

SUPPLEMENTARY INFORMATION

Identification of novel susceptibility loci for Guam neurodegenerative disease: Challenges of genome scans in genetic isolates

Weiva Sieh^{1,2}, Yoonha Choi³, Nicola H. Chapman¹, Ulla-Katrina Craig⁴, Ellen J. Steinbart^{5,8}, Joseph H. Rothstein¹, Kiyomitsu Oyanagi⁶, Ralph M. Garruto⁷, Thomas D. Bird^{1,5,8}, Douglas R. Galasko¹⁰, Gerard D. Schellenberg^{5,8,9,11,12}, Ellen M. Wijsman^{1,3,13}

1. Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA 98195, USA
2. Department of Health Research and Policy, Division of Epidemiology, Stanford University, Stanford, CA 94305, USA
3. Department of Biostatistics, University of Washington, Seattle, WA 98195, USA
4. Micronesian Health and Aging Study, University of Guam, Mangilao, Guam 96923
5. Geriatric Research Education and Clinical Center, Veterans Affairs Puget Sound Health Care System, Seattle, WA 98108, USA
6. Department of Neuropathology, Tokyo Metropolitan Institute for Neuroscience, Tokyo, Japan
7. Department of Anthropology, Laboratory of Biomedical Anthropology and Neurosciences, Binghamton University, Binghamton, NY 13902, USA
8. Department of Neurology, University of Washington, Seattle, WA 98195, USA
9. Department of Medicine, Division of Gerontology and Geriatric Medicine, University of Washington, Seattle, WA 98195, USA
10. Department of Neurosciences, University of California, San Diego, La Jolla CA 92093, USA
11. Department of Pharmacology, University of Washington, Seattle, WA 98195, USA
12. Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
13. Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA

22	D22S280	31.30	0.00	0.50	0.00	0.50	0.00	0.50	0.12	0.15	0.00	0.50
22	D22S283	38.62	0.45	0.10	0.08	0.30	0.14	0.20	0.00	0.50	0.00	0.50
22	D22S423	46.42	0.19	0.20	0.11	0.30	0.00	0.50	0.00	0.50	0.00	0.50
22	D22S274	51.54	0.00	0.50	0.00	0.50	0.00	0.50	0.15	0.15	0.00	0.50

Quality Control (QC)

Pedigree structures and genotype data were checked for errors. Relpair [ver 2.0](#) (<http://csg.sph.umich.edu/boehnke/relpair.php>) (1) was used to identify misspecified pedigree relationships based on the genome-wide marker data (see case-control analyses) with marker allele frequencies estimated from the U pedigree and the case-control sample. Mendelian-inconsistent genotypes were identified using Loki (<http://www.stat.washington.edu/thompson/Genepi/pangaea.shtml>) version 2.4.5(2). The rate of detectable pedigree and genotype errors was low, with only 2 misspecified relationships and 9/400 markers with a genotype inconsistency identified. Pedigree and genotype inconsistencies were removed prior to linkage analysis.

Deleted: (1)

We took several additional QC measures to verify the genotype data and meiotic map on chr 12 because of the complicated linkage signals found. Somatic cell hybrids were used to verify that the chr 12 primers used were on the correct chromosome. Merlin (<http://www.sph.umich.edu/csg/abecasis/Merlin>) (3) was used to check for Mendelian consistent genotype errors, and replicate genotyping was performed for all 26 markers on chr 12 to validate all genotypes. To investigate the possibility of chromosomal rearrangements, we checked for differences between the likelihood given the observed data and the inter-marker recombination fractions based upon the standard Marshfield genetic map versus the maximum likelihood estimates calculated from the data, described below.

Deleted: (2)

Deleted: s

Comparison of observed vs. expected map distances on chromosome 12.

Comparison of expected vs. estimated map distances was obtained with likelihood ratio tests. For a pair of markers i, j , the recombination fraction $\hat{\theta}_{i,j}$ that maximizes the likelihood, L ,

over all families was found with Merlin by grid search over $\theta_{i,j} = 0.01, 0.02, \dots 0.49$. We compare this to the published recombination fraction $\theta_{i,j}$. Define a likelihood-ratio statistic

$$X_{i,j}^2 = \begin{cases} \ln \left[\frac{L(\hat{\theta}_{i,j})}{L(\theta_{i,j})} \right] & \hat{\theta}_{i,j} \geq \theta_{i,j} \\ -\ln \left[\frac{L(\hat{\theta}_{i,j})}{L(\theta_{i,j})} \right] & \hat{\theta}_{i,j} < \theta_{i,j} \end{cases}$$

Large absolute values of this statistic indicate a difference between the map-based and data-based recombination fractions between the markers, with positive values indicating more and negative values less recombination, than expected, based on the map. Under standard asymptotic conditions and the null hypothesis that the map distance between markers i and j is correct, $2 \cdot X_{i,j}^2$ has a χ_1^2 distribution.

The data used to investigate the genetic map consisted of the S dataset, and five nuclear families extracted from the U pedigree in order to minimize the impact of misspecified extended pedigree structures and allele frequencies. No qualitative differences were found using the Rutgers map (<http://compugen.rutgers.edu/mapomat>) as the referent instead of the Marshfield map (not shown).

Analysis of marker-marker recombination patterns identified map/marker inconsistencies, consistent with possible map distortion in the sample (Figure S1). D12S99 showed more recombination with proximal markers D12S336 ($X^2 = 6.14$, $\chi_1^2 = 12.28$, $p = 0.00045$), D12S364 ($X^2 = 2.61$, $\chi_1^2 = 5.22$, $p = 0.022$) and D12S310 ($X^2 = 2.85$, $\chi_1^2 = 5.70$, $p = 0.017$) than expected. There was also evidence of excess recombination, relative to the map, between markers in two blocks: block 1 (36-49 cM) includes D12S1617 and spans 6 markers between D12S310 and D12S1640, and block 2 (46-75 cM) spans 6 markers between D12S345 and D12S83. In the 36 marker pairs

formed by choosing one marker from block 1 and one from block 2, the maximum X^2 statistic was 3.75 (range = 0.02-3.75, maximum $\chi_1^2 = 7.50$, $p=0.0062$). In contrast, the X^2 statistics were much closer to zero for marker pairs within block 1 (range = -0.90-0.22) and within block 2 (range = -0.25-1.34, maximum $\chi_1^2 = 2.68$, $p=0.1$).

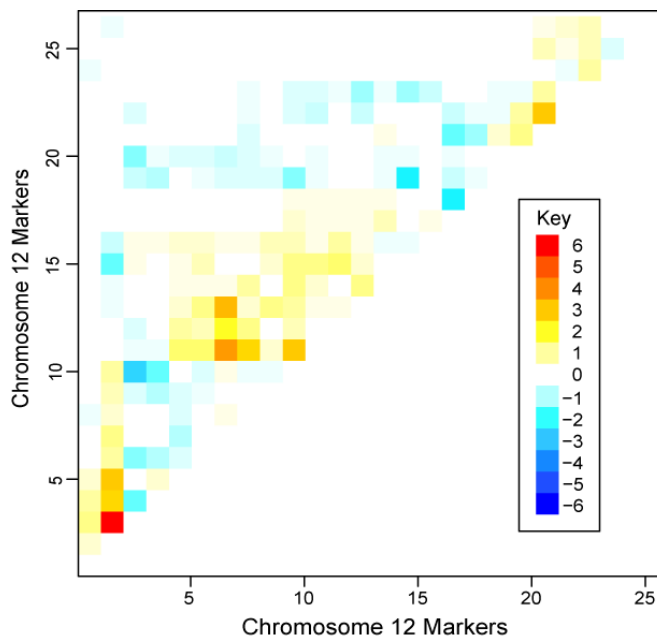


Figure S1 Values of X^2 , the signed log-likelihood ratio, for the 325 pairs of markers on chromosome 12. Shades of blue indicate log likelihood ratios below zero, for pairs where less recombination is observed than expected using the Marshfield map. Shades of orange/red indicate pairs where more recombination is observed than expected. The 26 markers in chromosome order were: D12S352, D12S99, D12S336, D12S364, D12S310, D12S1606, D12S1057, D12S1617, D12S1596, D12S1640, D12S345, D12S2080, D12S1048, D12S85, D12S368, D12S83, D12S326, D12S351, D12S346, D12S78, D12S1646, D12S79, D12S86, D12S324, D12S1659, and D12S1723.

References

1. [Epstein, M.P., Duren, W.L. and Boehnke, M. \(2000\) Improved inference of relationship for pairs of individuals. *American Journal of Human Genetics*, **67**, 1219-31.](#)

Formatted: Font: Italic

Formatted: Font: Bold

2. [Heath, S.C. \(1997\) Markov Chain Monte Carlo Segregation and linkage analysis for oligogenic models. *The American Journal of Human Genetics*, **61**, 748-760.](#)

3. [Abecasis, G.R., Cherny, S.S., Cookson, W.O. and Cardon, L.R. \(2002\) Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics*, **30**, 97-101.](#)

Formatted: Font: Italic

Formatted: Font: Bold

Formatted: Font: Italic

Formatted: Font: Bold

Formatted: Indent: Left: 0 pi, First line: 0 pi

Deleted: 1. . Heath, S.C. (1997) Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet*, **61**, 748-60.¶
2. . Abecasis, G.R., Cherny, S.S., Cookson, W.O. and Cardon, L.R. (2002) Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet*, **30**, 97-101.¶

Formatted: Indent: Left: 0 pi, Hanging: 3 pi