Multidisciplinary Approaches in Genetic Studies of Human Aging and Longevity

Q. Tan[1,*], A.I. Yashin[2], K. Christensen[3], B. Jeune[3], G. De Benedictis[4], T.A. Kruse[1] and J.W. Vaupel[5]

[1]Department of Clinical Biochemistry and Genetics, Odense University Hospital, Odense, Denmark; [2]Center for Demographic Studies, Duke University, Durham, North Carolina, USA; [3]Epidemiology, Institute of Public Health, University of Southern Denmark, Odense, Denmark and [4]Cell Biology Department, University of Calabria, Rende, Italy and [5]Max-Planck Institute for Demograhic Research, Rostock, Germany

Abstract: The amount of research on human aging and longevity has been growing rapidly in recent years. Multidisciplinary approaches, which integrate classic population genetics methods with the principles of epidemiological and demographic investigation, are emerging as powerful tools for disentangling the complex gene network which modulates human lifespan. We try to summarize the different approaches and discuss the various aspects concerning their applications in studies of human aging and longevity. We also discuss the significance of the newly emerging DNA chip technology and its implications by highlighting new research topics. In fact, with the entering of the post-genomic era, hints given by observational studies, and thus founded on statistical evidence, can be exploited to cast light upon biological pathways crucial in aging and longevity.

Received on: 15-01-2004 - Accepted on: 19-05-2004

Key Words: Statistical genetics, aging, longevity.

INTRODUCTION

The quest for longevity has been accompanying the human history of civilization. Although it may take the form of a belief or religion [1], experiments for alchemical immortality were already started in the T'ang Dynasty (618-907 AD), the most prosperous time of Chinese civilization. While some advancement of knowledge concerning naturally-occurring drugs had been made, such practice led to little lasting medical innovation. On the other hand, records of extreme human longevity have been frequent dating back even from the remote past in the history [2-4]. Most interestingly, a recent case study on the verified world extraordinary longevity of Jeanne Calment (122 years) suggested an exceptional genetic inheritance due to a high concentration of long-lived ancestors in her family [5].

The evidence of a genetic involvement in human survival actually first came from the studies on related individuals using population genetics methods. The earliest literature can be traced back to 1899 when Beeton and Pearson [6] reported lifespan correlation within families. After that, family studies on lifespan correlation continued through the last century [7-15]. Although some of the findings are controversial, data from well-defined populations support the notion that there exist transmittable familiar attributes affecting lifespan, and a certain portion of them can be of a genetic nature. One step further, studies on twins have explicitly ascertained the genetic contribution to human survival by separating the genetic component in lifespan variation from the component of environment [16-20].

*Address correspondence to this author at the Department of Clinical Biochemistry and Genetics (KKA), Odense University Hospital, Sdr. Boulevard 29, DK-5000 Odense C, Denmark; Tel: +45 65412822; Fax: +45 65411911; E-mail: qihua.tan@ouh.fyns-amt.dk

With the confirmation of genetic influence in lifespan, a more challenging task is to find out the genes that are involved. Taking advantage of the rapid development in molecular genetics, the past few decades have seen intensified genetic study of human aging and longevity using individual genotype data [21]. Because lifespan as a survival trait is different from other binary or quantitative disease traits, efficient data analyzing techniques are crucial in helping to interpret the results. Various statistical approaches, both borrowed from the genetic epidemiology of diseases and particularly developed using survival analysis, have been applied. In this paper, we try to summarize the different methods in use and discuss their strength and weakness in practical applications. In addition, we also highlight some new topics concerning both experimental and analytical techniques that are going to evolve in the near feature.

THE GENETIC ASSOCIATION ANALYSIS

Association analysis or the candidate gene approach is a very popular tool for studying complex genetic traits [22]. The idea is simply that, as a result of haploid disequilibrium between gene variants, tight linkage between disease and marker loci will lead to an excessive occurrence of certain combinations of alleles or haplotypes in the disease population (cases) with respect to the normal population (controls) [23]. Any gene frequency discrepancy so observed conditional on the disease status is accounted by linkage disequilibrium (LD) occurring between the disease variant and the marker variant. In the context of aging and longevity, the association approach is applied in the same way but the phenotype is age instead, i.e. studying the changes in gene frequency conditional on age. In the following text we summarize the different analytical methods derived in the

1389-2029/04 $45.00+.00

framework of association studies applied to the aging/ longevity phenotype.

1. The Case-Control Approach

As described above, if a gene marker is in LD with a causal variant affecting individual survival, then different gene (allele or genotype) frequency can be observed in the aged and young populations. Studying gene frequency difference between the two groups will help make inference on the genetic association with human survival. It is this simple design that has been predominating the marker-longevity association studies [21]. Popularity of the case-control design in aging study can be due to its relatively lower expenses and quick outcome as compared with cohort studies. In the sample collection, extremely long-lived subjects (usually centenarians) are assigned as cases and young subjects (usually middle aged) as controls. For each gene of interest, a χ^2 statistic is calculated and statistical significance in gene frequency difference between the two groups assessed by testing the null hypothesis of no association.

At a highly polymorphic locus hosting n alleles, there will be $n(n+1)/2$ possible genotypes to be observed. In this case, some of the cell counts in the 2 by $n(n+1)/2$ contingency table may become too small. Such a situation can invalidate the asymptotic sampling distribution of the χ^2 statistic and affects the test of significance. When Hardy-Weinberg equilibrium (HWE) holds in the data, one can choose to work with alleles rather than with genotypes thus largely reduce the dimension of the contingency table to 2 by n while maintaining asymptotical equivalence [24]. Sham and Curtis [25] presented simulation based Monte Carlo tests which can be applied to a more liberal situation (less stringent on HWE) to cope with the polymorphic situation.

Although popular in use, there are several drawbacks in applying the traditional case-control design in aging research. Lifespan is a continuous trait instead of a binary one. Simply or roughly grouping the subjects into old cases and young controls fails to make full use of the individual information contained in the data and thus reduces its statistical power. Moreover, the approach ignores the cohort effect in mortality improvement [26] in making the inferences. Other innovative approaches are thus appealing.

2. The Case-Only Approach

The case-only approach was first designed for assessing gene-environment interactions in the etiology of complex genetic diseases [27, 28]. By treating centenarians as cases, it was shown that the approach is applicable in studying gene-sex and gene-gene interactions in human longevity [29, 30]. To illustrate the simple idea of the case-only approach, we take gene-sex interaction for example. If there is a sex-dependent effect that contributes to female survival, one should expect to observe in the centenarians, a higher proportion of carriers of the allele or genotype in females than that in males. Based on this idea, one calculates the case-only odds ratio

$$OR_{ca} = \frac{n_{female(+)} / n_{female(-)}}{n_{male(+)} / n_{male(-)}}$$

and tests the null hypothesis $H_0 : OR_{ca} = 1$. Any statistically significant deviation of OR_{ca} from unity indicates that there is a sex-dependent effect that contributes to and modifies the probability of achieving longevity. Assuming the susceptible gene and sex are independent and that the event the interaction is associated with (i.e. longevity in our case) is rare, it can be easily shown that the case-only odds ratio measures the departure from a multiplicative joint effect of gene and sex. Piegorsch *et al.* [27] proved that the case-only design has increased precision in estimating interactions due to reduced variance in the estimation compared with using the traditional case-control design. Statistical test for the null hypothesis can be conducted by employing the log likelihood ratio (LLR) test by calculating twice the difference between the log likelihood at $OR_{ca} = 1$ and at OR_{ca} estimated. When sample size is large, the LLR is approximately distributed as χ^2 on 1 degree of freedom. Alternatively, we can apply the standard χ^2 statistic to test the null hypothesis.

It is necessary to point out that the case-only approach makes sense only when our primary interest is in interactions. No effect of the genetic variant alone is measured. However, by only looking at the extremely old cohort (the centenarians), the method avoids the confounding cohort effect in mortality improvement.

3. The Logistic Regression Approach

In the case-control approach, we look at the change in gene frequency between two age groups (aged cases and young controls). A natural extension would be to study the change by each age. Such a situation can be modeled by the popular logistic regression model [31]. In the basic form of a logistic regression,

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

the odd of frequency of the allele or genotype of interest $p(x)$ is modeled as a function of age x. When β_1 is significantly different from zero, frequency of the allele or genotype goes up ($\beta_1 > 0$) or down ($\beta_1 < 0$) with advancing age. The odds ratio for frequency change between two adjacent ages can be measured as e^{β_1}. By applying the logistic regression approach, one does not need to roughly group the individuals into cases and controls and thus obtain more power in the analysis [32]. Moreover, assuming Hardy-Weinberg equilibrium, the logistic regression model with polytomous responses can be introduced to handle highly polymorphic genes [33]. Genotype and allele-based parameterization can be used to investigate the modes of gene action and to reduce the number of parameters so that

the power is increased while the number of multiple testing is minimized.

Another very important feature of logistic regression is the feasibility of modeling the non-monotonous pattern of gene frequency resulted from the antagonistic pleiotropic effect in gene action during the aging process as reported in longevity studies [34, 35]. This is achieved by modeling the allele or genotype frequency as a non-linear function of age x [33], for example, the fractional polynomials [36]. The partial likelihood ratio test can be applied to choose the best fitting model and to make inferences on the statistical significance of the age-dependent pattern as compared with a linear model. Confirmation of any non-monotonous trajectory in gene frequency can help to further understand or infer the biological roles that the gene plays during the process of aging.

The logistic regression model is easily accessible from standard statistical packages popular in use and thus it is conveniently applicable. However, since the model only fits the observed frequency pattern, it is very sensitive to sampling bias especially in small scale investigations which is usually the case in longevity study due to difficulty in collecting the long-lived individuals.

4. Survival Analysis

As lifespan is a survival trait in nature, survival analysis is undoubtedly an innate approach. However, the subject's age at death or lifespan is unknown at the time when the biological sample was taken. Unless one is determined to undergo the tedious follow-up which takes time and money, the individual survival information so obtained is usually censored. Instead of working directly with lifespan as one usually does in survival analysis, an alternative is to model genotype or allele frequency changes across the ages similar to that by the logistic regression approach but within the framework of survival analysis. Assuming at one locus, we are interested in an allele with frequency p. If survival rates at age x for carriers of 0, 1 and 2 copies of the allele are $s_0(x)$, $s_1(x)$ and $s_2(x)$ respectively, then the mean survival at age x over the three genotypes is

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

Using (2), the genotype frequencies at age x can be calculated as

$$p_0(x) = (1-p)^2 s_0(x)/\bar{s}(x),$$
$$p_1(x) = 2p(1-p)s_1(x)/\bar{s}(x) \quad \text{and}$$
$$p_2(x) = p^2 s_2(x)/\bar{s}(x).$$

Based on the genotype frequency distribution, a likelihood function can be constructed as

$$L \prod_x p_0(x)^{n_0(x)} p_1(x)^{n_1(x)} p_2(x)^{n_2(x)}, \quad (3)$$

where $n_0(x)$, $n_1(x)$ and $n_2(x)$ are the number of individuals in the three genotypes at age x. In practical application, $\bar{s}(x)$ is approximated by the observed population survival function available from population life table data. Based on the specification of genotype specific survival functions, different approaches have been proposed.

4.1. The Relative Risk Model

Yashin *et al*. [37] proposed a proportional hazard approach in which hazard of death at age x for genotype i ($i = 0, 1, 2$) is defined as $\mu_i(x) = r^i \mu_o(x)$ where $\mu_o(x)$ is the baseline hazard function for a non-carrier and r is relative risk of carrying one copy of the allele. The genotype specific survival function in (2) is now

$$S_i(x) = e^{-\int_0^x \mu i(t)dt} = e^{-r^i H_0(x)} = S_o(x)^{r^i}$$

where $s_o(x)$ is the survival distribution corresponding to $\mu_o(x)$. Tan *et al.* [38] applied an EM algorithm to estimate the baseline hazard function in which a non-parametric $s_o(x)$ is obtained numerically without specifying any parametric form. This non-parametric approach is advantageous because, in reality, the ideal parametric form of the baseline hazard function is unknown. Moreover, the relative risk approach also facilitates measuring of the interaction terms (for example, gene-sex and gene-environment interactions) by assuming multiplicative effects of the risk factors [38].

4.2. The Parametric Approach

Although the relative risk model is distribution free, it relies heavily on the proportional hazard assumption. Such an assumption is unrealistic for antagonistic genes [39]. Toupance *et al*. [40] tried to handle the situation by parametrically fitting the genotype specific survival functions (Gompertz-Makeham) and claimed that the antagonistic effect can be captured by the intersection of mortality curves for different genotypes. However, there are several difficulties in their approach. First of all, it seems unfair to impose a specific form of survival distribution on a subgroup in a limited sample. Indeed both the choice of a parametric form and the sample size limitation could result in considerable error in estimating the genotype-specific survival functions. Consequently, the age-dependent frequency pattern resulting from differential survival will be unreliable. This may be a problem chiefly at extreme ages, when sample collection becomes very difficult. In addition, at old ages, the validity of the Gompertz-Makeham model becomes more questionable. Driver [41] showed that the Gompertz or its hybridization, Gompertz-Makeham curves does not measure aging at all. In this situation, any pattern based on these uncertainties can be doubtable if not arbitrary. Antagonistic pleiotropic effects can also be modeled by applying frailty modeling [37, 42]. However, estimation of the heterogeneity parameter is problematic in small scale studies. In this case, a very small change in the heterogeneity parameter can lead to a big difference in the fitted frequency

pattern. After all, the application of a parametric survival model and the interpretation of the results should be done with caution.

In the frailty modeling, one assumes that the unobserved frailty z follows a gamma-distribution with mean 1 and variance σ^2 [42, 43], so that the genotype specific survival functions in (2) become

$$S_i(x) = E[s(x)] = E\left[e^{-r^i z \int_0^x \mu_o(s)ds}\right] = [1-r^i\sigma^2 \ln(S_o(x))]^{\sigma^{-2}} . \quad i = 0, 1, 2 \tag{4}$$

σ^2 has been estimated when applying the above model to genotype data in marker-longevity association studies [38, 44, 45] 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 individual survival, even if they carry the same genotype. As survival is a complex trait, frailty modeling offers a nice way to account for the compositional effects due to the unobserved risk factors (both genetic and environmental) [26]. However, more work is needed in the specification of frailty distribution and in parameter estimation.

Similar to the genetic association study on human diseases, the candidate gene approach on aging and longevity also faces a problem of multiple testing. This problem is actually twofold, i.e. (1) the multiple testing occurring at a single polymorphic locus, and (2) the multiple testing on genes or markers spanning multiple loci. Tan *et al.* [46] extend the survival model to multiallelic locus to estimate allele and genotype relative risks (ARR and GRR) and largely reduce the number of parameters to be tested. A log likelihood ratio (LLR) test is used to infer an overall statistical significance at the locus tested. The method helps circumvent the multiple testing problem at a polymorphic locus. When multiple loci are tested in one analysis and all the genes are independent in terms of both interaction and LD, the Bonferroni procedure can be used to adjust the significance level. However, in the case of whole-genome association, such correction is definitely unacceptable. Further progress has been made in capturing the simultaneous effects of multiple loci to localize complex disease genes [47, 48]. Efforts are needed to be made in introducing the methods to longevity studies.

Population stratification can result in spurious conclusions in associating candidate-genes with human diseases [49, 50]. This is also true in longevity studies. A family based transmission disequilibrium test (TDT) has been proposed to overcome the population stratification problem [51]. Unfortunately, this does not work in the context of longevity study because parental genotype information necessary for inferring the allele transmission is unavailable when one achieves longevity. However, it is possible to check for stratification by including, in the study, marker loci unlinked to the candidate locus. It has been estimated that > 15-20 unlinked microsatellites are sufficient to test for population stratification, in the situation in which there is no prior reason to suspect population structure [52]. Thus the bias due to population substructure can be minimized in well-designed association studies [53].

LINKAGE ANALYSIS

Linkage analysis is a technique to localize disease genes by looking for evidence of co-segregation between the disease and the gene markers whose location are already known. 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 [54]. Although in longevity study, 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 (ASP) to infer the location of genes that are linked to the trait of interest [55]. In the context of longevity study, the idea is that if long-lived siblings share more alleles IBD at some marker locus than randomly expected among siblings, then that locus might be near the locus of a beneficial gene. The likelihood of observing a set of genotypes at a marker locus can be calculated conditional on different values for the probabilities of sharing 0 (z_0), 1 (z_1) and 2 (z_2) alleles. A maximized likelihood score (MLS) is calculated by taking the difference between the log base 10 for the maximum-likelihood and for the likelihood of the null-hypothesis sharing probabilities (z_0=0.25, z_1=0.5, z_2=0.25),

$$MLS = \log_{10}[L(z_0, z_1, z_2)/L(z_0 = 0.25, z_1 = 0.5, z_2 = 0.25)] . \tag{5}$$

In order to map the genes, a multipoint linkage analysis [23] can be conducted and MLS plotted against the map position so that causal genes are localized to some interval in the region.

The higher probability for achieving longevity in centenarians' siblings reported by recent studies [56, 57] suggests the feasibility of sampling long-lived, sib-pair data for non-parametric linkage analysis. As the first application, Puca *et al.* [58] scanned the whole genome by applying the 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. However, as lifespan is a quantitative trait, the non-parametric linkage analysis on long-lived sib-pairs requires that a certain age-cut be set as a threshold for defining longevity. In this case, an immediate concern is its statistical power. Nemani *et al*. [59] studied the efficiency of the ASP approach and found that the method is powerful in mapping rare recessive genes but has very low power in locating dominant genes. Tan *et al.* [60] reevaluated the power issue of the method in longevity study and reported that a sample size of over 600 long-lived pairs (centenarians and their siblings over age 90)

can have acceptable power in mapping a low frequency dominant allele that cuts down the hazard of death by half.

When applied to longevity study, the ASPs so collected are of different ages. The small proportion of the extremely long-lived ASPs is more genetically selected than the majority who are just above the threshold. This is also indicated by the increased IBD sharing at older ages in the region harboring a beneficial gene [60]. It should be possible to increase the power of non-parametric linkage by introducing an optimal weighing scheme to the sharing for each pair based on a function of their ages. Such a weighing scheme is more important for small-scale investigations when power consideration is critical.

LINKAGE VERSUS ASSOCIATION

Risch [61] showed that although the ASPs for linkage can provide statistical evidence for localizing high risk genes with intermediate allele frequency, it is unable to map a gene of modest relative risk (genotype relative risk of less than 2) except in unrealistic large samples. Tan *et al*. [60] reported that, even with full marker information content, the non-parametric linkage approach is unable to map a gene that accounts for less than 1% of the lifespan variations. By examining normal quantitative trait, Sham *et al*. [62] also concluded that linkage analysis is invalid for mapping a QTL that accounts for less than 1% of the phenotypic variance equivalent to a recombination fraction of 0.01 which is about 1 megabase (Mb). On the other hand, the association approach is often capable of narrowing down a region containing a putative susceptibility gene to less than 1 Mb [63]. As a complex trait, longevity does not seem to be resulted from some major genes but rather from numerous genes functioning interactively together with non-genetic factors [64-66]. In this case, one should not rely too much on the linkage approaches. Perhaps, a more feasible way would be to follow the regions identified through linkage mapping by association studies and conduct linkage disequilibrium fine mapping [61, 62].

With the newly emerging microarray or DNA chip technology, the association based LD mapping is gaining popularity [67]. Besides its use in gene expression analysis to be discussed later, the high-throughput technology also enables us to use single nucleotide polymorphisms (SNPs) as biallelic markers for whole-genome screening of complex trait genes. It is estimated that as many as 500,000 evenly spaced SNPs are required to detect LD for mapping purposes [68]. The situation challenges the traditional locus-by-locus approach in association studies. Multi-locus statistical approaches [48] that take into account of interdependence of gene function important in the genetic modulation of human survival [29] are appealing. Because particular DNA variants may remain together on ancestral haplotypes (set of ordered markers) for many generations, groups of neighboring SNPs can form haplotypic diversity with distinctive patterns of LD and which can be exploited in both genetic linkage and association studies [69]. Haplotype analysis is more efficient than the gene-by-gene association test because it makes use of the LD information contained in the flanking markers [70]. Haplotype estimating technique has helped the association approach to gain power in case-control studies [71, 72]. The same idea can be introduced to survival analysis to estimate haplotype relative risks (HRRs) on hazard of death and to infer the genes that affect human survival.

GENE EXPRESSION ANALYSIS

As the aging process depends on complex interactions of numerous genes and gene products, the recently emerging microarray technology [73, 74] capable of measuring the expression levels of thousands of genes is a useful tool to help us to deepen our understanding in the genetics of aging and longevity [75-77]. The abundant information provided by microarray experiments enables us to study gene activities as indicated by the messenger RNA levels under various biological and environmental conditions including aging [78, 79]. Using related and unrelated individuals, the technique can be used (1) to find genes showing dynamic profiles during the process of aging and (2) to address the question concerning the age dependence in the genetic regulation of the genome function.

Direct evidence for a connection between the regulation of gene expression and lifespan in *Drosophila* was actually reported before the application of microarray [80]. The analysis of gene transcriptional activity also showed relatively stable variance in gene expression as age increases [81]. These conclusions were confirmed later by using DNA chips [82] and moreover, the application of chips also identified the well-known metabolic pathways impacted by aging in *Drosophila melanogaster*. Unfortunately gene expression study on human aging using a microarray setup has been rare [83, 84]. One big difficulty with studying human subjects is, unlike in animal experiments, the long life course. An alternative would be to take random subjects from different age groups [85]. But this will bring in individual variations in the observed expression profile for one particular gene across ages [86]. However, if enough replicates are affordable at each age group, severity of the problem can be alleviated by studying the mean of gene expression. Such a setup not only enables the study on age-dependent changes in gene expression but also facilitate the study on individual differences in gene expression within each age group. By measuring mRNA levels in actively dividing fibroblasts isolated from young (one 7-year-old and one 9-year-old), middle-age (two 37-year-old), and old-age (three in their 90s) humans, Ly *et al*. [85] identified genes whose expression is associated with age-related phenotypes. Unfortunately, the samples size in this pioneering study was too small (7 individuals) to produce convincing results [86]. The second, but not less important, difficulty in microarray study on aging is that gene expression is highly tissue specific. For the purpose of aging study, postmitotic tissues or non-dividing cells can be targeted. By studying gene expression in the aging human retina, Yoshida *et al*. [83] reported a decreased use of biosynthetic pathways and increased reliance on genes involved in stress response, which is consistence with the conclusions from studies on

transcriptional profiling of aging in mouse skeletal muscles and brain [87, 88]. Such findings could suggest that common molecular mechanisms may regulate aging in different tissues and cell types [83].

For the genes displaying significantly age-dependent expression profiles, correlation on their expressions can be calculated between related individuals (twins or siblings) to see how the genetic component in the pattern regulation changes with age. Such an approach can help answer whether the importance of the genetic regulation in lifespan is going up or down or remained unchanged with advancing age [26, 57]. One obvious difference from the other gene expression analyses is that, here gene expression becomes a phenotype or trait. One can study the correlation on the gene expression phenotype between related individuals by applying methods in population genetics. Such an approach sets up a bridge between molecular and population genetics. Our pilot study on old Danish twins has shown significant difference between gene expression correlation in monozygotic (MZ) and dizygotic (DZ) twins suggesting epigenetic regulation of gene expression [89] at high ages. However, the crucial point concerning the aging pattern of the regulation remains unknown.

END REMARKS

The identification of the genetic factors which modulate rate and quality of human aging, and therefore human lifespan, is a challenge which requires integration among several disciplines. Hints given by studies in model organisms suggest that aging is a lethal side effect of adaptive processes earlier in life [90]. Several gene patterns involved in these processes are conserved along evolution, therefore candidate genes can be picked out in the human genome and checked as susceptibility factors in modulating the aging process. Then, by applying epidemiological guidelines in sampling procedures and statistic tools in genetic data analysis, the complex genetic network which modulates human lifespan could be disentangled. Moreover, gene expression analyses carried out for groups of genes entering in the age-related patterns so identified could help us understand how genes control rate and quality of human aging and, possibly, give us the environmental tools for modulating this complex trait.

AKNOWLEDGEMENT

This work was supported by the US National Institute on Aging (NIA) research grant NIA-P01-AG08761 and by "The MicroArray Center" project under the Biotechnological Research Program financed by the Danish Research Agency.

REFERENCES

[1] Holliday, R. Human ageing and the origins of religion. *Biogerontology* **2001**, *2*: 73-77.

[2] Zhao, Z. Record longevity in Chinese history-evidence from the Wang genealogy. In *Exceptional Longevity: From prehistory to the present*, Jeune, B; Vaupel, J.W.; Eds. Odense University Press: Denmark, **1995**, pp.93-108.

[3] Jeune, B. In search of the first centenarians. In *Exceptional Longevity: From prehistory to the present*, Jeune, B; Vaupel, J.W.; Eds. Odense University Press: Denmark, **1995**, pp.11-24.

[4] Oeppen, J.; Vaupel, J.W. Demography. Broken limits to life expectancy. *Science* **2002**, *296:*1029-1031.

[5] Robine, J.M.; Allard, M. The oldest human. *Science* **1998**, *279:* 834-835.

[6] Beeton, M.; Pearson, K. Data for the problem of evolution in man, II, A first study of the inheritance of longevity and the selective death rate in man. *Proceedings of the Royal Society of London.* **1899**, *65:* 290-305.

[7] Bell, A.G. The duration of life and conditions associated with longevity, a study of the Hyde genealogy. Genealogical Records Office, Washington, **1918**.

[8] Pearl, R. Studies on human longevity IV. The inheritnce of longevity preliminary report. *Hum. Biol.* **1931**, *3:* 245-269.

[9] Jalavisto, E. Inheritance of longevity according to Finnish and Swedish genealogies. *Ann. Med. Intern. Fenn*. **1951**, *40:* 263-274.

[10] Hawkins, M.R., Murphy, E.A. and Abbey, H. The familial component of longevity. A study of the offspring of nonagenarians. I. Methods and preliminary report. *Bull. Johns Hopkins Hospital,* **1965**, *117:* 24-36.

[11] Abbott, M.H., Murphy, E.A., Bolling, D.R. and Abbey, H. The familial component in longevity. A study of offspring of nonagenarians. II. Preliminary analysis of the completed study. *Hopkins Med. J.* **1974**, *134:* 1-16.

[12] Wyshak, G. Fertility and longevity of twins, sibs, and parents of twins. *Soc. Biol.* **1978**, *25:* 315-330.

[13] Sørensen, T.I.; Nielsen, G.G.; Andersen, P.K.; Teasdale, T.W. Genetic and environmental influences on premature death in adult adoptees. *N. Engl. J. Med*. **1988**, *318:* 727-732

[14] Gudmundsson, H.; Gudbjartsson, D.F.; Frigge, M.; Gulcher, J.R.; Stefansson, K. Inheritance of human longevity in Iceland. *Eur. J. Hum. Genet.* **2000**, *8:* 743-749.

[15] Cournil, A.; Legay, J.M.; Schachter, F. Evidence of sex-linked effects on the inheritance of human longevity: a population-based study in the Valserine valley (French Jura), 18-20th centuries. *Proc. R. Soc. Lond. B. Biol. Sci.* **2000**, *267:* 1021-1025.

[16] Carmelli, D. Intrapair comparisons of total lifespan in twins and pairs of sibs. *Hum. Biol.* **1982**, *54:* 525-537.

[17] Harris, J. Age differences in genetic and environmental influences for health from the Swedish adoption/twin study of aging. *J. Gerontol., Psychological Science* **1992**, *47:* 213- 220.

[18] Hayakawa, K., Shimizu, T., Ohba, Y., Tomioka, S., Takahasi, S., Amano, K., Yura, A., Yokoyama, Y. and Hayakata, Y. Intrapair differences of phisical aging and longevity in identical twins. *Acta Genet. Med. Gemellol.* **1992**, *41:* 177-185.

[19] McGue, M., Vaupel, J.W., Holm, N. and Harvald, B. Longevity is moderately heritable in a sample of Danish twins born 1870-1880. *J. Gerontol.* **1993**, *48:* B237-B244.

[20] Herskind, A.M., McGue, M., Holm, N.V., Sorensen, T.I., Harvald, B. and Vaupel, J.W. The heritability of human longevity, a population-based study of 2872 Danish twin pairs born 1870-1900. *Hum. Genet.* **1996**, *97:* 319-323.

[21] De Benedictis, G.; Tan, Q.; Jeune, B.; Christensen, K.; Ukraintseva, S.V.; Bonafe, M.; Franceschi, C.; Vaupel, J.W.; Yashin, A.I. Recent advances in human gene-longevity association studies. *Mech. Ageing Dev.* **2001**, *122:* 909-920.

[22] Tabor H.K.; Risch N.J.; Myers R.M. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Review Genetics* **2002**, *3:* 391-397.

[23] Sham, P. Statistical in Human Genetics, Arnold Applications of Statistics. Edward Arnold: London **1998**,

[24] Sasieni, P.D. From genotypes to genes: doubling the sample size. *Biometrics* **1997**, *53:* 1253-1261.

[25] Sham, P.C.; Curtis, D. Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Ann. Hum. Genet.* **1995**, *59:* 97-105

[26] Vaupel, J.W.; Carey, J.R.; Christensen, K.; Johnson, T.E.; Yashin, A.I.; Holm, N.V.; Iachine, I.A.; Kannisto, V.; Khazaeli, A.A.; Liedo, P.; Longo, V.D.; Zeng, Y.; Manton, K.G. and Curtsinger, J.W. Biodemographic trajectories of longevity. *Science* **1998**, *280:* 855-860.

[27] Piegorsch, W.W.; Weinberg, C.R.; Taylor, J.A. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Stat. Med.* **1994**, *13:* 153-162.

[28] Khoury, M.J.; Flanders, W.D. Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: case-

control studies with no controls *Am. J. Epidemiol*. **1996**, *144:* 207-213.

[29] Tan, Q.; De Benedictis, G.; Ukraintseva, S.V.; Franceschi, C.; Vaupel, J.W.; Yashin, A.I. A centenarian-only approach for assessing gene-gene interaction in human longevity. *Eur. J. Hum. Genet.* **2002**, *10:* 119-124.

[30] Tan, Q.; Yashin, A.I.; Bladbjerg, E.M,; de Maat, M.P.; Andersen-Ranberg, K.; Jeune, B.; Christensen, K.; Vaupel, J.W. A case-only approach for assessing gene by sex interaction in human longevity. *J. Gerontol. A. Biol. Sci. Med. Sci.* **2002**, *57:* B129-133.

[31] Tan, Q.; Yashin, A.I.; De Benedictis, G.; Cintolesi, F.; Rose, G.; Bonafe, M.; Franceschi, C.; Vach, W.; Vaupel, J.W. A logistic regression model for measuring gene-longevity associations. *Clin. Genet.* **2001**, *60:* 463-469.

[32] Pletcher, S.D.; Stumpf, M.P. Population genomics: ageing by association. *Curr. Biol.* **2002**, *12:* 328-330.

[33] Tan, Q.; Bathum, L.; Christiansen, L.; De Benedictis, G.; Bellizzi, D.; Rose, G.; Frizner, N.; Dahlgaard, J.; Vach, W.; Vaupel, J.W.; Yashin, A.I.; Christensen, K.; Kruse, T.A. Logistic regression models for polymorphic and antagonistic pleiotropic gene action on human aging and longevity. *Annals of Human Genetics* **2003**, *67:* 598-607.

[34] Schachter, F.; Faure-Delaneff, L.; Guenot, F.; Rouger, H.; Froguel, P.; Lesueur-Ginot, L.; Cohen, D. Genetic associations with human longevity at the APOE and ACE loci. *Nature Genetics* **1994**, *6:* 29-32.

[35] De Benedictis, G.; Carotenuto, L.; Carrieri, G.; De Luca, M.; Falcone, E.; Rose, G.; Yashin, A.I.; Bonafe, M.; Franceschi, C. Age-related changes of the 3'APOB-VNTR genotype pool in aging cohorts. *Ann. Hum. Genet.* **1998**, *62:* 115-122.

[36] Royston, P.; Altman, D. Regression using fractional polynomials of continuous: Parsimonious parametric modeling. *Applied Statistics* **1994**, *43:* 429-467.

[37] Yashin, A.I.; De Benedictis, G.; Vaupel, J.W.; Tan, Q.; Andreev, K.F.; Iachine, I.A.; Bonafe, M.; DeLuca, M.; Valensin, S.; Carotenuto, L.; Franceschi, C. Genes, demography, and lifespan: The contribution of demographic data in genetic studies on aging and longevity. *Am. J. Hum. Genet.* **1999**, *65:* 1178-1193.

[38] Tan, Q.; De Benedictis, G.; Yashin, A.I.; Bonafe, M.; DeLuca, M.; Valensin, S.; Vaupel, J.W.; Franceschi, C. Measuring the genetic influence in modulating human lifespan: Gene-environment and gene-sex interactions. *Biogerontology* **2001**, *2:* 141-153.

[39] Rose, M.R. *Evolutionary biology of aging*, London: Oxford University Press, **1991**.

[40] Toupance, B.; Godelle, B.; Gouyon, P.H.; Schachter, F. A model for antagonistic pleiotropic gene action for mortality and advanced age. *Am. J. Hum. Genet.* **1998**, *62:* 1525-1534.

[41] Driver, C. The Gompertz function does not measure ageing. *Biogerontology*. **2001**, *2:* 61-65.

[42] Vaupel, J.W.; Yashin, A.I. Heterogeneity's ruses: some surpring effects of selection on population dynamics. *Am. Stat.* **1985**, *39:* 176-185.

[43] Vaupel, J.W.; Manton, K.G.; Stallard, E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **1979**, *16:* 439-454.

[44] Yashin, A.I.; De Benedictis, G.; Vaupel, J.W.; Tan, Q.; Andreev, K.F.; Iachine, I.A.; Bonafe, M.; DeLuca, M.; Valensin, S.; Carotenuto, L.; Franceschi, C. Genes and longevity: Lessons from studies on centenarians. *Journal of Gerontology* **2000**, *55a*, B1-B10

[45] Ewbank, D.C. Mortality differences by APOE genotype estimated from demographic synthesis. *Genet. Epidemiol.* **2002**, *22:* 146-155.

[46] Tan, Q.; De Benedictis, G.; Yashin, A.I.; Bathum, L.; Christiansen, L.; Dahlgaard, J.; Frizner, N.; Vach, W.; Vaupel, J.W.; Christensen, K.; Kruse, T.A. Assessing genetic association with human survival at multi-allelic loci. *Biogerontology* **2004**, *5:* 89-97.

[47] Hoh, J.; Wille, A.; Ott, J. Trimming, weighting, and grouping SNPs in human case-control association studies. *Genome Res.* **2001**, *11:* 2115-2119.

[48] Hoh, J.; Ott, J. Mathematical multi-locus approaches to localizing complex human trait genes. *Nat. Rev. Genet*. **2003**, *4:* 701-709.

[49] Cardon, L.R.; Bell, J.I. Association study designs for complex diseases. *Nat. Rev. Genet.* **2001**, *2:* 91-99.

[50] Thomas, D.C.; Witte, J.S. Population stratification: a problem for case-control studies of candidate-gene associations? *Cancer Epidemiol. Biomarkers Prev*. **2002**, *11:* 505-512.

[51] Ewens, W.J.; Spielman, R.S. The Transmission/disequilibrium Test. In *Handbook of statistical genetics*. Balding, D.J.; Bishop, M.; Cannings, C. Eds. John Wiley and Sons **2001**.

[52] Pritchard, J.K.; and Rosenbreg, N.A. Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum. Genet*. **1999**, *65:* 220-228.

[53] Wacholder, S.; Rothman, N.; Caporaso, N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J. Natl. Cancer Inst*. **2000**, *92:* 1151-1158.

[54] Ott, J. Analysis of Human Genetic Linkage, 3rd ed., John Hopkins Univ. Press **1999**.

[55] Kruglyak, L.; Lander, E.S. Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am. J. Hum. Genet. 57:* **1995**, 439-454.

[56] Perls, T.T.; Bubrick, E.; Wager, C.G.; Vijg, J.; Kruglyak, L. Siblings of centenarians live longer. *Lancet* **1998**, *23:* 1560.

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

[58] Puca, A.A.; Daly, M.J.; Brewster, S.J.; Matise, T.C.; Barrett, J.; Shea-Drinkwater, M.; Kang, S.; Joyce, E.; Nicoli, J.; Benson, E.; Kunkel, L.M.; Perls, T. A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4. *Proc. Natl. Acad. Sci. USA* **2001**, *98:* 10505-10508.

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

[60] Tan, Q.; Zhao, J.H.; Iachine, I.; Hjelmborg, J.; Vach, W.; Vaupel, J.W.; Christensen, K.; Kruse, T.A. Power of non-parametric linkage analysis in mapping genes contributing to human longevity in long-lived sib-pairs. *Genetic Epidemiology* **2004**, *26:* 245-253.

[61] Risch, N.J. Searching for genetic determinants in the new millennium. *Nature* **2000**, *405:* 847-856.

[62] Sham, P.C.; Lin, M.W.; Zhao, J.H.; Curtis, D. Power comparison of parametric and nonparametric linkage tests in small pedigrees. *Am. J. Hum. Genet*. **2000**, *66:* 1661-1668.

[63] Morton, N.E. Significance levels in complex inheritance. *Am. J. Hum. Genet.* **1998**, *62:* 690-697.

[64] Christensen, K.; Vaupel, J.W. Determinants of longevity: genetic, environmental and medical factors. *J. Intern. Med*. **1996**, *240:* 333-341.

[65] Vauple, J.W.; Tan, Q. How many longevity genes are there? Paper presented at annual meeting of Population Association of America, Chicago. **1998**.

[66] Johnson, T.E. A personal retrospective on the genetics of aging. *Biogerontology*. **2002**, *3:* 7-12.

[67] Gray, I.C.; Campbell, D.A.; Spurr, N.K. Single nucleotide polymorphisms as tools in human genetics. *Hum. Mol. Genet*. **2000**, *9:* 2403-2408.

[68] Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet*. **1999**, *22:* 139-144.

[69] Schork, N.J.; Fallin, D.; Lanchbury, J.S. Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin. Genet*. **2000**, *58:* 250-264.

[70] Akey, J.; Jin, L.; Xiong, M. Haplotypes vs single marker linkage disequilibrium tests: what do we gain? *Eur. J. Hum. Genet.* **2001**, *9:* 291-300.

[71] Terwilliger, J.D.; Ott, J. A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. *Hum. Hered*. **1992**, *42:* 337-346.

[72] Zhao, J.H.; Sham, P.C. Faster haplotype frequency estimation using unrelated subjects. *Hum. Hered.* **2002**, *53:* 36-41.

[73] Brown, P.O.; Botstein, D. Exploring the new world of the genome with DNA microarrays. *Nat. Genet*. **1999**, *21:* 33-37.

[74] Lander, E.S. Array of hope. *Nat. Genet*. **1999**, *21:* 3-4.

[75] Han, E.; Hilsenbeck, S.G. Array-based gene expression profiling to study aging. *Mech. Ageing Dev.* **2001**, *122:* 999-1018.

[76] Helmberg, A. DNA-microarrays: novel techniques to study aging and guide gerontologic medicine. *Exp. Gerontol*. **2001**, *36:* 1189-98.

[77] Weindruch, R.; Kayo, T.; Lee, C.K.; Prolla, T.A. Gene expression profiling of aging using DNA microarrays. *Mech. Ageing Dev*. **2002**, *123:* 177-193.

[78] Weindruch, R. Caloric restriction, gene expression, and aging. *Alzheimer Dis. Assoc. Disord.* **2003**, **17** Suppl. *2:* S58-59.
[79] Case, C.C. Transcriptional tools for aging research. *Mech. Ageing Dev.* **2003**, *124:* 103-108.
[80] Rogina, B.; Helfand, S.L. Regulation of gene expression is linked to lifespan in adult Drosophila. *Genetics.* **1995**, *141:* 1043-1048.
[81] Rogina, B.; Vaupel, J.W.; Partridge, L.; Helfand, S.L. Regulation of gene expression is preserved in aging Drosophila melanogaster. *Curr. Biol.* **1998**, *8:* 475-478.
[82] Pletcher, S.D.; Macdonald, S.J.; Marguerie, R.; Certa, U.; Stearns, S.C.; Goldstein, D.B.; Partridge, L. Genome-wide transcript profiles in aging and calorically restricted Drosophila melanogaster. *Curr. Biol*. **2002**, *12:* 712-723.
[83] Yoshida, S.; Yashar, B.M.; Hiriyanna, S.; Swaroop, A. Microarray analysis of gene expression in the aging human retina. *Invest. Ophthalmol. Vis Sci.* **2002**, *43:* 2554-2560.
[84] Shelton, D.N.; Chang, E.; Whittier, P.S.; Choi, D.; Funk, W.D. Microarray analysis of replicative senescence. *Curr. Biol*. **1999**, *9:* 939-945.
[85] Ly, D.H.; Lockhart, D.J.; Lerner, R.A.; Schultz, P.G. Mitotic misregulation and human aging. *Science*. **2000**, *287:* 2486-2492.
[86] Cristofalo V.J. A DNA chip off the aging block. *Nat. Med*. **2000**, *6:* 507.
[87] Lee, C.K.; Weindruch, R.; Prolla, T.A. Gene-expression profile of the ageing brain in mice. *Nat. Genet.* **2000**, *25:* 294-297.
[88] Weindruch, R.; Prolla, T.A. Gene expression profile of the aging brain. *Arch. Neurol*. **2002**, *59:* 1712-1714.
[89] Jaenisch, R.; Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet*. **2003**, **33** Suppl: 245-254.
[90] Partridge, L. and Gems, D. Ageing: a lethal side-effect. *Nature* **2002**, *418:* 921.