.S. Statistical analyses: statistical analyses were primarily conducted by G.W.B.; additional analyses conducted by J.C.B., A.C.N., E.R.M., S.-H.C., A.D., and S.S. (affiliations noted above, all academic). Study supervision and coordination: primary study supervision and coordination was by G.S., R.M., E.B., M.A.P.-V., J.L.H., S.S., A.G., L.A.F., and E.W.

Acknowledgment: The Alzheimer's Disease Sequencing Project (ADSP) comprises 2 Alzheimer's Disease (AD) genetics consortia and 3 National Human Genome Research Institute (NHGRI)-funded Large Scale Sequencing and Analysis Centers (LSAC). The 2 AD genetics consortia are the Alzheimer's Disease Genetics Consortium (ADGC) funded by the NIA (U01 AG032984), and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) funded by the NIA (R01 AG033193), the National Heart, Lung, and Blood Institute (NHLBI), other NIH institutes, and other foreign governmental and nongovernmental organizations. The Discovery Phase analysis of sequence data is supported through UF1AG047133 (to G. Schellenberg, L.A. Farrer, M.A. Pericak-Vance, R. Mayeux, and J.L. Haines); U01AG049505 to S. Seshadri; U01AG049506 to E. Boerwinkle; U01AG049507 to E. Wijsman; and U01AG049508 to A. Goate. Data generation and harmonization in the Follow-up Phases is supported by U54AG052427 (to G. Schellenberg and Wang). The ADGC cohorts include Adult Changes in Thought (ACT), the Alzheimer's Disease Centers (ADC), the Chicago Health and Aging Project (CHAP), the Memory and Aging Project (MAP), Mayo Clinic (MAYO), Mayo Parkinson's Disease controls, the University of Miami, the Multi-Institutional Research in Alzheimer's Genetic Epidemiology Study (MIRAGE), the National Cell Repository for Alzheimer's Disease (NCRAD), the National Institute on Aging Late Onset Alzheimer's Disease Family Study (NIA-LOAD), the Religious Orders Study (ROS), the Texas Alzheimer's Research and Care Consortium (TARC), Vanderbilt University/Case Western Reserve University (VAN/CWRU), the Washington Heights-Inwood Columbia Aging Project (WHICAP) and the

Washington University Sequencing Project (WUSP), the Columbia University Hispanic–Estudio Familiar de Influencia Genetica de Alzheimer (EFIGA), the University of Toronto (UT), and Genetic Differences (GD). The CHARGE cohorts with funding provided by 5RC2HL102419 and HL105756, include the following: the Atherosclerosis Risk in Communities (ARIC) Study which is conducted as a collaborative study supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), the Austrian Stroke Prevention Study (ASPS), the Cardiovascular Health Study (CHS), the Erasmus Rucphen Family Study (ERF), the Framingham Heart Study (FHS), and the Rotterdam Study (RS). The 3 LSACs are the Human Genome Sequencing Center at the Baylor College of Medicine (U54 HG003273), the Broad Institute Genome Center (U54HG003067), 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 the 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 the NIA and at the Database for Genotypes and Phenotypes (dbGaP) funded by the NIH. This research was supported in part by the Intramural Research Program of the NIH and the National Library of Medicine. Contributors to the Genetic Analysis Data included Study Investigators on projects that were individually funded by the NIA and other NIH institutes, and by private U.S. organizations, or foreign governmental or nongovernmental organizations.

Study funding: Supported by the NIH, primarily the NIA, NHLBI, and NHGRI. Primary support includes the Alzheimer's Disease Genetics Consortium (ADGC) funded by the NIA (U01 AG032984), and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) funded by the NIA (R01 AG033193), the Human Genome Sequencing Center at the Baylor College of Medicine (U54 HG003273), the Broad Institute Genome Center (U54HG003067), and the Washington University Genome Institute (U54HG003079). Additional funding of contributing sites is noted in the Acknowledgment section.

Disclosures: G.W. Beecham receives funding from the NIH and the Department of Defense. J.C. Bis reports no disclosures. E.R. Martin has served on the editorial board of Frontiers in Statistical Genetics and Methodology *and holds US Patent No. 6697739 Test for Linkage and Association in General Pedigrees: The Pedigree Disequilibrium Test. S.-H. Choi reports no disclosures. A. DeStefano has received research support from the NIH. C.M. van Duijn and M. Fornage report no disclosures. S.B. Gabriel is an employee of a nonprofit entity and has been a consultant for WilmerHale Guidepoint Global. D.C. Koboldt receives a coinventor's share of license revenue for VarScan (a software tool for next-generation sequencing analysis), with licensing and disbursements handled by his former institution, Washington University in St. Louis. In the past 2 years, paying licensees included Bina Technologies, Janssen, Fera Science, Philips Electronics, and WuXi NextCODE. D.E. Larson has received research support from the NIH and St. Jude Children's Research Hospital. A.C. Naj has received speaker honoraria from Pfizer; has served on the editorial board of* PLoS One*; and has received research support from the NIA, the BrightFocus Foundation, and Penn Institute on Aging. B.M. Psaty serves on the DSMB of a clinical trial for a device funded by the manufacturer (Zoll Lifecor) and on the Steering Committee for the Yale Open Data Access project funded by Johnson & Johnson; is a contributing writer for JAMA; and has received research support from an entity/entities listed in the Acknowledgment section. W. Salerno has been a consultant for Lasergen. W.S. Bush serves on the editorial boards of* BMC BioData Mining *and* PLoS One*; and has received research support from the NIA. T.M. Foroud has served on the scientific advisory boards of the National Advisory Council on Alcohol Abuse and Alcoholism, the Washington University Alzheimer's Disease Research Center, and the NIA Genetics of Alzheimer's Disease Data Storage Site; has received travel funding from the Michael J. Fox Foundation for Parkinson's Research, the NIH, the University of Pittsburgh, and the University of Chicago; has received travel funding and speaker honoraria from the University of Texas at Austin; and has received research support from the NIH, the US Department of Defense, Columbia University, San Diego State University, the University of California, San Diego, the University of Massachusetts, the University of Pennsylvania, the State University of New York, and the Michael J. Fox Foundation for Parkinson's Research. E. Wijsman has served on the scientific advisory board of NIH NHLBI National Registry of Genetically Triggered Thoracic Aortic Aneurysms and Cardiovascular Conditions; has served on the editorial boards of* BMC Proceedings *and* Bioinformatics*; and has received research support from the NIH and the Metropolitan Life Foundation Award for Medical Research. L.A. Farrer has served on the editorial boards of the* American Journal of Alzheimer's Disease & Other Dementias *and* Clinical Genetics*; has 1 patent pending for the use of PLXNA4 as a drug target and biomarker for Alzheimer disease; has been a consultant for Novartis Pharmaceuticals, Gerson Lehrman, Guidepoint Global, and Finnegan & Associates, LLP; and has received research support from the NIH, the Fidelity Foundation, and the Thome Memorial Foundation. A. Goate has served on the scientific advisory board of Denali Therapeutics; has received travel funding from the Rainwater Foundation; has served on the editorial board of* eLife*; holds patents for PSEN mutations in AD, Tau mutations in FTD, and TDP43 mutations in ALS\FTD; has been a consultant for Cognition Therapeutics and AbbVie; has received research support from F-Prime, the NIA, the Rainwater Charitable Foundation, and the JPB Foundation; and receives royalty payments from Taconic Industries for tau mutation patent. J.L. Haines has served on the editorial boards of* Neurogenetics, Current Protocols in Human Genetics, *and* Human Molecular Genetics*; receives publishing royalties from John Wiley & Sons; and has received research support from the NIH. M.A. Pericak-Vance serves on the editorial boards of* Genetic Epidemiology, Molecular Autism, *and* Advances in Genomics and Gene Expression*; her immediate family member Dr. Jeffery Vance has served on the editorial boards of* Neurology Genetics *and* American Journal of Neurodegenerative Disease*; and she has received research support from the NIH and the JJ Vance Foundation. E. Boerwinkle has received a speaker honorarium from the American Society for Bone and Mineral Research; is a Scientific Officer at Codified Genomics, LLC; and has received research support from the NIH. R. Mayeux has received research support from the NIH. S. Seshadri serves on the editorial boards of* Journal of Alzheimer's Disease, Stroke, *and* Neurology *and has received research support from the NIA. G. Schellenberg has served on the scientific advisory boards of Alzheimer's Association, the Society of Progressive Supranuclear Palsy, the Alzheimer Research Consortium, the Peebler PSP Research Foundation, the United Kingdom Parkinson Disease Center, University College London, the Alzheimer's Disease Sequence Project, the Structural Variant Work Group, Mayo Clinic, Rochester, Udall Center, the University of Miami, and the Oxford Parkinson's Disease Centre; has received travel funding/speaker honoraria from the Alzheimer's Disease Center, CurePSP, the University of California, San Diego, Keystone Symposia, the University of California, the Institute for Memory Impairment and Neurological Disorders, Biomarkers in Neuropsychiatric Disorders (Toronto, Canada), the NIH, Novartis, the McKnight Brain Institute, the University of Florida, the NIA, the Keep Memory Alive Center (Cleveland Clinic), the Lou Ruvo Center for Brain Health, PSP/Lewy Body Disease Think-Tank, the American Association of Neuropathologists, the Fusion Conference, "What does the future hold?" (Tucson, AZ), "Progressive supranuclear palsy genetics—update" (La Jolla, CA), the Center for Public Health*

Genomics, Genome Sciences Seminar, the University of Virginia, Neurology Grand Rounds, and Columbia University; has served on the editorial boards of the Journal of Neural Transmission*,* Alzheimer's Research*, the* American Journal of Alzheimer's Disease and other Dementias*,* Neurodegenerative Diseases*,* Current Alzheimer Research*, and* Pathology and Laboratory Medicine International*; is a professor at the University of Pennsylvania; and has received research support from the NIA/NIH, CurePSP, and CBD Solutions. Go to Neurology.org/ng for full disclosure forms. The Article Processing Charge was funded by the authors.*

Received May 1, 2017. Accepted in final form August 17, 2017.

Correspondence to Dr. Beecham: gbeecham@med.miami.edu

1. Prince M, Jackson J. Alzheimer's Disease International, World Alzheimer Report 2009: The Global Prevalence of Dementia; 2009. Available at: alz.co.uk/research/files/WorldAlzheimerReport.pdf. Accessed October, 2016.
2. Bateman RJ, Xiong C, Benzinger TL, et al; Dominantly Inherited Alzheimer Network. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. N Engl J Med 2012;367:795–804.
3. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 1993;261: 921–923.
4. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013;45: 1452–1458.
5. Liu F, Arias-Vásquez A, Sleegers K, et al. A genomewide screen for late-onset Alzheimer disease in a genetically isolated Dutch population. Am J Hum Genet 2007;81:17–31.
6. Barral S, Cheng R, Reitz C, et al. Linkage analyses in Caribbean Hispanic families identify novel loci associated with familial late-onset Alzheimer's disease. Alzheimers Dement 2015;11:1397–1406.
7. Kunkle BW, Jaworski J, Barral S, et al. Genome-wide linkage analyses of non-Hispanic white families identify novel loci for familial late-onset Alzheimer's disease. Alzheimers Dement 2016;12:2–10.