Evidence of Recessive Alzheimer Disease Loci in a Caribbean Hispanic Data Set
Genome-wide Survey of Runs of Homozygosity

Mahdi Ghani, PhD; Christine Sato, MSc; Joseph H. Lee, PhD; Christiane Reitz, PhD; Danielle Moreno, BSc; Richard Mayeux, MD; Peter St George-Hyslop, MD; Ekaterina Rogaeva, PhD

**IMPORTANCE** The search for novel Alzheimer disease (AD) genes or pathologic mutations within known AD loci is ongoing. The development of array technologies has helped to identify rare recessive mutations among long runs of homozygosity (ROHs), in which both parental alleles are identical. Caribbean Hispanics are known to have an elevated risk for AD and tend to have large families with evidence of inbreeding.

**OBJECTIVE** To test the hypothesis that the late-onset AD in a Caribbean Hispanic population might be explained in part by the homozygosity of unknown loci that could harbor recessive AD risk haplotypes or pathologic mutations.

**DESIGN** We used genome-wide array data to identify ROHs (>1 megabase) and conducted global burden and locus-specific ROH analyses.

**SETTING** A whole-genome case-control ROH study.

**PARTICIPANTS** A Caribbean Hispanic data set of 547 unrelated cases (48.8% with familial AD) and 542 controls collected from a population known to have a 3-fold higher risk of AD vs non-Hispanics in the same community. Based on a Structure program analysis, our data set consisted of African Hispanic (207 cases and 192 controls) and European Hispanic (329 cases and 326 controls) participants.

**EXPOSURE** Alzheimer disease risk genes.

**MAIN OUTCOMES AND MEASURES** We calculated the total and mean lengths of the ROHs per sample. Global burden measurements among autosomal chromosomes were investigated in cases vs controls. Pools of overlapping ROH segments (consensus regions) were identified, and the case to control ratio was calculated for each consensus region. We formulated the tested hypothesis before data collection.

**RESULTS** In total, we identified 17 137 autosomal regions with ROHs. The mean length of the ROH per person was significantly greater in cases vs controls ($P = .0039$), and this association was stronger with familial AD ($P = .0005$). Among the European Hispanics, a consensus region at the EXOC4 locus was significantly associated with AD even after correction for multiple testing (empirical $P$ value $1 [\text{EMP}1]$, $0.001$; $\text{EMP}2$, $0.002$; 21 AD cases vs 2 controls). Among the African Hispanic subset, the most significant but nominal association was observed for CTNNA3, a well-known AD gene candidate ($\text{EMP}1$, $0.002$; 10 AD cases vs 0 controls).

**CONCLUSIONS AND RELEVANCE** Our results show that ROHs could significantly contribute to the etiology of AD. Future studies would require the analysis of larger, relatively inbred data sets that might reveal novel recessive AD genes. The next step is to conduct sequencing of top significant loci in a subset of samples with overlapping ROHs.


Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Authors: Ekaterina Rogaeva, PhD, and Peter St George-Hyslop, MD, Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Tanz Neuroscience Building, 6 Queen’s Park Crescent W, Room 227, Toronto, ON M5S 3H2, Canada (ekaterina.rogaeva@utoronto.ca and p.hyslop@utoronto.ca).

Supplemental content at jamanetwork.com
Alzheimer disease (AD) is the most common form of dementia, characterized by neurotoxic aggregates of tau and β-amyloid peptides in the brain. Known genetic factors account for only up to 40% of risk for AD, half of which is attributable to APOE and approximately 3% to dominant mutations in APP, PSEN1, and PSEN2. Genome-wide association studies (GWASs) have identified associations between late-onset AD (age >65 years) and single-nucleotide polymorphisms (SNPs) in 9 loci (CLU, PICALM, BIN1, MS4A4A/MS4A6E, CR1, CD2AP, CD33, EPHA1, and ABCA7). In which the pathogenic variants have not yet been revealed apart from CR1. In addition, a candidate gene approach discovered a reproducible association between AD and SORL1, and a recent exome study revealed AD risk variants in TREM2. Finally, the G4C2-repeat expansion in C9orf72 has been recently reported in a few patients with clinical AD; however, this association remains to be validated.

Intriguingly, a recent study on the concordance of AD among parents and offspring suggested that approximately 90% of early-onset AD cases are likely the result of autosomal recessive inheritance; however, the p.A673V substitution in APP is the only known recessive AD mutation. Recessive inheritance has not been widely investigated for complex traits. The lack of inbred families in most North American or European data sets has made mapping of recessive loci challenging; however, the development of array technologies has recently helped to identify rare recessive mutations among long runs of homozygosity (ROHs), in which both parental alleles are identical. Runs of homozygosity greater than 1 megabase (Mb) are relatively frequent in the general population and could arise without inbreeding as a result of common extended haplotypes at loci with rare recombination events. Small ROHs (<1 Mb) are too frequent (especially in inbred populations) to search for rare recessive loci, and therefore most ROH studies use a 1-Mb cutoff.

At present, ROHs have been associated with a risk for rheumatoid arthritis, Parkinson disease, and schizophrenia. Recently, genome-wide measurements of ROHs (>1 Mb) were studied in 2 outbred AD data sets of North American and European origin. In both studies, the global burden analysis of ROHs did not reveal a significant association with AD. The only significant result in locus-specific analyses was obtained for a consensus ROH region on chr8p12 in the North American data set (p = .017; 40 AD cases vs 9 controls), but no loci survived correction for multiple testing in the European data set. Surprisingly, homozygosity mapping of a small data set from an isolated Arab community in Israel (Wadi Ara) detected that the controls were more inbred than the AD cases. In addition, a whole-genome study of 2 affected siblings from a consanguineous AD family revealed several shared ROHs; however, lack of data from unaffected family members complicated the result interpretation.

In the present study, we investigated a late-onset AD data set of a Caribbean Hispanic population for global burden measurements of ROHs and the specific loci that could harbor recessive AD mutations or risk haplotypes. This data set was previously analyzed in an SNP-based GWAS that replicated several AD genes (APOE, CLU, PICALM, and BIN1). We also evaluated the data set for the incidence of rare (>100 kilobase [kb]) copy number variations, but only a nominal association with AD was detected for a duplication on chr15q11.

The Caribbean Hispanic population in our study originates mainly from the Dominican Republic or from Puerto Rico, which share a complex heritage. The native populations of both islands dwindled owing to disease, warfare, and forced labor after the Spanish conquered them in the 15th century, and large numbers of slaves were later imported from West Africa. At present, the population of the Dominican Republic is 9.8 million; that of Puerto Rico, 3.7 million.

Caribbean Hispanics are known to have a 3-fold-higher risk for AD compared with non-Hispanics in the same community and tend to have large families (>6 siblings) with evidence of inbreeding, which makes this population very appropriate for an ROH study. A previous study found a single-founder PSEN1 mutation (p.G206A) that was unique to Caribbean Hispanics and responsible for 42% of early-onset AD families in this population. However, the high incidence of late-onset AD in this population is not explained by PSEN1. Hence, we tested the hypothesis that the late-onset AD in the Caribbean Hispanic population might be in part explained by the homozygosity of unknown loci.

**Methods**

**Sample Collection and Genotyping**

This study was approved by the institutional review boards of Columbia University and the University of Toronto. We examined a Caribbean Hispanic case-control data set. Briefly, the data set was composed of 542 unrelated control subjects (68.5% female) and 547 unrelated AD cases (70.7% female), including 267 cases who had at least 1 relative with AD, representing familial AD (FAD). Family history of AD was also recorded for 47 controls (8.7%); however, none had any AD symptoms at the time of examination (mean age, 82 years). The mean (SD) age at onset of AD was 80 (8) years, and the mean (SD) age at the last examination of the controls was 79 (6) years. Analysis with the STRUCTURE program (http://pritch.bsd.uchicago.edu/software.html) had previously determined that 207 AD cases were of African ancestry and 329 cases were of European ancestry; 192 controls were of African ancestry and 326 controls were of European ancestry (11 cases and 24 controls were of unknown ancestry). The proportion of both subgroups was similar in cases and controls: African Hispanics made up 37.8% of cases and 35.4% of controls, whereas European Hispanics made up 60.1% of cases and 60.1% of controls.

All DNA samples were isolated from whole blood and genotyped on the Illumina HumanHap 650Y array (Illumina, Inc.). The genotyping success rate was greater than 99% for all samples, providing a reliable source to estimate the ROHs. We
included in this study only samples that previously passed stringent SNP-based quality control. The raw genotypedatahavebeensubmittedtotheGeneExpressionOmnibus,apub-
lic functional genomics data repository at the National Cen-
ter for Biotechnology Information (GSE33528).

Identification and Analysis of ROHs
Runs of homozygosity were identified using a whole-genome association analysis toolset (PLINK; http://pngu.mgh.harvard.
edu/~purcell/plink/), which has a reliable algorithm for ROH detection. As previously described, we used a sliding win-
dow of 50 SNPs across the genome with no more than 1 heter-
zygous SNP allowed in the window, and the minimal ROH size was defined as 1 Mb with an SNP density of at least 1 SNP per 50 kb. We calculated the total and mean lengths of the ROHs per sample and the total number of ROHs for each sample. Global burden measurements among autosomal chromosomes were investigated in cases vs controls with a 1-tailed test (10 000 permutations) for the number of ROHs and for the total and mean ROH lengths per individual (as in published ROH studies). We used a 1-tailed test because the Caribbean Hispanic population has a high incidence of AD and is more suitable for the detection of risk and not of protective alleles.

Pools of overlapping ROH segments were identified for the entire data set, and we calculated the ratio of cases to controls for each consensus region, defined as a shared segment of greater than 100 kb (>3 SNPs) among cases and controls. Runs of homozygosity were checked globally among the entire data set and each subgroup using the PLINK regional test against the consensus regions. The PLINK grouping function was also applied to cluster the ROH segments if they were at least 99% identical by genotype.

The standardized, pairwise Lewontin disequilibrium co-
efficient was used to estimate the strength of linkage disequi-
librium between SNPs using the Haplovie software package. In addition, we estimated inbreeding (F) using the SNPs from the Illumina HumanHap 650Y array to determine identity by descent as estimated by the genetic relationship matrix, which was implemented in the Genome-wide Complex Trait Analysis.

Results
Global Burden Analysis of ROHs
In total, we identified 17 137 autosomal regions with ROHs greater than 1 Mb (8912 in 542 controls and 8225 in 547 AD cases), including 4041 ROHs in 267 FAD cases (Table). The global burden measurements of rate and total size of ROHs were not associated with AD. However, AD was significantly associated with a larger mean ROH size per person: 2133 kb in cases vs 1934 kb in controls (P = .0039). This association was stronger with FAD (P = .0005). Also, the total size of ROHs per person was marginally longer in FAD cases (40 Mb) vs controls (34 Mb) (P = .05), which could be the result of consanguineous mar-

<table>
<thead>
<tr>
<th>Table. Global Burden Analyses of ROHs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Entire Data Set</strong></td>
</tr>
<tr>
<td>No. of ROHs</td>
</tr>
<tr>
<td>AD Cases (n = 547)</td>
</tr>
<tr>
<td>Controls (n = 542)</td>
</tr>
<tr>
<td>Rate, ROHs per person</td>
</tr>
<tr>
<td>15.04</td>
</tr>
<tr>
<td>16.44</td>
</tr>
<tr>
<td>Total distance spanned by ROHs per person, kb</td>
</tr>
<tr>
<td>36 200</td>
</tr>
<tr>
<td>34 300</td>
</tr>
<tr>
<td>Mean ROH size per person, kb</td>
</tr>
<tr>
<td>2133</td>
</tr>
<tr>
<td>1934</td>
</tr>
</tbody>
</table>

| **Familial AD Subgroup**            |
| No. of ROHs                         |
| AD Cases (n = 267)                  |
| Controls (n = 542)                  |
| Rate, ROHs per person               |
| 15.13                              |
| 16.44                              |
| Total distance spanned by ROHs per person, kb |
| 39 600                             |
| 34 300                             |
| Mean ROH size per person, kb        |
| 2260                               |
| 1934                               |

| **African Hispanic Subgroup**       |
| No. of ROHs                         |
| AD Cases (n = 207)                  |
| Controls (n = 192)                  |
| Rate, ROHs per person               |
| 8.98                               |
| 8.85                               |
| Total distance spanned by ROHs per person, kb |
| 21 180                             |
| 17 640                             |
| Mean ROH size per person, kb        |
| 2014                               |
| 1875                               |

| **European Hispanic Subgroup**      |
| No. of ROHs                         |
| AD Cases (n = 329)                  |
| Controls (n = 326)                  |
| Rate, ROHs per person               |
| 18.08                              |
| 18.30                              |
| Total distance spanned by ROHs per person, kb |
| 44 550                             |
| 40 190                             |
| Mean ROH size per person, kb        |
| 2221                               |
| 1994                               |

Abbreviations: AD, Alzheimer disease; kb, kilobase; ROHs, runs of homozygosity.
rages in this island population. Indeed, in our cohort, we observed a noticeable inbreeding comparable to second-cousin marriages (mean $F = 0.020$). The mean degree of inbreeding was even higher in the 2 subgroups, which was more evident in the European Hispanics ($F = 0.055$) than in the African Hispanics ($F = 0.046$).

The global burden analysis was extended to the population subgroups, as defined by the Structure program. Alzheimer disease was significantly associated with larger mean lengths of ROH in the European Hispanic subgroup ($P = .013$), which also had a 2-fold greater total ROH size per individual compared with the African Hispanic subgroup (Table).

### Association Analysis of Gene-Intersecting ROHs
We estimated which genes were intersected by ROHs more frequently in cases vs controls. This locus-specific analysis of the entire data set did not reveal significant findings after correction for multiple testing. The most significant nominal association with AD was observed for the ROHs at chr7q33 intersecting EXOC4 (NM_021807.3) (empirical $P$ value 1 [EMP1], .0006; 28 cases vs 8 controls). In the European Hispanic population, we observed a significant association between AD and EXOC4-intersecting ROHs even after correction for multiple tests (EMP1, .0001; EMP2, .009; 21 AD cases vs 2 controls) (Figure, A). The presence of EXOC4-intersecting ROHs was confirmed by manual inspection of the b-allele frequency and log R ratio plots in GenomeStudio Data Analysis Software (Illumina, Inc) (Supplement [eFigure 1]). Although no locus survived correction for multiple tests in the African Hispanic population, CTNNA3 (NM_013266.2) with its nested gene (LRRTM3 [NM_178011.3]) on chr10q21.3 revealed the strongest nominal association with AD (EMP1, .002). Runs of homozygosity-intersecting CTNNA3 were detected in 10 cases but not in controls (Figure, B). Based on previous copy number variation analysis, the homozygosity at the EXOC4 and CTNNA3 loci cannot be explained by the presence of large deletions.34

We did not detect any significant findings for ROHs intersecting SORL1, TREM2, APOE, or the 9 AD loci identified by recent large GWASs.45-50 Among known AD loci, the only nominal association with AD was detected in the European Hispanic subgroup for CLU (EMP1, .038; 7 cases vs 1 control). Of note, although AD in our cohort was strongly associated with the APOE ε4/ε4 genotype, we found no difference in the frequencies of APOE-intersecting ROHs between cases (5 individuals, including 3 with 3/3, 1 with 2/2, and 1 with 4/4 genotypes) and controls (4 individuals, all with genotype 3/3). Hence, in a Caribbean Hispanic population, the frequency of extended APOE haplotypes is low, as in other reported populations.27,28

### Analyses of ROH Consensus Regions
In the entire data set, we identified 1415 consensus regions (defined as the region shared by all individuals carrying an ROH at a given locus), but none of them was associated with AD after correction for multiple testing (Supplement [eTable 1]). However, in the European Hispanic population, a homozygous segment overlapping EXOC4 was significantly associated with AD even after correction for multiple testing (EMP1, .0001; EMP2, .002; 21 cases vs 2 controls). This consensus region was flanked by rs7793621 and rs7792010 located at chr7:132988570-13268110 (hg19) (Supplement [Figure 2] and [eTable 2]). In the African Hispanic population, the most significant nominal association was observed for the consensus region overlapping CTNNA3 (EMP1, .004), but no region survived correction for multiple testing (Supplement [Table 3]). Consensus regions of CTNNA3 and EXOC4 have several rare, potentially functional coding variations, including missense and nonsense SNPs, according to dbSNP137 (Supplement [eTable 4]).

In addition, we evaluated the samples with ROHs at EXOC4 for genotypic identity using the PLINK grouping function. The largest subgroup with more than 99% identical genotypes consisted of 11 cases (including 9 European Hispanics) and 1 control of African Hispanic ancestry (Supplement [eTable 5]). Manual inspection of the genotypes of these 12 individuals (Supplement [eTable 6]) confirmed the shared haplotype that extends from the Sec8-exocyst domain (PF04048). The controls of European Hispanic origin (n = 326) were used to study the linkage disequilibrium structure around EXOC4. The shared haplotype observed in the 9 European Hispanics belongs to a linkage disequilibrium block that contains only EXOC4 (Supplement [eFigure 2]).

### Discussion
Our results suggest the existence of recessive AD risk loci in the Caribbean Hispanic population. We detected an association between AD and a larger genome-wide mean ROH size ($P = .0039$), which was stronger with FAD ($P = .0005$). The previous studies on global burden measurements of ROHs did not identify any association with AD, likely owing to the outbred nature of the investigated North American and European data sets.27,28 In contrast, our study had greater power to detect association with ROHs because we studied a more homogeneous population with noticeable inbreeding. Of note, total ROH size was 2 times longer in the European Hispanic subset than in the African Hispanic subset, indicating closer relatedness in the European Hispanic population, in which we detected the association between AD and a larger mean ROH size ($P = .013$).

The association of ROHs with AD could reflect the cumulative effects of multiple ROHs widely distributed throughout the genome (as in schizophrenia44) or the contribution of specific loci harboring recessive mutations and/or risk haplotypes in a subset of patients with AD. In the gene-based approach, the most significant result (EMP2, .009) was observed in the European Hispanics for EXOC4, encoding a component of the exocyst complex involved in the trafficking of the N-methyl-D-aspartate receptor.45-47 In the African Hispanic population, the strongest but nominal association was found for CTNNA3, which is a previously reported functional and positional AD gene candidate (also known as VR22 or α-3-catenin). The CTNNA3 protein is a binding partner of β-catenin, and the complex of αβ-catenin interacts with presenilin 1, promoting cadherin processing.48-52 The CTNNA3 gene is mapped to 10q22.2 within an AD locus (OMIM 605526).53-54 Further-
more, CTNNA3 was suggested as an APOE-dependent AD gene and reported to influence the level of β-amyloid 42 in late-onset AD families. Finally, an association between AD and a block of CTNNA3 SNPs in our Caribbean Hispanic data set was recently described. Hence, the present study further supports the role of CTNNA3 in AD.

The EXOC4 and CTNNA3 proteins are part of a pathway responsible for the “stabilization and expansion of the E-cadherin adherens junction,” according to the Pathway Interaction Database (http://pid.nci.nih.gov/index.shtml). Also, both proteins are implicated in the “Arf6 trafficking events” pathway, involved in endocytosis and vesicular transport by acting on membrane lipid composition and actin organization.

In the analysis of ROH consensus regions, only the EXOC4 locus among European Hispanics survived correction for multiple testing. In total, we observed 61 consensus regions nominally associated with AD, which could not be explained by the presence of large copy number variations. Some of these re-
regions are mapped near known GWAS-significant SNPs (<1 Mb). For instance, the consensus region at chr8p21.1 (11 cases vs 1 control) is near the CLU gene. 5,6,10 Two other consensus regions are close to the significant SNPs detected in the African American GWAS: rs10850408 (chr12q24.21) and rs2221154 (chr3p24.1). 6,13 The regions on chr3p26.1 and chr4p15.1 were close to significant SNPs identified in a meta-analysis of AD age at onset (rs271066 and rs10517270, respectively). 6,14 Finally, among the largest consensus regions (>1 Mb) on chr21q21.1, g01 are closest to the significant SNPs identified in an African American GWAS (DrStGeorge-Hyslop).

For instance, the consensus region at chr8p21.1 (11 cases vs 1 control) is near the CLU gene. 5,6,10 Two other consensus regions are close to the significant SNPs detected in the African American GWAS: rs10850408 (chr12q24.21) and rs2221154 (chr3p24.1). 6,13 The regions on chr3p26.1 and chr4p15.1 were close to significant SNPs identified in a meta-analysis of AD age at onset (rs271066 and rs10517270, respectively). 6,14 Finally, among the largest consensus regions (>1 Mb) on chr21q21.1, g01 are closest to the significant SNPs identified in an African American GWAS (DrStGeorge-Hyslop).

In summary, we found that ROHs could significantly contribute to the etiology of AD in a population with noticeable inbreeding. Future studies would require the analysis of larger data sets, including association tests with regression analysis incorporating the principal components of a population (eg, age and sex). To characterize the molecular defects underlying AD, the next step is to conduct deep sequencing of top significant loci in a subset of samples with overlapping ROHs, followed by segregation studies in families affected by potential pathologic mutations.

ARTICLE INFORMATION
Accepted for Publication: May 15, 2013.
Published Online: August 26, 2013.

Author Affiliations: Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, Ontario, Canada (Ghani, Sat0, Moreno, St George-Hyslop, Rogaeava); Taub Institute for Research on Alzheimer’s Disease and the Aging Brain, Gertrude H. Sergievsky Center, Departments of Neurology, Psychiatry, and Medicine, College of Physicians and Surgeons, Columbia University, New York, New York (Lee, Reitz, Mayeux); Department of Epidemiology, School of Public Health, Columbia University, New York, New York (Lee, Reitz, Mayeux); Department of Medicine, University of Toronto, Toronto, Ontario, Canada (St George-Hyslop, Rogaeava); Cambridge Institute for Medical Research and Department of Clinical Neurosciences, University of Cambridge, Cambridge, England (St George-Hyslop).

Author Contributions: Study concept and design: St George-Hyslop, Rogaeava.
Acquisition of data: Ghani, Sat0, Moreno, Mayeux, Rogaeava.
Analysis and interpretation of data: Ghani, Lee, Reitz, Rogaeava.
Drafting of the manuscript: Ghani, Lee, Reitz, St George-Hyslop, Rogaeava.
Critical revision of the manuscript for important intellectual content: Sat0, Lee, Reitz, Moreno, Mayeux, St George-Hyslop.
Statistical analysis: Lee, Reitz.
Obtained funding: Ghani, Sat0, Moreno, Mayeux, St George-Hyslop, Rogaeava.
Administrative, technical, and material support: Mayeux, Rogaeava.
Study supervision: St George-Hyslop, Rogaeava.

Conflict of Interest Disclosures: None reported.

Funding/Support: This work was supported by grants R37 AG15473 (Dr Mayeux) and PO1 AG07232 from the National Institute on Aging, National Institutes of Health; by the Blanchett Hooker Rockefeller Foundation; by the Charles S. Robertson Gift from the Banbury Fund (Dr Mayeux); by the Garfield Weston Foundation (Dr Rogaeava); by the Canadian Institutes of Health Research; by the Wellcome Trust; by the Medical Research Council; by the National Institutes of Health, by the National Institute of Health Research; by the Ontario Research Fund; and by the Alzheimer Society of Ontario (Dr St George-Hyslop).

REFERENCES


