Short Report

A logistic regression model for measuring gene–longevity associations

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

The logistic regression model is a popular model for data analysis in epidemiological research. In this paper, we use this model to analyze genetic data collected from gene–longevity association studies. This new approach models the probability of observing one genotype as a function of the age of investigated individuals. Applying the model to genotype data on the TH and 3′ApoB-VNTR loci collected from an Italian centenarian study, we show how it can be used to model the different ways that genes affect survival, including sex- and age-specific influences. We highlight the advantages of this application over other available models. The application of the model to empirical data indicates that it is an efficient and easily applicable approach for determining the influences of genes on human longevity.

Q Tan[a], AI Yashin[a,b], G De Benedictis[c], F Cintolesi[d], G Rose[c], M Bonafe[e], C Franceschi[e], W Vach[f] and JW Vaupel[a]

[a] Max Planck Institute for Demographic Research, Rostock, Germany, [b] Center for Demographic Studies and Sanford Institute, Duke University, Durham, NC, USA, [c] Cell Biology Department, University of Calabria, Rende, Italy, [d] Physical and Theoretical Chemistry Laboratory, University of Oxford, UK, [e] Department of Experimental Pathology, University of Bologna, Italy, [f] Department of Statistics and Demography, University of Southern Denmark, Denmark

Key words: gene – logistic regression – longevity

Corresponding author: Prof. Anatoli I Yashin, Max Planck Institute for Demographic Research, Doberaner Str. 114, 18057 Rostock, Germany. Tel.: +49 381 2081106; fax: +49 381 2081169; e-mail. yashin@demogr.mpg.de

Received 9 August 2001, revised and accepted for publication 28 September 2001

Over the last few decades, interest in studying the association between genes and human longevity has grown (1). There have been reports on important genes that contribute to the process of human aging and longevity, such as the ApoE gene (2–6). As for statistical methods, most of the studies are based on a case–control design by which allelic or genotypic frequencies are compared between cases (usually centenarians) and controls (younger people from the same population) to see if significant differences exist between the allelic or genotypic pools. A new procedure that combines individual genotypic information with demographic information has recently been proposed and applied to data collected from Italian (7–10) and Danish (11) centenarian studies. Based on the proportional hazard assumption, the new procedure can, not only estimate the relative risk of the gene or genotype, but can also incorporate gene–environment

(9, 10) and gene–sex (11) interactions that contribute to the modulation of individual survival. In this new procedure, subjects are no longer divided into cases and controls, since individual survival information is fully utilized in the parameter estimation, thereby increasing the efficiency of the model (10, 11). However, users of the new procedure require training in computer programming and mathematical statistics in order to ensure a proper application. It is important to note that the new procedure is not capable of modeling genotype frequency that changes non-monotonously with age, since it is limited by its proportional hazard assumption.

The logistic regression model is popular in epidemiological studies since it can model dichotomous-dependent variables as a function of a set of continuous or categorical predictor variables. In this paper, we shall explore the feasibility of apply-

ing the logistic regression model to estimate the probability of observing one genotype as a function of age, assuming that frequency of the genotype that affects individual survival should change in the genotypic pool with advancing age. Empirical data taken from an Italian centenarian study (12, 13) will be used to show how the model can be implemented to cope with different situations concerning the observed age-related pattern of the genotype frequency. In the Discussion section we compare this model with other models that are in use in gene–longevity association studies and highlight the important features of the present application.

Materials and methods

Data

The individual genotype information is taken from a multicentric longevity study that was started in 1995 in Italy. A total of 12 polymorphic loci were analyzed (10, 12, 13). In this paper, we select two highly polymorphic markers, HUMTHO.1-STR at the tyrosine hydroxylase (TH) locus and 3'APOB-VNTR at the apolipoprotein B (APOB) locus, in order to demonstrate how the model can be manipulated to deal with different patterns of gene action. The HUMTHO.1-STR dataset comprised 555 genotypic records from two groups of subjects: 197 centenarians and 358 individuals aged 10–84. The 3'APOB-VNTR dataset comprised 787 genotypic records from two groups of subjects: 190 centenarians and 597 individuals aged 10–84. Details on the criteria of recruitment and on the molecular genotyping procedures are provided in De Benedictis et al. (12–14).

The HUMTHO.1-STR polymorphism included six alleles varying from 6 to 11 repeats of the (*AAGT*) core sequence (alleles 6, 7, 8, 9, 10, 11). The 3'APOB-VNTR polymorphism included 15 alleles varying from 26 to 55 repeats of a dimeric AT-rich core sequence (basic repeat unit of 15 bp). According to size and frequency patterns (13), 3'APOB-VNTR alleles were grouped into three categories, Small (S; less than 35 repeats), Medium (M; 35–39 repeats), and Large (L; more than 39 repeats).

Method

The basic model. If a genotype is associated with individual survival, the probability of observing it in a population in Hardy–Weinberg equilibrium should change with increasing age as a result of survival selection. In this case, we can introduce the logistic regression to model the probability of

observing a genotype in the population (the genotype frequency) as a function of the participants' ages. That is,

$$Pr(G = 1) = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 x)]} \qquad (1)$$

In [1], G denotes the genotype in question. For an individual carrying the genotype $G = 1$, otherwise $G = 0$. β_0 and β_1 are parameters to be estimated, x is the age of the subject at the time of the study. In [1], the gene longevity association is determined by testing H_0: $\beta_1 = 0$. Once the coefficients in [1], β_0 and β_1, are estimated one can calculate the probability that carriers of the genotype can be found in a population aged x, i.e., the genotype frequency at age x. If the gene does not affect individual survival, i.e., $\beta_1 = 0$, then [1] becomes

$$Pr(G = 1) = \frac{1}{1 + \exp(-\beta_0)} \qquad (2)$$

a constant that represents the initial frequency of the gene and that does not depend on x. [2] can be used to calculate the gene frequency at birth since it is a special case of [1] when x is zero. When β_0 is zero in [2], $Pr(G = 1) = 0.5$. This means that a test of the constant $\beta_0 = 0$ is equivalent to a test of $Pr(G = 1) = 0.5$. Thus, the significant test on the constant usually ignored by most traditional regression analysis is of some special meaning in this application. Rewriting [1] in the logit form and substituting $Pr(G = 1)$ with p for simplicity, we have

$$\ln \frac{p}{1 - p} = \beta_0 + \beta_1 x \qquad (3)$$

In [3], a significantly positive β_1 increases the probability of observing carriers of the genotype as age x increases. Likewise, a significantly negative β_1 reduces such probability with advancing age x.

Modeling sex-specific effect. By specifying two different β_1s for males ($\beta_{1,m}$) and for females ($\beta_{1,f}$), one can try to find sex-specific effects or gene–sex interactions for the gene of interest. When we are dealing with autosomal genes, β_0 should be the same for both sexes since it represents gene frequency at birth. Then we have

$$\ln \frac{p}{1 - p} = \beta_0 + \beta_{1,m} x U + \beta_{1,f} x(1 - U) \qquad (4)$$

In [4], U is an indicator for sex, $U = 1$ for males and 0 for females. After fitting the model, one needs to check if the two sex-specific parameters, $\beta_{1,m}$ and $\beta_{1,f}$ are statistically different or not. This can be done by comparing the parameter estimates with a consideration of their standard errors.

When $\beta_{1,m}$ and $\beta_{1,f}$ are significantly different, we know that the effect of the gene is sex-dependent. Otherwise, we say that, as revealed by the data, there is not enough evidence to show that the gene has a sex-specific influence on survival. In this case, we can simply combine data for the two sexes and fit [3].

Modeling age-specific effect. In case of a non-linear relationship in [3], new function forms for x can be added to the left-hand side of the equation. Since only one independent variable is considered, age x, the non-linear relationship can be approached by a polynomial model, in that [3] can be rewritten as

$$\ln \frac{p}{1-p} = \beta_0 + \sum_{i=1}^{k} \beta_i x^i \qquad (5)$$

However, the k-order polynomial model in [5] means that there will be $k + 1$ parameters to be estimated. When sample size is small, an effort needs to be made to limit the number of parameters. In our application, we add to [3] a new term with coefficient β_2 and non-linear transformation of x, i.e.

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 x + \beta_2 x^k \qquad (6)$$

In [6], k transforms the variable x to satisfy the non-linear relationship. A proper k has to be chosen to ensure the maximum likelihood of the observed data. One has to notice that when such a k is chosen, the remaining parameters are then maximized via a maximum likelihood estimate (MLE). In this case, the standard errors on the remaining parameters would be underestimated. One needs to be careful when making a conclusion on a parameter concerning its significance level. In principle, all of the parameters should be estimated in the likelihood framework, especially when a large sample size is available.

By examining the statistical significant level of β_2, one can decide if the model with an age-specific effect is necessary. When k is an MLE, the likelihood ratio tests with 2 degrees of freedom could also be used to determine if the model with age-specific effects is appropriate, since as one can see model [6] and model [3] are fully nested.

The odds ratio. An important parameter in the logistic regression is the odds ratio, which provides information about the relationship of the predictor variable to the dependent variable. Note that the left-hand sides of [3] and [6] are natural logs of the odds $p/(1-p)$. The odds by definition are the ratio of the probability of observing the genotype divided by the probability that the genotype is not observed. In [3], the odds ratio of age x to age $x - 1$ can be calculated as

$$OR_{x/x-1} = \frac{e^{\beta_0 + \beta_1 x}}{e^{\beta_0 + \beta_1(x-1)}} = e^{\beta_1} \qquad (7)$$

The odds ratio calculated from [7] is independent of age x, which means that the genotype has a constant influence on life span over all ages. This is similar to the situation with the proportional hazard model, which assumes that the relative risk of one observed genotype is proportional to the baseline hazard function (7–10). In the same manner, we can calculate the odds ratio for [6] as

$$OR_{x/x-1} = \frac{e^{\beta_0 + \beta_1 x + \beta_2 x^k}}{e^{\beta_0 + \beta_1(x-1) + \beta_2(x-1)^k}} = e^{\beta_1 + \beta_2(x^k - (x-1)^k)} \qquad (8)$$

This time it depends not only on the coefficient β_1, but it is also a function of age x. Such a relationship is important since there is evidence in longevity studies indicating that some genes can function differently at different ages (2, 13). The age-specific gene action is in accordance with the evolutionary theory of aging, cf. antagonistic pleiotropy. With the relationship in [8] it is possible to model the non-monotonous age dependence of genotype frequency. This important feature of the application will be demonstrated later in data analysis of the ApoB gene.

We used standard SPSS9.0 software for the logistic regression and Axum5.0 (15) for graphic presentations.

Results

Before fitting the model, we plotted genotype frequency by age to see if a non-monotonous pattern of age dependence exits. This can help to decide whether model [3] or model [6] is the proper choice. Grouping individuals according to their age at the time of participation and calculating the gene frequency for each group, no extraordinary pattern was found for polymophisms at the TH locus. This indicates that a linear model could be applied. We first fit [4] to the data in order to see if there is any sex-dependent effect from each of the alleles. In fitting the model we define the event as a carrier of the allele. The results in Table 1 show that there is only one allele, TH10, with significant influence on females (p = 0.002) but not on males (p = 0.288). However, when we compare $\beta_{1,f}$ with $\beta_{1,m}$ from the fit, they are not statistically different. At this point, we cannot conclude that there is a sex-specific effect from this allele. In Table 1, we only show the estimates with standard errors for β_0 since it makes no sense to test its statistical signifi-

Table 1. Parameter estimates for TH gene allele carriers by sex

Allele	β_0	SE	$\beta_{1,m}$	SE	p	$\beta_{1,f}$	SE	p
THO6	−0.132	0.175	0.001	0.003	0.680	−0.001	0.003	0.715
THO7	−0.536	0.185	−0.004	0.004	0.315	−0.004	0.003	0.191
THO8	−1.384	0.215	0.005	0.004	0.218	−0.001	0.003	0.728
THO9	−0.212	0.178	−0.005	0.003	0.184	−0.005	0.003	0.057
THO10	−0.867	0.185	0.004	0.003	0.288	0.008	0.003	0.002
THO11	−2.788	0.398	−0.009	0.009	0.336	−0.001	0.006	0.883

cance. Knowing there is no sex-dependent effect, we then fit [3] to the combined data. The results are presented in Table 2. Again, we see that only allele 10 manifests a significant affect that favors the carrier's survival (p = 0.006). By redefining the event as homozygote of one allele, we can use [3] to detect if there is recessive effect from that allele. However, before fitting the model, one has to make sure that the corresponding heterozygous genotype has no effect. This is true for all the other alleles except TH10 in Tables 1 and 2. We fitted [3] to alleles 6, 7, 8, 9, and 11 homozygotes but only allele 9 showed $\beta_1 = -0.020$ with a p value of 0.024, which means allele 9 homozygote tends to be an unfavorable genotype. Unfortunately, when multiple comparisons are considered, this p value is beyond significance.

According to [7], the odds ratio for two adjacent ages can be calculated as $e^{7.2e-3} = 1.007$ for allele 10. One can see that the odds ratio is only trivial since it is very close to one. However, one has to bear in mind that it only stands for an age interval of 1 year. If one considers an interval of 30 years, we then have

$$\text{OR}_{x/x-30} = \frac{e^{\beta_0 + \beta_1 x}}{e^{\beta_0 + \beta_1(x-30)}} = e^{30\beta_1}$$

in accordance with [7]. In this case, the odds ratios for allele 10 carriers will be 1.241, which represents quite remarkable changes in its frequencies.

In Table 2, we also calculated the frequencies of carriers at birth (p_0) concerning each allele using [2]. Note that it does not represent the proportion of carriers in the population as a whole when the corresponding allele is associated with survival. In this case, the carrier's proportion in a given population depends on the age structure of the population. With known frequency of carriers at birth, one can calculate the allele frequency easily, using the relationship between the frequency of carriers (p) and frequency of the allele (p'). From Table 2, we have $p = 2p'(1-p') + p'^2 = 1 - (1-p')^2$. Rearranging it, we have $p' = 1 - \sqrt{1-p}$. Consider allele 10, for example, the allele frequency is $1 - \sqrt{1-0.289} = 0.157$. By introducing the

parameter estimates into [1], one can calculate the proportion of carriers as a function of age x, as shown in Fig. 1. For comparison, we plotted the frequencies for allele 8 carriers in Fig. 1a, which does not show a significant association with longevity (Table 2). The estimated frequency is almost constant over age. In contrast, both the observed and the estimated frequencies of carriers of allele 10 increase with age in Fig. 1b.

In Tables 1 and 2, we observe only one significant allele (allele 10). One could argue that it could be a result of chance since there are many tests conducted in each table. In this case, the significance level should be adjusted using Bonferroni's correction in order to avoid false positive results. In Table 1, there are a total of 12 tests concerning the effects of the six alleles in both sexes. The new significance level is adjusted as $1 - (1 - 0.05)^{1/12} = 0.004$. The p value for allele 10 in females is 0.002, which means that it is still significant even after adjustment. In Table 2, we have conducted six tests each for one allele. Then we have $1 - (1 - 0.05)^{1/6} = 0.009$ as the new significance level. Again, we see that the p value obtained from Table 2 (0.006) is smaller than the adjusted threshold meaning a statistical significance.

An analysis of the 3′APOB-VNTR polymorphism by De Benedictis et al. (13) revealed a significant convex trajectory of the S/S genotype frequency by age. Here, we fitted model [6] to the 3′APOB-VNTR S/S genotype by choosing different values for parameter k in order to ensure the best fit of the model. The highest likelihood was achieved when k was set to 1.1 with a log likelihood of − 139.2695. The estimated coefficients are

Table 2. Parameter estimates for TH gene alleles for combined data

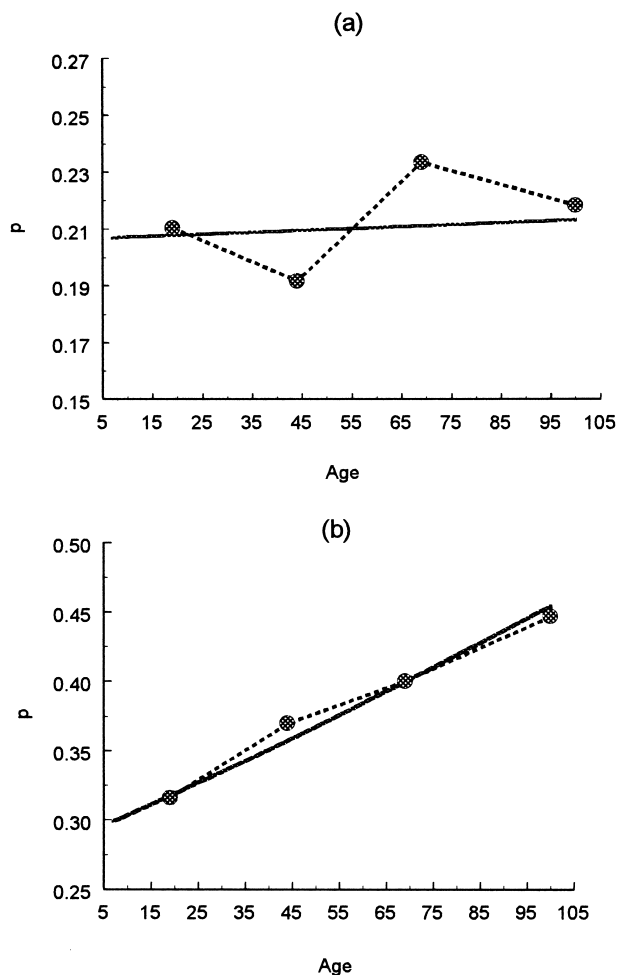| Allele	β_0	p	β_1	p	p_0
THO6	−0.031	0.841	−9.6e−5	0.403	0.492
THO7	−0.535	0.004	−3.7e−3	0.181	0.369
THO8	−1.345	0.000	4.0e−4	0.895	0.207
THO9	−0.208	0.241	−5.1e−3	0.057	0.448
THO10	−0.902	0.000	7.2e−3	0.006	0.289
THO11	−2.852	0.000	−2.1e−3	0.727	0.055

Fig. 1. The observed (dash-dotted) and the estimated (solid) frequencies for (a) TH8 and (b) TH10 carriers. The estimated genotype frequency for TH8 carriers does not show significant increase. However, there is a substantially linearly increase in the frequency for TH10 carriers.

$\beta_0 = -4.864$ (SE $= 1.226$, p $= 0.00$), $\beta_1 = 0.549$ (SE $= 0.272$, p $= 0.043$), and $\beta_2 = -0.342$ (SE $= 0.166$, p $= 0.040$). All the coefficients are significant, but the linear and non-linear parameters have opposite signs (β_1 positive and β_2 negative). In Fig. 2, we plotted the observed and estimated frequencies of the S/S genotype carriers by age. One can see that the genotype frequency increases at early ages, reaches a peak at middle age, and then begins to decrease after about the age of 50. The observed non-monotonous trajectory is well captured by the fitted convex curve.

Discussion

The application of the logistic regression model has shown that this model can be a very useful tool for analyzing genetic data for the study of human longevity. However, there are some special features of the model that must be taken into account when applying it. First, the strategy of modeling genotype frequency as a function of age requires only individual genotype and age at the time of participation, perhaps together with other covariates. The model can be applied to data collected from cross-sectional investigations. No follow-up is necessary. Second, the logistic regression model in this application is consequently merely an association test. In this situation, the result can be affected by intrapopulation heterogeneity in allele frequency (16, 17). As was the case for the case–control design, ethnic origin should be carefully controlled in the sampling process in order to avoid spurious conclusions.

A comparison of the differences among the major models that are in use in gene longevity studies is called for. Similar to the recently proposed relative risk model (7–11), the logistic regression model makes full use of individual survival information and thus achieves a higher level of efficiency than the popular gene frequency method, which relies on a simple case–control design. Like the relative risk model, no specific age concentration in the sampling is necessary, although extremely long-lived individuals are essential in order to achieve the goal of the study. Both the logistic regression and the relative risk models can estimate initial gene frequency. Instead of directly estimating a frequency parameter, however, the initial frequency in the logistic approach is calculated using [3] and setting age x to zero.

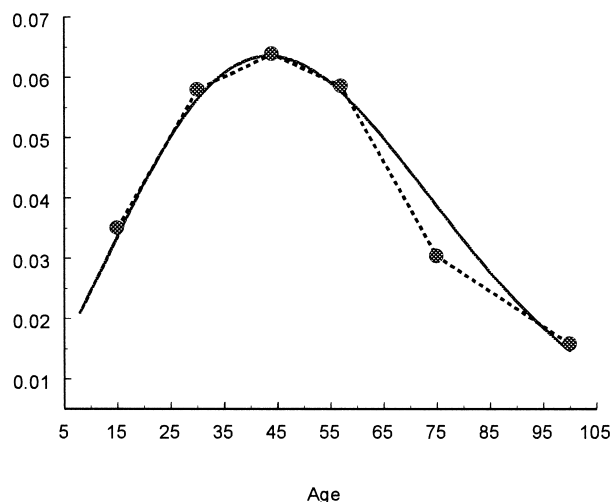Instead of modeling frequency as a dependent variable, one can approach the task the other way

Fig. 2. The observed (dash-dotted) and the estimated (solid) frequencies for 3'ApoB-VNTR S/S genotype. Instead of a monotonous pattern, the frequency for 3'ApoB-VNTR S/S genotype increases at early ages but decreases after about age 45.

around and model the age of participants as a dependent variable by assigning a value of '1' to centenarians '0' to the control group and setting genotype and other covariates as independent variables. However, as was the case for the gene frequency method based on case–control design, this approach does not make full use of the survival information since it divides individuals into two groups although life span is a continuous trait. Consequently, one cannot explore the frequency trajectory of the genotype by age and the important pattern as revealed in Fig. 2 could be missed.

Indeed, one striking advantage of the logistic regression approach is the capacity to model the non-monotonous pattern of the observed genotype frequency as illustrated in the analysis of the 3′ApoB-VNTR S/S genotype. As is shown by [7], the odds ratio estimated from the logistic approach can be compared to a risk estimate from the relative risk model. The odds ratio can, however, also be derived as a function of individual age [8], depending on the model specification. The 3′ApoB-VNTR S/S genotype in this application provides a very good example of this. In accordance with [8], the odds ratio for comparing age 45 with age 15 can be written as

$$OR_{45/15} = \frac{e^{\beta_0 + \beta_1 \times 45 + \beta_2 \times 45^{1.1}}}{e^{\beta_0 + \beta_1 \times 15 + \beta_2 \times 15^{1.1}}} = e^{\beta_1 \times 30 + \beta_2(45^{1.1} - 15^{1.1})}$$
$$= 3.373$$

Likewise, the odds ratio comparing age 100 with age 45 is

$$OR_{100/45} = \frac{e^{\beta_0 + \beta_1 \times 100 + \beta_2 \times 100^{1.1}}}{e^{\beta_0 + \beta_1 \times 45 + \beta_2 \times 45^{1.1}}} = e^{\beta_1 \times 56 + \beta_2(100^{1.1} - 45^{1.1})}$$
$$= 0.383$$

Obviously, there is more than a threefold increase in the probability of observing carriers versus non-carriers of the genotype from age 15 to 45. In contrast, there was a 260% decrease in this probability from age 45 to 100. The numbers indicate that the same genotype conveys beneficial effects at early ages only to exhibit harmful affects after the age of reproduction. The age-dependent odds ratio is of sound biological significance. The important evolutionary theory of aging based on antagonistic pleiotrapy predicts age-dependent genetic influences on survival as a result of the weak selection of late-acting deleterious genes due to the termination of reproduction. In our example, the central role played by the 3′ApoB-VNTR gene in cell cholesterol homeostasis (which is functional in membrane synthesis as well as in steroid hormonogenesis) may explain the age-related effect of the 3′ApoB-VNTR locus on survival.

One has to be aware that when the model is applied to cross-sectional data, the effect of certain genotype could be overestimated due to genotype frequency change in the population as selection goes on especially for an allele that is not fully recessive. However, it is unlikely that a big change in frequency could happen in a few generations.

The logistic regression model can easily be employed using standard statistical packages in popular use. The implementation requires no special training in computer programming, which makes the approach highly convenient. We believe that such an approach should definitely replace the conventional gene frequency method. It will help researchers to work more efficiently in their data analysis and thus promote progress in the genetic study of human longevity.

References

1. De Benedictis G, Tan Q, Jeune B et al. Recent advances in human gene-longevity association studies. Mech Aging Develop 2001: 122: 909–920.
2. Schachter F, Faure-Delaneff L, Guenot F et al. Genetic associations with human longevity at the APOE and ACE loci. Nat Genet 1994: 6: 29–32.
3. Kervinen K, Savolainen MJ, Salokannel J et al. Apolipoprotein E and B polymorphisms – longevity factors assessed in nonagenarians. Atherosclerosis 1994: 105 (1): 89–95.
4. Louhija J, Miettinen HE, Kontula K, Tikkanen MJ, Miettinen TA, Tilvis RS. Aging and genetic variation of plasma apolipoproteins. Relative loss of the apolipoprotein E4 phenotype in centenarians. Arterioscler Thromb 1994: 14 (7): 1084–1089.
5. Zhang JG, Ma YX, Wang CF et al. Apolipoprotein E and longevity among Han Chinese population. Mech Aging Develop 1998: 104: 159–167.
6. Gerdes LU, Jeune B, Ranberg KA, Nybo H, Vaupel JW. 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 2000: 19 (3): 202–210.
7. Yashin AI, Vaupel JW, Andreev KF et al. Combining genetic and demographic information in population studies of aging and longevity. J Epidemiol Biostat 1998: 3: 289–294.

8. Yashin AI, De Benedictis G, Vaupel JW et al. Genes, demography, and life span: the contribution of demographic data in genetic studies on aging and longevity. Am J Hum Genet 1999: 65: 1178–1193.

9. Yashin AI, De Benedictis G, Vaupel JW et al. Genes and longevity: lessons from studies on centenarians. J Gerontol 2000: 55a: B1–B10.

10. Tan Q, De Benedictis G, Yashin AI et al. Measuring the genetic influence in modulating human life span: gene–environment and gene–sex interactions. Biogerontology 2001: 2 (3): 141–153.

11. Tan Q, Yashin AI, Bladbjerg EM et al. Variations of cardiovascular disease associated genes exhibit sex-dependent influence on human longevity. Exp Gerontol 2001: 36 (8): 1303–1315.

12. De Benedictis G, Carotenuto L, Carrieri G et al. Gene/longevity association studies at four autosomal loci (REN, THO, PARP, SOD2). Eur J Hum Genet 1998: 6: 534–541.

13. De Benedictis G, Carotenuto L, Carrieri G et al. Age-related changes of the 3′APOB-VNTR genotype pool in ageing cohorts. Ann Hum Genet 1998: 62: 115–122.

14. De Benedictis G, Falcone E, Rose G et al. DNA multiallelic systems reveal gene/longevity associations not detected by diallelic systems: the APOB locus. Hum Genet 1997: 99: 312–318.

15. MatchSoft. Axum technical graphics and data analysis. Cambridge, MA, 1996.

16. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993: 52 (3): 506–516.

17. Ewens WJ, Spielman RS. The transmission/disequilibrium test: history, subdivision, and admixture. Am J Hum Genet 1995: 57 (2): 455–464.