HOW DO GENES AFFECT AGING:
THE BIODEMOGRAPHY OF HUMAN HEALTH AND SURVIVAL

Anatoliy I. Yashin[1], Deqing Wu[1], Konstantin G. Arbeev[1], Eric Stallard[1],
Kenneth C. Land[1], Svetlana V. Ukraintseva[1]

[1] Center for Population Health and Aging, Duke University, Durham, NC, 27708-0408, USA

Word Count: 3840

Address for correspondence and reprints:

Anatoliy I. Yashin, ScD, Center for Population Health and Aging, Duke University, 002 Trent Hall, Box 90408, Durham, NC, 27708-0408, USA. Tel.: (+1) 919-668-2713; Fax: (+1) 919-684-3861; E-mail: aiy@duke.edu

Key words: polygenic score, genetics of mortality, GWAS, stress resistance, resilience, additive genetic component

Abbreviated title: BIODEMOGRAPHY OF HUMAN SURVIVAL

Abstract

Background. In genome wide association studies (GWAS) of human life span, none of the genetic variants have reached the level of genome-wide statistical significance. The roles of such variants in life span regulation remain unclear.
Data and Method. Biodemographic analyses of genetic regulation of life span using data on low-significance longevity alleles selected in the earlier GWAS of the original Framingham cohort.
Results. Age-specific survival curves considered as functions of the number of longevity alleles exhibit regularities known in demography as "rectangularization" of survival curves. The presence of such pattern confirms observations from experimental studies that regulation of life span involves genes responsible for stress resistance.
Conclusion. Biodemographic analyses could provide important information about the properties of genes affecting phenotypic traits.

Introduction

Genome wide association studies (GWAS) of complex traits have been developed to perform intensive analyses of genetic influences on such traits. These studies have helped identify hundreds of genetic variants and have provided valuable information about their roles in such traits [1]. Despite this evident progress, GWAS have not entirely met the expectations of many researchers. Most genetic variants identified so far confer relatively small effects on the traits of interest. Many of the detected effects remain below the levels of statistical significance established to correct the analyses for multiple comparisons. These small-effects-low-significance single-nucleotide polymorphism (SNP) alleles conventionally have been excluded from further analyses of their roles in molecular metabolic pathways. The small contribution of selected SNP alleles to the variability of complex traits has generated debates about "missing heritability" [1-5]. The use of data from the whole genome scan combined with an intensive search for rare alleles has been suggested as an alternative to existing GWA approaches. However, genetic data for populations of appropriate sizes with such levels of genetic details are not yet available to researchers.

GWAS of human longevity share all these limitations. The candidate-gene approach used in genetic association studies of longevity has resulted in finding a number of genes whose connection to long life can also be associated with the roles they play in metabolic pathways. The effects of a number of such genes have been replicated in independent studies [6-9]. Surprisingly, however, these genes have not shown significant effects in genome wide association studies of human longevity [10, 11].

Lunetta et al. [10] performed GWAS using genetic data on 100K SNPs collected for participants of the original and offspring cohorts of the Framingham Heart Study (FHS). The authors concluded that longevity and aging traits are associated with SNPs on the Affymetrix 100K GeneChip. However, none of the associations achieved genome-wide significance.

Newman et al. [11] performed a meta-analysis of GWAS in Caucasians from four prospective cohort studies. The authors found 273 SNP associations with $p < 0.0001$, but none reached the pre-specified significance level of 5×10^{-8}. Another recent meta-analysis of GWAS from nine studies also found no genome-wide significant SNPs for all-cause mortality and survival free of major disease or death [12].

Two recent GWA studies confirmed the role of the apolipoprotein E (APOE) gene as the major genetic determinant of survival into old age. Deelen et al. [13] found one SNP located in TOMM40 at chromosome 19q13.32 close to the APOE gene ($p = 3.39 \times 10^{-17}$). Nebel et al. [14] found one SNP near the APOC1 gene ($p = 1.8 \times 10^{-10}$) and this association was fully explicable by linkage disequilibrium with the APOE allele e4. No other SNPs reached the genome-wide significance level in these studies.

Yashin et al. [15] found that genetic variants, individually selected using methods of GWAS, may jointly influence life span. The joint effect of polygenic score ("genetic dose") on life span was substantial, and highly statistically significant. The relationship was replicated using data on an independent population. The set of selected genetic variants was able to predict a similar relationship in the replicate population. In this paper, we evaluate age patterns of mortality and survival for subgroups of individuals having different numbers of longevity SNP alleles in their genomes. We show that survival functions in the subgroups differ substantially,

and the difference in age patterns is similar to that in population-level survival observed in the distinct time points of the last century. We explain these difference and similarity using biodemographical methods and show that these explanations are consistent with recent findings from studies of aging and longevity. Note that the construction of polygenic score (genetic dose) which counts the number of longevity alleles contained in person's genome resembles that of frailty (or cumulative deficits) index [16-18] which counts accumulation of deficits in individual during his/her life course.

Data and Methods

Data and General Approach. To reduce possible effects of missing data on survival functions of individuals having different genetic background we use data which passed high level of quality control procedure with SNP call rates 97% with subsequent imputation of missing genotypes. These procedures resulted in 954 individuals from the original FHS cohort. Note that this number is smaller than that used in Yashin et al. [15]. For each of these individuals we calculated the number of longevity alleles contained in their genomes (out of 39 longevity alleles selected from 550,000 SNPs in Yashin et al. [15]). The detailed description of the Framingham Heart Study and the FHS genome-wide genotyping data can be found on the *dbGaP* website (*phs000007.v3.p2*).

Using 39 longevity alleles selected from data on the original FHS cohort we constructed two polygenic score indices: one index measuring the additive genetic component of life span and another index counting the number of genetic variants contained in each subject's genomes (see details in **Supplemental Data**). Each index was used for evaluating the joint influence of subsets of genetic variants on survival. We showed that the two indices explain about the same percentage of life span variance and are able to predict life spans in individuals from the offspring FHS cohort using genetic variants detected from the data on the original FHS cohort. We divided participants of the original FHS cohort into sub-cohorts of individuals having different numbers of longevity alleles in their genomes. Then we evaluated the joint influence of subsets of genetic variants on the age patterns of mortality and survival in these sub-cohorts. We separately fitted the Gompertz model to the mortality data at available age ranges in the sub-cohorts, constructed the corresponding survival functions, and compared them. We used the Strehler-Mildvan mortality model [19] to guide our analysis of how the Gompertz parameters are affected by the genetic factors and "life saving" mortality model [20] to get insights about functional roles of detected genes in aging, health and life span.

Results

Regularities of genetic influence on survival. Fig. 1A illustrates the pattern of survival improvement observed in developed countries during the 20th century (sex-specific patterns are shown in Fig. S5 in Supplemental Data). Demographers have characterized such population survival curve improvements as "rectangularization," or "compression of mortality," by which is meant a shift towards very low mortality levels through the childhood and younger adult ages followed by steep decreases in survivorship at the older ages.

Using survival data on the participants of the FHS cohort carrying different numbers of 39 "longevity alleles" (i.e., alleles, having positive effects on survival) selected in Yashin et al. [15] from the original FHS cohort we evaluated mortality rates and survival curve patterns for the three groups of individuals having different numbers of such alleles in their genomes. We divided the entire sample of 954 individuals into these three groups so that the number of individuals in each group is approximately the same (344, 306, and 304, respectively). The number of longevity alleles carried by each individual in the first group is less than 23. The second group consists of individuals each having 23 to 25 longevity alleles in their genomes. The third group contains individuals with 26 or more longevity alleles. Fig. 1B shows that survival function for individuals from the second and third groups look more "rectangular" than that in the first group resembling difference in survival patterns observed in distinct time periods (see Fig. 1A).

Fig. 1 about here

Another striking similarity is between an almost linear increase of average life span as a function of the number of longevity alleles [15] and an almost linear increase in the life expectancy at birth over time [21]. Fig. 2 shows an increase in average life span for groups of individuals whose number of longevity alleles varies from 0 to 39 (Fig. 2A) and the historical data on life expectancy at birth in the U.S. in 1980-2007 (Fig. 2B; sex-specific patterns are shown in Fig. S6 in Supplemental Data).

Fig. 2 about here

What mechanisms might be responsible for such similarities in survival and life expectancy changes caused by two evidently different causes? To address this question the use of two biodemographic models could be helpful. The first one [19] represents human mortality rates as a result of interplay between the process of external disturbances or stresses and an age-dependent decline in "vitality" – an index, characterizing individual's ability to withstand stresses of life. The second model [20, 22] shows that observed trends in mortality decline can be explained by a process of "saving lives" resulting from improvement in economic and living conditions, as well as advances in health care and medical treatment.

The Strehler-Mildvan model of aging and mortality. More than 50 years ago, *Science* published the seminal Strehler and Mildvan [19] paper, in which the exponential increase with age, x, in the Gompertz mortality rate, $\mu(x) = R_0 \exp(\alpha x)$, was represented as a result of interplay between external disturbances (stresses of/challenges to life) and the decline in the "vitality" variable describing individuals' resistance to stresses. The model explained a striking regularity detected in comparisons of the Gompertz mortality rates across different populations: the parameters R_0 and α of this curve were not changing independently from one population to the next, as one might expect, but showed a strong negative correlation, later called the Strehler-Mildvan (SM) correlation. This model has been applied to explaining differences in mortality rates among different populations [19, 23, 24] and to differences in mortality rates in the same country at different time periods, or in subsequent sub-cohorts [24-26], as well as in cause-specific mortality rates [27].

Fig. 3 shows the logarithms of the Gompertz mortality rates evaluated for groups of individuals in the original FHS cohort having different numbers of longevity alleles in their genomes.

Fig. 3 about here

The pattern of changes across the trajectories is typical of that for rectangularization of survival curves. The evaluation of the dependence of life span on the number of "longevity" alleles individual possess may shed more light on genetic nature of this trait.

How the SM model explains observed patterns in survival improvements over time. The SM model represents the age-specific Gompertz mortality function $\mu(x) = R_0 \exp(\alpha x)$ (which typically gives a good fit to population patterns of human mortality rates between ages 30 and 85 years) in terms of two sets of parameters: one describes the age-dependent decline in vitality and the second characterizes external stresses. In the framework of the SM model, the observed rectangularization pattern of survival improvement over time (first time period in Fig. 1A) can be explained by a decline in the average magnitude of external stresses. The parallel shift of the entire survival curve to the right over time (second time period in Fig. 1A) can be explained by the decline in the frequency of external disturbances (see **Supplemental Data** for details). These are parameters in the model characterizing the properties of external disturbances. Note that external factors could also affect other two parameters describing initial vitality and its rate of its decline with age. Although possibility of such influence was discussed by Strehler and Mildvan [19], it was not represented explicitly in terms of parameters of external disturbances. Thus in the framework of SM model observed trends can be explained by changes in parameters of external disturbances.

How the SM model explains differences in survival for groups of individuals with different genetic backgrounds. The explanations given above are not valid for the patterns shown in Fig. 1B. This is because, instead of considering of how changes in external conditions over time influence human survival, we consider how such survival is affected by differences in genetic parameters of individuals taken from the same population cohort (original FHS cohort), and exposed to the same external conditions. Therefore, different age patterns of survival (mortality rates) for these sub-cohorts are likely to be associated with differences in the parameters of the vitality function, which are likely to depend on the genetic backgrounds of the individuals comprising the respective sub-cohorts. The representation of the parameters of the Gompertz mortality curve (see **Supplemental Data**) together with the distinct patterns of survival functions shown in Fig. 1B indicate that the rectangularization pattern of changes in survival, in this case, can be observed if the initial value of vitality increases with an increase in the number of longevity alleles contained in individual genomes. In populations with such genetic backgrounds, the relative rate of decline in vitality remains unchanged, so the absolute rate of decline increases (see **Supplemental Data**).

This connection between genetic changes and modifications of the hypothetical vitality curve, estimated from real data, indicates that changes in the genetic background of individuals may affect dynamic parameters of aging-related changes in physiological indices measured in longitudinal data. The use of the SM model shows what types of effects on the dynamic parameters of the age trajectories of physiological indices can be expected (e.g., improvement in survival may take place with and without changes in the rate of aging-related changes in relevant

biomarkers) when the genetic backgrounds of the study participants change. A better understanding of the roles of such genetic factors in biomarkers of aging may also shed light on the role of gene-environment interactions in survival changes over time (Fig. 1A). New environmental conditions may activate new genes, which may modulate parameters of the vitality curve.

The presence of the SM correlation in the Gompertz parameters is associated with the "rectangularization" pattern of survival improvement [28]. The corresponding decline in mortality rates can be represented by counter clock-wise rotation of the logarithms of the mortality curves around some point, so the logarithm of the parameter *a* declined and the parameter *b* increased. The use of the SM model allows for interesting interpretation of the roles of the longevity alleles in mechanisms responsible for changes in the shape of mortality rate. The effect of counter-clock-wise rotation takes place if the number of such alleles determines the initial value of the vitality function (see Fig. S2). This may indicate that each such allele contributes to an increase in robustness and in resistance to stresses.

Additional insights about possible biological mechanisms responsible for such patterns of changes can be gained by comparing these curves with those resulting from the "saving lives" model [20, 29], where the rectangularization pattern of survival changes corresponds to an increase in the number of times "individuals' lives have been saved." "Saving lives" can be achieved not only by direct life-saving interventions but also by providing living organisms with additional resilience, redundancy, and robustness, which increases their ability to withstand stresses. Analyses of data on factors and conditions experienced by centenarians have led researchers to the same conclusion: High resilience makes a substantial contribution to exceptional longevity in humans [30]. Fig. 4 shows that interventions that save lives once hypothetically applied to the U.S. population in 1950 would transform the 1950-mortality rate close to the U.S. 2007-pattern (sex-specific curves are shown in Fig. S7 in Supplemental Data).

Fig. 4 about here

Note that additional phenotypic traits relevant to our analyses include information on causes of death: cancer, CVD, or other/unknown cause and ages at onset of diseases (cancer, CVD, and diabetes mellitus). The occurrence of CVD, cancer and death in FHS cohorts has been followed through continuous surveillance of hospital admissions, death registries, clinical exams, and other sources, so that all the respective events are included in the study. We used the following data to calculate ages at onset of cancer, CVD, and diabetes mellitus: 1) dates of the first diagnosis of cancer (all sites but skin) from the follow-up data; 2) dates of the first CVD-related event, as defined by the FHS event codes corresponding to myocardial infarction, angina pectoris, stroke (definite cerebrovascular accident (CVA), atherothrombotic infarction, cerebral embolism, intracerebral hemorrhage, subarachnoid hemorrhage), coronary heart disease- or CVA-related death, and congestive heart failure) from the follow-up data; and 3) dates of the first exam (offspring FHS cohort) with diabetic status (a person was defined by the FHS investigators as diabetic in a specific exam if his/her level of fasting blood glucose exceeded 126 mg/dl or he/she indicated diabetes treatment in this exam). We also calculated age at onset of "unhealthy life" as the minimum of ages at onset of these three diseases.

We calculated the number of selected longevity alleles contained in the genomes of genotyped individuals in the original and offspring FHS cohorts. Then we divided genotyped

individuals from the original and offspring FHS cohorts into sub-cohorts corresponding to (i) genotyped individuals carrying between 0 and 22 longevity alleles in their genomes ((≤22)-sub-cohort), and (ii) genotyped individuals carrying between 23 and 39 longevity alleles in their genomes (>22) –sub-cohort). Then for obtained sub-cohorts of individuals and for each age we evaluated probability to survive to this age, probabilities of stay free of cancer, and free of CVD. Calculations of these characteristics for the sub-cohorts of the original FHS cohort were conditional on survival to age 80 years. This is because (>22)-sub-cohort has small number of individuals younger than 80 years. For the members of the offspring FHS cohort the data allow for evaluating corresponding characteristics starting from age 60.

Figure 5A shows graphs of survival functions corresponding to mortality rates from cancer for the two sub-cohorts of individuals from the original FHS cohort conditional on survival to age 80. One can see from this figure that individuals having larger numbers of longevity alleles in their genomes (>22) have substantially better survival than members of the other sub-cohort having smaller number of such alleles (≤22). Figure 5B shows age patterns of survival functions corresponding to mortality from CVD for the two sub-cohorts of individuals from the original FHS cohort conditional on survival to age 80. The figure illustrates that individuals having larger numbers of longevity alleles in their genomes (>22) have substantially better survival than members of the other sub-cohort having smaller number of such alleles (≤22). Figure 5C displays age patterns of survival functions corresponding to total mortality rates for the two sub-cohorts of individuals from the original FHS cohort conditional on survival to age 80 showing that individuals with larger numbers of longevity alleles in their genomes (>22) have substantially better overall survival than members of the other sub-cohort having smaller number of such alleles (≤22).

Fig 5 is about here

Figure 6 shows that genetic variants detected in the original FHS cohort jointly influence total mortality as well as mortality by cause in the offspring FHS cohort (FHSO). Figure 6A shows age patterns of survival functions associated with mortality rate from cancer in the two FHSO subcohorts conditional on survival to age 60. One can see from this panel that survial function in the (>22) sub-cohort is higher than in individuals having smaller number of longevity alleles in their genomes (≤22). Figure 6B shows a similar effect for CVD, and Figure 6C displays survival finctions corresponding to total mortality. Figure 6C reveals that overall survival among individuals from the (>22)-sub-cohort is higher than that from the (0≤22) subcohort having smaller number of longevity alleles in their genomes.

Fig. 6 is about here

Figure 7A illustrates how the number of longevity alleles contained in persons' genomes influences probability of staying free of cancer of all sites but skin in the FHSO cohort. One can see from this panel that study participants having larger number (>22) of longevity alleles in their genomes have better chances to stay free of cancer of all sites but skin than those having smaller number (≤22) of such alleles. Figure 7B displays how the number of longevity genes contained in persons' genomes influences probability of staying free of CVD in the FHSO cohort. It shows that study participants having larger number (>22) of longevity alleles in their genomes have better chances to stay free of CVD than those having smaller number (≤22) of such alleles. Figure 7C shows how the number of longevity genes contained in persons' genomes

influences probability of staying free of diabetes in the FHSO cohort. One can see from this panel that study participants having larger number (>22) of longevity alleles in their genomes have better chances to stay free of diabetes than those having smaller number (≤22) of such alleles. Figure 7D illustrates how the number of longevity genes contained in persons' genomes influences probability of staying free of three major aging related diseases: cancer of all sites but skin, CVD, and diabetes in the FHSO cohort, revealing that study participants having larger number (>22) of longevity alleles in their genomes have better chances to stay free of three major diseases than those having smaller number (≤22) of such alleles.

Fig. 7 is about here

Discussion

Demographers, studying trends in mortality and survival in populations of developed countries, have long had debates about "mortality compression" or "rectangularization of survival curve" – the process which took place in the first half of the 20th century. Although these changes were always linked to improvements in environmental and living conditions as well as advances in health care and medicine, there were no biological explanation of how these improvements got "under the skin" of the individuals in the populations to produce such rather specific trends in observed age patterns of mortality and survival. The biodemographic models described above provide useful insights: biological systems responsible for the human body's resistance to stresses (e.g., heat shock proteins) and resilience (e.g., DNA repair) are likely to be involved. Myers and Manton [28] showed that in the second half of the 20th century the tail of the survival curve in the United States tended to increase across the older years of age. Horiuchi and Wilmoth [31, 32] confirmed an increase of the tail of the life span distribution in the population of the United States. Wilmoth and Horiuchi [33] found that the decline in variability of life span, associated with a rectangularization pattern of changes in survival curves, came to an end around 1950 in Sweden and the United States so the "compression of mortality" concept lost its ability to describe mortality trends. These findings were summarized in papers by Yashin et al. [25, 26] which found that the process of rectangularization of the survival curve, which took place in the first half of the 20th century, was later replaced by an almost parallel shift of the entire survival curve to the right (Fig. 1A).

The use of the SM model in analyses of genetic data shows that genetic factors may modify values and dynamic properties of variables describing aging related transformations in the human body, and these modifications influence life span. Evidence for such influences is provided in a number of epidemiological studies. Port et al. [34] showed that the level of blood glucose affects mortality risk among subjects with cardiovascular disease. Yashin et al. [35] found associations between values of physiological indices at ages between 40 and 60 years and life span. Benetos et al. [36] found that dynamic properties of blood pressure affect mortality risk. Extending these analyses, Yashin et al. [37] also found that not only the values of these variables, but also their dynamic characteristics (e.g., the rate of changes), are associated with life span, and healthy life span. These findings together with the insights from the SM analyses suggest that at least some of the detected associations may be caused by the joint influence of the number of genetic variants individually selected for their effects on health and survival outcomes.

The observed patterns in survival/mortality changes have important interpretations from the point of view of reliability theory. Indeed, the parallel shift of the mortality curve to the right

corresponds to a proportional modification of the corresponding hazard rate. Such changes are expected in the series connection of *N* sub-systems with similar age patterns of failure rates when one or several sub-systems becomes invulnerable (e.g., by providing them with high levels of redundancy, or repair capacity). The rectangularization (mortality counter clock-wise rotation) pattern corresponds to providing a limited redundancy (or limited additional repair capacity) to one or more sub-systems [38]. This analogy is consistent with systems biology approaches [39] to studying aging and longevity with identification of corresponding systems blocks, connections, reserve capacities and repair mechanisms at different levels of an organism's biological organization. The presence of an "almost parallel shift" of survival curve to the right in Fig. 1A (showing changes in survival over time), and its absence in Fig. 1B (showing dependence of survival on genetic background), may indicate higher plasticity of mortality and survival in response to changes in external factors compared to changes in genetic background: The relatively low estimates of narrow sense heritability in life span (about 25%) support this view, and encourage the search for external factors capable of further improvement in human health and life span.

Improvements in environmental conditions (including changes in nutritional and living conditions, the use of new drugs, and new treatment procedures) are likely to activate new genes and modify metabolic pathways which contribute increases in life span. Some of these genes may have large effects on life span; evidence to date from GWAS indicates that many others are likely to have small individual contributions. Accordingly, it could be hypothesized that the patterns shown in Fig. 1A reflect the contribution of genetic changes as well, resulting from the activation of a large number of small-effects longevity alleles in response to the changing environmental conditions across the past two centuries. Strehler and Mildvan [19] envisaged such possibility by allowing the relative rate of vitality decline to depend on environmental factors as well. Being confirmed, this hypothesis will improve our understanding of how life span is regulated by external factors, and genetic analyses will help identify multiple genes which are likely to be collectively involved in such regulation. So the dependence of the parameters of the Gompertz mortality curves on characteristics of external stresses in the SM model is a useful simplification which was relevant for explaining SM correlation in the pre-genomic era. New models capturing specific effects of gene/environment interactions and describing genetic mechanisms mediating external influences on aging related changes, health and survival outcomes are needed to properly explain the response of the body and phenotypic traits to changes in external factors.

Note that even if the patterns of differences in survival functions in the two panels in Fig. 1 look similar, and can be explained in terms of improved robustness and resilience, the mechanisms responsible for such changes are not necessarily the same. An improvement in survival *over time* involves the influence of advancing health care and medical technology (e.g., proper access to emergency care, implantation of pacemakers, performing by-pass surgery, etc.), which could extend life without affecting genetic mechanisms, for example, by increasing the reliability of functioning in certain biological organs, or subsystems. Currently, more than 500,000 people in the U.S. have implanted pacemakers and over 100,000 people get new pacemakers implanted per year. The rate of implanting is higher in the elderly with over 85% of implants received by those over age 65. The results of this paper provide researchers with new insights about the roles of advances in health care, medical technology, and medicine in survival

improvements: medical interventions today may substantially compensate for limitations of genetically-based vitality mechanisms and increase life span.

The existence of additive genetic effect on lifespan explored in our study is in agreement with basic principles of quantitative genetics considering complex trait as a function of many genetic and non-genetic factors. The additive genetic component of phenotypic trait is responsible for transmission of a trait through generations and plays the key role in evolutionary theory of complex traits. It is also responsible for the narrow sense heritability of such traits, which in case of lifespan is about 25% [40, 41].

The additive joint influences of genetic variants on the risk of various diseases have been tested in several other studies using an aggregated index called "genetic or polygenic score" [42-47]. Similar to our indices, the construction of genetic score functions involves weighted and un-weighted sums of indicator functions of the genetic variants. These studies did not provide biological justifications for their score functions, and they used genetic variants detected in different studies and often in different populations. The results of such "meta-analyses", however, should be used with care, because the genes affecting the studied traits might be sensitive to external conditions. In GWA studies of schizophrenia and a number of other psychiatric disorders often none of genetic variants reached genome-wide significance. In such cases, substantially relaxed p-value thresholds are often used in allele selection procedures [48]. Selected small-effect-low-significant genetic variants are used to construct genetic scores, and test their influence on disease traits. Accordingly, more work is needed on the selection of influential alleles and the evaluation of regularities of their joint influence on health-related traits.

The results of our analyses show that genetic factors jointly affecting life span also jointly influence onsets of several major age-associated diseases. The index used for quantifying joint genetic influence called "genetic dose" (the non-weighted sum of genetic variants contained in human genome) is a surrogate measure of additive genetic component (AGC) of phenotypic trait, which is a weighted sum of genetic variants contained in persons' genomes with estimated effect sizes used as weights. The normalized (i.e., having values between 0 and 1) indices of genetic dose and AGC explain about the same portion of life span variance. It is natural to expect that the additive genetic component evaluated from the genome wide data influences life span, because life span is a heritable trait. Moreover, additive genetic component could be evaluated for any heritable trait when SNP map is dense enough.

The fact that SNP alleles selected for their effect on life span jointly influence major diseases suggests an idea of using this property in the two-step GWAS of complex diseases when data on life span and health histories are available for the same individuals. The first step, selecting genetic variants for their effect on longevity, could be considered as a procedure for prioritizing SNP alleles for studying genetics of complex disorders. Since this procedure dramatically reduces the number of SNPs compared to the initial set of SNPs, the p-value threshold corrected for multiple testing at the second stage of selecting SNP alleles affecting disease trait could be relaxed.

The fact that the identified SNPs jointly influence several major diseases suggests that respective genes might affect these diseases through non-disease-specific, aging related, mechanisms (e.g., they could influence overall stress resistance) and thus they might provide a

link between genetic regulation of aging, vulnerability to diseases, and lifespan. One should note that some of the loci previously identified in GWAS of diseases also appear to be involved in several diseases (10). For example, Zhernakova et al. (11) provided evidence for shared genetics and pathogenesis for autoimmune and inflammatory diseases. Easton and Eeles (12), Rafnar et al. (13), Sakamoto et al. (14), and Wu et al. (15) showed that the variants in a number of loci are linked to several cancers. Gudmundsson et al. (16) provided evidence on opposite effects of some genetic variants on prostate cancer and type 2 diabetes (T2D). Several SNPs in 9p21 locus were found to be associated with a number of diseases including various cancers, cardiovascular diseases (CVDs), and T2D (17-19).

Our analyses showed that the joint influence of longevity alleles on age at disease onset is smaller than that on life span. This may indicate that the identified genes (or some of them) are likely to contribute to life span beyond pathways dealing with specific disease development. Some of such genes might affect aging rate, and modulate age at onset of clinical frailty instead of disease. Some others might make major contribution to body's resilience, i.e., to the ability of coping with disease when it strikes, and have much smaller effects (if at all) on body's "robustness", i.e., protection against disease development.

In our 2010 paper (9) we provided summary of functions (known or suggested) for genes closest to the 39 longevity SNPs and found enrichment with genes related to cell adhesion (see Table 1 of respective paper). Further analysis of literature revealed that relevant genes are also broadly involved in cancer, proliferation, response to damage, and brain signaling, with individual genes often involved in several processes (e.g., CDH4, TGF alpha, STK24), so that their effects on longevity may involve both aging and disease related mechanisms. It is in principle possible that most of the effects of longevity genes on disease risks are realized through non-specific aging related mechanisms (e.g., by affecting speed of stress response, which depends on rates of proliferation, metabolism, and information processing in body) that similarly affect development of many diseases, rather than through disease-specific mechanisms. It is also useful to check whether genetic variants which are primarily associated with specific diseases will also jointly influence life span.

Few existing studies actually addressed this question. Beekman et al. (20) investigated joint effect of 30 SNP disease-associated alleles on human longevity in data collected in Leiden Longevity Study (LLS) and Leiden 85 Plus Study. Note that these alleles were detected in other (not in Leiden's) GWA studies. The authors found that the average number of disease alleles contained in the genomes of individuals of different age does not depend on age, which was interpreted as that the number of selected disease-alleles contained in persons' genomes does not compromise longevity. They also found that none of the selected alleles significantly affected life span when accounting for multiple testing. One reason could be that these same alleles (or some of them) might not be the risk alleles in the study population. The assumed effects of some of the selected alleles on disease risks were not confirmed in other studies (e.g., rs564398 was not confirmed in meta-analysis by Bao et al. (21); and rs10497721 was not replicated by the same team that originally found it (22,23)). This controversy raises an important question on replicability of genetic findings. Is the failure to replicate an indicator of false positive discovery, or it is a fundamental property of genetic compensatory and adaptive mechanisms which could activate different genes and some time different pathways from genetic network involved in trait regulation in response to difference in conditions? In this latter case the

requirement to replicate the association of genetic variant with a trait using data on independent population does not have solid theoretical background. It is probably more reasonable to consider replication of the pathways leading to the same outcome (e.g., a particular aging or longevity phenotype), rather than individual SNPs or genes.

The results of our analyses indicate the presence of genetic mechanisms responsible for connection among different diseases. Analyses of health related data on incidence, prevalence, and cause specific mortality indicate that genetic effects on disease development can be modulated by external factors, resulting in distinct patterns of disease dependence observed in different study populations. Using multiple causes of death data Yashin et al. (24) showed significant time trend in negative correlations between cancer and a number of other diseases. Ukraintseva et al. (25) showed that manifestation of trade-offs between risks of cancer and other diseases depends on sex, age, and study population. Understanding such mechanisms as well as causes of disease dependence could substantially improve demographic predictions of life expectancy at birth resulted from eradication, or reduction of mortality from selected causes (24). Traditional approaches to calculating consequences of such interventions assume that chronic disorders are independent. The results of our analyses confirm that dependence between diseases has also genetic background.

Experimental studies of aging and longevity using laboratory animals show that genetic and dietary interventions can simultaneously improve health and increase life span. Positive association between health and longevity has also been detected in a number of human studies. Newman et al. (26) used data from Long Life Family Study (LLFS) to test whether the recruitment targeting longevity resulted in a cohort of individuals with better health and function. The authors found that diabetes, chronic pulmonary disease and peripheral artery disease tended to be less common in LLFS probands and offspring compared to similar aged persons in the other cohorts. Physiological risk factors and functioning were also better in LLFS population.

Barral et al. (27) assessed cognitive performance in the combined offspring of the LLFS probands and their siblings as well as in their spouses. The results indicate that LLFS family members in the offspring generation demonstrate significantly better performance on multiple tasks requiring attention, working memory, and semantic processing when compared with individuals without a family history of exceptional survival.

Kulminski et al. (28) found a pleiotropic effect of the APOE e4 allele. It predisposes carriers of this allele to early onset of CVD but postpones cancers to older ages. Barzilai and Gabriely (29) showed that genetic factors that are associated with human longevity are heritable and may contribute not only to quantitative longevity but also to protection from age-dependent disease and promotion of exceptional health.

The results of studying genetic mechanisms, which increase life span by protecting from major aging related human diseases, will provide important insights for strategies of improvement of population health. Better understanding of factors and mechanisms responsible for body's robustness will play the key role in developing preventive measures. Investment in disease prevention will contribute to reduction of medical (e.g., Medicare) costs. The new

knowledge about genetic mechanisms of body's resilience will provide useful insights on development of more efficient medications and treatment procedures aiming to reach complete recovery from disease.

Acknowledgements

The FHS project is conducted and supported by the NHLBI in collaboration with Boston University (N01 HC25195). The FHS data used for the analyses were obtained through dbGaP (phs000007.v3.p2). The authors acknowledge the investigators that contributed the phenotype and genotype data for this study. This manuscript was not prepared in collaboration with investigators of the FHS and does not necessarily reflect the opinions or views of the FHS, Boston University, or the NHLBI. This work was partly supported by NIH/NIA grant R01AG030612. The authors acknowledge the University of California, Berkeley (USA), and the Max Planck Institute for Demographic Research (Germany) for developing and maintaining Human Mortality Database (available at www.mortality.org or www.humanmortality.de; data downloaded on 07/06/2011).

Author Disclosure Statement

No competing financial interests exist.

References:

1. Hardy J, Singleton A. Genomewide Association Studies and Human Disease. *New Engl. J. Med.* 2009;360(17):1759-1768.

2. Maher B. Personal genomes: The case of the missing heritability. *Nature.* 2008;456(7218):18-21.

3. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009;461(7265):747-753.

4. Slatkin M. Epigenetic Inheritance and the Missing Heritability Problem. *Genetics.* 2009;182(3):845-850.

5. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era - concepts and misconceptions. *Nat. Rev. Genet.* 2008;9(4):255-266.

6. Anselmi CV, Malovini A, Roncarati R, et al. Association of the FOXO3A Locus with Extreme Longevity in a Southern Italian Centenarian Study. *Rejuvenation Res.* 2009;12(2):95-103.

7. Flachsbart F, Caliebeb A, Kleindorp R, et al. Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc. Natl. Acad. Sci. U. S. A.* 2009;106(8):2700-2705.

8. Willcox BJ, Donlon TA, He Q, et al. FOXO3A genotype is strongly associated with human longevity. *Proc. Natl. Acad. Sci. U. S. A.* 2008;105(37):13987-13992.

9. Zeng Y, Cheng LG, Chen HSA, et al. Effects of FOXO Genotypes on Longevity: A Biodemographic Analysis. *J. Gerontol. A. Biol. Sci. Med. Sci.* 2010;65(12):1285-1299.

10. Lunetta KL, D'Agostino RB, Sr., Karasik D, et al. Genetic correlates of longevity and selected age-related phenotypes: a genome-wide association study in the Framingham Study. *BMC Med. Genet.* 2007;8(Suppl. 1):S13.

11. Newman AB, Walter S, Lunetta KL, et al. A Meta-analysis of Four Genome-Wide Association Studies of Survival to Age 90 Years or Older: The Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. *J. Gerontol. A. Biol. Sci. Med. Sci.* 2010;65(5):478-487.

12. Walter S, Atzmon G, Demerath EW, et al. A genome-wide association study of aging. *Neurobiol. Aging.* 2011;32(11):2109.e2115-2109.e2128

13. Deelen J, Beekman M, Uh H-W, et al. Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell.* 2011;10(4):686-698.

14. Nebel A, Kleindorp R, Caliebe A, et al. A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. *Mech. Ageing Dev.* 2011;132(6-7):324-330.

15. Yashin AI, Wu DQ, Arbeev KG, Ukraintseva SV. Joint influence of small-effect genetic variants on human longevity. *Aging.* 2010;2(9):612-620.

16. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. *ScientificWorldJournal.* 2001;1:323-336.

17. Kulminski A, Yashin A, Ukraintseva S, et al. Accumulation of health disorders as a systemic measure of aging: Findings from the NLTCS data. *Mech. Ageing Dev.* 2006;127(11):840-848.

18. Yashin AI, Arbeev KG, Kulminski A, et al. Health decline, aging and mortality: how are they related? *Biogerontology.* 2007;8(3):291-302.

19. Strehler BL, Mildvan AS. General theory of mortality and aging. *Science.* 1960;132(3418):14-21.

20. Vaupel JW, Yashin AI. Repeated resuscitation: how lifesaving alters life tables. *Demography.* 1987;24(1):123-135.

21. Oeppen J, Vaupel JW. Broken limits to life expectancy. *Science.* 2002;296(5570):1029-1031.

22. Finkelstein MS. Lifesaving explains mortality decline with time. *Math. Biosci.* 2005;196(2):187-197.

23. Gavrilov LA, Gavrilova NS. *The Biology of Life Span: A Quantitative Approach*. New York: Harwood Academic Publisher, 1991.

24. Zheng H, Yang Y, Land KC. Heterogeneity in the Strehler-Mildvan General Theory of Mortality and Aging. *Demography.* 2011;48(1):267-290.

25. Yashin AI, Begun AS, Boiko SI, et al. The new trends in survival improvement require a revision of traditional gerontological concepts. *Exp. Gerontol.* 2001;37(1):157-167.

26. Yashin AI, Begun AS, Boiko SI, et al. New age patterns of survival improvement in Sweden: do they characterize changes in individual aging? *Mech. Ageing Dev.* 2002;123(6):637-647.

27. Riggs JE, Millecchia RJ. Using the Gompertz-Strehler model of aging and mortality to explain mortality trends in industrialized countries. *Mech. Ageing Dev.* 1992;65(2-3):217-228.

28. Myers GC, Manton KG. Compression of mortality: myth or reality. *Gerontologist.* 1984;24(4):346-353.

29. Yashin AI, Iachine IA, Begun AS. Mortality modeling: A review. *Mathematical Population Studies.* 2000;8(4):305–332.

30. Zeng Y, Shen K. Resilience Significantly Contributes to Exceptional Longevity. *Current Gerontology and Geriatrics Research.* 2010;2010:525693.
31. Horiuchi S, Wilmoth JR. Age patterns of the life table aging rate for major causes of death in Japan, 1951-1990. *J. Gerontol. A. Biol. Sci. Med. Sci.* 1997;52(1):B67-B77.
32. Horiuchi S, Wilmoth JR. Deceleration in the age pattern of mortality at older ages. *Demography.* 1998;35(4):391-412.
33. Wilmoth JR, Horiuchi S. Rectangularization revisited: Variability of age at death within human populations. *Demography.* 1999;36(4):475-495.
34. Port SC, Boyle NG, Hsueh WA, et al. The predictive role of blood glucose for mortality in subjects with cardiovascular disease. *Am. J. Epidemiol.* 2006;163(4):342-351.
35. Yashin AI, Akushevich IV, Arbeev KG, et al. Insights on aging and exceptional longevity from longitudinal data: novel findings from the Framingham Heart Study. *Age.* 2006;28(4):363-374.
36. Benetos A, Zureik M, Morcet J, et al. A decrease in diastolic blood pressure combined with an increase in systolic blood pressure is associated with a higher cardiovascular mortality in men. *J. Am. Coll. Cardiol.* 2000;35(3):673-680.
37. Yashin AI, Arbeev KG, Akushevich I, et al. Dynamic determinants of longevity and exceptional health. *Current Gerontology and Geriatrics Research.* 2010;2010:381637.
38. Barlow RE, Proschan F. *Mathematical Theory of Reliability*. New York, NY: John Wiley and Sons, Inc., 1996.
39. Fuellen G, Adjaye J, de Grey A, et al. Bioinformatics in Aging Research: A Workshop Report. *Rejuvenation Res.* 2010;13(6):763-767.
40. McGue M, Vaupel JW, Holm N, Harvald B. Longevity is moderately heritable in a sample of Danish twins born 1870-1880. *J. Gerontol.* 1993;48(6):B237-B244.
41. Herskind AM, McGue M, Holm NV, et al. The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870-1900. *Hum. Genet.* 1996;97(3):319-323.
42. Meigs JB, Shrader P, Sullivan LM, et al. Genotype Score in Addition to Common Risk Factors for Prediction of Type 2 Diabetes. *New Engl. J. Med.* 2008;359(21):2208-2219.
43. Paynter NP, Chasman DI, Pare G, et al. Association Between a Literature-Based Genetic Risk Score and Cardiovascular Events in Women. *JAMA.* 2010;303(7):631-637.
44. Reeves GK, Travis RC, Green J, et al. Incidence of Breast Cancer and Its Subtypes in Relation to Individual and Multiple Low-Penetrance Genetic Susceptibility Loci. *JAMA.* 2010;304(4):426-434.
45. Ruiz JR, Gomez-Gallego F, Santiago C, et al. Is there an optimum endurance polygenic profile? *J. Physiol. (Lond).* 2009;587(7):1527-1534.
46. Talmud PJ, Hingorani AD, Cooper JA, et al. Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *Br. Med. J.* 2010;340:b4838.
47. Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum. Mol. Genet.* 2009;18(18):3525-3531.
48. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 2009;460(7256):748-752.

References:

1. Ku CS, Loy EY, Pawitan Y, Chia KS. The pursuit of genome-wide association studies: where are we now? *J Hum Genet*. 2010;55:195-206.
2. Goldstein DB. Common Genetic Variation and Human Traits. *New Engl J Med*. 2009;360:1696-1698.
3. Hirschhorn JN. Genomewide Association Studies - Illuminating Biologic Pathways. *New Engl J Med*. 2009;360:1699-1701.
4. Kraft P, Hunter DJ. Genetic Risk Prediction - Are We There Yet? *New Engl J Med*. 2009;360:1701-1703.
5. Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum Mol Genet*. 2009;18:3525-3531.
6. Peterson RE, Maes HH, Holmans P, et al. Genetic risk sum score comprised of common polygenic variation is associated with body mass index. *Hum Genet*. 2011;129:221-230.
7. Demirkan A, Amin N, Isaacs A, et al. Genetic architecture of circulating lipid levels. *Europ J Hum Genet*. 2011;19:813-819.
8. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748-752.
9. Yashin AI, Wu DQ, Arbeev KG, Ukraintseva SV. Joint influence of small-effect genetic variants on human longevity. *Aging*. 2010;2:612-620.
10. Cho JH. The genetics and immunopathogenesis of inflammatory bowel disease. *Nature Reviews Immunology*. 2008;8:458-466.
11. Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat Rev Genet*. 2009;10:43-55.
12. Easton DF, Eeles RA. Genome-wide association studies in cancer. *Hum Mol Genet*. 2008;17:R109-R115.
13. Rafnar T, Sulem P, Stacey SN, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat Genet*. 2009;41:221-227.
14. Sakamoto H, Yoshimura K, Saeki N, et al. Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat Genet*. 2008;40:730-740.
15. Wu XF, Ye YQ, Kiemeney LA, et al. Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer. *Nat Genet*. 2009;41:991-995.
16. Gudmundsson J, Sulem P, Steinthorsdottir V, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet*. 2007;39:977-983.
17. Shete S, Hosking FJ, Robertson LB, et al. Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet*. 2009;41:899-904.
18. Bishop DT, Demenais F, Iles MM, et al. Genome-wide association study identifies three loci associated with melanoma risk. *Nat Genet*. 2009;41:920-925.
19. Stacey SN, Sulem P, Masson G, et al. New common variants affecting susceptibility to basal cell carcinoma. *Nat Genet*. 2009;41:909-914.
20. Beekman M, Nederstigt C, Suchiman HED, et al. Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity. *Proc Natl Acad Sci U S A*. 2010;107:18046-18049.
21. Bao XY, Xie C, Yang MS. Association between Type 2 Diabetes and CDKN2A/B: a meta-analysis study. *Mol Biol Rep*. 2011:in press.

22. Meigs JB, Manning AK, Fox CS, et al. Genome-wide association with diabetes-related traits in the Framingham Heart Study. *BMC Med Genet*. 2007;8:Article No.: S16.
23. Florez JC, Manning AK, Dupuis J, et al. A 100K genome-wide association scan for diabetes and related traits in the Framingham Heart Study. *Diabetes*. 2007;56:3063-3074.
24. Yashin AI, Ukraintseva SV, Akushevich IV, Arbeev KG, Kulminski A, Akushevich L. Trade-off between cancer and aging: What role do other diseases play? Evidence from experimental and human population studies. *Mech Ageing Dev*. 2009;130:98-104.
25. Ukraintseva SV, Arbeev KG, Akushevich I, et al. Trade-Offs Between Cancer and Other Diseases: Do They Exist and Influence Longevity? *Rejuvenation Research*. 2010;13:387-396.
26. Newman AB, Walter S, Lunetta KL, et al. A Meta-analysis of Four Genome-Wide Association Studies of Survival to Age 90 Years or Older: The Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. *J Gerontol A Biol Sci Med Sci*. 2010;65:478-487.
27. Barral S, Cosentino S, Costa R, et al. Cognitive function in families with exceptional survival. *Neurobiol Aging*. 2011:in press, doi:10.1016/j.neurobiolaging.2011.1002.1004.
28. Kulminski AM, Culminskaya I, Ukraintseva SV, et al. Trade-off in the effects of the apolipoprotein E polymorphism on the ages at onset of CVD and cancer influences human lifespan. *Aging Cell*. 2011;10:533-541.
29. Barzilai N, Gabriely I. Genetic Studies Reveal the Role of the Endocrine and Metabolic Systems in Aging. *J Clin Endocrinol Metab*. 2010;95:4493-4500.

30. Hindorff LA, Junkins HA, Hall PN, Mehta JP, and Manolio TA. A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies. Accessed June 25, 2011 (1212 GWA at p<5x10-8 for 210 traits published through 12/2010).

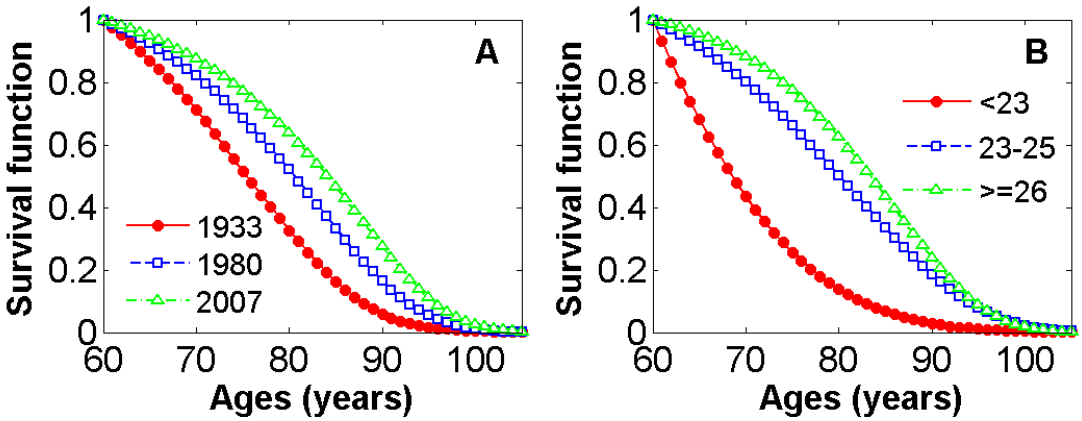

Fig. 1: Different factors produce similar patterns of changes in survival. **(A)** Survival curves (conditional at age 60) in the U.S. total population (both sexes) in years 1933-2007 (data source: Human Mortality Database). **(B)** Patterns of changes in survival of carriers of different numbers of longevity alleles detected in our GWAS of the original FHS cohort corresponding to Gompertz approximations of corresponding mortality curves.

.

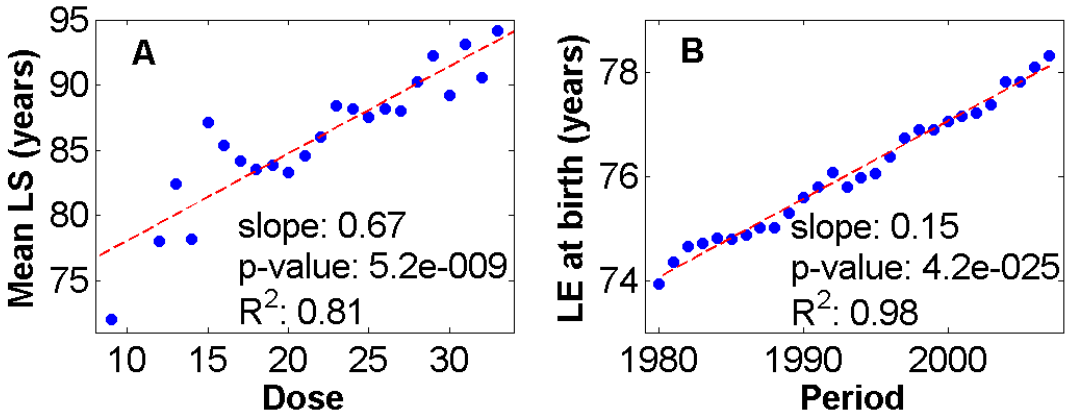

Fig. 2: Different factors produce similar changes in mean life span. **(A)** The "genetic dose – phenotypic response" relationship between the numbers of selected "low-effect longevity" alleles (39 total) contained in individuals' genome and mean life span of individuals carrying a given number of longevity SNPs in their genomes (analyses of 500K SNP data, original FHS cohort). Dots represent observed data; dashed line represents the fitted linear regression. Longevity alleles were selected using a linear regression procedure, which involved comparison of characteristics of life span distributions among carriers and non-carriers of each of 500K genetic variants. **(B)** Life expectancy at birth in the U.S. total population (both sexes), 1980-2007 (data source: Human Mortality Database). Dots represent observed data; dashed line represents the fitted linear regression.

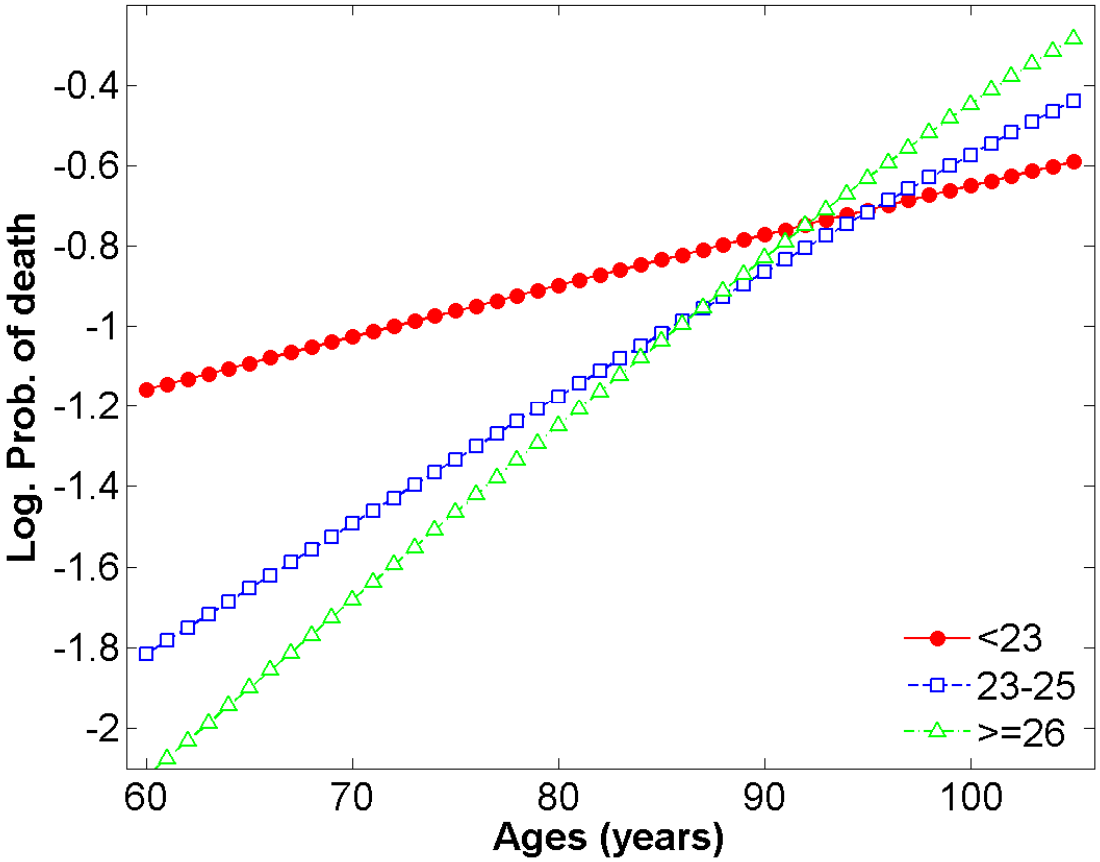

Fig. 3: Mortality rates for populations with different genetic background show SM-correlation. The logarithms of estimated mortality rates approximated by the Gompertz curves in groups of FHS original cohort members having different numbers of low-effect longevity alleles in their genomes

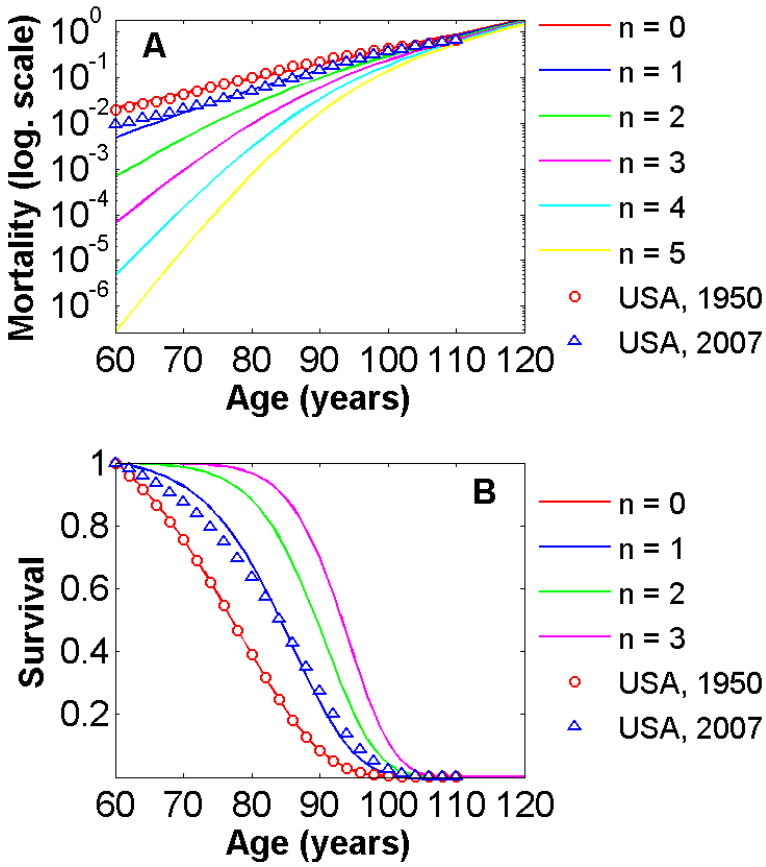

Fig. 4: Saving lives contributes to explanations of mortality and survival improvement. **(A)** Logarithms of mortality rates for U.S. total population (females and males) in 1950 (open circles) and in 2007 (open triangles). The 1950 mortality rate is approximated by the Gompertz function with $n = 0$, where n is the number of times individual's life has been saved. The other curves correspond to mortality rates obtained by transforming the 1950 Gompertz mortality rate using the life saving equation with $n = 1$, 2, 3, 4, and 5. **(B)** Survival functions for U.S. total population (females and males) in 1950 (open circles) and in 2007 (open triangles). Three other survival curves correspond to mortality rates obtained by transforming the 1950 Gompertz curve using the life saving equation with $n = 1$, 2, and 3.

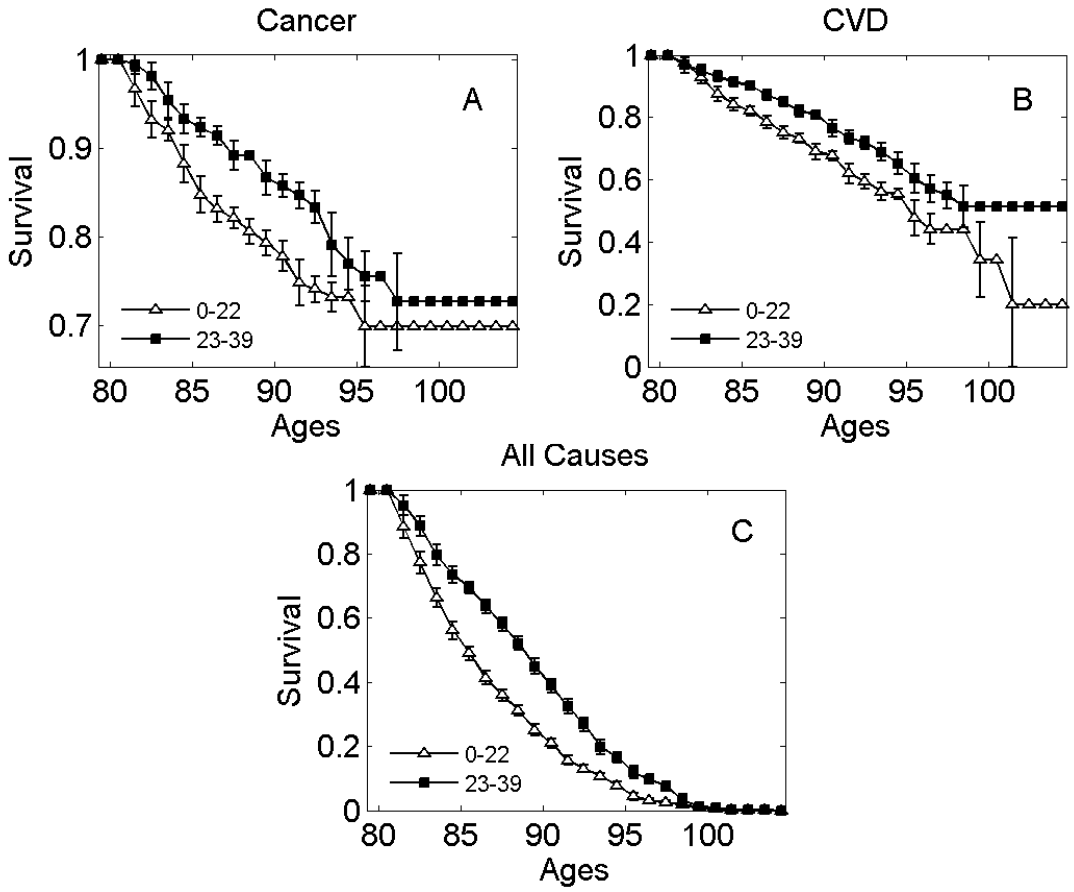

Fig. 5: Survival functions corresponding to cause-specific and total mortality for two sub-cohorts of individuals from the *original* FHS cohort (those having 0-22 longevity alleles vs. carriers of 23-39 longevity alleles), conditional on survival to age 80: A) Survival functions for mortality from cancer; B) Survival functions for mortality from CVD; C) Survival functions for total mortality

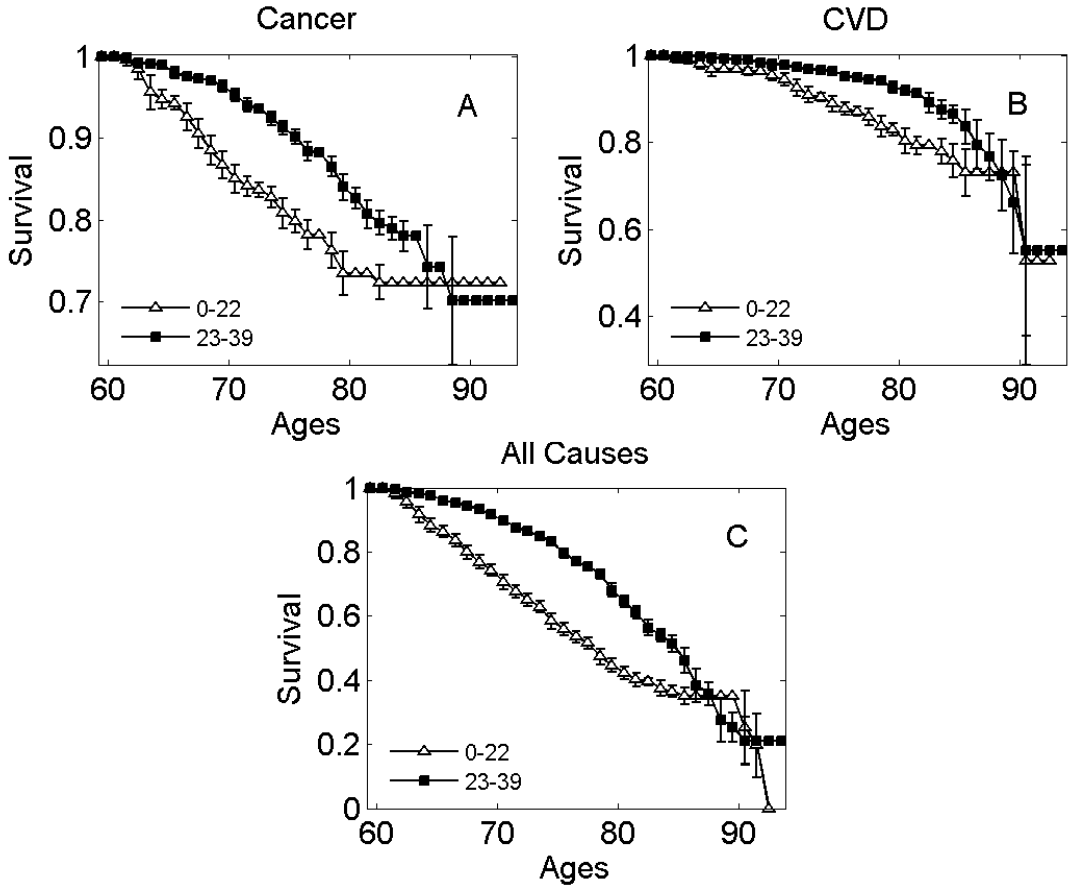

Fig 6: Survival functions corresponding to cause-specific and total mortality for two sub-cohorts of individuals from the *offspring* FHS cohort (those having 0-22 longevity alleles vs. carriers of 23-39 longevity alleles, identified in the *original* cohort), conditional on survival to age 60: A) Survival functions for mortality from cancer; B) Survival functions for mortality from CVD; C) Survival functions for total mortality

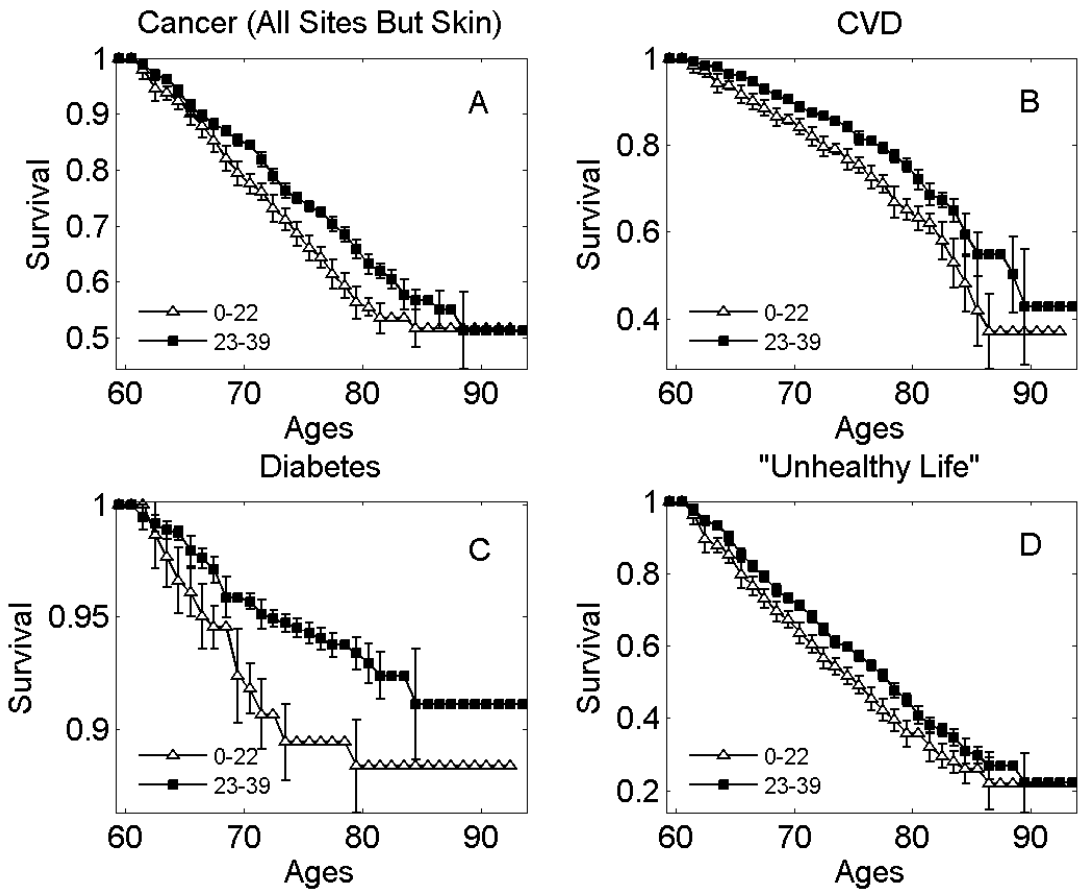

Fig. 7: Probabilities of staying free of different diseases for two sub-cohorts of individuals from the *offspring* FHS cohort (those having 0-22 longevity alleles vs. carriers of 23-39 longevity alleles), conditional on staying free of the disease to age 60: A) Probabilities of staying free of cancer (all sites but skin); B) Probabilities of staying free of CVD; C) Probabilities of staying free of diabetes; D) Probabilities of staying free of "unhealthy life" (defined as the onset of cancer (all sites but skin), CVD, or diabetes)

SUPPLEMENTAL DATA

Index for measuring additive genetic contributions to life span. The additive genetic component of phenotypic trait is a well established notion. Evolutionary models of phenotypic

traits, theoretical principles of quantitative genetics, breeding experiments, as well as many other aspects related to the transmission of genetic effects through generations involve this notion. In the pre-genomic era, the effects of additive genetic components of phenotypic traits were estimated indirectly using data on related individuals. The availability of genome wide data nowadays allows for direct evaluation of the respective effects. To do this, denote by B the set of 39 SNP alleles, i.e., $SNP_i \in B$ if SNP_i was selected in the allele selection procedure in Yashin et al. [1] ; and let $\hat{\beta}_i$ be the effect size of SNP_i, i = 1, 2, …, 39, estimated in that procedure. Denote by $B_j \subseteq B$ the subset of B consisting of SNP alleles contained in the genome of the j^{th} individual, j=1, 2,…, 954. The additive genetic component (AGC-index) of the life span of the j^{th} individual, G_j, can be represented as a weighted sum of indicators, $I(SNP_i \in B_j)$, $SNP_i \in B$, with normalized weights, $\hat{\beta}_i \Big/ \sum_{k=1}^{N} \hat{\beta}_k$, $N = 39$:

$$G_j = \sum_{i \in B} \left(\hat{\beta}_i \Big/ \sum_{k \in B} \hat{\beta}_k \right) I(SNP_i \in B_j) . \qquad \text{(S1)}$$

This function is sometimes called the "genetic or polygenic score" [2-6]. Thus, by the definition of the additive genetic component, the life span of the j^{th} individual (LS_j) can be represented as

$$LS_j = \alpha_0 + \alpha_1 G_j + E_j . \qquad \text{(S2)}$$

Here E_j denotes the value of a random variable for individual j representing the environmental component of life span with an expected value of zero. The function G_j is constructed from genetic data, and is treated as an observed covariate. The parameters α_0 and α_1 have to be estimated from the data. The percent of phenotypic variance explained by the estimated relationship, $\hat{LS}_j = \hat{\alpha}_0 + \hat{\alpha}_1 G_j$, can be used as the measure of goodness of fit.

Note that the values of the effect sizes for individual SNP alleles are never known exactly. Their estimates as well as the *p*-values of these estimates may depend substantially on the statistical model describing the connection between the genetic variant and the phenotype of interest, and used in the allele selection procedure. We verified this statement empirically by selecting longevity alleles using Cox, logistic, and linear regressions, the GEE method, and mixed effects models, in the allele selection procedure. The use of these methods resulted in different sets of SNP alleles and in different estimates of the effect sizes of SNP alleles from the overlapping sets. Therefore, the values of the genetic scores, the estimates of α_0 and α_1, as well as the percentages of explained phenotypic variance characterizing genetic contributions to life span, also differed from one model to the next. This dependence on the methods used in the allele selection procedure may jeopardize the validity of the research results concerning the strengths of genetic influences on a trait.

In view of this finding, we proceeded as follows. First, we used the set *B* of 39 SNP longevity alleles selected in Yashin et al. [1]. Second, we extended the notion of "genetic dose" widely used in the one-locus-models of quantitative genetics to quantify additive genetic influence, to the multi-loci situation. Note that formally the genetic dose index can be obtained by summing nearly equal contributions of different alleles to life span (i.e., for longevity alleles *i* and *k*, this implies $\hat{\beta}_i \approx \hat{\beta}_k$). In this case, the genetic score function AGC, for the j^{th} individual simplifies to $G_j = n_j = \frac{N_j}{N} = \frac{1}{N}\sum_{i \in B} I(SNP_i \in B_j)$, i.e., becomes proportional to the numbers of longevity variants contained in person's genome (NLV-index). In this case the life span of the j^{th} individual can be represented as follows:

$$LS_j = \alpha_0 + \alpha_1 n_j + E_j . \qquad \text{(S3)}$$

Surprisingly, the percentages of phenotypic variance explained by representations (S2) and (S3) were about 15% in each case (see Fig. S1). This similarity supports use of the proportion of genetic variants contained in each subject's genome as convenient characteristics for quantifying the person's genetic background.

Polygenic score indices in the analyses of complex traits. Note that versions of additive genetic components (also called polygenic score indices, or genetic risk scores in some papers) have been intensively tested in a number of genetic studies of complex traits. In some studies, including Machiela et al. [7], effects were not found. In many others effects of polygenic scores on the traits of interest were clearly demonstrated, and its usefulness for genetic analyses was emphasized. For example, Reeves et al. [8] studied how the risk of breast cancer and its subtypes depend on low-penetrance susceptibility loci, individually or in combination. The authors used 14 SNPs previously linked to the disease to construct several polygenic risk scores. The analyses showed that the polygenic risk score was substantially more predictive of estrogen receptor (ER)-positive than of ER-negative breast cancer, particularly for absolute risk. Witte and Hoffman [9] used a polygenic model in GWAS of prostate and breast cancer. The authors showed that the polygenic model can explain an increasing-albeit low-amount of heritability for both of these cancers, even when excluding the most statistically significant associations. In addition, nonaggressive prostate cancer and breast cancer appear to share a common polygenic model, potentially reflecting a similar underlying biology. The authors concluded that their results support the further development and application of polygenic models to genomic data. Qi et al. [10] constructed a genetic risk score to investigate genetic influence on myocardial infarction (MI). The authors found that discrimination of MI was significantly improved when the genetic risk score was added to a model including clinical predictors. Chen et al. [11] investigated the effects of a genetic risk score on psoriasis. The authors found that such a score combining 10 psoriasis risk loci captured significantly more risk than any individual SNP and was associated with early onset of disease and a positive family history. Hivert et al. [12] found that in multivariate-adjusted models, the genetic risk score was significantly associated with increased risk of progression to diabetes.

Effect of polygenic score on lifespan in the presence of information on smoking. In order to test whether joint influence of genetic variants on lifespan will remain significant in the presence of observed covariates we performed regression analyses of lifespan considered as function of our polygenic score constructed from 39 longevity SNPs and smoking status (never smoked or ever smoked). The results are shown in Table S1. One can see from this table that the effect of polygenic score remains highly significant.

How the SM model explains observed patterns in survival improvements over time. One of the key variables in the SM model is "vitality," $V(x)$, where x denotes age. The decline in vitality with age is described by a linear function and interpreted as a reduction in capacity to withstand stresses associated with aging:

$$V(x) = V_0(1 - Bx) . \qquad \text{(S4)}$$

External disturbances or challenges to survival in the SM model are described by a Poisson-like stochastic process, which is characterized by two parameters: the frequency, K, and the average magnitude of stresses, εD, respectively (here we follow the original notation by Strehler and Mildvan). The function $V(x)$ is characterized by the intercept, or initial value of this index, V_0, and the negative slope, V_0B. The SM model represents parameters of the Gompertz mortality curve $\mu(x) = R_0 \exp(\alpha x)$ (which typically gives a good fit to population patterns of human mortality rates between ages 30 and 85 years) in terms of V_0, B, K, and εD:

$$R_0 = K \exp(-V_0/\varepsilon D); \qquad \alpha = V_0 B/\varepsilon D. \tag{S5}$$

In the framework of the SM model, the observed rectangularization pattern of survival improvement over time (first time period in Fig. 1A) can be explained by a decline in average magnitude of external stresses, εD. The parallel shift of the entire survival curve to the right over time (second time period in Fig. 1A) can be explained by the decline in the frequency of external disturbances, K. This is because εD, and K are the only parameters characterizing properties of external disturbances. Changes in V_0 and B are not expected, because these parameters represent properties of individual genetic backgrounds, which require evolutionary time for significant changes.

How the SM model explains differences in survival for groups of individuals with different genetic backgrounds. The explanations given above are not valid for the patterns shown in Fig. 1B. This is because, instead of considering how changes in external conditions over time influence human survival, we consider how such survival is affected by differences in genetic parameters of individuals taken from the same population cohort (the original FHS cohort), and exposed to the same external conditions. Therefore, different age patterns of survival (mortality rates) for these sub-cohorts are likely to be associated with differences in the parameters V_0 and B of the vitality function, $V(x)$, which may reflect differences in the genetic backgrounds of the individuals comprising the respective sub-cohorts. The representation of the parameters of the Gompertz mortality curve given by eq. (S2) together with survival curves shown in Fig. 1B indicate that the rectangularization pattern of changes in survival, in this case, can be observed if the initial value of vitality, V_0, increases when the number of longevity alleles carried by each study participants increases (see Figs. S2 and S3). In populations with such genetic backgrounds, the parameter B remains unchanged, so the rate of vitality decline (which is characterized by the product, V_0B) increases.

It is important to note that the SM correlation could be an artifact due to negligence of the Makeham baseline term in the Gompertz-Makeham model of mortality, when its value is not small compared to other component of mortality [13]. To test whether this is the case in our situation we approximated mortality rates for each sub-population of individuals with different genetic background using the Gompertz-Makeham and Gompertz curves and compared parameters using the likelihood ratio test. The test showed that the Gompertz model provides better fit to the data from both populations. This means that the Makeham term is not responsible for SM correlation between the Gompertz parameters in case of genetic model of mortality.

How saving lives transforms mortality rates. If, in some population experiencing mortality rate $\mu(x)$, it would be possible to save individuals' lives n times for everybody, then the mortality rate, $\mu(x)$, would be transformed and become [14],

$$\mu^{(n)}(x) = \mu(x)\frac{H^n(x)}{n!\sum_{i=0}^{n}\frac{H^i(x)}{i!}} \tag{S6}$$

where

$$H^n(x) = \left[\int_0^x \mu(t)dt\right]^n .$$

Characteristics of 39 SNPs. As Fig. S4 shows, call rate of most of 39 SNPs is over 95%, and there are only four SNPs with call rate below 90%. Although a general tendency in the literature is to make quality control (QC) procedure more stringent, e.g., to increase call rates, a number of studies used rather liberal values of these characteristics. For example, Cupples et al. [15] and Lunetta et al. [16] published GWAS analyses on longevity based on call rate >= 80%. A genome-wide association study of breast and prostate cancer in the Framingham Heart Study was also based on call rate >= 80% [17]. In our paper Yashin et al. [1] (where these 39 SNPs were selected) we adopted similar criteria to make our findings comparable with those of Lunetta et al. [16]. Note that p-values of HWE test of 39 SNPs are between 0.0001 and 0.8549, and minor allele frequencies are between 14%-49%, which are within the current standards of the QC procedure.

The presence of population stratification may contribute to errors in GWAS. Therefore we carefully investigated the situation with population stratification in the FHS data. Sebro et al. [18] found evidence for ancestry-related assortative mating in the Framingham Heart Study participants and suggested the possibility of population stratification. However, Benjamin et al. [19] concluded that evidence for considerable population stratification in the FHS data has not been detected. In earlier publication Wilk et al. [20] found no evidence for the presence of population stratification within FHS. Later Wilk et al. [21] confirmed this conclusion using the method based on principal components analysis in the EIGENSTRAT software. The authors found that none of the first ten components were statistically significant ($p<0.05$) suggesting that population substructure was unlikely to confound the population-based association analysis. Cupples et al. [15] suspected that there is little population stratification in the FHS sample and Seshadri et al. [22] concluded that population stratification is not a major concern in the FHS sample due to the high homogeneity of ancestry (European).

We performed Linkage Disequilibrium test of 39 SNPs with software WGAViewer [23] and plotted distribution of LD values (r^2) between any two SNPs (see Fig. S4). The maximum r^2 of 0.03 shows that 39 SNPs can be considered independent. Therefore, bias due to dependence between SNPs is not a major concern in our analyses.

Note that GWAS allows for selecting genetic variants with both positive and negative effects on lifespan ("longevity" and "frailty" SNPs). Strictly speaking, both polygenic score indices constructed from "longevity" and "frailty" SNPs have to be used in evaluating joint effect of selected SNPs on lifespan. It turned out, however, that the numbers of detected genetic variants with positive and negative effects contained in persons' genomes were highly negatively correlated in our analyses. This correlation can be explained by the allele selection process: individuals who have a large number of "frailty" alleles and a small number of "longevity" alleles (and vice-versa) are more likely to contribute to a negative (positive) estimate of genetic

influence on lifespan than those who have about equal numbers of such alleles. Because of such correlation, we evaluated joint effect of "longevity" alleles only.

Supplemental References:

1. Yashin AI, Wu DQ, Arbeev KG, Ukraintseva SV. Joint influence of small-effect genetic variants on human longevity. *Aging.* 2010;2(9):612-620.

2. Meigs JB, Shrader P, Sullivan LM, et al. Genotype Score in Addition to Common Risk Factors for Prediction of Type 2 Diabetes. *New Engl. J. Med.* 2008;359(21):2208-2219.

3. Paynter NP, Chasman DI, Pare G, et al. Association Between a Literature-Based Genetic Risk Score and Cardiovascular Events in Women. *JAMA.* 2010;303(7):631-637.

4. Reeves GK, Travis RC, Green J, et al. Incidence of Breast Cancer and Its Subtypes in Relation to Individual and Multiple Low-Penetrance Genetic Susceptibility Loci. *JAMA.* 2010;304(4):426-434.

5. Ruiz JR, Gomez-Gallego F, Santiago C, et al. Is there an optimum endurance polygenic profile? *J. Physiol. (Lond).* 2009;587(7):1527-1534.

6. Talmud PJ, Hingorani AD, Cooper JA, et al. Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *Br. Med. J.* 2010;340:b4838.

7. Machiela MJ, Chen C-Y, Chen C, et al. Evaluation of Polygenic Risk Scores for Predicting Breast and Prostate Cancer Risk. *Genet. Epidemiol.* 2011;35(6):506-514.

8. Reeves GK, Travis RC, Green J, et al. Incidence of Breast Cancer and Its Subtypes in Relation to Individual and Multiple Low-Penetrance Genetic Susceptibility Loci. *J. Am. Med. Assoc.* 2010;304(4):426-434.

9. Witte JS, Hoffmann TJ. Polygenic Modeling of Genome-Wide Association Studies: An Application to Prostate and Breast Cancer. *OMICS: J. Integrative Biol.* 2011;15(6):393-398.

10. Qi L, Ma J, Qi Q, et al. Genetic Risk Score and Risk of Myocardial Infarction in Hispanics. *Circulation.* 2011;123(4):374-380.

11. Chen H, Poon A, Yeung C, et al. A Genetic Risk Score Combining Ten Psoriasis Risk Loci Improves Disease Prediction. *PLoS ONE.* 2011;6(4):e19454.

12. Hivert M-F, Jablonski KA, Perreault L, et al. Updated Genetic Score Based on 34 Confirmed Type 2 Diabetes Loci Is Associated With Diabetes Incidence and Regression to Normoglycemia in the Diabetes Prevention Program. *Diabetes.* 2011;60(4):1340-1348.

13. Gavrilov LA, Gavrilova NS. *The Biology of Life Span: A Quantitative Approach*. New York: Harwood Academic Publisher, 1991.

14. Yashin AI, Iachine IA, Begun AS. Mortality modeling: A review. *Mathematical Population Studies.* 2000;8(4):305–332.

15. Cupples LA, Arruda HT, Benjamin EJ, et al. The Framingham Heart Study 100K SNP genome-wide association study resource: overview of 17 phenotype working group reports. *BMC Med. Genet.* 2007;8(Suppl. 1):S1.

16. Lunetta KL, D'Agostino RB, Sr., Karasik D, et al. Genetic correlates of longevity and selected age-related phenotypes: a genome-wide association study in the Framingham Study. *BMC Med. Genet.* 2007;8(Suppl. 1):S13.

17. Murabito JM, Rosenberg CL, Finger D, et al. A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. *BMC Med. Genet.* 2007;8(Suppl. 1):S6.
18. Sebro R, Hoffman TJ, Lange C, et al. Testing for Non-Random Mating: Evidence for Ancestry-Related Assortative Mating in the Framingham Heart Study. *Genet. Epidemiol.* 2010;34(7):674-679.
19. Benjamin AM, Suchindran S, Pearce K, et al. Gene by sex interaction for measures of obesity in the framingham heart study. *Journal of Obesity.* 2011;2011:329038.
20. Wilk JB, Manning AK, Dupuis J, et al. No evidence of major population substructure in the Framingham Heart Study. *Genet. Epidemiol.* 2005;29(3):286-286.
21. Wilk JB, Chen T-H, Gottlieb DJ, et al. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet.* 2009;5(3):e1000429.
22. Seshadri S, DeStefano AL, Au R, et al. Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham study. *BMC Med. Genet.* 2007 2007;8(Suppl. 1):S15.
23. Ge D, Zhang K, Need AC, et al. WGAViewer: Software for genomic annotation of whole genome association studies. *Genome Res.* 2008;18(4):640-643.

Supplemental Tables:

Table S1: The effect of polygenic score constructed from 39 SNPs on lifespan after controlling for smoking

Variable	Estimate	S.E.	t Value	P-value
Intercept	80.36752	0.51957	154.68	<.0001
Polygenic score (39 SNPs)	0.39365	0.02412	16.32	<.0001
Smoking	-3.17152	0.40033	-7.92	<.0001

Supplemental Figures:

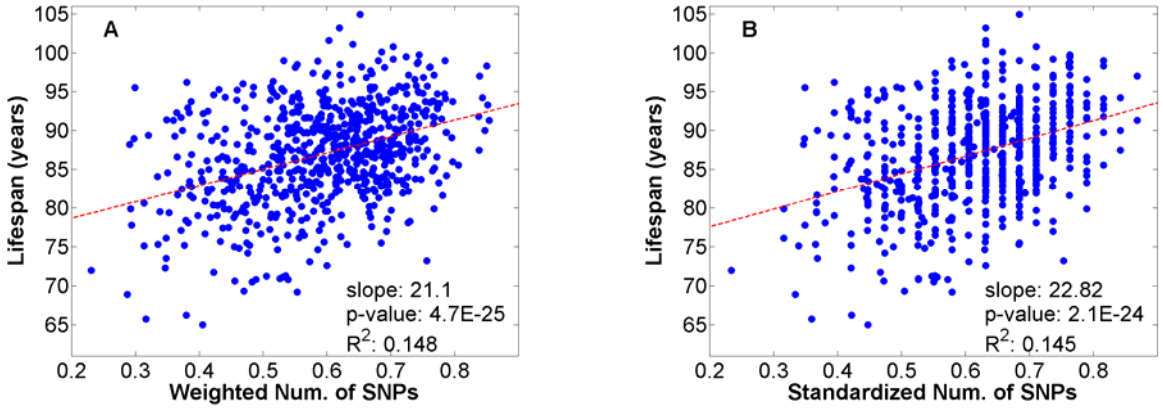

Fig. S1: Two indices explain about the same proportion of life span variance. (A) Life span as a function of additive genetic component (AGC index) evaluated for individuals from the original FHS cohort. Linear regression explains about 15% of phenotypic variance. (B) Life span as a function of the number of individually selected genetic variants (NLV index) contained in individuals' genomes evaluated for individuals from the original FHS cohort. Linear regression shows "genetic dose – life span response" relationship and explains about 15% of life span variance.

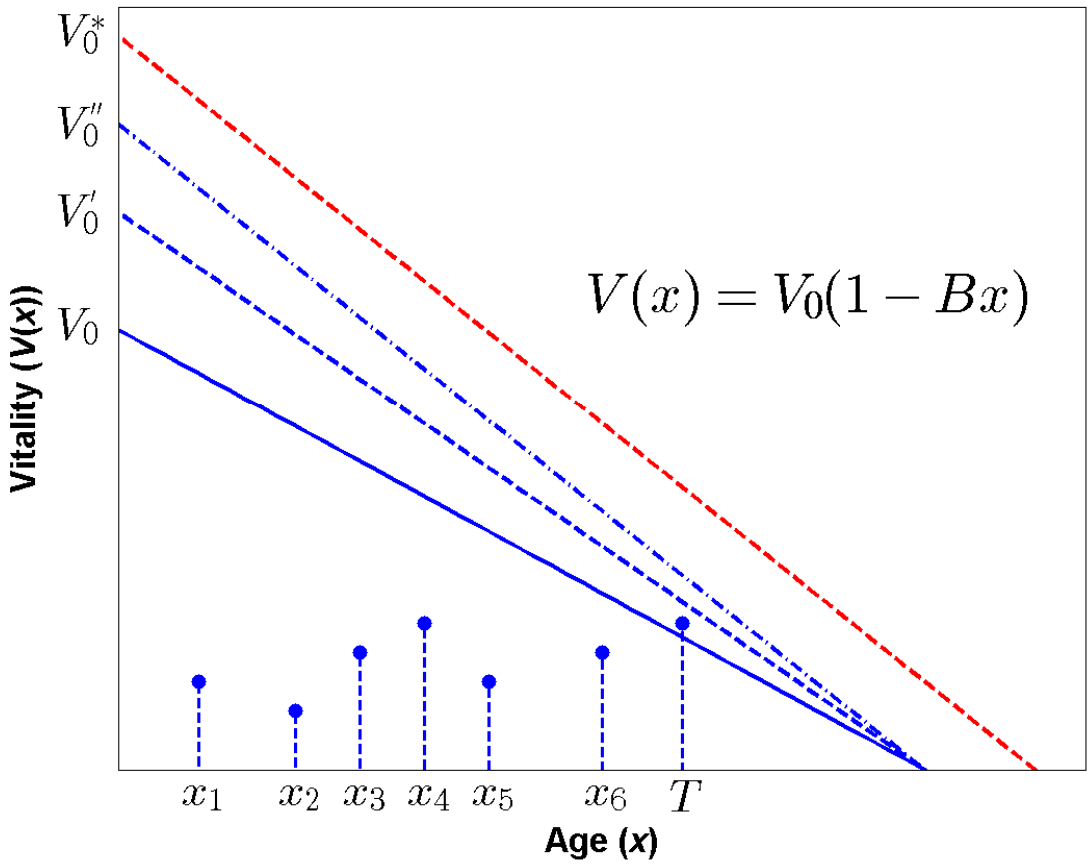

Fig. S2: Vitality functions in the Strehler and Mildvan model. The blue lines correspond to changes in initial vitality, V_0. When V_0 increases and B is constant the absolute rate of vitality decline V_0B increases. The vertical lines characterize external disturbances of different magnitudes occurring at random times $x_1, x_2, \ldots x_6, T$. Death occurs at age T when the magnitude of disturbance exceeds the value of vitality function. The death at age T can be avoided when initial vitality increases. The red line corresponds to simultaneous increase in V_0 and decrease in B, so the slope of vitality function V_0B does not change.

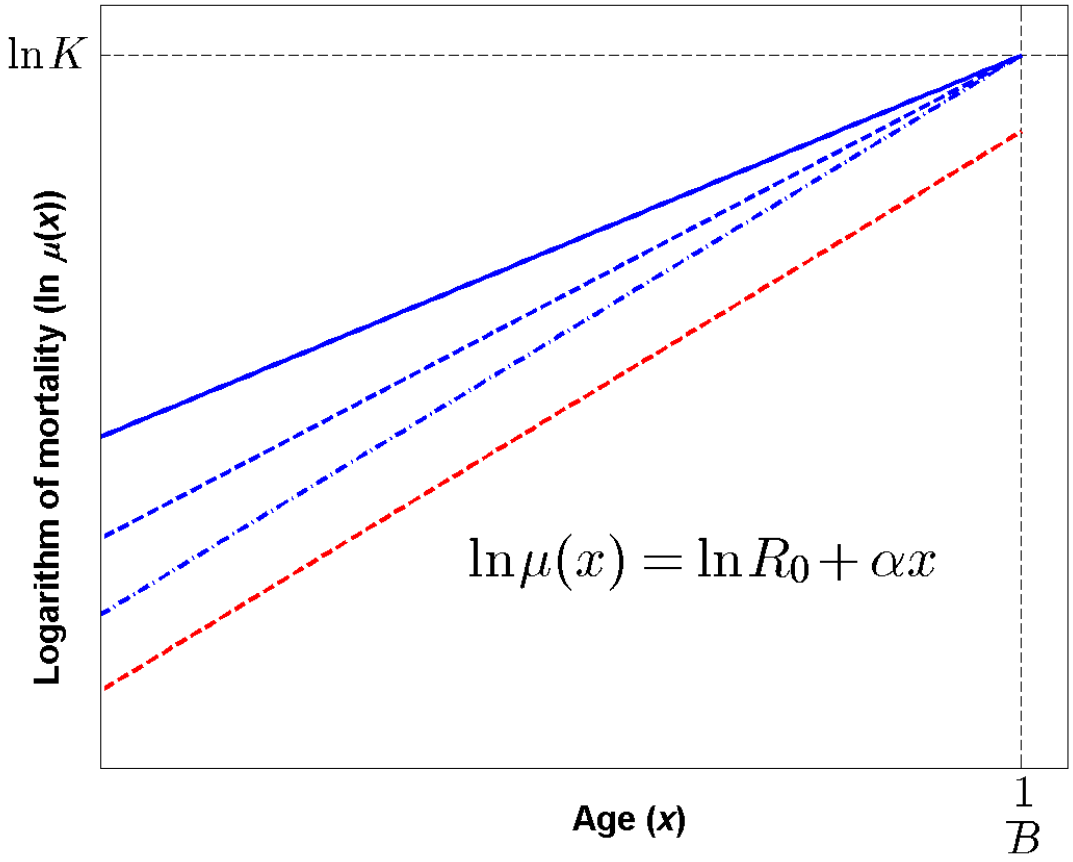

Fig. S3: Changes in the age patterns of the logarithms of mortality rates in response to modification of vitality. The blue lines correspond to changes in initial vitality, V_0, when parameter B remains unchanged. An increase in V_0 results in decline of the Gompertz parameter R_0 and in an increase of α. Respective changes correspond to counter clockwise rotation of the logarithm of mortality rate $\mu(x)$ around the point $(1/B, \ln K)$. The red line corresponds to the logarithm of the mortality rate resulted from an increase in V_0 and simultaneous decline in B, so the absolute rate of vitality decline, V_0B, remains unchanged.

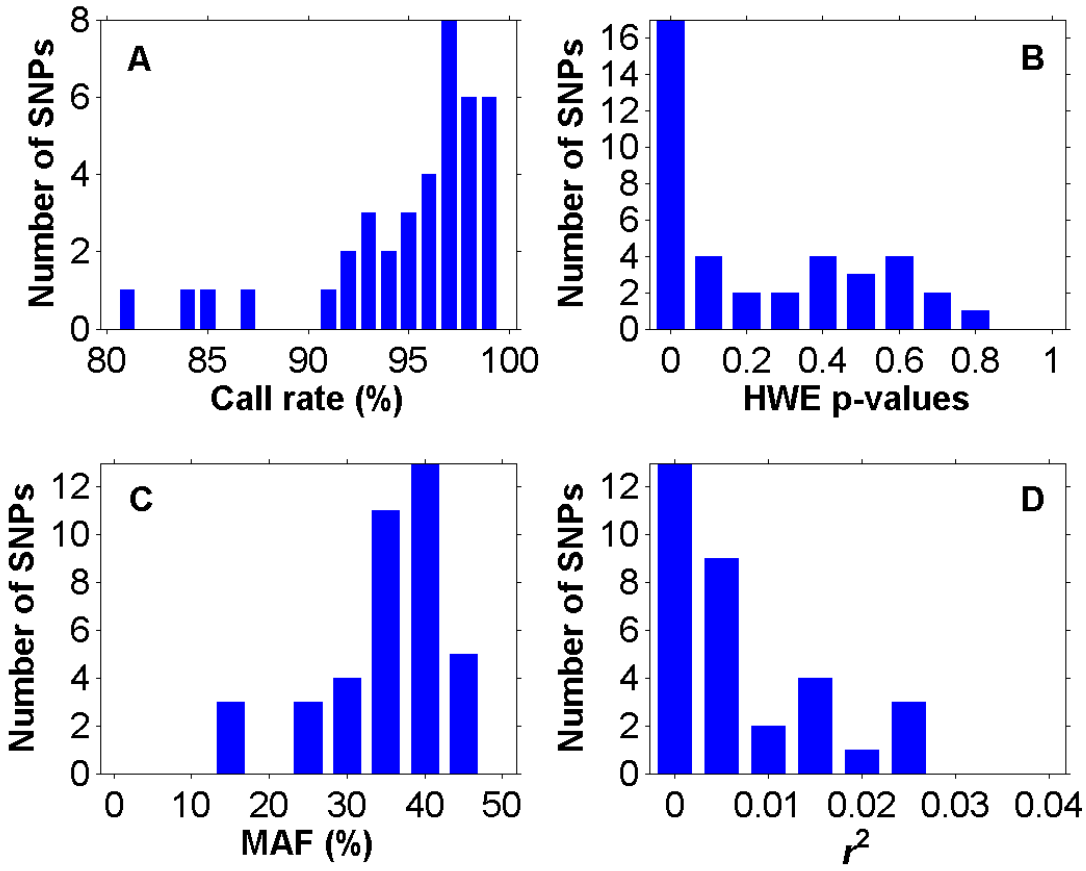

Fig. S4: Distributions of different characteristics of 39 SNPs: (**A**) call rates; (**B**) Hardy-Weinberg equilibrium (HWE) p-values; (**C**) minor allele frequencies (MAF); (**D**) linkage disequilibrium (LD) values (r^2).

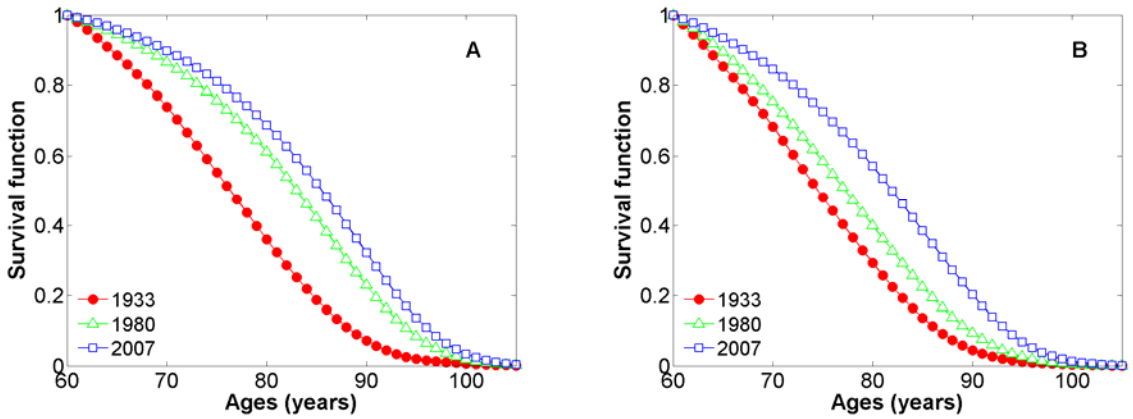

Fig. S5: Survival curves (conditional at age 60) in the U.S. in 1933-2007: (**A**) females; (**B**) males (data source: Human Mortality Database).

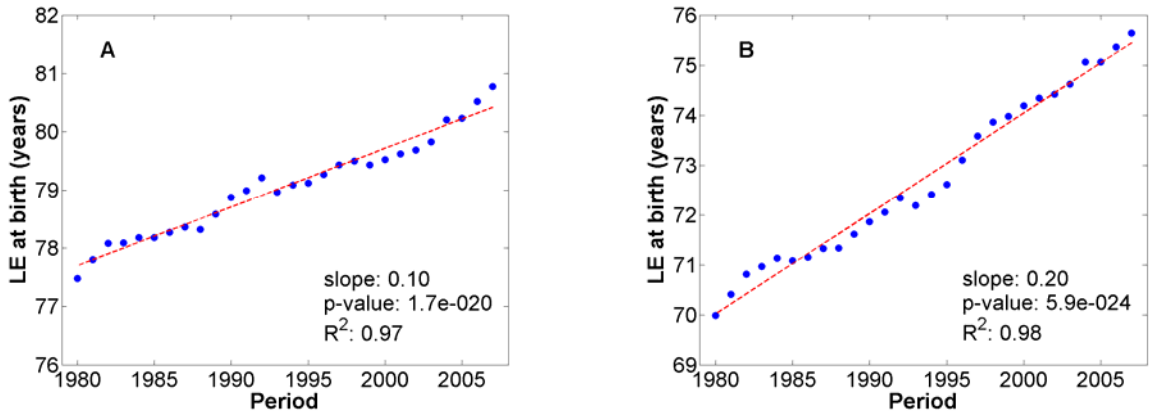

Fig. S6: Life expectancy at birth in the U.S. in 1980-2007: (**A**) females; (**B**) males (data source: Human Mortality Database). Dots represent observed data; lines represent the fitted linear regressions.

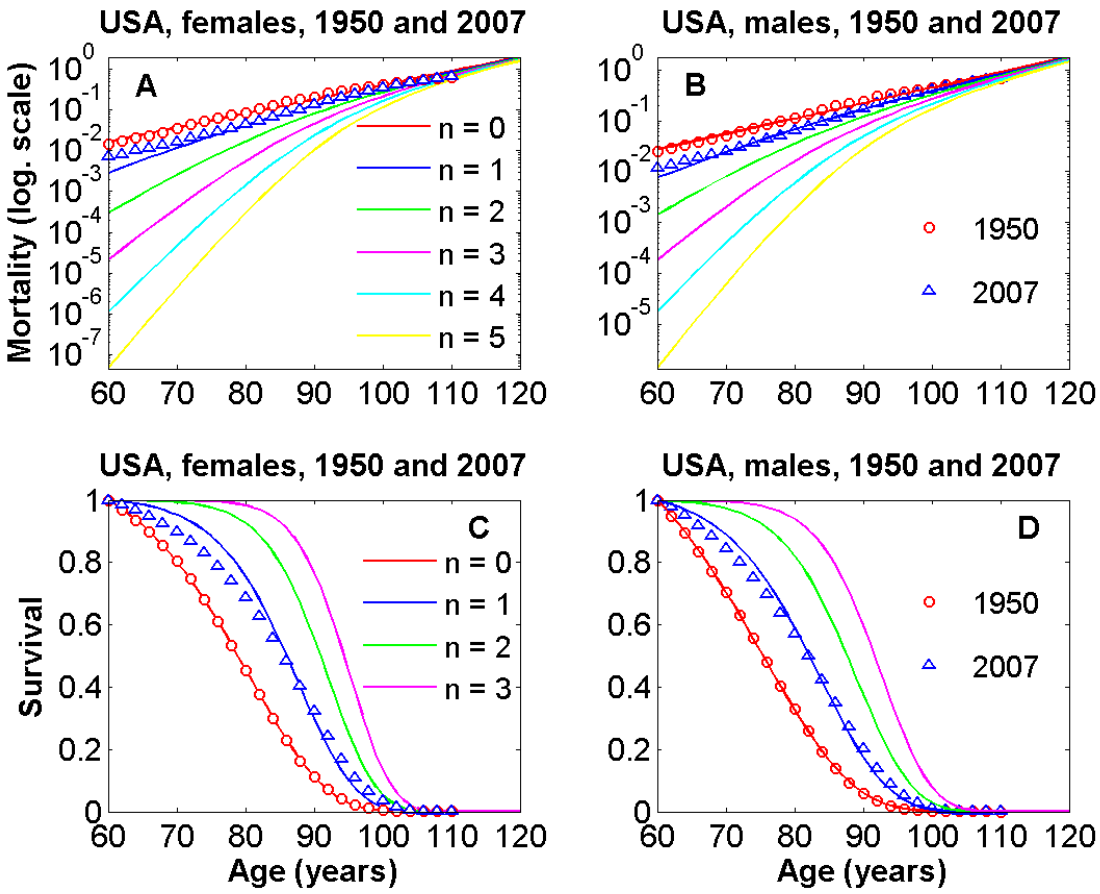

Fig. S7: Sex-specific curves illustrating how saving lives contributes to explanations of mortality and survival improvement (see Fig. 4 for total population). **(A)** Logarithms of mortality rates for U.S. females in 1950 (open circles) and in 2007 (open triangles). The 1950 mortality rate is approximated by the Gompertz function with $n = 0$, where n is the number of times individual's life has been saved. The other curves correspond to mortality rates obtained by transforming the 1950 Gompertz mortality rate using the life saving equation with n = 1, 2, 3, 4, and 5. **(B)** Survival functions for U.S. females in 1950 (open circles) and in 2007 (open triangles). Three other survival curves correspond to mortality rates obtained by transforming the 1950 Gompertz curve using the life saving equation with n = 1, 2, and 3. (**C**) Logarithms of mortality rates for

U.S. males in 1950 and in 2007 (all notations are similar to (**A**)). **(D)** Survival functions for U.S. males in 1950 and in 2007 (all notations are similar to (**B**)).