A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4

Annibale A. Puca*†, Mark J. Daly‡, Stephanie J. Brewster§, Tara C. Matise¶, Jeffrey Barrett‡, Maureen Shea-Drinkwater‖, Sammy Kang¶, Erin Joyce§, Julie Nicoli*, Erica Benson§, Louis M. Kunkel*, and Thomas Perls‖

*Howard Hughes Medical Institute, Genetics Division, Children's Hospital, and Harvard Medical School, Boston, MA 02215; ‡Center for Genome Research at the Whitehead Institute for Biomedical Research, Cambridge, MA 02214; ¶Department of Genetics, Nelson Biological Laboratories, Rutgers University, Piscataway, NJ 08854; §Centagenetix, Boston, MA 02119; and ‖Gerontology Division and Biometrics Center, Beth Israel Deaconess Medical Center and Division on Aging, Harvard Medical School, Boston, MA 02215

Contributed by Louis M. Kunkel, July 2, 2001

Substantial evidence supports the familial aggregation of exceptional longevity. The existence of rare families demonstrating clustering for this phenotype suggests that a genetic etiology may be an important component. Previous attempts at localizing loci predisposing for exceptional longevity have been limited to association studies of candidate gene polymorphisms. In this study, a genome-wide scan for such predisposing loci was conducted by using 308 individuals belonging to 137 sibships demonstrating exceptional longevity. By using nonparametric analysis, significant evidence for linkage was noted for chromosome 4 at D4S1564 with a MLS of 3.65 ($P = 0.044$). The analysis was corroborated by a parametric analysis ($P = 0.052$). These linkage results indicate the likelihood that there exists a gene, or genes, that exerts a substantial influence on the ability to achieve exceptional old age. Identification of the genes in humans that allow certain individuals to live to extreme old age should lead to insights on cellular pathways that are important to the aging process.

Centenarians have lived ≈20 years longer than the average life expectancy of those in their birth cohort who survived beyond the age of 65 years (1). For most centenarians, the majority of those additional years are spent in good health (2). We hypothesize that, relative to the general population, centenarians have a history of aging slowly, and they either markedly delay or escape age-associated diseases such as cardiovascular disease, stroke, diabetes, cancer, and Alzheimer's disease (3). While studying the phenotypic characteristics of centenarians in an attempt to discern characteristics that might explain this disease resistance and survival advantage, we frequently encountered subjects who reported similarly long-lived siblings. We therefore suspected that these siblings had genetic and environmental factors in common that conferred an unexpected survival advantage. Comparing the longevity of siblings of centenarians and siblings of a control group who were from a similar birth cohort, the siblings of the centenarians had a substantially greater chance of surviving to extreme old age compared with the siblings of the controls. This relative risk (λ_s) of survival steadily increased with age for siblings of the centenarians to the point that the siblings had a 4-fold greater probability of survival to age 91 (4). Moreover, in the course of recruiting subjects, we discovered four families with many exceptionally old individuals that also contributed to our suspicion that there is a significant genetic component to the ability to achieve exceptional old age (5). Finally, the dramatic degree of demographic selection that occurs in the nonagenarian years might indicate that relatively few polymorphisms/genes are significantly responsible for the survival of a relatively few individuals achieving even older age (6). These observations, along with our ability to enroll a substantial number of sibships, prompted us to conduct a sibling pair linkage study.

Methods

Subjects. Written informed consent was obtained from all participants and/or their legal proxies, in accordance with the requirements of the Institutional Review Boards of Beth Israel Deaconess Medical Center and Children's Hospital of Boston. Proof of birth, in the form of a birth certificate, was obtained from 90% of subjects. Other acceptable proof was obtained from the remaining subjects; otherwise, a potential subject was not enrolled (7). The collection of centenarians or near-centenarians and their siblings began in 1997, well in advance of the specific planning for a genome scan analysis.

A minimum age of 98 years for at least one member of the sibship (the proband) was chosen because preliminary sibship recruitment efforts indicated that lowering the age from 100 to 98 was necessary to enroll enough sibships for a potential genetic study. All siblings age 90 years and older were recruited. Before the conduct of the genome scan and subsequent analysis, it was decided that all sibships with a proband, at least one male sibling age 91 years old or older, and/or a female sibling age 95 years old or older would constitute the sample set to be analyzed. These sibling ages were chosen because they represented the 5% oldest individuals in the birth cohort based on U.S. and Canadian life tables. This selection resulted in 143 sibships and 322 individuals predominantly of European descent.

Linkage Analysis. DNA was extracted from whole blood with QIAamp DNA Blood Maxi Kit from Qiagen (Chatsworth, CA) and dispensed into 384-well plates with a Hydra 9600 robot (Robbins, Sunnyvale, CA). The ABI Prism Linkage Mapping Set, Version 2 (Applied Biosystems), and True Allele PCR Premix (Applied Biosystems, Foster City, CA) were used for the genome-wide scan reactions on affected individuals as well as on control DNA from Centre d'Étude du Polymorphisme Humain (Paris) (no. 1347/02), to standardize the allele sizes during the editing process. The ABI Prism Linkage Mapping Set, Version 2, True Allele PCR Premix (Applied Biosystems, Foster City, CA), and PCR conditions are described in Applied Biosystems user manuals. Reactions were generated in an ABI 9700 PCR machine. The reactions were pooled with a Genesis RSP 150 robot (Tecan, Durham, NC), and products were separated on an ABI Prism 377 sequencer under standard conditions (gels were poured with Long Ranger Single Packs from BMA Biomedi-

Abbreviations: cM, centimorgans; apo, apolipoprotein; lod, logarithm of odds; hlod, heterogeneity, lod; MLS, maximum lod score.

†To whom reprint requests should be addressed at: Howard Hughes Medical Institute, Genetics Division, Children's Hospital and Harvard Medical School, 300 Longwood Avenue, Boston, MA 02115. E-mail: apuca@rascal.med.harvard.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

POPULATION BIOLOGY

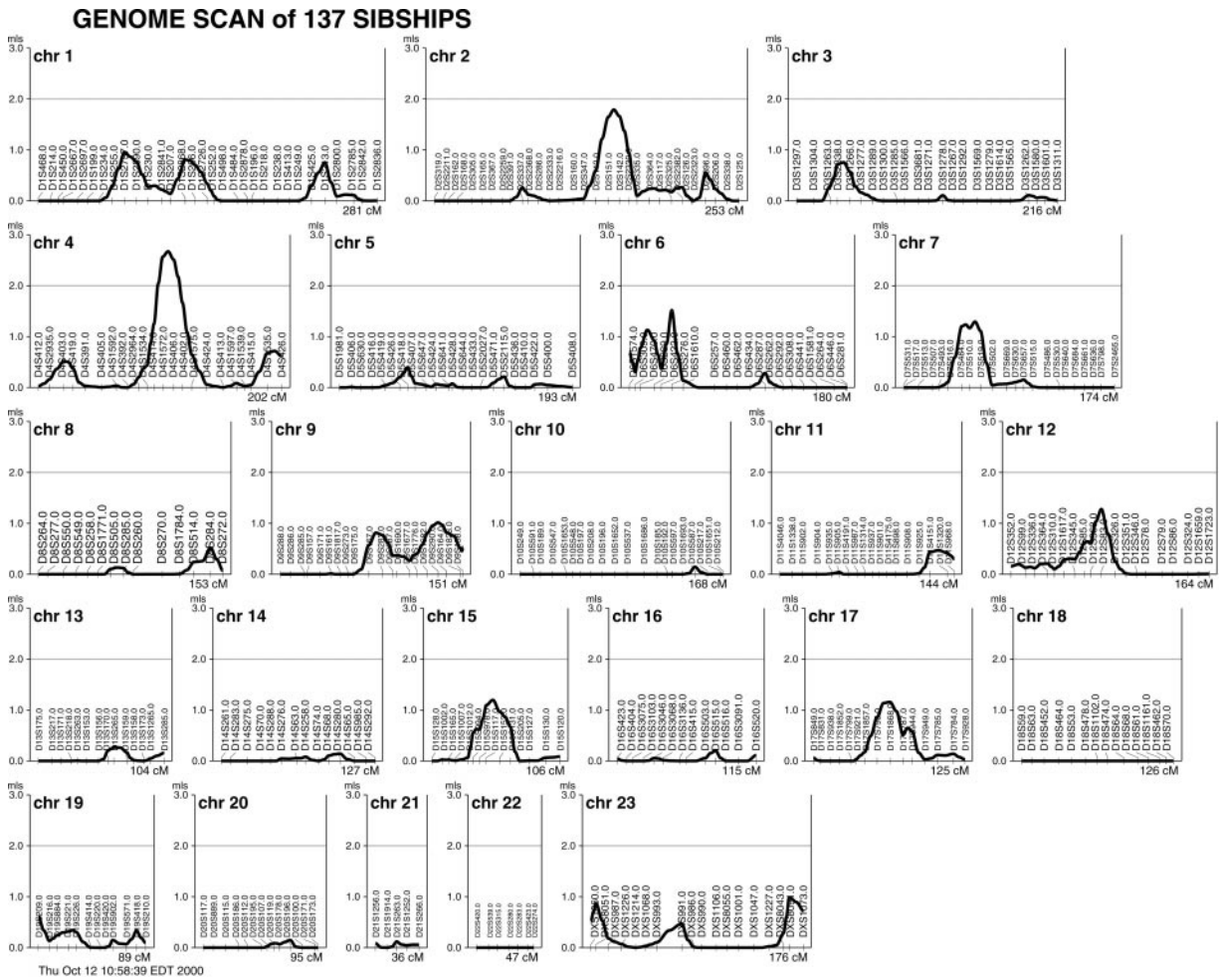

Fig. 1. Multipoint lod scores for the genome-wide scan. Individuals in 137 sibships demonstrating exceptional longevity were genotyped at 400 marker loci throughout the genome. Scores are plotted as a function of specific markers. Chromosome number is designated at the top of each plot. The horizontal threshold lines on each graph represent MLS = 2.0, a score slightly higher than the average maximum score expected by chance once in a genome scan.

cals). GENESCAN ANALYSIS Software 3.1 (Applied Biosystems) was used for size processing together with ROX 400 size standard (Applied Biosystems). The allele-calling software was GENOTYPER 2.5, from Applied Biosystems.

DNA samples of individuals were initially genotyped with 400 simple tandem repeat markers with an average heterozygosity of 0.70 and an average marker density of 10 centimorgans (cM). Subsequently, loci suggestive for linkage were genotyped with additional markers (8). Sibships in which the observed genotype data were 10^5-fold more likely to have arisen if the individuals were unrelated, or half-siblings were excluded from the genetic analysis. In addition, those families with aberrantly high numbers of Mendelian inheritance errors were excluded (six families and one individual, a half-sib of a larger sibship, were excluded), thus resulting in 137 sibships and 303 individuals in the final analysis. Size standards and alleles were determined by using Applied Biosystems software. Allele frequencies were computed by using PEDMANAGER software (M.J.D., unpublished work), which estimates them from alleles present in founder individuals in each pedigree. Calculated allele frequencies did not differ from the Centre d'Étude du Polymorphisme Humain (Paris) frequencies database. In suggestively linked regions, a higher-density map was analyzed, genotypes that implied double recombinants were identified, and the raw data were reexamined or retyped.

Results

An obvious consequence of studying the exceptional longevity phenotype is that neither unaffected family members nor parents are available for the study. Because no prior hypothesis of a specific genetic mechanism for extreme longevity existed, and the limited information from most of these families precluded a meaningful estimate of such a model, an affected-only multipoint nonparametric analysis was performed (9). This analysis tested for excess sharing of chromosomal regions that were identical by descent in the sibships. The nonparametric analysis [maximum lod score (MLS)] was performed by using the GENEHUNTER 2.0 program (10).

The genome screen of 308 individuals belonging to 137 families identified a region on chromosome 4 between D4S1572 and D4S406 as highly suggestive of linkage (MLS = 2.67) (see Fig. 1). Fine mapping at an average of 1 marker every 3 cM was performed in a 20-cM region around this peak, resulting in an increased MLS (3.65) at marker D4S1564 (see Table 1).

To estimate the significance of the observation of linkage to chromosome 4, simulations under the hypothesis of no linkage were designed that match the family structure and marker heterozygosity of our study. One thousand genome scans were simulated with a genome-wide map density of 1 marker every 3 cM (matching the fine-mapping density performed in the region of maximum linkage), and in only 44 of those 1,000 genome scans was the observed MLS of 3.65 exceeded ($P = 0.044$).

Previous studies indicate that parametric linkage analyses by using both a dominant and recessive inheritance model may be more powerful in some cases than nonparametric studies for

Table 1. Nonparametric MLS and parametric hlod scores within the exceptional longevity susceptibility locus relative to markers in the chromosome 4 region

Marker	Position	MLS*	$hlod^{*}_{dom}$
D4S1534	95.0	0.57	0.54
D4S414	100.8	1.53	1.30
D4S2986	105.3	2.78	2.26
D4S1572	108.0	3.07	2.57
D4S411	109.0	3.07	2.60
D4S1564	112.6	3.65	3.26
D4S406	117.1	2.55	2.15
D4S1611	121.6	1.70	1.56
D4S402	124.5	1.39	1.06
D4S2975	126.7	0.94	0.89

*Nonparametric MLS and parametric hlod scores were calculated for within the exceptional longevity susceptibility locus relative to markers in the chromosome 4 region. Positions are from the Marshfield map, (http://research.marshfieldclinic.org/genetics/Map_Markers/maps/IndexMapFrames.html). Fine mapping at an average of 1 marker every 3 cM around the peak noted in Fig. 1 resulted in an increased MLS (3.65) at marker D4S1564. A dropoff of 1.5 in the MLS score on either side of the peak MLS defines the area in which we can be 95% confident the gene resides. A dropoff in the MLS of 2 on either side of the peak is observed in a 20-cM region encompassed by D4S414 and D4S1611. The MLS scores in this region and the hlod scores under the dominant model are shown.

detecting genes contributing to complex traits (11). In this case, a positive result from an independent parametric analysis could serve as a validity check of the nonparametric findings. In addition, nonparametric methods do not easily allow for the presence of locus heterogeneity. Therefore, multipoint genome-wide affected-only parametric linkage analyses were conducted by using the ALLEGRO program under both a dominant and recessive model with reduced penetrance and allowing for heterogeneity (12).

The disease allele frequency under each model was determined such that the penetrance model fit a population prevalence of 5% (0.047 for dominant and 0.30 for recessive). Because we analyzed only affecteds, the penetrance of the susceptible genotypes was set to 1 and to 0.005 for the unsusceptible genotypes (phenocopies). As these penetrance parameters are necessarily based on assumptions about the frequency of nongenetic cases, after performing our analysis with this predefined model, we repeated the analysis by varying the penetrance parameters (penetrance ratio varied from 200:1 to 50:1). These variations on our model had a negligible effect on our results.

The finely mapped region of chromosome 4 provided the strongest results, with a maximum heterogeneity logarithm of odds (hlod) of 3.3 under the dominant model (with 57% of families linked) and a maximum hlod of 2.9 under the recessive model (with 45% of families linked), both occurring at marker D4S1564. Empiric *P* values were determined by genome-wide simulations. One thousand unlinked replicates were generated to match the actual study. These replicates were analyzed exactly as was the original dataset. Of 1,000 replicates, 26 had a lod score >3.3 under the dominant model ($P = 0.026$), and 77 had a lod score >2.9 under the recessive model ($P = 0.077$). Because two inheritance models were evaluated, these parametric results are appropriately summarized and corrected by multiplying the more significant associated *P* value by 2, thus producing a final parametric *P* value of 0.052. These analyses provide significant evidence for heterogeneity, with odds in favor of heterogeneity of 1,800:1. For this dataset and trait, the parametric and nonparametric analyses corroborate each other and demonstrate that the results are robust to different sets of assumptions.

Discussion

Of the 4-fold risk to siblings of centenarians (λ_s) to achieve at least their early nineties (4), the degree of excess allele sharing indicates that a locus in the D4S1564 region could explain $\approx$1.65-fold of that risk, and thus a more modest common polymorphism might be inferred. A number of genetic mechanisms are likely to be at play in effecting such an influence. A relative lack of polymorphisms predisposing to age-associated diseases appears to be one prerequisite to achieving exceptional old age (13). The marked decreased frequency of the apolipoprotein (apo)E ε-4 allele among centenarians exemplifies such a mechanism (14). Another prerequisite may be some degree of genetic control over how fast a person ages and the consequent impact on susceptibility to a myriad of diseases normally associated with aging (3). The possibility that we are tracing a locus effecting a single common disease is less likely given recent findings by scientists working with the French Centenarian Study sample set. Simulations performed by Nemani *et al.* indicate that the allele-sharing method is unable to detect polymorphisms that predispose young individuals to significant mortality and occur at low frequency, as in the case of carriers of the apoE ε-4 allele (15). Consistent with this observation, we have not detected any linkage at the apoE or other known disease loci.

Alternatives to this linkage study are case-control studies performed on *a priori*-selected candidate genes hypothesized to effect fundamental mechanisms of aging (14, 16). The drawbacks of such studies include the improbability of picking the right gene to study the myriad of known and unknown genes affecting the process of interest (17). The linkage study described here markedly improves the efficiency of such association studies by defining a region likely to contain polymorphism(s) with significant influence on life span.

Additional association studies with these families and replication of these results with an independent data set should facilitate the positional cloning of a gene that influences the ability to age well and achieve exceptional longevity. Identification of the genes in humans that allow certain individuals to live to extreme old age should lead to insights on cellular pathways that are important to the aging process.

We thank Drs. Dale Nyholt, Eric Lander, and Jurg Ott for their helpful remarks. We also thank the remarkable family members who participated in this study for their time and assistance in our phenotypic data and genetic material gathering efforts. L.M.K. is a Howard Hughes Institute Investigator and an Ellison Medical Foundation Senior Scholar. T.P. is supported by grants from the Alzheimer's Association Darrell and Jane Phillippi Faculty Scholar Award, the Paul Beeson Faculty Scholar in Aging Research Award (the American Federation for Aging Research and the Alliance for Aging Research), the Institute for the Study of Aging, the National Institute on Aging (1R01AG18721, 1R21AG16916), and the Retirement Research Foundation.

1. Bell, F., Wade, A. & Goff, S. (1997) *Cohort Life Tables* (U.S. Dept. of Health and Human Services, Social Security Administration, Washington, DC).
2. Hitt, R., Young-Xu, Y., Silver, M. & Perls, T. (1999) *Lancet* **354,** 652.
3. Perls, T. T. (1995) *Sci. Am.* **272,** 70–75.
4. Perls, T. T., Bubrick, E., Wager, C. G., Vijg, J. & Kruglyak, L. (1998) *Lancet* **351,** 560.
5. Perls, T., Shea-Drinkwater, M., Bowen-Flynn, J., Ridge, S. B., Kang, S., Joyce, E., Daly, M., Brewster, S. J., Kunkel, L. & Puca, A. A. (2000) *J. Am. Geriatr. Soc.* **48,** 1483–1485.
6. Wachter, K. W. (1997) *In Between Zeus and the Salmon: The Biodemography of Longevity*, eds. Wachter, K. W. & Finch, C. E. (Natl. Acad. Press, Washington, DC), pp. 1–16.
7. Perls, T. T., Bochen, K., Freeman, M., Alpert, L. & Silver, M. H. (1999) *Age Ageing* **28,** 193–197.
8. Lander, E. & Kruglyak, L. (1995) *Nat. Genet.* **11,** 241–247.
9. Risch, N. (1988) *Am. J. Hum. Genet.* **42,** 353–364.

10. Kruglyak, L., Daly, M. J., Reeve-Daly, M. P. & Lander, E. S. (1996) *Am. J. Hum. Genet.* **58,** 1347–1363.
11. Durner, M., Vieland, V. J. & Greenberg, D. A. (1999) *Am. J. Hum. Genet.* **64,** 281–289.
12. Gudbjartsson, D. F., Jonasson, K., Frigge, M. L. & Kong, A. (2000) *Nat. Genet.* **25,** 12–13.
13. Tomita-Mitchell, A., Muniappan, B. P., Herrero-Jimenez, P., Zarbl, H. & Thilly, W. G. (1998) *Gene* **223,** 381–391.
14. Schächter, F., Faure-Delanef, L., Guenot, F., Rouger, H., Froguel, P., Lesueur-Ginot, L. & Cohen, D. (1994) *Nat. Genet.* **6,** 29–32.
15. Nemani, M., Sahbatou, M., Blanche, H., Thomas, G. & Pascoe, L. (2000) *Mech. Aging Dev.* **119,** 25–39.
16. Rebeck, G. W., Perls, T. T., West, H. L., Sodhi, P., Lipsitz, L. A. & Hyman, B. T. (1994) *Neurology* **44,** 1513–1516.
17. Hadley, E. C., Albert, S. A., Bailey-Wilson, J. & NIA Aging and Genetic Epidemiology Working Group (2000) *Am. J. Epidemiol.* **152,** 1003–1008.