Power of Non-Parametric Linkage Analysis in Mapping Genes Contributing to Human Longevity in Long-Lived Sib-Pairs

Qihua Tan,[1]* J. H. Zhao,[2] I. Iachine,[3] J. Hjelmborg,[4] W. Vach,[3] J. W. Vaupel,[5] K. Christensen,[4] and T.A. Kruse[1]

[1]*Department of Clinical Biochemistry and Genetics, KKA, Odense University Hospital, Odense, Denmark.*
[2]*Department of Epidemiology and Public Health, University College London, London, United Kingdom*
[3]*Statistics, Institute of Public Health, University of Southern Denmark, Odense, Denmark*
[4]*Epidemiology, Institute of Public Health, University of Southern Denmark, Odense, Denmark*
[5]*Max-Planck Institute for Demograhic Research, Rostock, Germany*

This report investigates the power issue in applying the non-parametric linkage analysis of affected sib-pairs (ASP) [Kruglyak and Lander, 1995: Am J Hum Genet 57:439–454] to localize genes that contribute to human longevity using long-lived sib-pairs. Data were simulated by introducing a recently developed statistical model for measuring marker-longevity associations [Yashin et al., 1999: Am J Hum Genet 65:1178–1193], enabling direct power comparison between linkage and association approaches. The non-parametric linkage (NPL) scores estimated in the region harboring the causal allele are evaluated to assess the statistical power for different genetic (allele frequency and risk) and heterogeneity parameters under various sampling schemes (age-cut and sample size). Based on the genotype-specific survival function, we derived a heritability calculation as an overall measurement for the effect of causal genes with different parameter settings so that the power can be compared for different modes (dominant, recessive) of inheritance. Our results show that the ASP approach is a powerful tool in mapping very strong effect genes, both dominant and recessive. To map a rare dominant genetic variation that reduces hazard of death by half, a large sample (above 600 pairs) with at least one extremely long-lived (over age 99) sib in each pair is needed. Again, with large sample size and high age cut-off, the method is able to localize recessive genes with relatively small effects, but the power is very limited in case of a dominant effect. Although the power issue may depend heavily on the true genetic nature in maintaining survival, our study suggests that results from small-scale sib-pair investigations should be referred with caution, given the complexity of human longevity. © 2004 Wiley-Liss, Inc.

Key words: power; non-parametric linkage analysis; longevity

INTRODUCTION

The genetic studies on human longevity using DNA data have been dominated by the association approaches [De Benedictis et al., 2001]. The implementation of newly developed statistical methods [Yashin et al., 1999; Tan et al., 2001a,b, 2003] has helped the association approaches to gain preference due to increased statistical power [Pletcher and Stumpf, 2002]. In addition, important interaction terms (gene-sex, gene-environment, gene-gene) can also be considered within the framework of association [Tan et al., 2001a, 2002a]. There are, however, crucial issues that confront the approach. Similar to the other complex traits, longevity is likely modulated by multiple genes [Cournil and Kirkwood, 2001]. Selecting and testing the candidate genes or marker regions can be difficult and tedious work, accompanied by the eventual frequent false-positive outcomes due to multiple testing. This highlights the need for a genome-wide approach that takes into account the interdependence of genomic data. Linkage analysis that is popular in use in mapping human disease genes can be an important approach. Although linkage analysis has only low resolution in gene mapping [Cardon and Bell, 2001], a more feasible approach would be to follow the regions identified through linkage mapping by association studies and conduct

linkage disequilibrium fine mapping [Sham et al. 2000; Risch 2000].

The multi-point linkage analysis based on examining patterns of disease gene co-segregation within families has been a powerful tool to map both Mendelian and non-Mendelian disease genes [Ott, 1999]. Although in the context of longevity studies, parental genotype information is usually missing, the non-parametric linkage analysis suitable for studying genes linked to late-onset diseases could be a useful method to localize genes implicated in human longevity. The method makes use of the information on alleles shared identical by decent (IBD) between affected sib-pairs to infer the location of genes that are linked to the trait of interest [Kruglyak and Lander, 1995]. Higher probability for achieving longevity in the centenarians' siblings reported by recent studies [Perls et al., 1998, 2002] suggests the feasibility of sampling long-lived sib-pair data for non-parametric linkage analysis. As the first application, Puca et al. [2001] scanned the whole genome by applying non-parametric linkage analysis to long-lived sib-pairs and reported a region on chromosome 4 that could possibly harbor a gene affecting human longevity.

Nemani et al. [2000] studied the efficiency of the ASP approach in longevity and reported that the method has high power in detecting low-frequency recessive alleles that are beneficial for survival but is inefficient in mapping dominant genes. We think, however, that their conclusions are incomplete because all their dominant genes were specified as harmful so that they reduce the probability of achieving longevity. Such situation leads to little additional allele IBD sharing in the long-lived siblings even though selection against the harmful allele is strong. As the deleterious genes are usually disease-related and since our goal here is to map the genes that contribute to (rather than impede) human longevity, we think it is necessary to investigate the power issue for the non-parametric linkage in mapping rare beneficial genetic variations. Most importantly, as the phenotype we are interested in, i.e., life span, is continuous, it requires that a proper age-cut be chosen as a threshold for defining longevity before collecting the samples to ensure sufficient statistical power with consideration of affordable sample size. In addition, the samples collected are age structured following their survival distribution with higher ages more selected and more valuable in the analysis. A proper simulation should be done in accordance with the observed population survival and age structure at high ages in order to address the above questions.

We first introduce a recently developed survival model [Yashin et al., 1999; Tan et al., 2001a] to simulate siblings' life spans for given risk and frequency parameters. By dealing with survival models, we try to make the simulation sensible to the practical situation and the results instructive to empirical studies. This is followed by the method and strategy used to generate the genotype in the pedigrees and for power assessment. After presenting the results, we discuss the significance of, and the relevant problems in, applying non-parametric linkage to longevity studies. We also elaborate on some controversial issues in the genetics of human longevity using the results from our simulation.

MATERIALS AND METHODS

POPULATION AND GENOTYPE-SPECIFIC SURVIVALS

We assume that there is a biallelic locus hosting one causal allele. Then there are 3 genotypes in the population corresponding to individuals carrying 2 (homozygous), 1 (heterozygous), and 0 copies of the causal allele. Their genotype-specific probabilities of survival at age x are $s_2(x)$, $s_1(x)$ and $s_0(x)$, respectively. Under Hardy-Weinberg Equilibrium, the mean population survival at age x is the sum of the weighted genotype-specific survivals of the 3 genotypes,

$$\bar{s}(x) = p^2 s_2(x) + 2p(1-p)s_1(x) + (1-p)^2 s_0(x) \quad (1)$$

where p is frequency of the causal allele. For any individual, defining the hazard of death at age x (the risk of dying after surviving to age x) as $\mu(x) = r_i z \mu_0(x)$, where r_i is the relative risk for genotype i ($i = 0, 1$ or 2), z is an unobserved individual frailty that also affects survival, and $\mu_0(x)$ is the baseline hazard function for an individual with both frailty and genotype relative risk set to 1. The corresponding individual survival function is $s(x) = e^{-r_i z \int_0^x \mu_0(s)ds}$. To describe the relationship of the three genotype-specific survival functions, Yashin et al. [1999] introduced a model that assumes that the unobserved frailty follows a gamma-distribution with mean 1 and variance σ^2 [Vaupel et al., 1979; Vaupel and Yashin, 1985], so that mean survival for the subpopulation carrying

genotype i is

$$s_i(x) = E[s(x)] = E\left[e^{-r_i z \int_0^\kappa \mu_0(s)ds}\right]$$

$$= [1 - r_i\sigma^2 \ln(s_0(x))]^{\sigma^{-2}}. \quad (2)$$

Here $s_0(x)$ is the baseline survival corresponding to $\mu_0(x)$. Assuming risk of the allele is additive on the log hazard of death, we have $r_i = r^i$ when r is the relative risk for the heterozygous genotype [Yashin et al., 2000; Tan et al., 2002b; Ewbank, 2002]. σ^2 has been estimated when applying the above model to genotype data in marker-longevity association studies [Yashin et al., 2000; Tan et al., 2001a; Ewbank, 2002] but with large variations due to small sample sizes. A higher σ^2 is expected to bring about larger variation in the unobserved frailty and thus more uncertainty in the individuals' life spans, even if they carry the same genotype. We take the Danish period life table for 1990 (Befolkningens bevægelser 1990) as the population survival in expression (1). This period life table is used to approach the survival rate at old ages for our samples because their cohort life table is unavailable before they die. With the observed population survival function and the given allelic risk, frequency, and heterogeneity parameters, expression (1) can be solved numerically to get the baseline survival function $s_0(x)$. Once individual genotype at the locus is specified, expression (2) can be used to generate life spans for carriers of the different genotypes.

The genetic variation has been estimated to explain about 25% of the total human life span variation using the well-known Danish twin data [Herskind et al., 1996]. In order to show the effects on survival for the different genes in our simulation, we calculate, for each of the genes, a percentage of life span variation due to the gene σ_G^2 in the total life span variation σ_T^2 in the population,

age x for genotype i corresponding to $s_i(x)$ in expression (2) and e_0 is the life span expectancy at birth for the total population, which can be calculated from the observed life table data. In expression (3), heritability is a function of the risk, frequency, and effect of the gene (dominant or recessive). Expression (3) can be used as an overall measurement to compare effects of the genes with different risk and frequency parameters under different modes of gene action as shown later in the text.

DATA GENERATION

To simulate the data, we assume that we have 400 highly polymorphic markers (10 alleles for each marker with equal frequency) evenly distributed along the 22 autosomal chromosome pairs. We only take one chromosome with its corresponding marker map for the simulation. In addition, one extra biallelic locus is put in the middle of the simulated chromosome between two markers. With the marker distribution and the allele frequencies (including the biallelic locus), we use Allegro1.1 [Gudbjartsson et al., 2000] to simulate the pedigrees (each with 2 siblings). Individual life spans are simulated using the above mentioned method for a given genotype at the causal locus. We then select from them only sib-pairs who both survived to above a critical age-cut and form the pedigree files by dropping the genotype at the causal locus. Based on knowledge from modeling large population survival, we set σ^2 to 0.1 and 0.2 in our simulations (Table I).

POWER ASSESSMENT

We performed multipoint non-parametric linkage analysis on the data again using Allegro1.1 and record the NPL scores [Kruglyak et al., 1996] for power assessment. The peaks for NPL scores at

$$h^2 = \frac{\sigma_G^2}{\sigma_T^2} = \frac{\sum_{i=0}^{2}\binom{2}{i}p^i(1-p)^{2-i}\left(\int_0^\infty s(x)dx\right)^2 - e_0^2}{\int_0^\infty \sum_{i=0}^{2}\binom{2}{i}p^i(1-p)^{2-i}\mu_i(x)s_i(x)x^2dx - \left(\int_0^\infty \sum_{i=0}^{2}\binom{2}{i}p^i(1-p)^{2-i}\mu_i(x)s_i(x)xdx\right)^2}. \quad (3)$$

Expression (3) is simply derived using the density distribution function of life span for the different genotypes. It represents the percentage of life span variation explained by a specific gene, or heritability. Here, $\mu_i(x)$ is the hazard function at

the region harboring the causal allele (one marker distance, 8.7 cM, around the two adjacent markers) are counted to calculate the statistical power. We take an NPL score of 4.2 as the genome-wide significance threshold [Lander and Kruglyak,

TABLE I. Power for different settings[a]

Setting	NPL score ≥3 (pairs)			NPL score ≥4.2 (pairs)		
	300	600	1,000	300	600	1,000
Dominant						
r=0.35, P=0.10, $h^2 \approx 0.06$						
95/90	0.75	1.00	1.00	0.13	0.70	0.99
95/90[b]	0.53	0.88	1.00	0.05	0.47	0.91
r=0.50, P=0.10, $h^2 \approx 0.03$						
95/90	0.14	0.42	0.84	0.00	0.04	0.21
99/90	0.46	0.83	0.94	0.05	0.29	0.83
r=0.65, P=0.10, $h^2 \approx 0.01$						
95/90	0.01	0.03	0.08	0.00	0.00	0.00
99/90	0.05	0.08	0.16	0.00	0.01	0.02
Recessive						
r=0.35, P=0.25, $h^2 \approx 0.06$						
95/90	0.99	1.00	1.00	0.83	1.00	1.00
r=0.50, P=0.25, $h^2 \approx 0.03$						
95/90	0.58	0.95	1.00	0.17	0.55	0.95
r=0.65, P=0.25, $h^2 \approx 0.02$						
95/90	0.08	0.17	0.37	0.00	0.04	0.09
99/90	0.29	0.79	0.97	0.04	0.20	0.60

[a]Each cell is calculated from 100 replicates, σ^2 is set to 0.1.
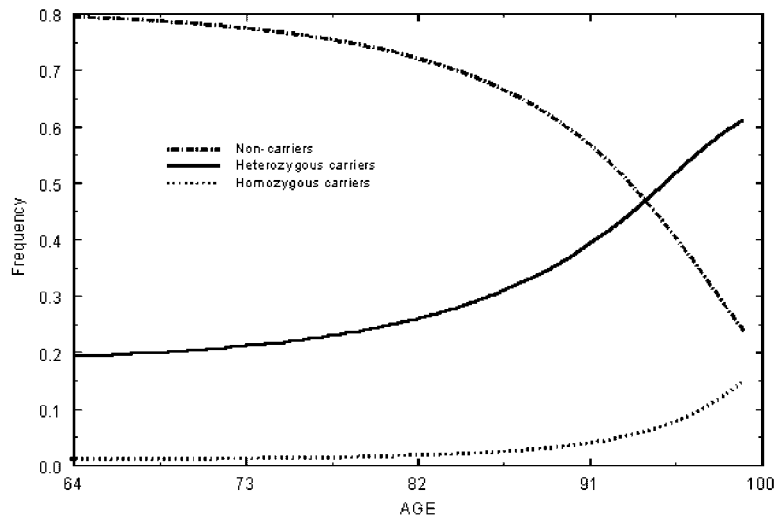[b]$\sigma^2 = 0.2$.

Fig. 1. Genotype specific frequency by age for heterozygous (solid line), homozygous (dotted line) and non-carrier (dash-dotted line) genotypes of the causal allele. As frequency for the non-carriers of the allele decreases with advancing age, that for the carriers of the allele increases as a result of survival selection. However, the frequency pattern at early ages is relatively stable because mortality, the force that drives the selection, is low.

1995] and pick up peaks with NPL scores above 3 as indicators of linkage. The statistical power is evaluated for different sample sizes (300, 600, and 1,000 pairs) and at different age cuts: 95/90 (one sib above age 95, the other above age 90), 99/90 (one sib above age 99, the other above age 90). The relative risk for the causal allele, r, is set to 0.35, 0.5, and 0.65, corresponding to 65, 50, and 35%

reduction in the hazard of death. For dominant alleles, allele frequency is set to 0.1 to simulate relatively rare and dominant genetic variations which affect longevity. Figure 1 is the genotype-specific frequencies by age at a biallelic locus hosting a dominant allele with frequency of 0.1 and relative risk of 0.5. Figure 1 shows that frequencies of the 3 genotypes change as a result

of survival selection at advanced ages (note the drastic change and intersection after age 90) but keep stable at younger ages because mortality, the force driving the selection, is low. Frequency of the recessive alleles is adjusted to 0.25 so that the heritability (h^2) can be comparable to that obtained for the dominant alleles with the same risk and heterogeneity parameters.

RESULTS

In Figure 2, we show the distribution of the estimated NPL scores at the region that harbors a dominant allele with a frequency of 0.1 and a relative risk of 0.5 using different age-cut and a sample size of 1,000 long-lived sib-pairs. The median for each 100 NPL scores by each age-cut shows an increasing trend due to increased IBD sharing at the extreme ages. The results from both dominant and recessive models are shown in Table I. For a strong dominant allele with relative risk of 0.35, which reduces the hazard of death by 65% at all ages, the power is very high when setting an NPL score of 3 as a threshold. A rare dominant allele that reduces a carrier's hazard of death by half can be mapped by a large sample size of 1,000 pairs with at least one sib above age 99 in each pair. However, a sample size of around 600 pairs can have sufficient power in mapping the gene if the critical NPL score is set to 3 to trade for a higher false-positive rate. The ASP approach has very low power in mapping rare dominant genetic variations with small effect (allelic risk of 0.65), even for the large samples.

The second row in Table I is the power for the same setting as the first row but with doubled variance of the unobserved frailty ($\sigma^2 = .2$). As a result, the power is clearly dropped. This is sensible because with increased heterogeneity, individuals carrying the beneficial allele have increased uncertainty in attaining longevity due to more competing risk factors that underline the unobserved frailty.

It is interesting to see that the statistical power is higher in all the settings for the recessive alleles, with heritability comparable to the dominant settings in Table I. The power can be acceptable even for a small effect allele using a large sample of over 1,000 pairs with an age-cut of 90/99, which is encouraging. The ASP approach is very powerful in mapping strong effect recessive alleles and has an acceptable power for modest effect alleles (allelic risk of 0.5), even if only a small sample size of low age-cut (90/95) is available.

In Table II, we show the power of nonparametric linkage by just single marker analysis when assuming complete information for linkage
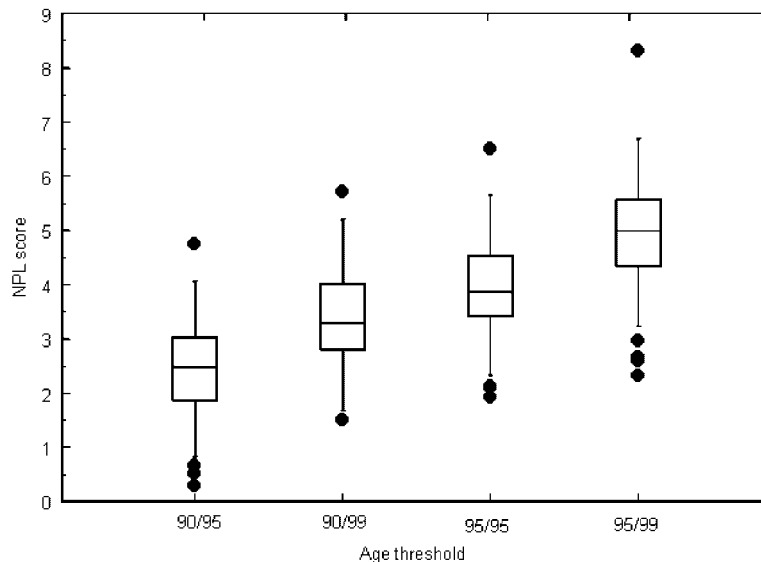

Fig. 2. Distribution of the estimated NPL scores obtained by different age-cut at the region that harbors a dominant allele with frequency of 0.1 and relative risk of 0.5 using a sample size of 1,000 long-lived sib-pairs. The median for each 100 NPL scores by each age-cut shows an increasing trend due to increased IBD sharing at the extreme ages. The step is set as one times the H-spread. Black circles: The end NPL values. To save computer time, we used the Danish 2002 period life table (Statistisk Årbog 2002) for the population survival. As the late-age survival is higher than that in the old cohort population, power applicable to current studies cannot be assessed using this data.

TABLE II. Power for different settings using a fully informative marker at the causal locus[a]

Setting	NPL score ≥ 3 (pairs)			NPL score ≥ 4.2 (pairs)		
	300	600	1,000	300	600	1,000
Dominant						
$r=0.50$, $P=0.10$, $h^2 \approx 0.03$						
95/90	0.27	0.65	0.94	0.04	0.21	0.62
99/90	0.61	0.96	1.00	0.21	0.68	0.98
$r=0.65$, $P=0.10$, $h^2 \approx 0.01$						
95/90	0.02	0.03	0.11	0.00	0.00	0.00
99/90	0.10	0.21	0.46	0.00	0.03	0.11
Recessive						
$r=0.50$, $P=0.25$, $h^2 \approx 0.03$						
95/90	0.69	0.97	1.00	0.29	0.86	1.00
$r=0.65$, $P=0.25$, $h^2 \approx 0.02$						
95/90	0.07	0.28	0.52	0.00	0.01	0.14
99/90	0.56	0.87	0.98	0.11	0.51	0.88

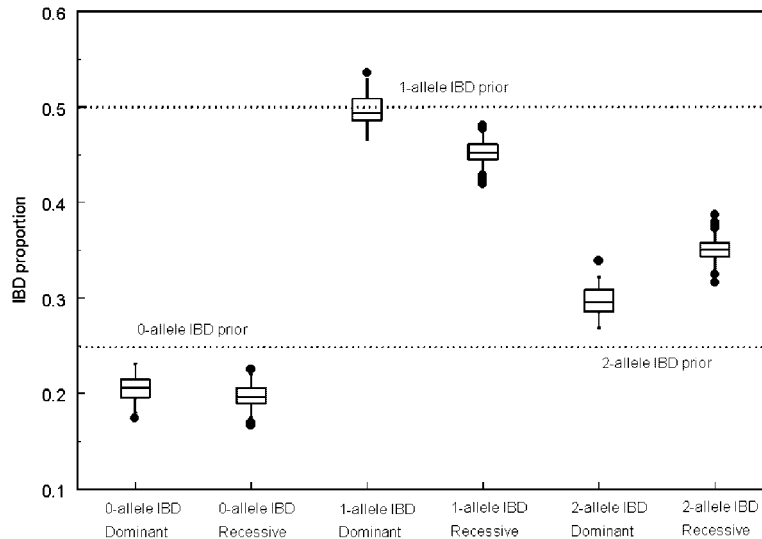[a]Each cell is calculated from 100 replicates, σ^2 is set to 0.1.

Fig. 3. Box-plot showing the distribution of mean IBD sharing at the gene locus with full information content. The step is set as one times the H-spread. Each mean is estimated from 100 replicates containing 1,000 long-lived sib-pairs over age-cut 90/95. Significant deviations from the prior IBD sharing proportion when assuming no linkage are observed for 2-allele (increased) and 0-allele (decreased) IBDs in both the dominant and recessive models and for 1-allele (decreased) IBD in the recessive model.

is available at the causal locus. Although the general trend is the same as in Table I, we observe increased power for all the settings. However, even with increased power, the affected sib-pair method is unable to map a dominant gene with relatively small effect ($r=0.65$, $p=0.10$). This is in contrast with the recessive case, for which the power is 0.88 when 1,000 long-lived sib-pairs over ages 99/90 are collected. In Figure 3, we show the distribution of mean IBD sharing at the gene locus. Each mean is estimated from 100 replicates containing 1,000 long-lived sib-pairs over age-cut 90/95. Significant deviations from the prior IBD sharing proportion when assuming no linkage are observed for 2-allele IBD sharing (increased) and 0-allele IBD sharing (decreased) in both the dominant and recessive models and for 1-allele IBD sharing (decreased) in the recessive model.

DISCUSSION

Our results suggest that the IBD sharing based ASP approach can be a powerful tool for mapping

large effect dominant or recessive genes that contribute to human longevity. When a large sample size with a high age-cut (90/99) is available, it has acceptable power to localize modest effect dominant genes or even small effect recessive genes. This method is unlikely to be able to detect rare dominant genetic variations with small effect even if the sample size is very large. Table I also shows that the results from small-scale studies could only be trusted if major genes that contribute to longevity do exist. We have tried to estimate the number of genes that contribute to human longevity by decomposing variances in life span using a survival model. Our results showed that the estimated genetic component from the family and twin data can be equally matched by the contribution from a couple of very strong effect genes or by that from several thousand small effect genes [Vaupel and Tan, 1998; Begun et al., 2000]. If the genetic component to longevity is due to one or a few genes [Gudmundsson et al., 2000], according to our simulation, the non-parametric method could be a useful tool to help mapping them. On the other hand, if multiple genes with only small effects are involved [Cournil and Kirkwood, 2001; Carnes and Olshansky, 2001], one will have the power problem, especially when only limited samples are available. Results from the candidate gene approach using marker-longevity association have suggested a high likelihood of the polygenic nature of longevity [De Benedictis et al., 2001]. Likewise, marker-longevity association studies using the above-mentioned survival model have not yet detected genetic variations that show strong effect matchable to that used in our simulation. However, one cannot rule out the possibility of an interplay of a handful of major genes plus many genes with minor effects.

As shown in Table I, heterogeneity has a profound effect on the power of a non-parametric approach. Several authors have tried to estimate the parameters using a gamma frailty model [Yashin et al., 2000; Tan et al., 2001a; Ewbank, 2002] in fitting survival functions to individual genotype data but found that the estimated variance of the unobserved frailty varies considerably in different studies, perhaps due to small sample sizes used. Based on our experience in fitting survival distributions to large population data, we set σ^2 to 0.1 in our simulation to ensure a reliable result. However, as just mentioned, since

such estimation has not yet been proved by large-scale genotype data, it could be either an optimistic or a pessimistic one. Another significance of heterogeneity is relevant to association studies. Given the existence of competing risk factors or heterogeneity, the risk for any particularly observed factor (including genetic factors) will tend to decline [Ewbank, 2002]. In our simulation, we set the risk of the allele constant at the individual level. Accordingly, the group mean of the risk should decline with advancing age because of the introduction of heterogeneity. On the other hand, we found in our linkage approach that the IBD sharing of the beneficial allele is increasing in the long-lived sib-pairs (Fig. 2) indicating a growing importance of the allele in maintaining survival. However, one should not conclude that the genetic influence is increasing at extreme ages [Perls et al., 2002], as the phenomenon is only the result of a proportional hazard model.

High age-cut is essential for the collected sample to merit a longevity study. However, raising the age-cut also adds difficulty to the sample collection, a situation one always has to balance. The two age-cuts in our simulation are selected with consideration of the sample collection as extremely long-lived sib pairs are very rare. Fortunately, our results show that regions harboring important genes can be localized using samples with lowest critical age set to 90.

Our results in Table II show that even with complete information content, the non-parametric linkage applied to long-lived sib-pairs is incapable of mapping dominant genes that account for a small proportion of life span variation ($h^2 \approx 0.01$). Although we are dealing with survival traits, our result is consistent with that from a power study on QTL mapping for sib-pair data by Sham et al. [2000] who suggested that a genome scan using association instead of linkage be performed to detect QTLs that account for only 1% of the phenotypic variance. In order to improve the power, a two-step strategy can be adopted by which promising genome regions are identified using lower level critical NPL scores in an initial linkage scan and then linkage disequilibrium mapping using very dense markers can be conducted within these regions.

Gerdes et al. [2000] estimated genotype-specific relative risk for APOE gene and concluded that the gene is a frailty gene, rather than

a "longevity gene." As we have mentioned at the beginning, if one genetic variation is harmful to longevity, survival selection will not bring about additional IBD sharing at the locus. In this case, the IBD-based non-parametric linkage approach is unlikely to be an efficient method for mapping the gene. We hope this helps to explain the negative result in mapping the APOE gene using long-lived sib-pair data [Nemani et al., 2000]. On the contrary, such genes can be mapped by using sib-pairs affected by relevant genetic disorders but not these long-lived siblings who are enjoying their lives.

ACKNOWLEDGMENTS

REFERENCES

Begun AZ, Iachine IA, Yashin AI. 2000. Genetic nature of individual frailty: comparison of two approaches. Twin Res 3:51–57.

Cardon LR, Bell JI. 2001. Association study designs for complex diseases. Nat Rev Genet 2:91–99.

Carnes BA, Olshansky SJ. 2001. Heterogeneity and its biodemographic implications for longevity and mortality. Exp Gerontol 36:419–430.

Cournil A, Kirkwood TB. 2001. If you would live long, choose your parents well. Trends Genet 17:233–235.

De Benedictis G, Tan Q, Jeune B, Christensen K, Ukraintseva SV, Bonafe M., Franceschi C, Vaupel JW, Yashin AI. 2001. Recent advances in human gene-longevity association studies. Mech Ageing Dev 1:909–920.

Ewbank DC. 2002. Mortality differences by APOE genotype estimated from demographic synthesis. Genet Epidemiol 22:146–155.

Gerdes LU, Jeune B, Ranberg KA, Nybo H., Vaupel JW. 2000. Estimation of apolipoprotein E genotype-specific relative mortality risks from the distribution of genotypes in centenarians and middle-aged men: apolipoprotein E gene is a "frailty gene," not a "longevity gene". Genet Epidemiol 19:202–210.

Gudbjartsson DF, Jonasson K, Frigge ML, Kong A. 2000. Allegro, a new computer program for multipoint linkage analysis. Nat Genet 25:12–13.

Gudmundsson H, Gudbjartsson DF, Frigge M, Gulcher JR, Stefansson K. 2000. Inheritance of human longevity in Iceland. Eur J Hum Genet 8:743–749.

Herskind AM, McGue M, Holm NV, Sorensen TI, Harvald B, Vaupel JW. 1996. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870–1900. Hum Genet 97:319–323.

Kruglyak L, Lander ES. 1995. Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet 57:439–454.

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. 1996. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363.

Lander E, Kruglyak L. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247.

Nemani M, Sahbatou M, Blanche H, Thomas G, Pascoe L. 2000. The efficiency of genetic analysis of DNA from aged siblings to detect chromosomal regions implicated in longevity. Mech Ageing Dev 119:25–39.

Ott J. 1999. Analysis of human genetic linkage, 3rd ed. Baltimore: John Hopkins University Press.

Perls TT, Bubrick E, Wager CG, Vijg J, Kruglyak L. 1998. Siblings of centenarians live longer. Lancet 23;351:1560.

Perls TT, Wilmoth J, Levenson R, Drinkwater M, Cohen M, Bogan H, Joyce E, Brewster S, Kunkel L, Puca A. 2002. Life-long sustained mortality advantage of siblings of centenarians. Proc Natl Acad Sci USA 99:8442–8447.

Pletcher, SD & Stumpf, MP. 2002. Population genomics: ageing by association. Curr Biol 12:328–330.

Puca AA, Daly MJ, Brewster SJ, Matise TC, Barrett J, Shea-Drinkwater M, Kang S, Joyce E, Nicoli J, Benson E, Kunkel LM, Perls T. 2001. A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4. Proc Natl Acad Sci USA 98:10505–10508.

Risch NJ. 2000. Searching for genetic determinants in the new millennium. Nature. 405:847–856.

Sham PC, Lin MW, Zhao JH, Curtis D. 2000 Power comparison of parametric and nonparametric linkage tests in small pedigrees. Am J Hum Genet 66:1661–1668.

Tan Q, De Benedictis G, Yashin AI, Bonafe M, DeLuca M, Valensin S, Vaupel JW, Franceschi C. 2001a. Measuring the genetic influence in modulating human life span: Gene-environment and gene-sex interactions. Biogerontology 2:141–153.

Tan Q, Yashin AI, De Benedictis G, Cintolesi F, Rose G, Bonafe M, Franceschi C, Vach W, Vaupel JW. 2001b. A logistic regression model for measuring gene-longevity associations. Clin Genet 60:463–469.

Tan Q, De Benedictis G, Ukraintseva SV, Franceschi C, Vaupel JW, Yashin AI. 2002a. A centenarian-only approach for assessing gene-gene interaction in human longevity. Eur J Hum Genet 10:119–124.

Tan Q, Bellizzi D, Rose G, Garasto S, Franceschi C, Kruse T, Vaupel J, De Benedictis G, Yashin AI. 2002b. The influences on human longevity by HUMTHO1.STR polymorphism (Tyrosine Hydroxylase gene). A relative risk approach. Mech Ageing Dev 123:1403–1410.

Tan Q, Bathum L, Christiansen L, De Benedictis G, Dahlgaard J, Frizner N, Vach W, Vaupel JW, Yashin AI, Christensen K, Kruse TA. 2003. Logistic regression models for polymorphic and antagonistic pleiotropic gene action on human aging and longevity. Ann Hum Genet 67:598–607.

Vauple JW, Tan Q. 1998. How many longevity genes are there? Presented at annual meeting of Population Association of America, Chicago.

Vaupel JW, Yashin AI. 1985. Heterogeneity's ruses: some surprising effects of selection on population dynamics. Am Stat 39:176–185.

Vaupel JW, Manton KG, Stallard E. 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16:439–454.

Yashin AI, De Benedictis G, Vaupel JW, Tan Q, Andreev KF, Iachine IA, Bonafe M, DeLuca M, Valensin S, Carotenuto L, Franceschi C. 1999. Genes, demography, and life span: The contribution of demographic data in genetic studies on aging and longevity. Am J Hum Genet 65:1178–1193.

Yashin AI, De Benedictis G, Vaupel JW, Tan Q, Andreev KF, Iachine IA, Bonafe M, DeLuca M, Valensin S, Carotenuto L, Franceschi C. 2000. Genes and longevity: Lessons from studies on centenarians. J Gerontol 55a:B1–B10.